UNION INTERSECTION TEST IN INTERPRETING SIGNAL FROM MULTIVARIATE CONTROL CHART

Siti Rahayu Mohd. Hashim

Thesis submitted to the University of Sheffield
for the degree of Doctor of Philosophy.

Department of Probability & Statistics,
School of Mathematics & Statistics

December 2014

ABSTRACT

Statistical Process Control (SPC) has been a very important discipline in quality control study since pioneered by Walter A. Shewhart in 1920s. Control charting is one of the important tools in SPC and has received wide attention from researchers as well as practitioners. The complexity and the impracticality in monitoring several univariate control charts for a multivariate process has made many practitioners use a multivariate control chart instead. Its usage gives a better control of the overall Type I error and the interdependency among variables is retained. Unfortunately, a multivariate control chart is not able to pinpoint the responsible variable(s) once an out-of-control (OOC) signal is triggered. Many diagnostic methods have been proposed to overcome this problem but all of them have their own limitations and drawbacks. The applicability of a diagnostic method for a limited number of variables, lack of physical interpretation, the complexity of the computation procedure and lack of location invariance are among the factors that have inhibited the implementation of multivariate charts. Lack of comparative studies for various diagnostic methods also makes it difficult for practitioners to choose an appropriate diagnostic method.

This study highlights some problems that might arise in a comparison of diagnostic methods and makes suggestions to overcome them, hence, making the results of a comparative study more relevant and reliable. The effects of several factors such as the size of the deviation in a mean vector, the combination of various sizes of shifts in a mean vector and the inter-correlation among the variables on the performance of diagnostic methods are studied and a summary of the suitability of certain diagnostic methods for certain situations is given. This study presents a new comparison involving two diagnostic methods adapted from the methods proposed by Doganaksoy, Faltin and Tucker (1991) and Maravelakis et al. (2000). A problem related to the usage of eigenvectors with similar eigenvalues is revealed in this study and suggestions from previous studies regarding this matter are presented.

Due to lack of multivariate approaches in dealing with the interpretation of a multivariate control chart signal, this study proposes a new method which embraces the principles of Union Intersection Test (UIT) in diagnosing an OOC signal. A thorough discussion of the UIT principle, the hypotheses, the test statistic and the application of the union intersection technique in the diagnosis problem is presented. An extension of the first comparison study is which includes the proposed method is carried out. The performance of the new diagnostic method is studied and its strengths and weaknesses are discussed. A simplified version for the new method, involving application of spectral decomposition, is also proposed. By using this simplified approach, the common practice of considering multiple types of covariance matrices in a comparison study of diagnostic methods can be avoided to some extent. This study is concluded with a few suggestions of potential further work.

To

Wilter @ Azwal Malandi

Muhammad Fiqri Wilter

Siti Nur Fathihah Wilter

Muhammad Fadlin Wilter

ACKNOWLEDGEMENTS

I am heartily thankful to my supervisor, Eleanor Stillman, whose guidance and support has made the completion of this thesis a reality. I am also grateful to my previous supervisor, Nick Fieller, whose encouragement from the initial level until the end has enabled me to develop an understanding of the subject.

I would also like to extend my deepest gratitude with love to all my family members especially to my beloved husband, Wilter @ Azwal Malandi and my dearest children, Fiqri, Fathihah and Fadlin, for their support and prayers for my success.

I am also grateful to have met many friends who have supported me throughout this journey. The special note of thanks goes to the kindest friend I've ever have, Suriani Hassan, a person who always finds a way to help me in any circumstances. Last but not least, huge thanks to Lu Zou, who is always available to help me in my R Programming and also to Nor Zainiyah Mokhtar, who is always there whenever I need her.

Lastly, I owe my deepest gratitude to Universiti Malaysia Sabah and the government of Malaysia for giving me the opportunity to further my study.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CUSUM	Cumulative Sum
EWMA	Exponentially Weighted Moving Average
LCL	Lower Control Limit
LD	Largest Deviation
MiniMax	Minimum Maximum
MP	Multivariate Profile
OOC	Out of Control
PCA	Principal Component Analysis
S.E.	Standard Error
SPC	Statistical Process Control
UCL	Upper Control Limit
UIT	Union Intersection Test

CHAPTER 1

INTRODUCTION

1.1 Introduction

Statistical Process Control (SPC) is very important in Quality Control studies. Even though it has been developed for almost 100 years, some issues are yet not fully resolved. One of the most important issues is handling multivariate data and diagnosis causes of any problems to aid in suitable control strategies. The diagnosis of causes of faults is the subject of this thesis. This chapter's structure begins with the historical development of Quality Control and the people who contributed to its development. Mass production and its consequences is discussed which led to the awareness and further knowledge on another topic on variation and its causes. Variation and its causes is the important and significant reason of the introduction of the first univariate control chart, one of the most important tools in SPC. Control charting also developed and led to the introduction of multivariate control charts and the diagnosis problem in the later sub section. Several review studies on SPC are presented next to highlight the importance of SPC in variability studies and several future research suggestion were given. This chapter concluded with reviews for the rest of the chapters in this thesis.

1.2 Development of Statistical Process Control

Statistical Process Control (SPC) is defined by Montgomery (1997) as "... a powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability...". Statistical Process Control (SPC) has been a very important discipline since pioneered by Walter A. Shewhart of Bell Telephone Laboratories in the 1920's. The industrial revolution has played a very important role in the expanding of the knowledge on Statistical Process Control (SPC) and the implementation of statistical techniques in productions. It began in the United Kingdom during the 18th century and later on extended to United States and other countries such as Europe and Japan. Bell Labs has been responsible in setting an international standard quality through the U.S. telecommunications industry in 1930s and Shewhart's contributions played a large part of it through his statistical techniques (Richard, 1992).

Pre-Industrial Revolution
Quality "controlled" by the individual craftsman/artisan who was involved in all aspects of the product life cycle.

1875 **Birth of "Taylorism"**
The concepts and methods of mass production and the notion of the division of labour begin to appear in the American industrial sectors.
F.W. Taylor develops the principles of scientific management.

1925 **Shewhart Introduces Statistical Process Control**
Walter Shewhart of Bell Labs develops a statistical approach to the study of manufacturing process variation for the purpose of improving the economic viability of the process. The methods are based on the continual on-line monitoring of process variation.

1930 **Dodge and Romig Introduce Accceptance Sampling Methods**
Dodge and Romig at Bell Labs develop a system of lot-by-lot sampling inspection of manufactured product for the purpose of determining its suitability for shipment to the customers. The methods are based on a probabilistic approach to the prediction of the lot character based on sampling results.

1950 **Deming Approach to Quality/Productivity Improvement**
W. Edwards Deming develops a statistically based approach to quality/productivity improvement patterned scientifically after the work of Shewhart and projected on an institutional basis. Central to this approach is emphasis on the responsibilities and obligations of top management.

1980 **United States Recognizes the Deming Approach and Taguchi Methods**
U.S industrial leaders begin to embrace the Deming philosophy of quality improvement and America begins to transform its industrial sector. The United States is introduced to the methods of Taguchi and the techniques of statistical design of experiments become well known. Emphasis begins to be placed on pushing the quality issue upstream into engineering design.

Figure 1.1 Historical Evolution of Quality Control (adapted from Richard, 1992)

The Second World War triggered an extensive use of statistical quality methods especially in America. Statistical quality methods were used to improve America's war time production. On the other hand, Japan implemented SPC successfully during post war years. The individuals who were responsible for the Japan's success were Joseph M. Juran who worked with Walter A. Shewhart at Bell Labs and W. Edwards Deming who was greatly influenced by Shewhart. In 1940s, they were already recognized as the world's foremost experts on quality (Leavengood and Reeb, 1999). Unlike the companies in the United State, the Japanese welcomed both of them and embraced their quality-management philosophies heartily. Juran was responsible for teaching the Japanese about quality management and Deming for their census. Later on, Deming became more interested and heavily involved in helping the Japanese rebuild their industry and by extension boosting up their economy (Lewis, 1994). The implementation of quality methods grew rapidly in Japan during the 1960s and 1970s and by the 1980s, Japan companies became strong competitors in America and in the world. In 1980's, United States begins to accept and embrace Deming's philosophy widely. A new approach introduced by Taguchi also accepted by the United States in which later on known as an engineering quality approach (Alwan, 2000). Taguchi also put his interest on variation as Shewhart. In fact, Taguchi gave a new terminology for variation which is known as "noise" and he has categorized noise factor into three which are external, deterioration and manufacturing noise (Alwan, 2000). Taguchi has actually developed the concept of robustness by taking the noise factor into account (Besterfield, 2004). A timeline of important events is shown in Figure 1.1 and these concepts are discussed in the following subsections.

1.2.1 Mass Production and its Consequences

Due to high demand in products during the industrial revolution, the industrial sectors began to transform from the individual craftsmen to a big group of workers involved in a mass production of products. It was during this time the principles of scientific management been introduced and F.W. Taylor pioneered the field of industrial management (Richard, 1992). Mass production became a common practice. In mass production, machinery has replaced humans in many production tasks and the labour has been split into divisions. Even though repetitive tasks by workers and the use of machinery accelerate the production time and result in other productivity gains, unfortunately it brings in other problems. It took away the pride of workmanship from workers as well as the pride of ownership of the production process since the process is shared by many people from various divisions as well as machinery involved in the production line.

Since there are many people and machines involved in a mass production line then it results in many factors and variables affecting the quality of a product and making product quality become harder to manage. The production supervisors or production engineers are forced to move from the old practice of searching for problems or a “product oriented” quality control system to a new practice of preventing the problems or a “process oriented” quality control system. In an attempt to prevent problems, production engineers were forced to look beyond the old practices in monitoring and controlling quality of products. They need to monitor and control the process of the production instead and to do so they need to use appropriate statistical methods and techniques. The statistical methods which are very popular among the production or quality engineers nowadays are known as Statistical Process Control (SPC) tools.

1.2.2 Variation and the Causes

Shewhart was the first person to introduce the use of statistical techniques for monitoring and controlling quality (Wadsworth, Stephens and Godfrey, 2002). During the industrial revolution, Bell Telephone Laboratories was trying to monitor and control the variation of the quality of their components and their finished products economically. Shewhart realized that the best way to achieve it is by monitoring and controlling variation throughout production. This monitoring and controlling is very useful in ensuring a process behaves in a predictable way. A process that behaves in a predictable way will result in a consistent product quality.

Attaining consistent product quality requires a good understanding of process variation and a good understanding of how to monitor and to control the variation. Shewhart himself laid the foundations for SPC and the most important one is by recognizing and emphasizing the two causes of variation in a process which are called chance causes and assignable causes. Chance causes are also known as common or random causes, whereas assignable causes are also known as special causes. Even though both of them create variation, the chance causes are considered as contributing to ‘controlled’ variation whereas assignable causes contribute to ‘uncontrolled’ variation. Shewhart (1931) gives a further explanation of the term of ‘controlled’ variation as follows

“A phenomenon will be said to controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that prediction within limit means that we can state, at least approximately, the probability that the observed phenomenon will fall within given limits.”

Shewhart (1931) also proposed the following three postulates:

1. “*All chance systems of cause are not alike in the sense that they enable us to predict the future in terms of the past*”.
2. “*Systems of chance causes do exist in nature such that we can predict the future in terms of the past even though the causes be unknown. Such a system of chance is termed constant.*”
3. “*It is physically possible to find and eliminate chance causes of variation not belonging to a constant system*.”

By relating to the concept of variation too, Deming (1982) defines control in SPC as “*A stable process, one with no indication of a special cause of variation*”. A process is said to be in statistical control if only chance causes of variation are present in the process (Deming, 1993; Montgomery, 2005). In order to monitor any special cause of variation in a process, Shewhart has invented a visual tool which came to be known as the control chart or famously known as Shewhart control chart in honour of him as its inventor. Shewhart introduced the sketch of a control chart the first time in an unpublished memorandum dated May 16, 1924 (Alwan, 2000). Shewhart continually refined the concept and the technique of the control chart which led to the publication of his book titled *Economic Control of Quality of Manufactured Product* in 1931 (Alwan, 2000).

1.2.3 Control Charts

The first control chart introduced by Shewhart was a univariate control chart. A univariate chart is used to monitor one process variable or quality characteristic. Normally, one or more variables in a process are continuously measured and plotted along with a specific range known as control limits, determined probabilistically. Measurements or observations are obtained in sequence and in the graphical presentation plotted against sequence number. If an observation falls outside the specified limits, this is regarded as an out-of-control (OOC) signal and the process is said to be statistically instable or statistically out-of-control. As long as all the observations remain within the specified limits, the process is regarded as statistically in-control or stable.

1.2.4 SPC Research

Woodhall (2000) emphasized the importance of SPC in any research on variability. SPC is important in any attempt to understand variability, model it and reduce it. There are many useful research areas waiting to be explored by SPC researchers and practitioners as discussed by Woodhall and Montgomery (1999) and one of them is multivariate methods. Woodhall (2000) stated that research activities on multivariate methods specifically on multivariate SPC has been “at its highest level” now. This happened due to the increased in measurement and the advancement in computing capability.

Another active research area in SPC nowadays is research on the effect of estimation errors (Woodhall and Montgomery, 1999). Woodhall and Montgomery (1999) pointed out that more research is needed on the evaluation of control charts, comparison studies on the performance of control chart especially in Phase II control charting, estimation effects on the performance of control chart and studies on control charts’ control limits. Kourti and MacGregor (1996) also emphasize on the importance of control chart assessment specifically in its ability to detect an event as well as “its robustness to false signals when any of the other event occurs”. SPC has become more important now and continues to be so with adaptation to the changes in manufacturing environments. The pressure for higher quality requirements, shorter production runs, and massive data available and an advancement and greater computing capability will require changes in SPC approach.

1.2.5 Multivariate Control Charts and the Diagnosis Problem

It is very easy to monitor one or two variables in a process since one shall know straight away which variable caused instability in a process. But a problem arises when practitioners are dealing with a multivariate process, a process which involves several variables. It is not easy to monitor several control charts at a time and in a real situation it is obviously impractical. Furthermore, the variables related to the process might have correlation between them and that in some ways might affect the process. By using multiple separate univariate control charts to monitor a process does not take into account any possible correlation between the process variables. Hotelling in 1947 was the first to introduce a multivariate statistic, named Hotelling's T^2, which can be plotted as multivariate control chart and soon after, the chart has been applied by Hotelling (1947, 1951) in bombsights study. This statistic combined the information on means and dispersions of multivariate observations.

In recent years, multivariate control charts have received wide attention among the researchers and practitioners in SPC. Being one of the most important tools in SPC and supported by the advances in computer programming, multivariate control charts have been used by many in monitoring multivariate processes. Unlike univariate control charts, multivariate control charts are capable in dealing with the issues on inter-correlation among process variables and controlling the overall type I error. Unfortunately, multivariate control charts have one major problem. Once the out-of-control (OOC) signals have appeared in a multivariate control chart, it is not easy to tell which process variables caused the signals.

Therefore, since 1985, a number of interpretation methods, which shall be referred to as diagnostic methods in later discussions, have been proposed to assist practitioners in finding the aberrant variables which are responsible for the OOC signals in multivariate control charts. However, performance of these methods has not been rigorously scrutinized and some have obvious deficiencies. In this thesis, we examine several prior proposals.

1.3 Thesis Overview

The role of the control chart as one of the basic tools in SPC will be discussed briefly in Chapter 2 to give some basic understanding of its importance in process monitoring. The discussion covers the phases in control charting, and the utilization of control charts in both univariate and multivariate process monitoring. Univariate control charts are discussed first and a few drawbacks in monitoring several univariate processes using a few univariate control charts are highlighted, which later brings multivariate process monitoring into the picture. The multivariate Hotelling's T^2 control chart will be presented as the most popular multivariate control chart. The application of the multivariate control chart will be discussed and another two popular multivariate control charts are introduced. It will become apparent on the following discussion that a multivariate control chart is not always easy to interpret. The interpretation problem will lead to another discussion on the methods available to assist in the interpretation of out-of-control signals. Several interpretation methods applicable to multivariate process control monitoring, hereafter called diagnostic methods, are also presented in this second chapter. One of the most recent studies in comparing several diagnostic methods is discussed at the end of the chapter.

Chapter 3 discusses in detail the result of a comparison study done by Das & Prakash (2008). The studied diagnostic methods are compared based on their power as defined by Das and Prakash (2008). The discussion is assisted by new tables and figures presenting Das and Prakash's results. It will become apparent later that several remarks and conclusions given by Das and Prakash are contradicted and this study offers additional new observations. The results are studied thoroughly and the effect of the size of shifts in mean vector, the combination of the shifts in mean and the correlation structure is observed. Finally, this chapter is concluded with suggestions for a new comparison study.

Chapter 4 presented a new comparison between one of the diagnostic methods in Das and Prakash (2008), proposed by Doganaksoy, Faltin and Tucker (1996) and a new method, called the Ratio method, which is adapted from the method proposed by Maravelakis et al. (2002). The simulation study presented in this chapter utilise the correlation matrices proposed by Doganaksoy, Faltin and Tucker (1996) and the performance of the methods is tested against different shifts in mean vector. The simulation results are presented in three different subsections, separating the findings for the cases of one aberrant variable, two aberrant variables in the same directions and two aberrant variables in opposite directions. Throughout the discussion, this study will highlight a peculiar result produced by the Ratio method and further study is carried out to investigate the cause of the peculiar result. Since the diagnostic method proposed by Maravelakis et al. (2002) utilizes eigenvectors from a principal component analysis, some research studies with regards to that matter are presented and discussed.

The Union Intersection test is introduced in Chapter 5. A new approach which adapts Union Intersection principles for interpreting the out-of-control signal triggered by a multivariate control chart is proposed. The underlying hypothesis testing and the theoretical background of the new approach are also given. A simulation study on the application of the new approach, called the Largest Deviation (LD) method, is carried out together with the two diagnostic methods in Chapter 4. The simulation study presented in this chapter again utilizes the correlation matrices proposed by Doganaksoy, Faltin and Tucker (1996) and by Das and Prakash (2008). The simulation results are presented in two parts with respect to the type of correlation matrices used in the simulation, non equi correlation matrices (Doganaksoy, Faltin and Tucker, 1996) and equi correlation matrices (Das and Prakash, 2008). For each part, the simulation results are presented in four different cases which separate the findings for the cases of one aberrant variable, two aberrant variables in the same positive directions, two aberrant variables in the same negative directions and two aberrant variables in opposite directions. The performance of the proposed method also tested under several randomly selected correlation matrices. Chapter 5 concludes with some discussion of the strengths and the weaknesses of the new approach.

Chapter 6 proposes a way to improve the efficiency of the new approach by transforming data with known covariance matrix or very well estimated into a standard data space. The square root of the covariance matrix is needed for the data transformation and spectral decomposition procedure is shown to be able to provide it. A few examples are given to show that no generality is lost if we consider the identity covariance matrix in dealing with data in standard data space. Chapter 7 will presents the proposed procedures in applying threshold values for the LD method. The threshold values are introduced in a standard as well as in the original data space. An estimated power assessment is shown for selected combinations of shifts in mean vector. A few examples are also given at the end of this chapter to illustrate the application of the extended method and the identification of the aberrant variable for the dimensions, p = 2 and p = 4. Chapter 8 presents the overall conclusions and discussions of this study where the summary of chapters 3, 4, 5, 6 and 7 is also given. Some suggestions on the future work are given in the final section of this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Industrial statistics consists of the areas of acceptance sampling, statistical process control (SPC), design of experiments and capability analysis. There are seven basic tools in SPC and control charts are the most important (Montgomery,1995). The univariate control chart was introduced by Walter Shewhart in the 1920s and it is basically a graphical presentation of a process measurement which depicts the behaviour of a process. The trend of the points in a control chart might tell one whether the process is stable statistically or in other words in an 'in-control' state. A stable process only exhibits variation from chance causes which is inherent in the process and always regarded as a part of the process. Normally, all the points in the control charts will be within the specified control limits. With some fluctuations, a not 'in-control' process has at least one out-of-control (OOC) signal, where at least one of the measurement points is located beyond the control limits or maybe a systematic pattern of points or trend exists which depicts a shift in the mean of the process variable. The unstable process consists of variation from assignable causes which need to be removed in order to bring the process back to the "in-control" state. Since the subject of this thesis is how to identify or diagnose causes of OOCs in multivariate control charts, we begin our study by discussing the underlying elements. That is, we firstly introduce univariate control charts then multivariate forms together with the conditions for triggering an OOC. Existing diagnostic method are then considered and their limitations outlined. Studies which have compared diagnostic performance are introduced at the end of the chapter before a more detailed examination of a particular study is given in Chapter 3.

2.2 Univariate Control Chart

There is quite a number of univariate control charts which depend on the level of measurement of the data and the feature of interest. Among the most popular ones for quantitative data are a Range control chart, $\overline{X}$ or Average control chart, Exponentially Moving Average (EWMA) control chart and Cumulative Sum (CUSUM) control chart.

2.2.1 Shewhart Control Chart

When a process has only one process variable, X, or only one to be assessed, where $X \sim N(\mu_0, \sigma_0^2)$ and the parameters are known, then a Shewhart control chart for the mean has μ_0 as the centre line on the control chart and the following upper control limit (UCL) and lower control limit (LCL) (Lowry and Montgomery, 1995) as given below

$$\mathrm{UCL} = \mu_0 + Z_{\frac{\alpha}{2}}\left(\frac{\sigma_0}{\sqrt{n}}\right)$$

$$\mathrm{LCL} = \mu_0 - Z_{\frac{\alpha}{2}}\left(\frac{\sigma_0}{\sqrt{n}}\right)$$

where $Z_{\frac{\alpha}{2}}$ is the critical value of a standard normal distribution (Alt, 1985) for a specified level of significance, α, whereas μ_0 and σ_0 is a known mean and standard deviation of variable X with *n* individual observations, respectively. The upper and the lower limits of the control can also be written (Alt, 1985) as $\mu_0 \pm A\sigma_0$ where *A* is depending on *n* and the tabulated values of *A* are given in Duncan (1974). The values given in Duncan (1974) is specifically for *n* = 2, 3, ... , 25. The μ_0 and σ_0 are assumed known in this case.

However, the 'parameters' are most of the time unknown. A good estimation is required for the parameters. Suppose that a process variable from a sample sized *n*, is normally distributed with mean, μ and standard deviation, σ, then the average of variable X is given by

$$\overline{\mathrm{X}} = \sum_{i=1}^{n} \frac{x_i}{n}$$

and $\overline{\mathrm{X}}$ is normally distributed with mean μ and standard deviation $\sigma_{\bar{\mathrm{X}}} = \frac{\sigma}{\sqrt{n}}$. Suppose that *m* samples of size *n* are drawn from a process and let $\overline{\mathrm{X}}_1, \overline{\mathrm{X}}_2, \ldots, \overline{\mathrm{X}}_{\mathrm{m}}$ be the average of each sample. The best estimator of the process average, μ is

$$\overline{\overline{\mathrm{X}}} = \sum_{\mathrm{j}=1}^{m} \frac{\overline{\mathrm{X}}_{\mathrm{j}}}{m}$$

The statistic $\overline{\overline{\mathrm{X}}}$ is called the 'grand average' and is used as the centre line on the $\overline{\mathrm{X}}$ control chart (Montgomery, 2005). Control limits for the control chart are needed to assess the in-control statistical state of a process. A process is said to be no longer in-control state when any points of $\overline{\overline{\mathrm{X}}}$ falls above UCL or below LCL. In practice, $\overline{\mathrm{X}}$ control chart is always used together with Range control chart.

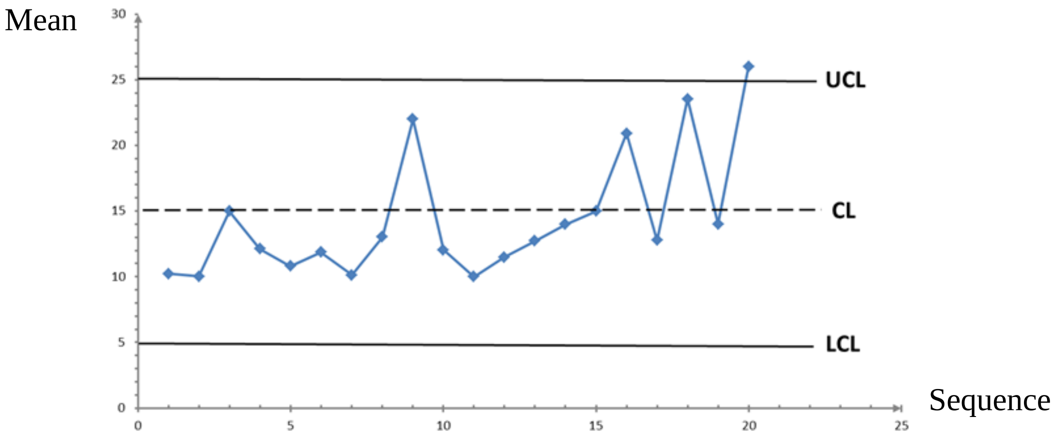


Figure 2.1: Univariate control chart

2.2.2 CUSUM Control Chart

The Cumulative Sum (CUSUM) control chart was first introduced by Page (1954) and it has been studied ever since. There are many authors who have contributed to the application development of CUSUM control charts (Ewan, 1963; Bissel, 1969; Lucas, 1973, 1976; Hawkins, 1981; Woodhall, 1985; Montgomery, 1996). The CUSUM chart is more capable in monitoring deviations in a process (Lucas, 1976; Lucas and Crosier, 1982; Woodall and

Ncube, 1985; Healy, 1987; Crosier, 1988). This chart makes use all the historical data by accumulating the difference of the successive observation and the target value. The condition of a process is monitored by analysing the slope of the chart.

Suppose a few samples of size *n* are collected, $\bar{x}_j$ is the average of the *j*th sample and μ_0 is the target value of a process. The CUSUM control chart statistic is given by

$$C_i = \sum_{j=1}^{i} (\bar{x}_j - \mu_0)$$

where C_i is the sum of cumulative discrepancy including the *i*th sample. The state of a process is assessed and monitored by the changes in C_i. If the process average changes to any value $\mu_1 > \mu_0$, C_i has an ascendant tendency, while any changes to some value below $\mu_1 < \mu_0$ indicates a negative direction of C_i. If one of these two tendencies appears, it is considered as a sufficient evidence that the process has changed due to an assignable cause (Vargas et al., 2004). The V-mask procedure on CUSUM chart became popular after being suggested by Barnard (1959) due to its usefulness in the interpretation of the CUSUM control chart.

2.2.3 EWMA Control Chart

The Exponentially Weighted Moving Average (EWMA) control chart was initially proposed by Roberts (1959). Subsequently many authors have given significant contributions to this chart (Vargas et al., 2004). Montgomery (1996) is one of the contributors and he has defined the EWMA as

$$Z_i = \lambda x_i + (1 - \lambda) Z_{i-1}$$

where $0 < \lambda < 1$ is a smoothing constant, *i* is the number of period and the starting value, $Z_0 = \mu_0$. This starting value is required for the first sample at *i* = 1. In many cases, the starting value assumed the value of the average of preliminary data, i.e. $Z_0 = \bar{x}$. The centre line is set at μ_0 and the control limits for the EWMA control chart are,

$$\text{UCL} = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2i})}.$$

and

$$\text{LCL} = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}(1-(1-\lambda)^{2i})}.$$

The L is representing the width of the control limits or in other words, the number of standard deviations to the control limits and $(1-\lambda)^{2i}$ is approximately 0 as i gets larger. Montgomery (1996) has proposed a different control limits after the control chart has been used for several time. It is said that the control limits approaching steady state values after running for several time periods where the smoothed average stabilizes and the proposed control limits are

$$\text{UCL} = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}}; \qquad \text{LCL} = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}}.$$

It was also suggested by Montgomery (1996) that the steady state values control limits be used for small values of i.

These control charts represent behaviour of a process of one quality measurement or process variable. Monitoring multiple individual variables with separate univariate control charts is very common in industry. It provides simple, clear and direct identification on the responsible variables when out of control signals occur in the control charts. But the practicability of this method is questionable, especially when the process involves a greater number of process variables, which is undeniably quite common nowadays. Furthermore, the usage of more than two univariate control charts at a time ignores the possibility of interdependency between variables, if it exists, and so might lead to unreliable conclusions in the end. These reasons make practitioners turn to multivariate techniques as a solution. The first publication on the application of the multivariate process control technique was by Harold Hotelling in 1947.

2.3 Multivariate Control Charts

A traditional univariate Shewhart control chart has an extended procedure which is applicable for the multivariate situation. The control chart is called Hotelling's multivariate control chart and it is among the most popular multivariate control charts (Alt, 1985). EWMA and CUSUM also have their own extension for multivariate problems which are known as Multivariate Exponentially Weighted Moving Average (MEWMA) and Multivariate Cumulative Sum (MCUSUM) control charts respectively. All of these extension versions of the univariate control charts are based on Hotelling T^2 statistic (MacGregor and Kourti, 1995). These multivariate charts have a number of advantages, such as being able to take into account the relationship between process variables, better control of Type I error, unlike the univariate control charts which suffer from multiple hypothesis testing problems, and being able to ease the process monitoring.

Hotelling's T^2 multivariate control charts have always been the most popular in multivariate quality control mainly for their simplicity (Das, 2006). This chart is good at detecting OOC signal when the mean shifts are big, however, it has little power when detecting small or moderate process shifts (Lowry and Montgomery, 1995). Unlike Hotelling's multivariate control chart, MEWMA and MCUSUM charts are more sensitive to small shifts in process mean (Lowry and Montgomery, 1995). In this thesis, we only consider the interpretation of OOC signals obtained from the signals triggered by Hotelling's T^2 multivariate control chart.

2.3.1 Multivariate Hotelling's T^2 Control Chart

Many studies in diagnostic methods (Maravelakis et. al,1992; Murphy, 1987; Jackson, 1980 & 1985; Tracy et.al,1995) are based heavily on a Hotelling's T^2 multivariate control chart. Originally, this chart was introduced by Hotelling (1947) but it has been discussed in more detail by many researchers since, such as Alt (1977, 1985), Alt and Smith (1988), Ryan (1989, 2000) and Jackson (1991). The multivariate Hotelling's T^2 control chart is sometimes called the multivariate Shewhart control chart (Crosier, 1988) and is stated by Lowry and Montgomery (1995) as "a natural multivariate extension to the univariate Shewhart chart".

Normally, it is assumed that p quality characteristics are jointly distributed as a p-variate normal distribution and that random samples of size n (individual observations) or m groups of size n (groups of observations) are collected sequentially along the process. In many applications, the data collected is not in the form subgroups but in individual observations instead. We concentrate firstly on this simpler case, the more general form is considered in Section 2.4.

Given that $\mathbf{x}_i$, i = 1,2,…, m are p x 1 vectors of m multivariate individual observations, normally distributed p-variables with known in-control mean vector μ_0 and known variance covariance matrix Σ_0. To test whether the vectors are at the desired target then a statistic

$$\chi_i^2 = (x_i - \mu_0)'\Sigma_0^{-1}(x_i - \mu_0) \qquad [2.1]$$

is computed and compared with the 100 (1-α) percentile of a central Chi-squared distribution with p degrees of freedom, where α is the specified level of significance for performing the test.

A multivariate Chi-squared control chart is constructed by plotting the χ^2 obtained from [2.1] versus time with $\chi^2_{\alpha,p}$ as its upper control limit (UCL). This chart will detect an assignable cause of variation in a process whenever a point falls beyond the UCL. No LCL is appropriate, since any deviation from μ_0 will result in an increases χ^2. If the in-control mean vector and the corresponding variance covariance matrix are unknown, then they must be estimated from a sample of size n drawn from the past multivariate observations. A more detailed explanation on this is explained in section 2.3.

Lowry and Montgomery (1995) stated that since Hotelling (1947) multivariate control chart procedure is based on only the most recent observation, it is insensitive to small and moderate shifts in the mean vector. So, other multivariate control charts are proposed which do use additional information from the recent history of the process.

2.4 Control Charting

The usage of control chart involves two distinct phases and each phase has different objectives. Each phase has a different control limit specification (Ryan, 1989, Lowry and Montgomery, 1995). The control limits are also different between a univariate and a multivariate control chart (Alt, 1984, Lowry and Montgomery, 1995). Alt (1984), Jackson (1985) and Nedumaran and Pignatiello (1998) discuss the computation of the control limits for subgroup and individual data. Some important issues in the choice of control limits for multivariate control chart are also discussed by Lowry and Montgomery (1995). The calculation of the control limits is presented for individual and subgroup data and the discussion of both phases in the following section is based on their discussions. Control charts can be used to monitor a process for any shift of mean or process dispersion, but, since this study is focusing on the problem of a shift in mean in multivariate observations, the following presentation will concentrate on the cases relevant to that matter.

2.4.1 Phase I

In this phase, a set of data from a process is gathered and analyzed retrospectively (Alt, 1984, Lowry and Montgomery, 1995). Woodhall (2000) describes control chart usage in Phase I as iterative. The focus is more on process understanding and process improvement. This phase is important and necessary in order to assist the operating personnel in bringing an out-of-control process into an in-control state (Montgomery, 2005). Trial control limits are constructed to determine whether the process has been statistically in-control over the period of time where the set of data was collected. The objective of this phase is to see whether reliable control limits can be established and hence can be used for future production monitoring (Lowry and Montgomery, 1995).

Subgroup data

The first step in establishing the initial control limits for subgroup data is by selecting *m* rational subgroups each with *n* observations. Let *p* be the number of process variables and $\mathbf{X}_{\mathrm{ij}} = (\mathbf{X}_{\mathrm{ij1}}, \mathbf{X}_{\mathrm{ij2}}, \dots, \mathbf{X}_{\mathrm{ijp}})'$ denote a *p* x 1 vector with *i* =1,2,…,*m* and *j* =1,2,…,*n*. It is usually assumed that X_{ij}'s are independent and identically distributed and X_{ij} follows $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ when the process is in-control. Also let $\bar{X}_i$ and S_i denote the unbiased estimate of the mean vector and the covariance matrix for the *i*th subgroup, respectively; that is,

$$\bar{\boldsymbol{X}}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij} \qquad \text{and} \qquad S_i = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij} - \bar{\boldsymbol{X}}_i\right)\left(X_{ij} - \bar{\boldsymbol{X}}_i\right)'$$

Then the procedure can be divided into 4 steps:

Step 1 Calculate unbiased estimates of the mean vector and the covariance matrix by pooling data from the *m* subgroups. The estimates are given by,

$$\bar{\bar{\mathbf{X}}} = \frac{1}{m}\sum_{i=1}^{m}\bar{\mathbf{X}}_{\mathrm{i}} \qquad \text{and} \qquad \bar{\mathbf{S}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{S}_{\mathrm{i}}$$

Step 2 Plot the statistic T_i^2 for the *i*th subgroup on a T^2 control chart. The T_i^2 is given as follows

$$T_i^2 = n\left(\bar{\mathbf{X}}_i - \bar{\bar{\mathbf{X}}}\right)'\bar{\mathbf{S}}^{-1}\left(\bar{\mathbf{X}}_i - \bar{\bar{\mathbf{X}}}\right) \qquad [2.2]$$

Step 3 Set the control limit for the T^2 control chart given by

$$\mathrm{UCL} = \mathrm{C(m,n,p)}F_{p,mn-m-p+1,\alpha} \qquad \text{and} \qquad \mathrm{LCL} = 0$$

where

$$C(\mathrm{m,n,p}) = \frac{\mathrm{p(m-1)(n-1)}}{\mathrm{mn-m-p+1}}$$

and $F_{p,mn-m-p+1,\alpha}$ is the (1- α)th percentile of the *F*-distribution with *p* and (*mn-m-p+1*) degrees of freedom with α the specified probability for each subgroup producing a false alarm on the chart. Since the **μ** and **Σ** are unknown and are estimated by $\bar{\bar{X}}$ and $\bar{S}$, respectively, then the UCL is taken from *F*-distribution instead of χ^2 distribution.

Step 4 Any point at which T_i^2 falls outside the UCL is investigated for any assignable causes. The control limit is revised once another T_i^2 falls outside the UCL and the new one is calculated once the assignable causes are removed. This step sometimes repeated several times until no more points falls outside the UCL. This process as stated by Montgomery (2005) "… *will require several cycles…assignable causes are detected and corrected, revised control limits are calculated, and the out-of-control action plan is up-dated and expanded…*".

Individual data

The first step in establishing the initial control limits for individual data is by selecting *m* subgroups with *n* = 1 observations. Let $\mathbf{X}_\mathrm{i} = \left(\mathbf{X}_{\mathrm{i1}}, \mathbf{X}_{\mathrm{i2}}, \dots, \mathbf{X}_{\mathrm{ip}}\right)'$ denote the *i*th of *m* multivariate observations. This time the mean and sample covariance matrix are given by

$$\bar{\mathbf{X}} = \frac{1}{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}} \bar{\mathbf{X}}_\mathrm{i} \qquad \text{and} \qquad \mathbf{S} = \frac{1}{\mathrm{m}-1}\sum_{\mathrm{i}=1}^{\mathrm{m}}(\mathbf{X}_\mathrm{i} - \bar{\mathbf{X}})\,(\mathbf{X}_\mathrm{i} - \bar{\mathbf{X}})'.$$

All the 4 steps explained in the subgroup data procedure are repeated with the obvious amendments on the calculation of the T_i^2 statistic and upper control limit. In *Step 2*, the statistic T_i^2 plotted is

$$\mathrm{T}_\mathrm{i}^2 = (\mathbf{X}_\mathrm{i} - \bar{\mathbf{X}})'\mathrm{S}^{-1}(\mathbf{X}_\mathrm{i} - \bar{\mathbf{X}}), \qquad i=1,2,\dots,m \qquad [2.3]$$

and in *Step 3*, the control limits are different from the subgroup data. Tracy, Young and Mason (1992) have shown the statistic follows a Beta distribution with degrees of freedom *p*/2 and (*m*-*p*-1)/2. Thus, the control limits are given by

$$\mathrm{UCL} = \left(\frac{(m-1)^2}{m}\right) B\left(\frac{\alpha}{2}; \frac{p}{2}; \frac{m-p-1}{2}\right) \quad \text{and} \quad \mathrm{LCL} = \left(\frac{(m-1)^2}{m}\right) B\left(1 - \frac{\alpha}{2}; \frac{p}{2}; \frac{m-p-1}{2}\right).$$

The next step is similar to Step 4 for subgroup data.

2.4.2 Phase II

Phase II in control charting received a lot of attention by many practitioners and authors. It is always assumed that the control limits have been established in the previous phase and all the retained observations are ready to be used in this phase. The retained observations from Phase I used to calculate the control limits in Phase II control charting (Mitra, 2012, p.348). The new control limits based on the retained observations at the end of Phase I is used to test whether the process is still in-control state whenever any future subgroups or individual data are drawn (Lowry and Montgomery, 1995). The objective of this stage is to monitor the process and assess it for any departure from the standard or in-control mean vector. Different control limits are proposed for this phase with respect to the data. In fact, many practitioners prefer to use a χ^2 approximation for the upper control limit in both phases especially when the number of samples or subgroups is greater than 25 (Lowry and Montgomery, 1995).

Subgroup data

The same statistic as in [2.2] is used in this phase. Ryan (1989) proposed an upper control limit

$$\frac{p(m+1)(n-1)}{mn-m-p+1}F_{\alpha,p,mn-m-p+1}$$

where

$F_{\alpha,p,m-p}$ = the *F*-percentile with *p* and (*mn-m-p+1*) degrees of freedom.

m = the number of subgroups

n = the subgroup size

p = the number of process variables

Individual data

There are a couple of options in determining the control limit for individual observations dataset in Phase II statistical process control monitoring. The first control limit proposed by Jackson (1985) and the second one, known as the exact control limit, proposed by Ryan

(1989). The same statistic [2.2] is used in Step 2. Jackson (1985) suggested a fair approximate control limit for a large m or to be precise m > 100 and given by

$$\left[\frac{p(m-1)}{(m-p)}\right] F_{\alpha,p,m-p}$$

and the control limit suggested by Ryan (1989) is

$$\frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p}$$

where,

$F_{\alpha,p,m-p}$ = the F-percentile with p and (m-p) degrees of freedom.

m = the number of subgroups taken in Phase I

p = the number of process variables

Lowry and Montgomery (1995) discussed the effect of the subgroup size and p upon the proposed control limits on both subgroup and individual data. The χ^2 approximation is found to be more accurate as n increases in subgroup data. Whereas for individual data, the χ^2 approximation produces inaccurate control limits with respect to the exact control limit with the same p when the p increases (Lowry and Montgomery, 1995).

2.5 Interpretation of a Signal for a Multivariate Control Chart

The interpretation of an OOC signal from a multivariate quality control chart has always been a major problem. Once a multivariate control chart produces an OOC signal, it is difficult to tell which process variable or variables might have triggered the signal. In univariate control charts, it is very easy to identify, as each process variable has its own control chart, but this is not the case for multivariate charts where all process variables are solely represented by a single statistic. In order to take the best corrective action, practitioners need to know the root cause of the OOC signal by interpreting or diagnosing the signal. In other words, one needs to identify the aberrant variable, or the combination of variables, which triggered the OOC signal. Practitioners and quality control researchers have came out with quite a number of approaches as to how to overcome the problem of interpretation in multivariate control charts.

Most of the approaches are based on the OOC signals produced by Hotelling's control chart. The most popular approach is by using principal component analysis (PCA) proposed by Jackson (1980, 1981 & 1985). Maravelakis et al. (2002) also used PCA in their ratio method. Other approaches are by decomposing the Hotelling's T^2 statistics (Mason, Tracy and Young, 1995; MacGregor and Kourti, 1995; Timm, 1996), based on discriminant analysis (Murphy, 1985) and regression methods (Hawkins, 1991 & 1993). Doganaksoy, Faltin and Tucker (1991) proposed a method which ranked process variables using univariate t statistics.

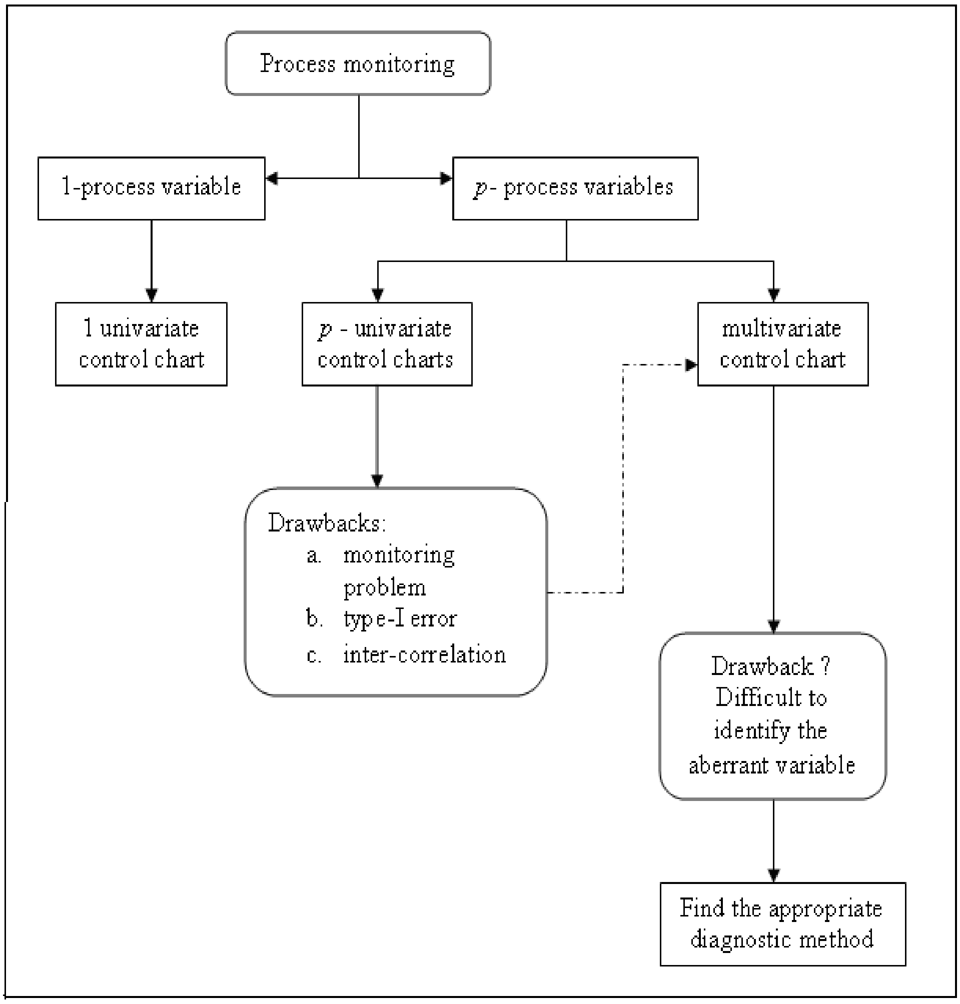


Figure 2.2: Process monitoring for univariate and multivariate processes.

In later chapters, we will show that each method mentioned above has its own flaws. Practitioners in multivariate process quality control are still struggling to find the most appropriate method to deal with the problem of mean (and scale) shifts as well as the effect of correlations among process variables in identifying the correct aberrant variables from OOC signals. An efficient method which is simple, easy to implement, practical, and reliable in identifying aberrant variables in multivariate processes is greatly needed in statistical process monitoring and improvement.

2.6 Graphical Diagnostic Methods

Several procedures have been developed for interpreting the OOC signals. Some of the popular approaches are by performing principal component analysis (Jackson, 1959, 1980, 1981 & 1985; Kourti and MacGregor, 1995; Maravelakis et al., 2002), decomposing the T^2 statistic (Kourti and MacGregor, 1995; Kourti and MacGregor, 1996; Mason, Tracy and Young, 1995 & 1997), discriminant analysis (Murphy, 1987), calculation of univariate *t*-statistics (Doganaksoy, Faltin and Tucker, 1991), and regression methods (Hawkins, 1991 & 1993). Among the most popular approaches is to decompose the T^2 statistic from the original X-space to principal component space in the shape of orthogonal components and interpret it. It is believed that the orthogonal decompositions of the T^2 statistics are easier to utilize because they enable direct interpretation. Furthermore, it allows practitioners to assess which of the components are important or warrant detailed investigation.

2.6.1 Elliptical Region

An elliptical region can be used when there are two variables involved in a process control monitoring. It can replace the role of a univariate control chart described in section 2.2.1. Unlike univariate control chart, an elliptical region is able to pinpoint the responsible variable that led a process to an out-of-control condition.

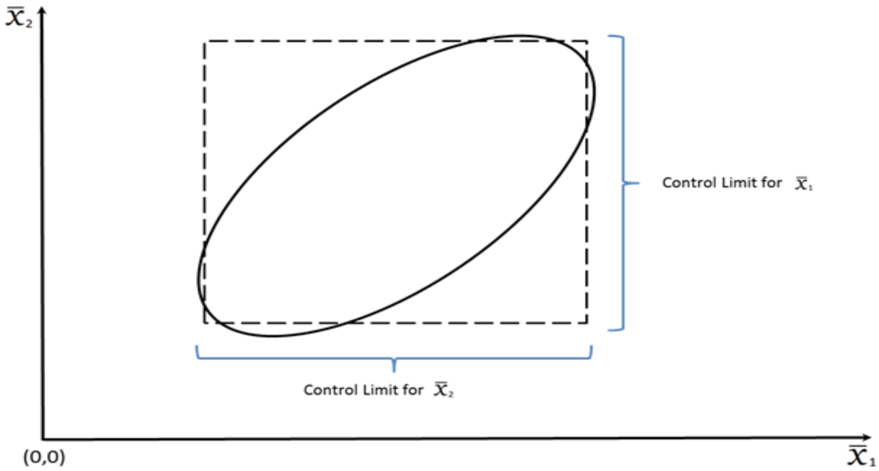


Figure 2.3: Elliptical region

Alt (1985) was among the earliest researchers to propose a graphical method to solve the interpretation problem of multivariate control chart signals. Elliptical control region was proposed by Alt (1985) and also discussed by Jackson (1991). Chua and Montgomery (1992) have extended the original elliptical control region proposed by Alt (1985).

2.6.2 MP Chart

Fuchs and Benjamini (1994) proposed a new type of control chart called the MP chart or multivariate profile chart. This chart uses symbols to represent the summaries of data and regarded as a "symbolic scatterplot" by Chambers et al. (1983). The univariate and multivariate statistics are displayed at the same time on this chart. Fuchs and Benjamini (1994) provide a guideline in detecting an out-of-control condition from the MP chart. A process is deemed to be out-of-control when the symbol is darker and the size of the symbols increases with the deviations. Fuchs and Kenett (1998) extended the development of this method by developing a programming in Minitab to create the MP chart.

2.6.3 MiniMax Control Chart

Sepulveda and Nachlas (1997) proposed a MiniMax control chart which can monitor a multivariate process and at the same time provide the information as to the aberrant variables whenever the multivariate process in a state of out-of-control. This chart monitors the maximum and the minimum standardized samples means as described in Sepulveda (1996). A process which is under in-control condition has its standardized samples means within the upper and lower control limits which are determined by simulation. This chart is different than the one proposed by Timm (1996) where the process monitoring is only on the maximum value.

2.6.4 Andrews Curves

Andrews (1972) proposed the curves as a multivariate data analysis tool. Andrews mapped $\mathbf{x} = \left(x_1, x_2, \dots, x_p\right)'$ into a form of a function given by,

$$f_{\mathbf{x}}(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \cdots. \quad -\pi < t < \pi$$

Maravelakis and Bersimis (2009) proposed the use of Andrews curves as a diagnostic method for the out-of-control signals produced by multivariate control chart. They used the same function as Andrews (1972) and stated that the method will produce an abnormal Andrews curve for the observation that responsible for the OOC signal on multivariate control chart. The main properties of Andrew curves are given by Andrews (1972) and are hold in a 5 step procedure of the method proposed by Maravelakis and Bersimis (2009). This 5 step procedure is briefly explained here.

Step 1: Look for the OOC signal by comparing the statistic of a multivariate control chart for the mean to the 100(1-α) percentage point of the χ^2 distribution.

Step 2: Set the control limits and plot them for every $t_r = -\pi + \frac{\pi r}{180}$, r = 0,1,…,360. The control limits are given as follows,

$$\text{UCL} = \mu'\mathbf{a}(t) + \sqrt{\frac{1}{n}c_a\mu'\Sigma\mu} \quad \text{and} \quad \text{LCL} = \mu'\mathbf{a}(t) - \sqrt{\frac{1}{n}c_a\mu'\Sigma\mu}$$

For all values of *t*, where

μ = the in-control mean

n = the subgroup size

Σ = the variance covariance matrix (assumed known)

c_a= the 1- *a* percentage point of the *p* – variate χ^2 distribution

$\mathbf{a}(t) \quad = (a_1(t), a_2(t), a_3(t), \dots)'$

$\quad\quad = \left(\frac{1}{\sqrt{2}}, sin(t), cos(t), sin(2t), \cos(2t), \dots\right)'$

Step 3: Calculate the value of $f_{\mathbf{x}}(t_r)$ for each of the t_r points that lie outside the control limits. Maravelakis and Bersimis (2009) defined that "*From all those $a_i(t_r)$, i=1,2,…,p with the same sign as $f_{\boldsymbol{x}}(t_r)$, the one with the largest contribution in $f_{\boldsymbol{x}}(t_r)$ pinpoints the out-of-control variable.*"

Step 4: The OOC variables identified with the highest frequency in *Step 3* is identified as the out-of-control variable.

Step 5: Repeat *Step 1* after the OOC variable identified in *Step 4* has been removed until there is no further OOC signal.

2.7 Analytical Diagnostic Methods

There are several analytical approaches proposed for interpreting signals in multivariate control charts. Alt (1985) was among the earliest researchers, proposing a solution by introducing the implementation of Bonferroni intervals for each process variable. Ten years later, Hayter and Tsui (1994) tried to extend the method by providing a procedure for exact simultaneous control intervals for every process variables' mean.

In the meantime, Murphy (1987) proposed discriminant analysis to distinguish the variables which caused the OOC signals and which did not. Later on, Doganaksoy et al. (1991) proposed the ranking of the univariate-t statistics in order to do so. In the same year, Jackson (1991) proposed the application of principal component analysis in interpreting signals from multivariate control charts. This approach has been the most popular one among and there has been much discussion regarding the implementation and the practicability of this method. Tracy et al. (1992) has provided a bivariate setting for the approach and it helps in giving a meaningful interpretation for the principal components since the principal components do not always have a physical interpretation. Kourti and MacGregor (1996) tried to improve the approach by proposing the implementation of normalized principal components.

The most recent diagnostic method using principal component analysis has been proposed by Maravelakis et al. (2002). They have tried another approach by introducing two methods of ratio calculation based on different type of covariance matrix. The ratio calculation uses the information from the loadings of selected principal components. They claim the ratio represents the contribution of each variable in the OOC signal in the multivariate control chart.

Mason et al. (1995, 1997) proposed a different approach to measure the contribution of an individual variable which is by decomposing the T^2 statistics into independent parts. Mason et al. (1996) again proposed a double T^2 decomposition but this time to reflect the contribution of individual process variables from a step processes. Timm (1996) used the same idea of interpreting OOC signals and a procedure known as a step down procedure has been proposed in which finite intersection tests were performed. All these methods are discussed further below.

Diagnostic Method with Principal Component Analysis

There are two procedures involved in monitoring multivariate processes using PCA. The first procedure introduced by Jackson (1980,1985) obtains the z-score of the principal components (PCs) for each observation vector and then computes the T^2 statistics. Jackson (1980, 1991) also discusses the identification of the out of control variable by decomposing T^2 statistic into a sum of p principal components and Jackson (1985) used both the individual variables and the principal components with the univariate charts to aid in the interpretation of an out of control signal without losing information about the correlation effect of the variables.

The second procedure introduced by Kourti and MacGregor (1996) is where T^2 statistics are expressed in terms of normalized principal components scores of multinormal variables and a contribution plot is used to identify the variable which caused the signal. Maravelakis et al. (2002) identifies the out of control variable by computing a ratio using the principal components for each variable from the signal.

One of the drawbacks of these approaches is that identifying a change in one or more of the constructs does not result in identification of which of the original variables or quality characteristics have changed. Sometimes, it is difficult to interpret the principal components and often no conclusion can be obtained from it (Doganaksoy, Tucker and Faltin ,1991; Mason, Tracy and Young, 1997). The attempt to reduce a p-dimensional data vector into a uni-dimensional statistics (in this case the Hotelling's T^2 statistic) often masks the primary causes of the signals (Mason, Tracy & Young, 1997). As a result, in many cases it is difficult and sometimes impossible to attach meaning to the principal components and to determine the characteristics which associated with the out of control signals (Mason, Tracy & Young, 1995).

2.7.1 Principal Components

Jackson (1985, 1991) stated that there are four conditions should be fulfilled by any multivariate quality control procedure, and they are

i. The procedure should be able to answer the question "Is the process in control?"

ii. The type I error should be specified.

iii. The relationships among the variables should be put into account.

iv. The procedure should be able to answer the question, “If the process is out of control, what is the problem?”

The first condition is quite easy to determine. The generalized t-test statistic is used to indicate the overall conformance of an individual observation vector to its mean or an established standard (Jackson, 1985). Jackson (1991) also stated that in multivariate quality control, the use of T^2 and PCA together adds some power to the control procedure and Jackson (1991) also provided a guideline for multivariate quality control using PCA. The guideline consists of two steps.

Step 1: For each observation vector, obtain the y-scores of the principal components to compute T^2.

Step 2: If T^2 is out of control, examine the *y*-scores. The *y*-scores are obtained from the transformation of principal components to another form of uncorrelated principal components but with variances equal to unity as defined in Jackson (1980, 1981) and given below

$$\mathbf{y} = \mathbf{w}'[x - \bar{x}]$$

It can be done by rescaling the characteristic vectors, **U** vectors, which are orthonormal and the scaled vector is given in Jackson (1980, 1981) as

$$w_i = \frac{u_i}{\sqrt{l_i}}$$

The quantities are called *y*-scores and due to its unit variances, it has been employed a great deal by Jackson (1991) in quality control. The *y*-scores are plotted together with the original variables and T^2 statistic to get the insight of the problem, if there is an out of control signal and likely may lead to the identification of the cause of the problem or the identification of the aberrant variable.

The principal component data may still be useful in detecting trends that will lead to an out of control condition even though T^2 remains in control. Practitioners are urged to diagnose T^2 if and only if T^2 is out of control. And this practice will help to fulfill the first three of the previous stated conditions required for the multivariate quality control procedure and whereas the second step of the multivariate quality control guideline will help to fulfill the fourth condition.

2.7.2 Normalized Principal Components

MacGregor and Kourti (1995) also refers to [2.3] in monitoring multivariate processes. The traditional Hotelling's T^2 in equation [2.3] is stated as equivalent (Kourti and MacGregor, 1996, MacGregor and Kourti, 1995) to:

$$T^2 = \sum_{i=1}^{q} \frac{t_i^2}{\lambda_i} = \sum_{i=1}^{q} \frac{t_i^2}{s_{t_i}^2} = \sum_{i=1}^{d} \frac{t_i^2}{s_{t_i}^2} + \sum_{i=d+1}^{q} \frac{t_i^2}{s_{t_i}^2}$$

with d as the number of chosen principal components and t_i a chosen principal component with λ_i the corresponding eigenvalue. By scaling each t_i^2 by the reciprocal of its variance, each PC term plays an equal role in the computation of T^2 irrespective

of the amount of variance it explains in the **Y** matrix where **Y** is the $n \times q$ matrix of mean centered and scaled measurements (MacGregor and Kourti, 1995).

Also this approach, in order to detect an occurrence of special events which were not present in the reference data, a squared prediction error (SPE_y) of the residuals of new observations (Kresta et al., 1991) has been introduced.

$$SPE_y = \sum_{i=1}^{q} \left(y_{new,i} - \hat{y}_{new,i}\right)^2$$

Jackson (1991) referred to this statistic as the Q-statistic or distance of the model. When the process is in control, SPE_y or Q should be small and the upper control limits for this statistic can be computed from historical data using approximate results for the distribution of quadratic forms (Jackson, 1991).

MacGregor and Kourti (1995) proposed the use of both the T^2 chart on the d dominant orthogonal PC's $(t_1,\dots,t_d)$ and SPE_y chart which may assist on the identification of the aberrant variables in multivariate processes.

2.7.3 Ratio Method

Maravelakis et al. (2002) proposed two new methods differently to identify the variable responsible for the out-of-control signal. Ratios are computed under two conditions, covariance matrix with positive correlation values and covariance matrix with positive and negative correlations. Say, X_{np} is a dataset with p variables following a multivariate normal distribution with mean equal to 0 and variance, Σ_0. The first condition has ratio as follows

$$r_{ki} = \frac{(u_{k1}+u_{k2}+\cdots+u_{kd})x_{ki}}{Y_{1i}+Y_{2i}+\cdots+Y_{di}} \qquad [2.4]$$

where,

x_{ki} = the i-th value of variable X_k

$(u_{1k}, u_{2k}, u_{3k}, \dots, u_{pk})'$ = the corresponding k-eigen vector

Y_{ji} = the score for vector $\mathbf{x}_i$ in PC-k

$= u_{1k}x_{1i} + u_{2k}x_{2i} + \cdots + u_{pk}x_{pi}$

d = the number of selected PC

j = 1,2,…,d.

The numerator represents the sum of the contributions of variable X_k in the first d PCs in observation (vector)-i whereas the denominator counts the sum of scores of observation (vector)-i in the first d PCs. The rationale of the methods is to compute the impact of each of the p variables on the out of control signal by using its contribution to the total score. The ratio for the second condition is also computed using the same formula as the first with a slight modification in the denominator utilizing a specified in-control value of means.

The responsible variable(s) are identified by comparing the value of the ratio with the control limits. For the first condition, as the numerator and denominator both follow a standard normal distribution so the control limits are the a and 1 - a percentage points of the following distribution with suitable parameters (Hinkley,1969).

$$F(r) = L\left\{\frac{\mu_1 - \mu_2 r}{\sigma_1\sigma_2\alpha(r)};\ -\frac{\mu_2}{\sigma_2};\ \frac{\sigma_2 r - \rho\sigma_1}{\sigma_1\sigma_2\alpha(r)}\right\} + L\left\{\frac{\mu_2 - \mu_1 r}{\sigma_1\sigma_2\alpha(r)};\ -\frac{\mu_2}{\sigma_2};\ \frac{\sigma_2 r - \rho\sigma_1}{\sigma_1\sigma_2\alpha(r)}\right\}$$

where

$$L(h;k;\gamma) = \left(\frac{1}{2\pi\sqrt{1-\gamma^2}}\right)\int_h^\infty \int_k^\infty exp\{-(x^2 - 2\gamma xy + y^2)/2(1-\gamma^2)\} \times dxdy$$

Meanwhile, for the second condition, since the denominator is only a constant, therefore the control limits for the ratio are taken from the standard normal distribution. Whenever the ratio of a certain variable(s) is not within the lower and upper control limits then it is identified as a responsible variable(s).

2.8 Other Diagnostic Methods

The other diagnostic methods discussed in this section do not involve principal component analysis. The approaches used for these methods are by decomposing the Hotelling's T^2 statistic, using the Bonferroni inequality, discriminant analysis, regression analysis and by ranking the *t*-statistic of the process variables.

2.8.1 Decomposition Method

Mason, Tracy and Young (1995, 1997) proposed and discussed a procedure for decomposing the T^2 statistic into orthogonal components to aid the interpretation effort. Mason, Tracy and Young (1995) have stated that the primary reason for partitioning the statistic is to obtain information on which variables significantly contribute to an out of control signal. The general decomposition of Hotelling's T^2 for p variables is as follows:

$$T^2 = T_1^2 + T_{2.1}^2 + T_{3.1.2}^2 + \cdots + T_{p.1,2,\ldots,p-1}^2$$

or

$$T^2 = T_1^2 + \sum_{j=1}^{p-1} T_{j+1.1,2,\ldots,j}^2$$

The statistic $T_{p.1,2,\ldots,p-1}$ is the pth component of vector X_i adjusted by the estimates of the mean and standard deviation of the conditional distribution of X_p given $X_1, X_2, \ldots, X_{p-1}$. It is given by

$$T_{p.1,2,\ldots,p-1} = \frac{X_{ip} - \bar{X}_{p.1,2,\ldots,p-1}}{s_{p.1,2,\ldots,p-1}}$$

where

$$\bar{X}_{p.1,2,\ldots,p-1} = \bar{X}_p + b_p' \left(X_i^{(p-1)} - \bar{X}^{(p-1)} \right)$$

$\bar{X}_p$ is the sample mean of the *n* observations on the *p*th variable and b_p is a (*p-1)*-dimensional vector estimating the regression coefficients of the *p*th variable regressed on the first *p-1* variables and is given by

$$b_p = S_{XX}^{-1} s_{xX}$$

where

$$S = \begin{bmatrix} S_{XX} & s_{xX} \\ s_{xX}' & s_x^2 \end{bmatrix}$$

and

$$s_{p.1,2,\ldots,p-1}^2 = s_x^2 - s_{xX}' S_{XX}^{-1} s_{xX}$$

Each of the terms is distributed as a constant times an F distribution having *1* and *n-1* degrees of freedom. Tracy, Young and Mason (1992) give the value of the constant

$\left((n+1)/2\right)^{1/2}$. Each of the terms then can be compared to the *F* distribution given below to determine if it is significant.

$$T^2_{j+1.1,2,\ldots,j} \sim \frac{n+1}{n} F_{(1,n-1)}$$

Mason, Tracy and Young (1995) highlighted two important points, which are the complete decomposition of the T^2 statistic into *p* independent T^2 components is not unique as *p!* distinct non-independent partitions are possible but that the *p* terms within a particular decomposition are independent of one another although the terms across the *p!* decompositions are not all independent.

Mason, Tracy and Young (1995) also devised a scheme to output only the significant values which gives solution to the problem of multiplicity of significance tests. So, each component in the decomposition can be compared to a critical value as a measure of the strength of the contribution to the signal rather than for statistical significance (Mason, Tracy & Young, 1995).

2.8.2 Bonferroni Inequality Approach

This method has been introduced by Alt (1985) and commonly referred to as "a Bonferroni type" method. The general idea of this approach is to construct *p* intervals (one for each quality characteristic) for each subgroup that produces an out of control signal in the multivariate control chart. Ryan (2000) has shown that for the *j*th subgroup, the interval for the *i*th characteristics, $\bar{x}_i^{(j)}$, would be

$$\bar{\bar{x}}_i \pm t_{\frac{\alpha}{2},p,k(n-1)} . s_{p_i} \sqrt{\frac{k-1}{kn}}$$

where s_{p_i} denotes the square root of the pooled variance for the *i*th characteristic and *k* is the number of subgroups. If the previous equation is not satisfied for the *i*th characteristic, the values of that characteristic would then be investigated for the *j*th subgroup. The entire subgroup would be deleted for all *p* characteristics once the assignable cause is detected and removed and the upper control limit recomputed.

Although the Bonferroni approach is frequently used, it is very difficult to determine the level of significance to be used (Hayter and Tsui, 1994). Doganaksoy, Faltin and Tucker (1991) felt that the Bonferroni intervals are much too wide when the T^2 statistics is significant so are seldom able to identify the responsible variables unless the properties show a sufficiently drastic change. Furthermore, this approach does not provide an alternative form of guidance to direct the search for which attribute has caused the change if it does not identify the variables when T^2 charts signal the process is out of control. This is a general feature of the Bonferroni

correction for multiplicity, they are excessively conservative for even moderate numbers of variables, for example more than three or four.

2.8.3 Discriminant Analysis

Murphy (1987) provides a simple test for selecting out of control variables and interpretation of T^2 values based on the concept of discriminant analysis. The quality control procedure is treated as an attempt to discriminate between the processes being in control Π_0 and out of control, Π. The true odds, Ω, Moran & Murphy (1979) in favor of an observed $\bar{x}$ being in Π to Π_0 been are defined as

$$\Omega = exp\left[-\frac{n}{2}(\mu-\bar{x})'\Sigma^{-1}(\mu-\bar{x})+\frac{n}{2}(\mu_0-\bar{x})'\Sigma^{-1}(\mu_0-\bar{x})\right]$$

The value of μ may be estimated by $\bar{x}$ and Ω become

$$\Omega \quad = \exp[(\mu_0 - \bar{x})'\Sigma^{-1}(\mu_0 - \bar{x})]$$

$$= exp\left[\frac{1}{2}T^2(\bar{x})\right]$$

Given that K is the cut-off point on the T^2 chart then, exp $\left[\frac{1}{2}K\right]$ is the quantity that must be exceeded by exp $\left[\frac{1}{2}T^2\right]$ before an 'out-of-control' signal is given. Given that a particular $\bar{x}^*$ signals 'out-of-control' on the T^2 control chart (i.e. $T^2(\bar{x}^*) > K$), the question of immediate interest is which of the p variables or subset p_1 of them where $p = p_1 + p_2$, caused the signal. An effective approach to answer the question is by partitioning the variables,

$$\bar{x}^* = (\bar{x}^{*(1)}, \bar{x}^{*(2)})$$

where $\bar{x}^{*(1)} = p_1$ subset of the p variables which we suspect caused the signal

$\bar{x}^{*(2)}$ = remaining p_2 variables.

With μ_0 and Σ partitioned as is $\bar{x}^*$, T_p^2 denoting the full squared distance

$$T_p^2 = T^2(\bar{x}^*) = n(\mu_0 - \bar{x}^*)'\Sigma^{-1}(\mu_0 - \bar{x}^*)$$

T_{p1}^2 denoting the reduced distance corresponding to the p_1 subset,

$$T_{p1}^2 = T^2\left(\bar{x}^{*(1)}\right) = n\left({\mu_0}^{(1)} - \bar{x}^{*(1)}\right)'\Sigma^{-1}\left({\mu_0}^{(1)} - \bar{x}^{*(1)}\right)$$

In discriminant analysis, the true squared distance between populations Π to Π_0 is defined as

$$\Delta_p^2 = n(\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)$$

and the reduced distance is,

$$\Delta_{p1}^2 = n\left(\mu^{(1)} - {\mu_0}^{(1)}\right)'\Sigma_{11}^{-1}\left(\mu^{(1)} - {\mu_0}^{(1)}\right).$$

The difference between the full squared and the reduced distances is used to reach the conclusion whether the p_1 subset discriminates just as well as the full set of p variables.

$$D = T_p^2 - T_{p1}^2$$

In other words, we have our hypothesis null as, $H_0: T_p^2 - T_{p1}^2 = 0$ (or $H_0: \Delta_p^2 - \Delta_{p1}^2 = 0$ in discriminant analysis). When D is large enough, we shall reject the hypothesis that the p_1 subset caused the signal, and we accept if otherwise. It is shown by Murphy (1987) that D ~ χ^2_{p2}. The appropriate F test is given in Seber (1984) for the case with estimated μ_0 and Σ.

The drawback of this method is the applicability of the approach. It is severely limited when the number of the quality variables involved is moderately large (about 20). The more variables there are in the process, the more ambiguity is introduced in the identification process and sometimes leading to erroneous conclusion (Lowry and Montgomery, 1995).

2.8.4 Regression Adjustment Techniques

Hawkins (1991, 1993) has suggested another approach which requires regression adjustment of variables in cascade processes which are commonly encountered in chemical plants and semiconductor manufacturing. In such processes, a shift in some quality characteristics in earlier step or stage may potentially affect the following step or process. The proposed method is an extension of the regression control chart proposed by Mandel (1969), who showed that by regressing a quality characteristic and control charting the regression was more effective than control charting the quality characteristic directly (Lowry and Montgomery, 1995).

A similar concept to the regression adjustments proposed by Hawkins (1991,1993) has been adopted by Zhang (1985) via his cause-selecting chart which has been thoroughly discussed by Wade and Woodall (1993). Wade and Woodall (1993) have proposed the use of prediction limits with cause-selecting charts to improve their statistical performance. The methods proposed by Zhang (1985) and Wade and Woodhall (1993) are not discussed here since there are not among the popular diagnostic methods used in comparison study.

2.8.5 Ranking Method

Doganaksoy, Faltin and Tucker (1991) proposed a combination approach of univariate ranking procedures and Bonferroni type simultaneous intervals (Alt, 1985). They claim the proposed approach is largely robust with respect to correlation structure and the nature of the shift in the mean vector. It is generically applicable and provides a priority ranking of attributes to be investigated even in instances where no unambiguous source identification is feasible. The *t*-statistic is calculated as;

$$t = \frac{(\bar{x}_{i,new} - \bar{x}_{i,ref})}{\left[s_{ii}\left(\frac{1}{n_{new}} + \frac{1}{n_{ref}}\right)\right]^{1/2}} \qquad [2.5]$$

where,

$\bar{x}_{i,new}$ = mean of new sample of variable-*i*

$\bar{x}_{i,ref}$ = mean of reference samples of variable-*i*

s_{ii} = variance of variable-*i*

n_{new} = new sample size

n_{ref} = reference sample size

For each variable, K_{ind} and K_{Bonf} will be computed and K_{ind} will be plotted first on a (0-1) scale. Variables with larger K_{ind} values are the ones with relatively larger univariate *t*-statistics values which possibly being among those components changed. Given the cumulative distribution function of the *t* distribution with (n_{ref} -1) degree of freedoms is T (t; n_{ref}-1) and K_{sim} is a specified nominal confidence level, then

$$K_{ind} = \left|2T(t; n_{ref} - 1) - 1\right| \qquad [2.6]$$

$$K_{Bonf} = \frac{(p + K_{Sim} - 1)}{p} \qquad [2.7]$$

Components having $K_{ind} > K_{Bonf}$ are classified as being those which are most likely to have changed.

2.8.6 Stepdown Finite Intersection Test

Timm (1996) proposed a Finite Intersection Test (FIT) called the stepdown FIT procedure for a given order of p variables. This method was originated by Krisnaiah (1965, 1979) and discussed in detail by Timm (1996). Timm (1996) also show that the method proposed by Hayter and Tsui (1994) is a single step FIT where the variance covariance matrix is known.

Under this method, the variables in the dataset are presumed to have some sort of order, either known or unknown. Each stepdown FIT procedure is applied on each of the p-orderings of the variables where the approach is quite similar to the approach proposed by Hawkins (1991, 1993) in which a regression of each variable is performed on the other p-1 variables. A stepdown FIT procedure could be performed at α^* in such a way that the family wise error rate is maintained at the nominal level α (Timm, 1996). The α^* level is given as

$$\alpha^* = \frac{\alpha}{p} \text{ or } \alpha^* = 1 - (1-\alpha)^{1/p}$$

For a given order of process variables, a stepdown FIT can be constructed by first defining the conditional distribution of x_{i+1}. Given that $\mathbf{x}' = (x_1, x_2, \dots, x_p) \sim N(\mu, \Sigma_0)$ and the conditional distribution of x_{i+1} given $x_1, x_2, \dots, x_i$ is also distributed normally with conditional variance σ_{i+1} and conditional mean with i = 1,2,…, p-1 as follows

$$E(x_{i+1} \,|\, x_1, x_2, \dots, x_i) = \eta_{i+1} + (x_1, x_2, \dots, x_i)'\beta_i$$

where,

$$\eta_{i+1} = \mu_{i+1} - \beta_i'\mu_i$$
$$\mu_i = (\mu_1, \mu_2, \dots, \mu_i)'$$

and

$$\beta_i' = (\sigma_{1,i+1}, \dots, \sigma_{i,i+1})\Sigma^{-1}$$
$$\sigma_{i+1} = \frac{|\Sigma_{i-1}|}{|\Sigma_i|}$$

where,

$|\Sigma_0| = 1$ and $\beta_0 = 0$.

With the conditional model given previously, the problem of testing $H_{0i}: \mu_i = \mu_{0i}$ versus $H_{1i}: \mu_i \neq \mu_{0i}$ is now equivalent to testing

$$H_0 = \cap_{i=1}^{p} H_{0i}^{*} \text{ where } H_{0i}^{*}: \eta_i = \eta_{0i}$$

and

$$H_1 = \cup_{i=1}^{p} H_{1i}^{*} \text{ where } H_{1i}^{*}: \eta_i \neq \eta_{0i}$$

The test statistic for testing H_{0i}^{*} is given by Timm (1996) as

$$F_i = \frac{(\hat{\eta}_i - \eta_{0i})^2 (n-i)}{d_i s_i^2}$$

where

$\hat{\eta}_i$ = the least square estimate of η_i

$d_i s_i^2$ = the variance of $\hat{\eta}_i$.

Timm (1996) pointed out, the statistic can alternatively be written in another expression that was developed by Mudholkar and Sabbaiah (1980a, 1980b)

$$F_i = \frac{(\hat{\eta}_i - \eta_{0i})^2 (n-i)}{\left[n^{-1} + \sum_{k=1}^{i-1} \frac{(\hat{\eta}_k - \eta_{0k})^2}{s_k^2}\right] s_i^2}$$

for i = 1,2,…,p where $s_i^2 = \frac{|\mathrm{S_i}|}{|\mathrm{S_{i-1}}|}$ and S is the unbiased estimate of Σ_0. The process is out-of-control if at the ith step, the F_i is larger than the critical value of the multivariate F distribution at the α^* level as defined earlier.

2.9 Limitation and Drawback

Although quite a number of diagnostic methods are proposed with various approaches all of them have their own limitations and sometime drawbacks. One is the complexity of the procedures such as in the methods proposed by Murphy (1987) and Mason et al.

(1995, 1997). The discriminant analysis approach proposed by Murphy (1987) still needs an individual control chart to tell the direction of the aberrant variables identified and the computation for the procedure is quite considerable. Whereas, the method by Mason et al. (1995, 1997) does not have a unique independent T^2 components to identify any variables responsible for the OOC signal.

Secondly, some of the methods lack physical interpretation. This particular drawback is quite obvious in any diagnostic methods which involve principal component analysis. If the selected principal components do not provide a meaningful result then it is quite hopeless to proceed with the diagnostic step. Thirdly, some of the graphical methods are only applicable for a process with only two process variables for example an elliptical control region. Hence, the application of this method is very limited and there are, in fact, very few options available for multivariate processes with the number of process variables greater than two.

The method proposed by Maravelakis et al. (2002) appears not to be location invariant and shows limitations. This limitation at some points might prevent practitioners to adopt it since it is not freely applicable to datasets with mixed sign correlations in which quite common in multivariate processes. This renders the method of little practical use. These authors have also ignored the very real possibility that sample eigenvalues may be in a different order from population eigenvalues, especially when two consecutives population eigenvalues are close together. The result is that the corresponding eigenvector is orthogonal to the one which is really required.

Murphy (1987) proposed a method based on a discriminant analysis but unfortunately does not provide clear criteria that make up the condition for an out-of-control signal. The ability of the proposed method to identify the out of control variables is very poor when the shifts in a mean vector are not in accordance with the correlation structure between the variables and impractical for a moderate number of variables i.e. 20 (Doganaksoy, Faltin and Tucker, 1991).

2.10 Performance of a Diagnostic Method

One of the most recent studies in the performance of a method in identifying the out-of-control variable(s) is done by Das and Prakash (2008) and it is the extension of a comparison study done by Das (2006). The study tried to compare the performance of four methods by Mason, Tracy and Young (1995), B.J Murphy (1987), Douglass Hawkins (1991), and Doganaksoy, Faltin and Tucker (1991). The scope of their study is restricted to the shifts in the mean vector under three assumptions which are;

i. $Y \sim N_p(\mu, \Sigma)$; Y is a p-dimensional vector
ii. Σ remains undisturbed, and
iii. Variables are equi-correlated

The estimated power of the diagnostic methods is then compared based on the formula below:

$$\widehat{Power} = \left(\frac{n(successful)}{n(OOC)}\right) 100\% \qquad [2.8]$$

where

n(OOC) = the number of times T^2 detects the shift or out-of-control

n(successful) = the number of times the particular method is 'successful' when the T^2 control chart gives alarm.

There are many features which contribute to a useful diagnosis method such as simplicity, speed of computation, interpretability etc., but at the core must be an assessment of the power in correctly diagnosing OOCs. As we will see in Chapter 4, even this power can be difficult to define when the shifts in the process mean are complex and multiple variables are involved, since then the concept of 'partially correct diagnosis' becomes relevant. However, at its simplest (following Das and Prakash, 2008), the power of the test actually measures the percentage of correct classification made by the diagnostic method concerned in detecting a particular shift. The higher the percentage value, the better diagnostic method would be.

2.11 Comparative Studies

Das and Prakash (2008) carry out a similar comparative study on the performance of four different methods by Mason, Tracy and Young (1995), Murphy (1987), Hawkins (1991) and Doganaksoy, Faltin and Tucker (1991). The performances of these techniques were compared for different correlation structures, strictly to the shifts in the mean vector with the assumptions the variance-covariance matrix remain undisturbed and the variables are equicorrelated. The assumptions and the correlation structures chosen are quite similar to those in Doganaksoy, Faltin and Tucker (1991). Das and Prakash (2008) showed that all the studied diagnostic methods lose their power to detect the shift as the magnitude of the shift decreases. Among their findings are that the methods by Hawkins (1991) and Murphy (1987) become effective when shifts are in a counter-correlation direction especially at low negative and all positive correlation whereas for small shifts, Murphy's method performs better than the others.

Maravelakis et al. (2002) performed a simulation study to compare their method with the method by Kourti and MacGregor (1996). They used Bonferroni limits on the normalized scores and calculated the contributions of the variables. They found that the method of Kourti and MacGregor (1996) has lower success in identifying the out of control variables compared to their method. They used the same covariance matrices and mean vectors proposed in Maravelakis et al. (2002).

Both of the comparative studies mentioned above generally calculate the number of times each method detects the aberrant variable for a given shift to estimate the power of a diagnostic test. It is necessary to explore further the power of the methods available, not only on the basis of the number of times each method detects the aberrant variable but in terms of partially correct (not all aberrant variables detected) performance for more complex observations.

The correlations among the variables are very important and this has been shown to have an impact on the out-of-control signal (Doganaksoy et al., 1991; Maravelakis, et al., 2002) as well as on the effort in identifying the cause of the signal. A thorough study of these correlational effects on the diagnostic methods is also very important in assessing the power of the diagnostic method itself. One of the values of a multivariate approach is that the combined power of all p-dimensions should permit more sensitive detection (i.e. detection of smaller shifts) than examination of p separate charts. The intervariable correlations are clearly the key here. The main obstacles would be in determining the lowest value of the correlation among variables that would spark a signal in multivariate control chart and whether it can be diagnosed by the methods discussed previously.

It would be very interesting to carry out an extended study on the comparative study by Das and Prakash (2008) with the additional method by Maravelakis et al. (2002), using the mean vectors and variance covariance matrices proposed by Maravelakis et al. (2002). An extended study on the method by Doganaksoy, Faltin and Tucker (1991) is also possible by adding more mean vectors to test the counter correlational effects as discussed previously. That is we might try to compare the performance of all the available and appropriate methods for the selected mean vectors and covariance matrices suggested in previous studies. An appropriate guideline in selecting the best diagnostic method for different kind of conditions such as different variables' mean shifts, shift magnitudes, and correlations is also needed to allow practitioners to choose the right one for the monitoring and process improvement purposes. We provide such a study in Chapter 3.

CHAPTER 3
COMPARATIVE STUDY

3.1 Introduction

This chapter will discuss a comparative study done by Das & Prakash (2008). A summary of the final results from Das & Prakash (2008) was given in the preceding chapter, but we review it in more detail here because it will be the basis of the new comparison study presented in Chapter 4. A new method named a Ratio method, adapted and based on the method proposed by Maravelakis et al. (2002), will be included in the comparison study. It is important to note that no new simulations are performed in this chapter. We simply review and discuss the findings of Das & Prakash (2008).

Section 3.2 will gives an overall review of the study performed by Das & Prakash (2008). The background of their simulation studies will be explained thoroughly to ensure a good understanding of its purpose. Section 3.3 will observe and summarize the findings of a special case when the mean of just one process variable deviated from the in-control mean vector. Section 3.4 will look at the cases with two aberrant variables. The discussion will focus on two situations, when shifts are in accordance to correlation structure and vice versa. The effect of correlation structure in which includes the sign and the strength of the correlation between variables are also discussed. Section 3.5 presents some conclusions of the discussions. It will also highlight a few important points that need to be taken into account in any future comparative study.

3.2 Das & Prakash Study

Not many studies have been carried out in comparing the performance of diagnostic methods. One of the most recent was carried out by Das & Prakash (2008). There are four diagnostic methods, proposed by Mason, Tracy & Young (1995), Murphy (1987), Hawkins (1991) and Doganaksoy, Faltin & Tucker (1991), included in this study. The methods are reviewed and their performance compared. A brief discussion of all the methods is given in previous chapter.

Assumptions

In Das and Prakash (2008), it is assumed that the process variables follow the multivariate normal probability distribution. This specific case considered three process variables whose mean, $\boldsymbol{\mu}$, and variance covariance matrix, $\boldsymbol{\Sigma}$, are assumed known. The in–control mean vector is assumed to be a zero vector and all variables are equi-correlated. So, the variance covariance matrix is assumed to be scaled, such that it has unit variance for all the process variables.

Performance Measure

All the studied diagnostic methods are compared based on their power. The estimated 'Power (P)' of a diagnostic method is initially defined in [2.8] by Das and Prakash (2008). The estimated power is in percentage for the number of times a multivariate control chart detects the shift for the number of times it successful in identifying correctly the aberrant variable which triggered the signal.

It is relatively easy to understand what "successful" means when the shift in mean happens in only one variable, but it is quite tricky to define it when we have, say, two shifted variables with different sized shifts. The questions one might need to ask themselves are

(i) How are we going to count m?

(ii) Does the size of the shifts matter?

(iii) Does the combination of shifts matters?

(iv) Does the shift or the combination of two or more shifts give the same power measure with respect to the structure of the variance covariance matrix?

(v) Are false identifications penalized?

Das and Prakash (2008) gave the interpretation of *P* as “an estimate of the percentage of correct classification made by the concerned diagnostic method to detect a particular shift”. Unfortunately, Das and Prakash didn’t elaborate further especially for the cases with two shifted means in the mean vector. It is quite unclear when it comes to the particular case mentioned in (i) earlier. We could presume *m* to be the number of correct identifications of both shifted variables as the aberrant variables or it could be defined as the number of correct classifications of at least one of them. The second question above now comes into play. For the first presumption diagnostic method must detect both variables as aberrant, regardless of the amount of the shift in mean from each of them even though the shifts may be very small. On the other hand, the second presumption only concerns the ability of a diagnostic method to detect at least one deviated variable as the aberrant one. We presume that the more contaminated a process variable will be detected more frequently than the lesser shifted process variable. We also presume that when two variables are contaminated equally, both of them will be detected approximately equally.

Combinations of the shifts in mean are considered by Das and Prakash (2008). Data was generated through simulation under different shift patterns in the mean structure in order to determine the sensitivity of a diagnostic method. By studying the patterns closely, one can tell whether there is a difference in performance in terms of the size of the contamination of a different variable or whether the direction of the shift in mean plays an important role in the performance of a diagnostic method.

Simulation

One thousand observations of 3-dimensional vectors representing out-of-control observations were generated by Das and Prakash for each shift in mean as listed in Table 3.1, or combination of shifts in mean. In other words, for each pattern of mean structures which representing out of control mean vector and correlation matrix, there are 1000 random observations generated for three process variables respectively, simulated from a multivariate normal distribution with specified mean (shifted) vector and variance covariance structure. Das and Prakash (2008) used Hotelling T^2 control chart in monitoring the multivariate process but no further explanation is given about the phase of the process monitoring. So here in this discussion we assumed that it was done in Phase II process control monitoring. Any disruption in the process is presumed to have been encountered and solved beforehand since this is normally done in the first phase of process control monitoring. By referring to the definition of the power in [2.8], we may expect that the denominator in [2.8] is not necessarily equal to 1000. Even though Das and Praksah (2008) stated that 1000 multivariate observations were generated for the comparison study but they didn't put it plainly that the value is not necessarily 1000. So here we explain that, 1000 observation of 3-dimensional vectors are generated for each pattern of mean structures which represent the out of control mean vector and correlation matrix, but only those which triggered the Hotelling T^2 control chart are counted in and diagnosed by a diagnostic method. As a result, we have different number of out-of-control observations for different mean structures. This feature affects the reliability of our power assessments, with different cases being based on different numbers of diagnoses. However, we do not make reference to this in detail while reporting the Das and Prakash results, but consider it further in our own study in Chapter 4.

Das and Prakash (2008) considered five different correlations between variables which are -0.45, -0.2, 0.2, 0.5 and 0.8. For each of the equi-correlation structures, a performance is calculated for every shift in mean shown in the table, where the descriptive categories given by Das and Prakash (2008) is tabulated here for easy understanding,

Table 3.1: The shift in mean from the in-control mean vector.

Shift	Shift's magnitude
Small	-1, 0.5
Intermediate	-2, 1.5
Large	-3, 2.5

The specified shift in mean is introduced to the mean of the first process variable of the in-control mean vector in order to simulate a dataset which contains one aberrant variable. The contaminants are the values of the shift in mean given in Table 3.1. If the shift in mean is small then it means the mean for variable 1 in the mean vector is either -0.5 or -1 or in other words, the shifted mean vector would be $(-0.5, 0, 0, 0)'$ or

(-1, 0, 0, 0)′. For cases with two aberrant variables in a dataset, the mean of the first and the second process variables were contaminated by introducing one of the contaminants into each of their means. Various contaminants and combination of contaminants provide different patterns of contaminated or shifted mean vectors. For a case of one aberrant variable, there are six patterns of (shifted) mean structure for each equi-correlated variance covariance matrix, whereas for a case with two aberrant variables, there will be 20 combinations of shifts in mean. An example of a shifted mean vector with small contaminant for variable 1 and large contaminant for variable 2 is (1, 3, 0, 0)′.

Results Presentation

The findings from the comparison study of Das and Prakash (2008) are quite detailed and extensive and can be divided into two sections. Section 3.3 will present the result of the performance of the diagnostic methods when a single process variable is shifted from the in-control mean. Section 3.4 will look at the performance when the mean of two process variables shifted in the same and opposite directions. These two sections produce a summary of the 15 tabulated results from the original paper. Since Das & Prakash (2008) did not give any specific guidelines in describing the performance of the tested diagnostic methods, a suggestion is given below.

Table 3.2: Performance based on the percentage of correct identification, Power (*P*)

$P = \left(\frac{m}{n}\right)100$	Performance
$P < 10\%$	Very poor
$10\% \leq P < 40\%$	Poor
$40\% \leq P < 60\%$	Fair
$60\% \leq P < 80\%$	Good
$P \geq 80\%$	Very good

The discussions of the results are focusing on the type of the shift in mean, the strength of the shift in mean and the inter-correlation among variables. In order to simplify the discussion, the method proposed by Mason, Tracy & Young (1995) will be referred to as method MTY. Method DFT refers to the method proposed by Doganaksoy, Faltin & Tucker (1991) while HAW and MUR refer to the methods of Hawkins (1991) and Murphy (1987) respectively.

3.3 One Aberrant Variable

This study would like to add some more findings based on our close observation of their results which are shown in Table 1-5 in Das and Prakash (2008). It was stated in general that MTY and DFT performed equally well for large shifts across different correlation structures. Unfortunately, Das and Prakash (2008) failed to mention that for large shifts, both methods have also shown a better and more consistent performance than MUR and HAW in most correlation structures. MTY and DFT are found to be superior in their estimated Power than MUR and HAW except for the case with $\rho = 0.5$ where HAW is quite close after DFT.

For low correlation, MTY and DFT still showed a good performance as well as MUR and HAW except in case $\rho = 0.2$ where HAW has the highest Power. On the other hand, MUR's power dropped to a fair performance. In most cases, MTY showed slightly higher performance than DFT except when the correlation between the process variables is positively strong.

Das and Prakash (2008) also stated that MUR and HAW performed far below average for high positive correlation. By studying the distribution of the estimated power in Table 5 in Das and Prakash (2008), MUR did show a very poor performance but this study would like to add that the poor performance is not actually for high positive correlation alone. The same poor performance can be seen for negative moderate correlation and in all the negative shifts for moderate positive correlation. Several cases have a power less than 10%. This indicates that MUR was often unable to pinpoint the correct aberrant variable when there is one mean value shifted from the in-control mean vector.

Das and Prakash (2008) summarized that both MUR and HAW follows the same pattern of performance with MUR showing less efficiency. HAW is said to have shown a satisfactory performance for large shifts at low and positive moderate correlation. This study is partially agrees with their conclusion. MUR did show generally less efficiency compared to HAW but not for all the correlations. Table 2 in Das and Prakash (2008) showed that MUR has slightly better efficiency than HAW for ρ = -0.2. MUR seems to perform differently for weak correlation of different signs, being particularly poor in low positive case. There are appreciable drops for both MUR and HAW when the correlation is moderate and strong. Whereas, when the correlation between the process variables is low negative, all the studied diagnostic methods tend to have a similar power.

It is important for this study to add, even though it is not surprising, the general pattern that can be seen is that the power of all methods increases when the shift in mean is bigger regardless of the direction of the shift. However, for larger correlations, it is true for methods MTY and DFT only and for HAW when ρ = 0.5, MUR (and HAW for other ρ) behave in a counterintuitive way, being slightly better able to pick out small shifts.

It is also worthwhile to mention another point here, since it is not stated clearly in Das and Prakash (2008), MUR and HAW did not show equal performances across different shifts in mean when the correlation is moderate positive and negative. MUR and HAW are nearly unresponsive when the shift is large and when the correlation is moderate negative. Their performances increase but are still below average when the shift goes from intermediate to small. HAW is able to perform nearly as well as MTY and DFT when the correlation is moderate positive. MUR also performed much better with moderate positive correlation but only between fair to poor level of performance. It is obviously not as we expected as we discussed in Section 3.2. MTY and DFT did not show the same performance as MUR and HAW for the moderate correlations. Both of the methods show similar performance for both moderate positive and moderate negative correlation unlike MUR and HAW.

If we are interested to see the power across the correlation structures, Figure 3.1 below can give a direct comparison. The first two graphs on the top of Figure 3.1 represent a large shift in mean, followed by another two graphs in the middle for intermediate shifts and, at the bottom of Figure 3.1, performance for the small shift in mean. Generally we can see that the performances of MTY and DFT are higher when the shifts in mean are bigger and dropped when the shifts get smaller. MUR and HAW also follow the same pattern except for moderate and strong positive correlation. The performance of MTY and DFT varied ±20% or less across different correlation structures. On the other hand, MUR and HAW show appreciable drops from their best performances at low negative correlation, few upward as well as downward changes in correlation.

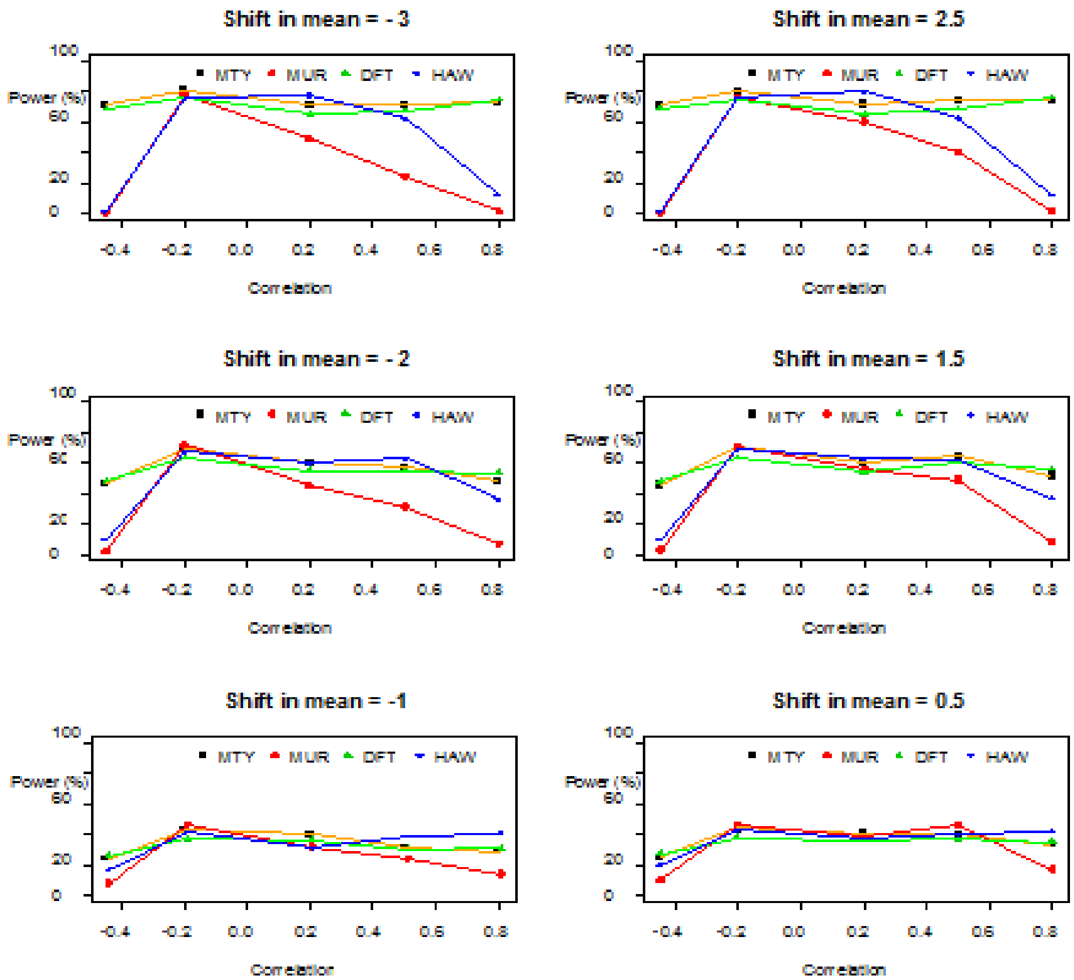


Figure 3.1: The performance of diagnostic methods when one variable deviates from the target mean through different equi-correlation structures with respect to various shifts in mean.

Generally, all the diagnostic methods performed lesser when the shift getting smaller for the small correlation regardless of the direction. The rule applies to MTY and DFT for moderate to high correlation but unfortunately not for MUR and HAW. MUR performed poorly when the correlation is moderate and high and same goes for HAW, except when the correlation is moderate positive. Overall, only two methods, which are MTY and DFT, give a consistent performance throughout all correlation structures and the changes in performance of these two methods with respect to the changes in the mean shift or contaminant is considered relatively predictable. Whereas, MUR and HAW can only detect the shifts in mean, reliably, when the correlation is small.

3.4 Two Aberrant Variables

There are a couple of issues arise when studying the summaries and conclusions given by Das and Prakash (2008) for cases with two aberrant variables. For instance, Das and Prakash (2008) summarized that MTY and DFT perform a "satisfactory level" for large and intermediate shifts across different correlation structure. Unfortunately, Das and Prakash (2008) did not explain in detail about the range of the percentage values of the performance for a "satisfactory level". Nor did they elaborate more on "large" and "intermediate" shifts. One might have problem to decide whether the combination of shifts of mean (-3,-1) is considered as "large" shift or whether (-1,-2) can be considered as "intermediate" shift. We presume that the large and intermediate shifts are meant for combinations of two large shifts or two intermediate shifts respectively. In that case, the power of satisfactory level will be as low as 13% (as shown in their Table 6 for shifts in mean (1.5, 1.5)) and that obviously does not seem right.

Due to the complexity in understanding the effect of the size of the shifts in mean, the correlation structure, and probably the sign of the correlation, upon the estimated power of the studied diagnostic methods, this study splits the discussion into two subsections. The first subsection will discuss the power of the diagnostic methods when the shifts of mean are in accordance to the correlation structure. Meanwhile, the second subsection will discuss the power when the shifts of mean are not in accordance with the correlation structure or in other words, is counter to the correlation structure. By this we mean, for example, that if the correlation is positive, then a ‘with correlation case’ will have both shifts of the same sign, but shifts of different signs will be described as ‘counter-correlational’. The shifts in mean vector such as $(-1,-2, 0, 0)'$ and $(1.5, 0.5, 0, 0)'$ with $\rho = 0.2$ between the variables are a couple of examples for shifts in mean in accordance with the correlation structure. Shifts in mean vector with mixed signs such as $(-3, 0.5, 0, 0)'$ and $(3,-1, 0, 0)'$ with correlation between variables is -0.2 are also said in accordance with the correlation structure. On the other hand, if $\rho = 0.2$ for the same shifts with mixed signs, then the shifts in mean is said counter-correlational.

3.4.1 Shifts in Accordance to Correlation Structure

Das and Prakash (2008) stated a few comments on the performance of MTY and DFT when the shifts of means are in accordance with the correlation structure. In general, MTY and DFT performance are claimed to be at satisfactory level for large and intermediate shifts. The performance is said to be significantly increased whenever the shift is in accordance with the correlation structure. Specifically, Das and Prakash (2008) lay out additional comments with respect to positive correlations for both of the methods which are:

i) $\rho = 0.8$

The power is high for large shifts and the performance is poor when one of the shifts is small.

ii) $\rho = 0.5$

Performance increases for large shifts and decreases one of the shifts reduces in magnitude.

iii) $\rho = 0.2$

Both methods performed well.

Das and Prakash (2008) didn’t discuss much the MUR and HAW performance for the cases with positive correlations except for $\rho = 0.5$. MUR performance is said to be much better than others when one of the shifts is small and HAW remains ineffective throughout different shifts.

Generally, this study agrees with the observations made by Das and Prakash (2008) except for a few important things that need some elaboration. This study noticed some inconsistencies in their comments. For example, it was said that for $\rho = 0.8$, the performance of MTY and DFT is poor when both or one of the shifts is small. By referring to their Table 10, the highest power for the said poor performances is 30% (shifts in mean (-1, -2)). But, Das and Prakash (2008) also said in general that for the large and intermediate shifts, MTY and DFT performed at a "satisfactory level" which gives 16% (Table 9) as the lowest power and 65% (Table 8) as the highest power for satisfactory level. It seems, there are overlapping levels of performance here and the limit between poor and satisfactory performances are not clear. However, this study will used the performance categories defined in Table 3.2 whenever necessary to avoid any confusion in later discussions.

Even though Das and Prakash (2008) only mentioned the increase in power for large shifts for $\rho = 0.5$, we would like to add that the same situation is also true for $\rho = 0.2$ and $\rho = 0.8$, i.e. the power decreases due to the reduced shifts is true for all the positive correlations. With regards to MUR, it is not entirely true that MUR is better than others when both or one of the shifts is small for $\rho = 0.5$. In fact, from Table 9, MTY and DFT performed better than MUR for shifts (-1,-2), (1, 0.5) and (1.5, 0.5). For the HAW method, even though it is ineffective throughout different shifts (we might add throughout positive correlations too), for $\rho = 0.2$ the power is 36% for shifts (-3,-3) which is much higher than MUR. MUR and HAW share a similar poor performance pattern for positive correlations when the shifts are in accordance with the correlation structure.

For negative correlations, Das and Prakash (2008) also give specific comments on the performance of all the diagnostic methods, as follows

MTY and DFT:

i) $\rho = -0.2$

A satisfactory performance is shown for large shifts.

ii) $\rho = -0.45$

Performance was at a satisfactory level for large shifts. The power dropped to an average level for a combination of large and intermediate shifts. For (both) intermediate shifts the performance dropped far below average as well as when one of the shifts is small.

MUR and HAW:

i) $\rho = -0.2$

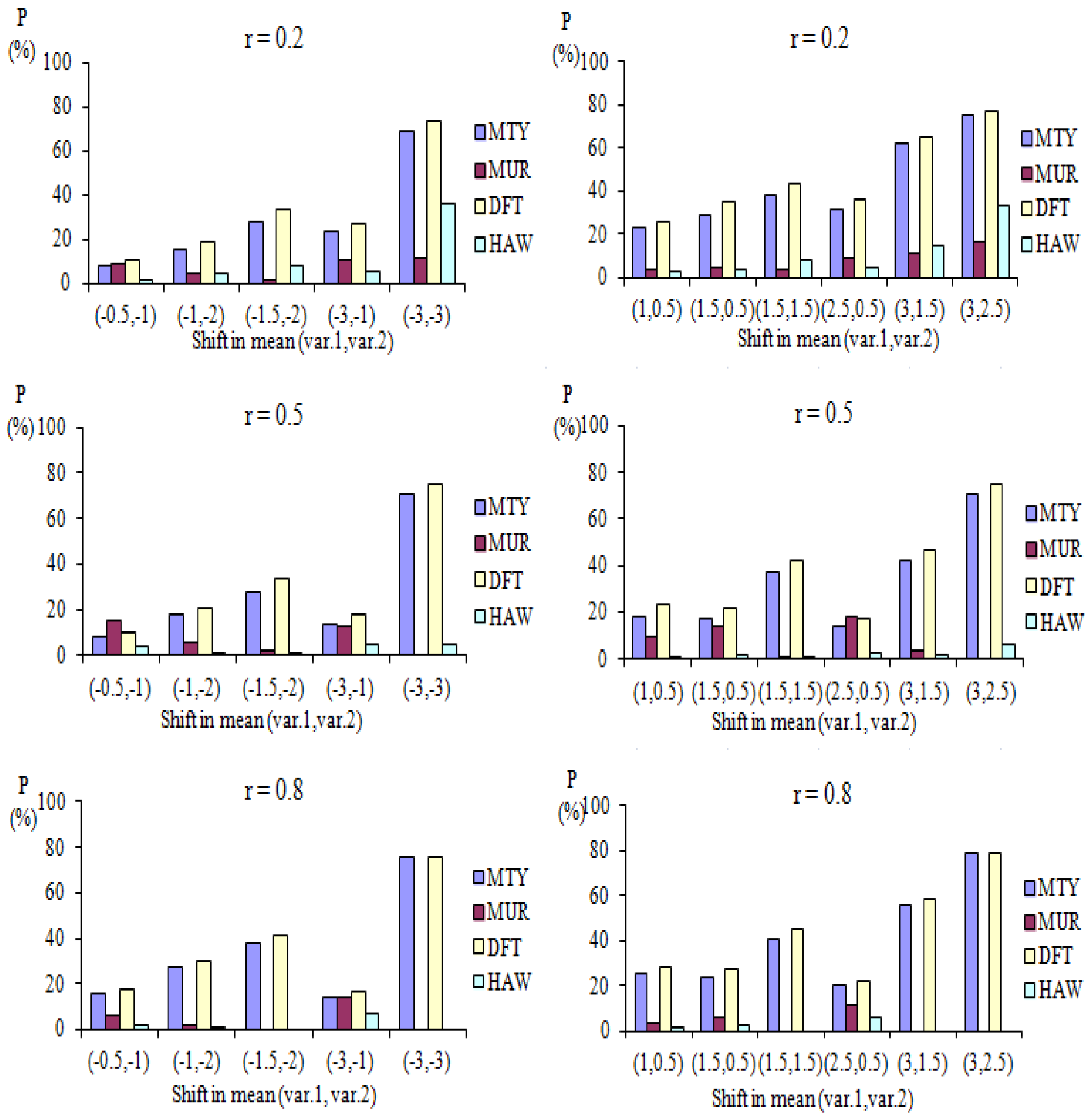
Gain power for large shifts.

ii) $\rho = -0.45$

Remain ineffective throughout different shifts.

This study would like to add a few more observations on the performance which are not given by Das and Prakash (2008). Generally, Table 6 and Table 7 showed that MTY and DFT performance is more or less similar for $\rho = -0.45$ and $\rho = -0.2$ throughout different combination of shifts. Interestingly, the power for small shifts is higher compared to the combination of large and small shifts for $\rho = -0.45$. Even so, it is difficult to conclude anything from these results especially when the n differs and very low for the small shifts compared to the n from the combination of large and small shifts.

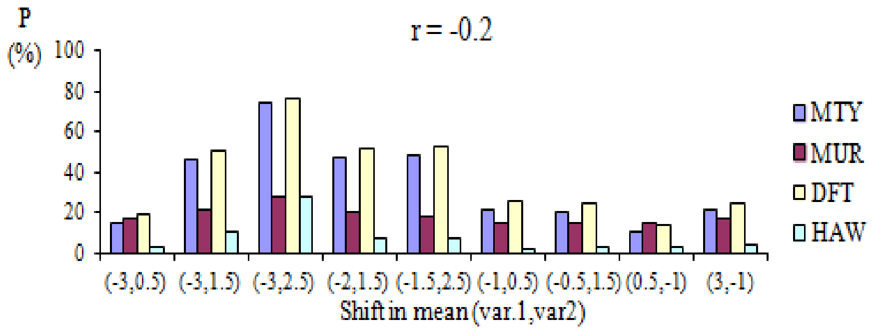
Das and Prakash (2008) also failed to mention the difference in power for the shifts with the same small magnitude but on different variables (i.e. (-1, 0.5) and (0.5, -1)) for $\rho = -0.2$ and $\rho = -0.45$ (Table 6 and Table 7). This peculiar feature showed by MTY and DFT but not by MUR and HAW. Figure 3.2 and 3.3 show the pattern of the performance across different shifts and correlations. Another general observation that can be added from Figure 3.2 that is for the intermediate shifts, the power for MTY and DFT methods is higher compared to the combination of large and small shifts. The power for both methods for the intermediate shifts and the difference of power between the type of shifts increases when the correlation increases in a positive direction.



(a) (b)

Figure 3.2: The performance of four diagnostic methods when (a) two variables deviate in the same negative direction and (b) two variables deviate in the same positive direction.

In Figure 3.3, one can see in general that the bars representing MUR and HAW methods for moderate negative correlation are mostly very low and almost “hidden” whereas for low negative correlation, most of the bars are taller especially for the MUR method. A close observation between Figure 3.2 and 3.3 showed that a similar correlation value but different sign does not give similar estimated power value.



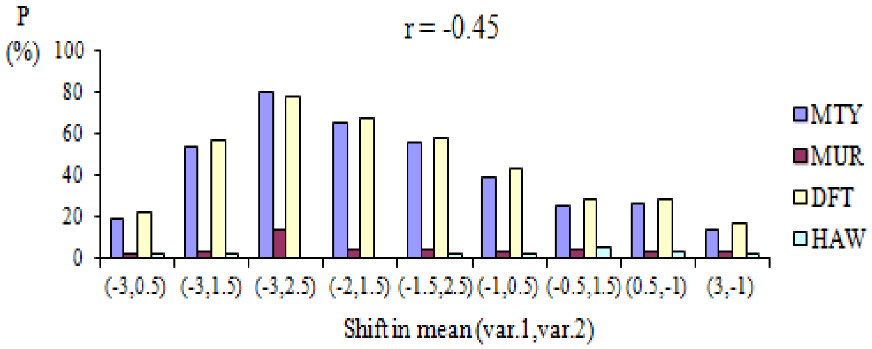


Figure 3.3: The performance of four diagnostic methods when the shifts in mean are in accordance with the correlation structure.

3.4.2 Shifts Not in Accordance with the Correlation Structure

The results for this case are shown in Tables 6, 7, 8, 9 and 10 in Das and Prakash (2008). In general, Das and Prakash (2008) concluded that the MTY and DFT performance is satisfactory for large and intermediate shifts irrespective of correlation structure. On the other hand, HAW and MUR are efficient when the correlation is low negative or positive. More detail is given below.

MTY and DFT

i) $\rho = 0.2$

Performance dropped to some degree for large shifts.

ii) $\rho = 0.5$

Lose power when the large shifts were in opposite directions and when one of the shifts reduced in magnitude. The performance was good whenever the shift is a combination of intermediate and large magnitude.

iii) $\rho = 0.8$

Perform well for large shifts but performed poorly when at least one of the shifts is small (in magnitude).

MUR and HAW

i) $\rho = 0.2$

Performance was above average for large shifts.

ii) $\rho = 0.5$

Power was extremely higher than the cases with shifts in accordance with the correlation structure.

iii) $\rho = 0.8$

Perform well for large shifts but performed at average level when at least one of the shifts is small.

Das and Prakash (2008) stated that "*for intermediate shifts in at least one of the variable and large shift in another, HAW works excellently followed by MUR whenever the shifts are in opposite directions*". It is true that HAW performed extremely well for these combinations but it is not entirely true for MUR. The same excellent performance showed by MUR only for shifts (-3, 1.5). For the other combinations of intermediates or large and intermediate shifts, MTY and DFT are better or more or less similar to MUR.

This study does not agree with Das and Prakash (2008) regarding the methods' performance for the combination of intermediate and large shift when $\rho = 0.5$. MTY and DFT is said showed a good performance with power 29% and 37% (Table 9) respectively for shifts (-3, 1.5). Based on earlier discussion in subsection 3.4.1, the highest limit for poor performance believed implied in one of Das and Prakash (2008)'s statements is 30%. Obviously, there is another contradiction in the assessment of the methods' performance.

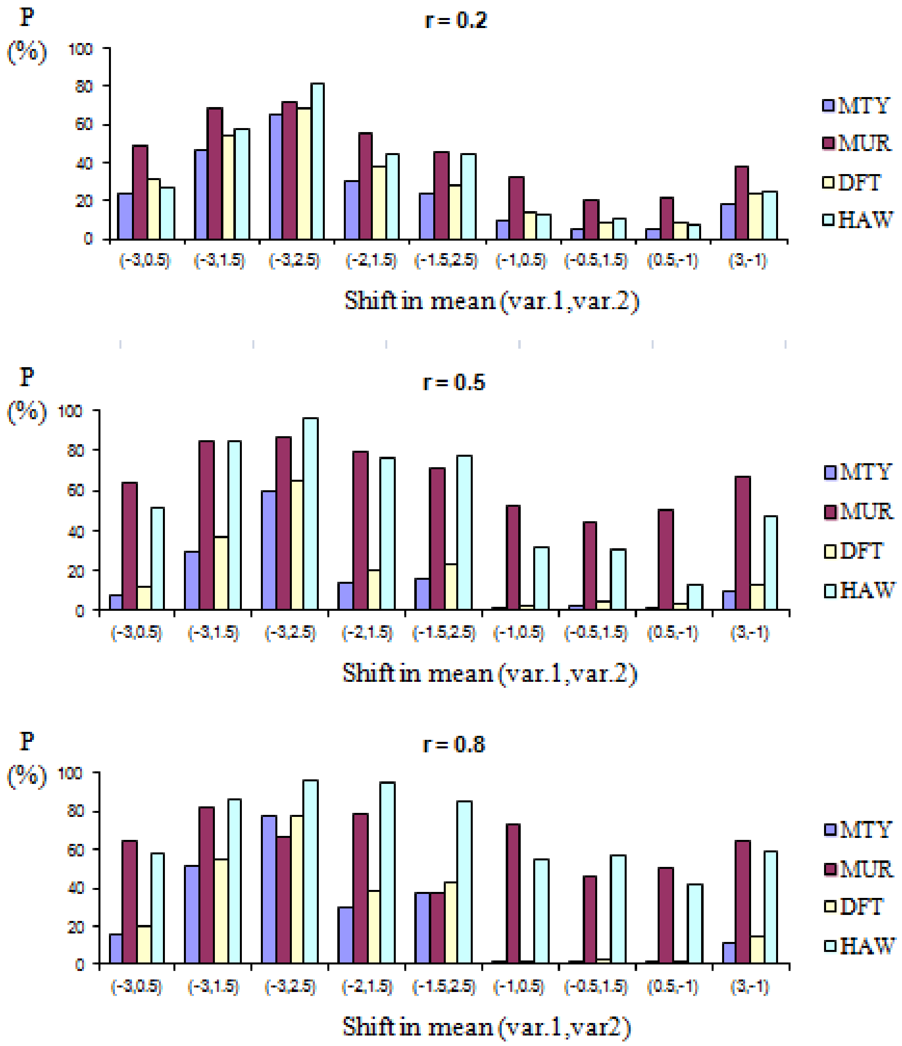


Figure 3.4: The performance of the diagnostic methods with mixed shifts in two deviated variables with low, medium and high correlation between them.

Figure 3.4 shows that in general, MUR and HAW performed much better than MTY and DFT for positive correlations. HAW especially consistently shows a good or a very good performance as long as no small shift in the shifts combination. The performance is very good when the shifts are both large or both intermediate (for strong positive correlation) or a combination of large and intermediate shifts (except for low positive correlation). MUR also follows the same pattern as HAW but with slightly lower power. MUR performed better than HAW in most of the cases when at least one of the shifts is small.

MTY and DFT performed lower or sometimes similar to MUR and HAW in most of the cases throughout the combinations of shifts in mean. Their performances are very poor when the shifts are small. A good performance is shown when both of the shifts in mean are large. The power is higher for the strong positive correlation between variables and drop lower for moderate correlation than low correlation

Some of the results in Tables 6 and 7 in Das and Prakash (2008) showed the performance for cases with shifts in mean in the same directions with low negative and moderate negative correlations between the variables. The observations on the performance for negative correlations are given by Das and Prakash (2008) as follows

MTY and DFT

i) $\rho = -0.45$

A satisfactory performance when the shifts in mean are large. The estimated power is far below average when at least one of the shifts is small.

ii) $\rho = -0.2$

Perform at satisfactory level when both shifts in mean are large. The estimated powers are at an average level and lower than MUR and HAW.

MUR and HAW

i) $\rho = -0.45$

MUR and HAW are not effective.

ii) $\rho = -0.2$

Both methods gained power when shifts in mean are large. Performance was better than MTY and DFT for an intermediate shift in at least one of the variables.

This study would like to add that for $\rho = -0.45$ and the shifts are large, the performance of MTY and DFT methods are relatively lower than the cases with the shifts in mean in accordance with the correlation structure. The estimated power is more or less similar to the one with $\rho = -0.2$ for the shifts combination with at least one large shift. The estimated power is slightly or sometimes notably higher for $\rho = -0.2$ when no large shift in the shifts combination.

Methods MUR and HAW showed an extremely higher estimated power for ρ = -0.2, on the other for ρ = -0.45, both methods are not responsive. Figure 3.5 shows that MTY and DFT performed much better than MUR and HAW for negative low correlation for shifts combination (-3,-3) and approximately equal to HAW for combination (3, 2.5). For the rest of the shifts combinations, the power for MTY and DFT methods is lower than the other two methods. MUR is more superior than the other methods when there is at least one small shifts in one of the two aberrant variables. For moderate negative correlation, MTY and DFT show a good performance with MUR and HAW perform very poorly and in some cases, both methods are unresponsive when at least one of the shifts is large.

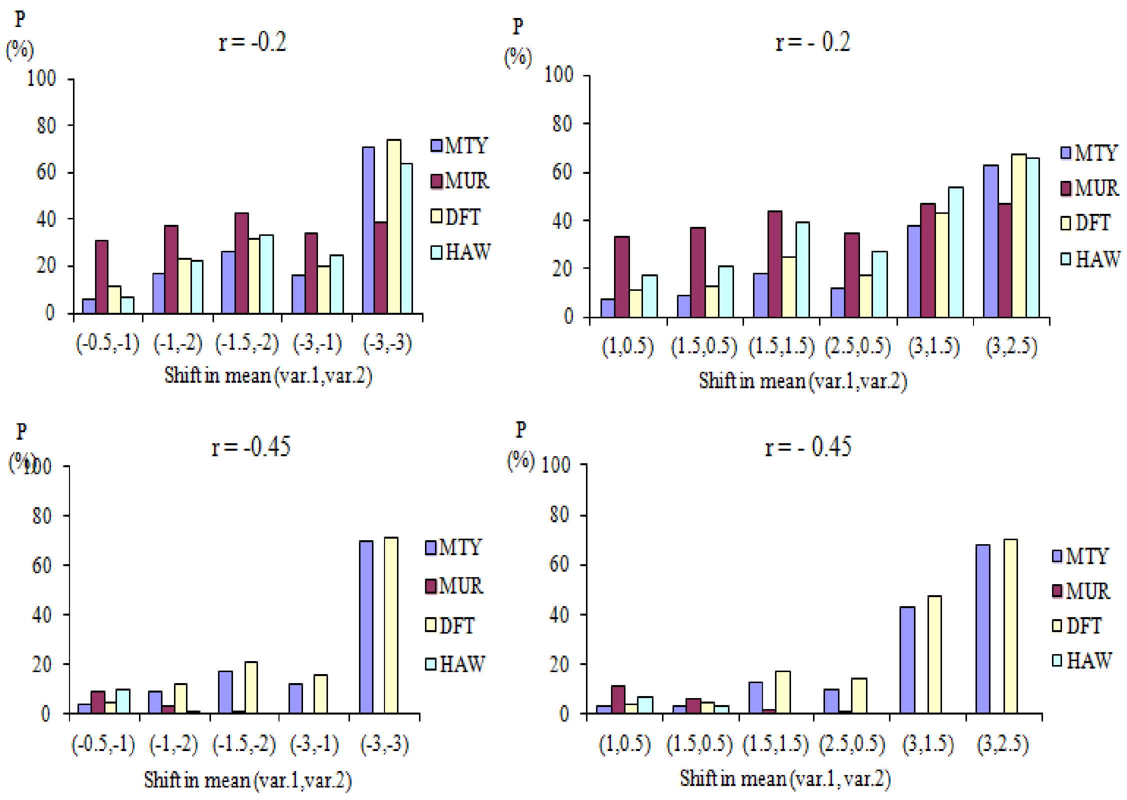


Figure 3.5: The performance of the diagnostic methods with shifts in mean in the same directions with negative correlation between them.

3.4.3 The Effect of Correlation Structure, Sign and Strength

Das and Prakash (2008) give a detail observation on the performance of the diagnostic methods with respect to the magnitude and the direction of the shifts. However, the authors didn't directly relate the changes in power to the changes in the correlation's strength and signs. This study took a step further to optimise the information available from the Das and Prakash (2008) simulation results. A few selected cases will be used to show the effect of correlation structure, sign and magnitude upon the power of the studied diagnostic methods. The discussion will focus on low correlation alone. There are six combinations of shifts in mean vector considered in this discussion. The shifts are (-3, 2.5), (3, 2.5), (-3, 1.5), (3, 1.5), (3, -1) and (-3, -1). The first two shifts intended for the investigation on the performance when the shifts are large and in accordance as well as not in accordance to the correlation structure. The third and the fourth shifts are for large and intermediate contaminants under the same two conditions and the last two shifts are when the shifts are large and small.

The effect of the sign of the correlation value

Based on close observation on each of the shifts (Table 7 and Table 8 in Das and Prakash (2008)), this study finds that MTY and DFT have similar power regardless of the sign of the correlation values except for combination of large and intermediate shifts. The power is appreciably higher for $\rho = 0.2$. This performance is similar for both cases, shifts in accordance and not in accordance with the correlation structure. On the other hand, MUR does not show any similar pattern of performance between the correlation signs. For HAW, the power is higher for $\rho = 0.2$ except for the case with combination of large and small shifts.

We may conclude here, for all the studied shifts, MTY and DFT are not affected by the sign of correlation value except for the combination large and intermediate shifts. MUR seems affected by the correlation values but even so, the correlation structure is also suspected to play a role in the inconsistency of the power shown by MUR. HAW also showed different power for different correlation signs and the power is higher when the correlation is low positive except for the combination of large and small shifts.

The effect of the correlation structures

MTY and DFT showed a consistent pattern of performances where the power is higher throughout all the studied shifts when the shifts are in accordance with the correlation structure. The powers are also consistent and similar between negative and positive correlations except for a combination of large and intermediate which has been noted in the previous discussion. On the other hand, MUR and HAW showed appreciably higher power when the shifts are not in accordance with the correlation structure.

We may also conclude here, for all the studied shifts, there is a difference in term of performance pattern between MTY and DFT with the other two studied methods, MUR and HAW. MTY and DFT showed slightly better performance when the shifts are in accordance with the correlation structure whereas MUR and HAW performed much better when the shifts are not accordance with the correlation structure. Generally, another pattern that is noticeable for all the studied diagnostic methods is that the power decreased when one of the shifts reduced in magnitude.

3.5 Discussions and Conclusion

Studying the findings of Das & Prakash (2008), it is not difficult to realise that the performance of the tested diagnostic methods differs with respect to the level of the shift in mean from the target value. Generally, the performance of the studied diagnostic method increases with the magnitude of the shift in mean. The comparison of performances between the diagnostic methods becomes more complicated when more than one variable deviated from its mean.

It is quite difficult to single out one diagnostic method as the best for all combinations of shifts in mean across different types of correlation among variables. Even so, based on the comparative studies discussed previously, an appropriate diagnostic method can be selected with respect to the consistency of a particular diagnostic method's performance across various shifts in mean combinations and correlation among variables. Even though HAW showed the best result when the shifts are not in accordance to the correlation structure (when the correlation between the variables are positive), its performance is inconsistent and often exceedingly poor. MUR and HAW especially seem very sensitive to the correlation between variables specifically to the sign of the correlation coefficient or the direction of the association between the process variables in which, in a real situations maybe unknown. This study suspects that the performance is not strictly depending on the direction itself, but on whether the shifts in mean vector agrees with the correlation structure between the variables.

Generally, MTY and DFT have shown a consistent and better performance in most of the cases. Even though both methods showed higher power when the shifts in mean are in accordance with correlation structure, the power of the two methods between the two cases didn't differ a lot unlike the powers shown by MUR and HAW. DFT has been chosen to be included in the extended study in later chapters simply because of the simplicity and the practicality of the method against the complexity of the MTY procedures.

The shifts in Das and Prakash (2008) study will be used for the new comparison study with a few precautions on several matters which arise when studying the authors' simulation results. Firstly, we must provide a clear definition of m and n on equation [2.8] especially for the cases with two aberrant variables. Secondly, we must provide a clear definition of performance for the diagnostic methods and making sure no ambiguities arise in describing the power intervals in determining the level of performance of the studied diagnostic methods. Thirdly, we aim to maintain the focus of the results observation with respect to the effect of the shifts in mean and the correlation structures.

CHAPTER 4

Extended Comparative Study

4.1 Introduction

This chapter provides an extended comparative study of some diagnosis methods for the multivariate control problem. Some of the ambiguities of the previous Das and Prakash (2008) are clarified and an additional method is included. Specifically, we include a new approach that utilising the ratio computation proposed by Maravelakis et al. (2002), hereafter referred to as the Ratio method. It is important to stress that, this study did not follow the approach proposed by Maravelakis et al. (2002) entirely. The focus is solely on the computation of the ratio. Since, the diagnostic method DFT has shown a consistent and good result in Das & Prakash (2008), it will be compared to the Ratio method with respect to their performance in identifying the correct aberrant variable(s). Beforehand, each diagnostic method will be discussed thoroughly in sections 4.2 and 4.3 to ensure a very good understanding of their procedures in implementing their approaches. Later sections will describe various stages of the comparative study, starting from the generation of the datasets, the identification of the out-of-control signal, and then performing both selected diagnostic methods to identify the aberrant variable(s) and finally comparing their performance. Section 4.5 presents the results of the comparison between the two methods. A further investigation of the Ratio method is carried out in section 4.6 to study the potential and the drawbacks of the method. Some discussion and conclusions are given in the final section of this chapter.

4.2 Doganaksoy, Faltin and Tucker Method

The diagnostic method introduced by Doganaksoy, Faltin & Tucker (1991) is a univariate approach to interpreting the OOC signals produced by multivariate control charts. The approach is based on the calculation of univariate *t*-statistics. Figure 4.1 illustrates the step by step procedure in implementing the method. Once the OOC signal is received, a primary diagnosis is carried out to see how likely each process variable is to have contributed to the OOC signal, then a secondary assessment decides whether contributions are significant.

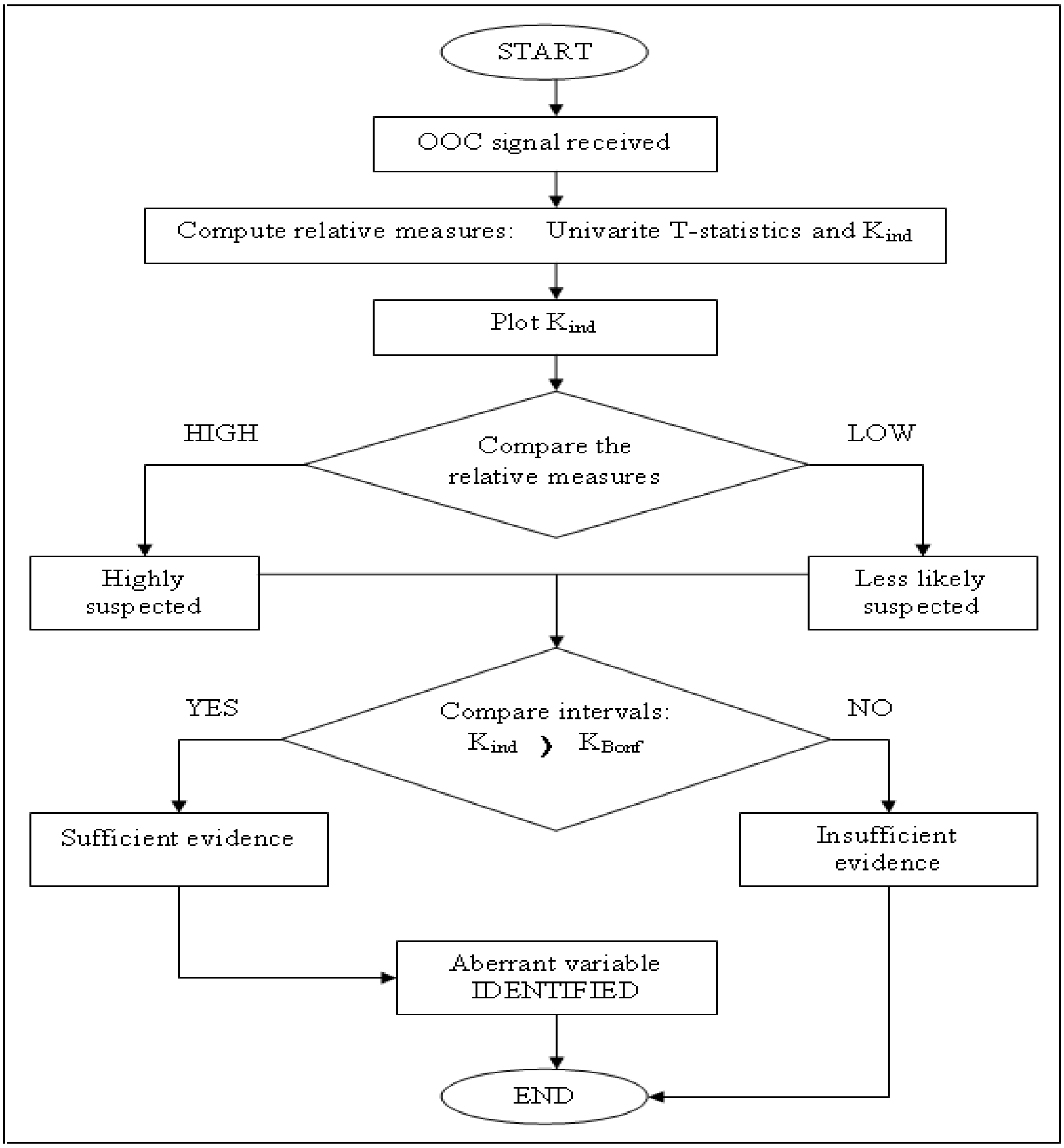


Figure 4.1: The flow chart of the process for the implementation of the DFT method.

A univariate *t*-statistic in [2.5],

$$t_i = \frac{\left(\bar{x}_{i,new} - \bar{x}_{i,ref}\right)}{\left[s_{ii}\left(\frac{1}{n_{new}} + \frac{1}{n_{ref}}\right)\right]^{1/2}}$$

is produced for each variable and a relative measure, $K_{ind} = |2T(t; n_1 - 1) - 1|$ in [2.6] is computed and plotted as suggested by Doganaksoy, Faltin and Tucker (1991). The higher the K_{ind} measure, the more probable the variable will be as being the source of the change in the mean vector, and so responsible for the OOC signal.

In order to determine whether there is sufficient evidence to pinpoint specific variable(s) as aberrant, Doganaksoy, Faltin and Tucker (1991) recommended the use of Bonferroni type simultaneous confidence intervals as a supplement to the K_{ind} measure. Another measure to represent the Bonferroni type confidence interval, $K_{Bonf} = \frac{(p+K_{Sim}-1)}{p}$ in [2.7], is computed and compared to K_{ind}.. Any component (one or more process variables) with its measure of $K_{ind} > K_{Bonf}$ is classified as being one whose mean is likely to have deviated from the in-control value.

Doganaksoy, Faltin and Tucker (1991) clearly defined two measures of Performance in their findings. The performance measures are the percentage of ranking correct (*Ranking Correct*) and the percentage of Bonferroni correct (*100K% Bonf. Correct*). The computation of both performance measures are given below.

$$\textit{Ranking Correct} = \frac{n(\max\{T\})}{n(OOC)} \text{ x } 100\% \qquad [4.1]$$

$$\textit{100K\% Bonf. Correct} = \frac{n(VI)}{n(VBonf.)} \text{ x } 100\% \qquad [4.2]$$

where,

$n(\max\{T\})=$ the number of the truly changed variable(s) having the largest absolute univariate t statistic value

$n(OOC)$ = the number of out-of-control signals triggered by a multivariate control chart

$n(VI)$ = the number of the truly changed variable(s) with violated intervals

$n(VBonf.)$ = the number of violation of Bonferroni type simultaneous interval

The performance is called a "selective power" in Doganaksoy, Faltin and Tucker (1991).

4.3 Ratio Method

The diagnostic approach proposed by Maravelakis et al. (2002) has a slightly different approach from the Ranking method. It is still a complementary method to the multivariate control charts in that is to be applied once OOC signal(s) are received. The similarity of this method with the Ranking method is that another measure is also to be computed for each of the process variable, which in this case a ratio [2.5]. Instead of using a univariate approach, a principal component analysis, which is one of the multivariate analysis methods, is performed on the in-control dataset. The coefficients of the first PC or the first eigenvector of the in-control dataset provides a specific weighting to each of the process variables and is utilised in the ratio calculation. The coefficients are regarded by Maravelakis et al. (2002) as the expected contribution of each process variable in a normal process, or statistically stable process, condition.

Figure 4.2 below illustrates the implementation of the Ratio method once the OOC signal is triggered by a multivariate control chart. As explained in section 2.4.3, Maravelakis et al. (2002) proposed a ratio calculation on equation [2.4] for covariance matrices with positive covariance matrix. The denominator of [2.4] is calculated by using the in-control mean vector and not by the observation vector, X_i for covariance matrix with mixed sign values. So it is not surprising when Maravelakis et al. (2002) does not recommend this method to be implemented on standardized values. It is very clear to users that for the cases with positive and negative correlation in covariance matrix, the denominator will be zero in value and the ratio will end up undefined. However, no further explanation or reasons given by Maravelakis et al. (2002) regarding the necessity to have two different ratio calculations for two types of covariance matrices.

This study would like to highlight that having two different methods for different type of covariance matrices shows the impracticality of the diagnostic method itself. The suggestion for not using the method for standardized data makes it difficult for the practitioners to use it since in many process control monitoring there are several process variables with different units involved. It is also indicates that the diagnostic method is not location invariant. Since this study used standardized values and there is no warning against the usage of [2.4] for simulated data from mixed sign variance covariance matrix, hence, this study used [2.4] for ratio calculation for both type of covariance matrices.

Generally, Maravelakis et al. (2002) proposed two steps of control charting. The first chart is a Hotelling's multivariate control chart which has been referred to by Maravelakis et al. (2002) as a Chi-Square multivariate control chart. This step as we already know is to identify any discrepancy in the process. Once the out of control condition detected then the ratio charting will follow. Unlike the Ranking method, the Ratio method is heavily depending on the control limits of the ratio charting. Since there are two Ratio computations to serve different type of covariance matrices, then it gives us two sets of control limits. The control limits for a Ratio chart of a multivariate process with positive correlations come from a Bivariate Normal probability distribution whereas the control limits for a Ratio chart of a multivariate process with positive and negative correlations come from a Normal probability distribution.

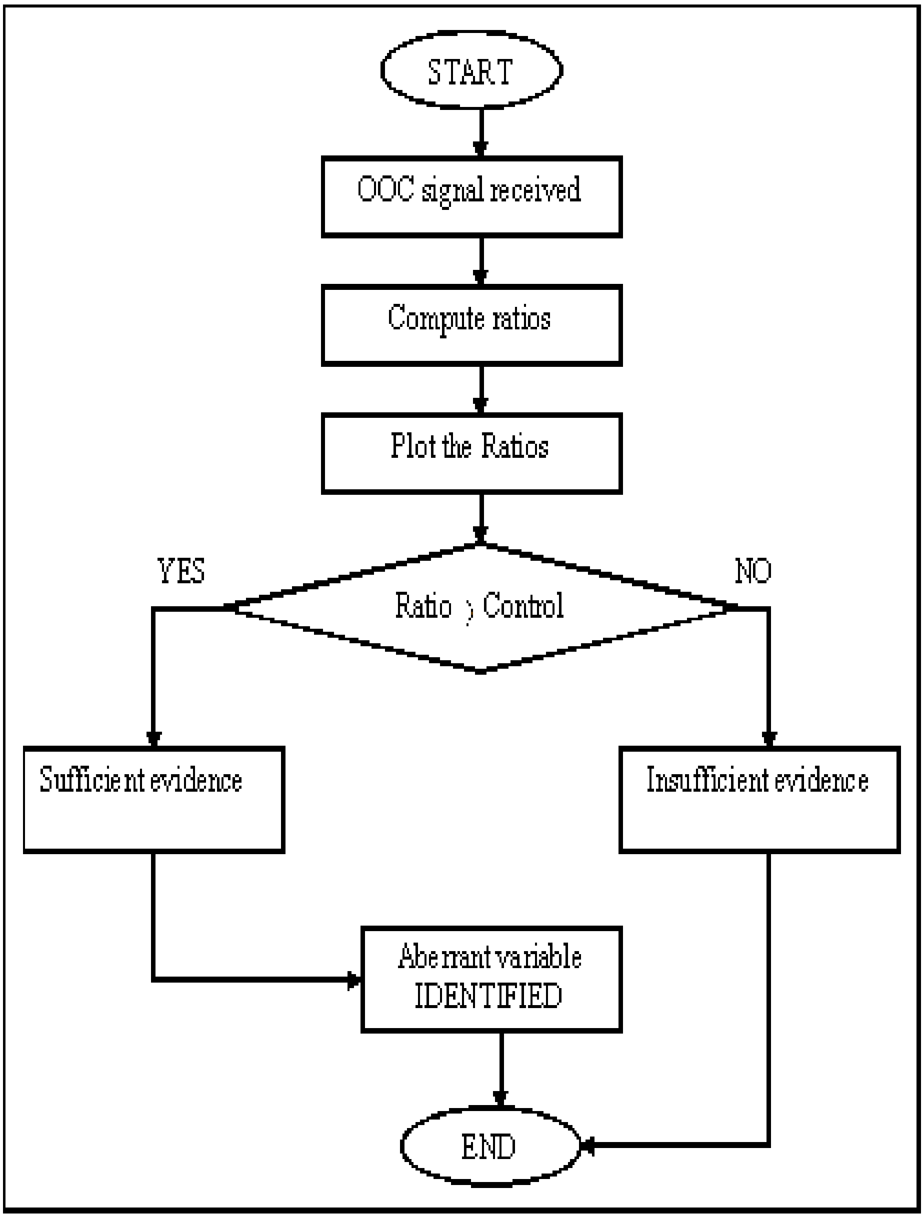


Figure 4.2: The flow chart on the process of the implementation of the Ratio method.

Maravelakis et al. (2002) rely on the Average Root method (Jackson, 1990) in determining the number of principal component in which the first principal component alone is considered in the ratio calculation based on the fact that the first principal component contains the most information about the data. Hence, the ratio for variable *k* in observation *i* will be

$$\mathrm{r}_i = \frac{(\mu_{k1} x_{ki})}{\mathrm{Y}_{1\mathrm{i}}} \qquad [4.3]$$

where,

x_{ki} = the i-th value of variable-k

u_{k1} = the coefficient of the eigen vector for variable-k

Y_{1i} = the score for vector $\mathbf{x}_\mathrm{i}$ in the first principal component

In this study, the ratios will not be plotted with the control limits to identify the aberrant variables but merely assessed against them numerically. As stated by Maravelakis et al. (2002), the ratio represents the contribution of variable k in observation i, therefore this study tried to adopt Doganaksoy, Faltin and Tucker (1991) approach by ranking the contributions of each variable in an observation. The ratios are treated as the weight of a variable in an observation (Maravelakis et al., 2002). The higher the weight of the variable, the more likely it is aberrant. Instead of studying whether the ratios are within the control limits, we are now ranking the ratio of each variable based on its contribution in an observation.

4.4 Simulation Study

Datasets

Diagnostic methods are studied by simulating data from multivariate normal distribution. The in-control mean vector, $\boldsymbol{\mu}$ and variance covariance matrix, $\boldsymbol{\Sigma}$ are assumed known. There are two covariance matrices considered in this simulation study. The covariance matrices, $\mathbf{c_1}$ and $\mathbf{c_2}$, are two of the four covariance matrices introduced in Doganaksoy, Faltin and Tucker (1991). Doganaksoy, Faltin and Tucker (1991) stated that "*the covariance matrices were carefully chosen so as to cover a wide range of possible situations*".

$$\mathbf{c}_1 = \begin{pmatrix} 1 & 0.8 & 0.55 & 0.65 \\ 0.8 & 1 & 0.65 & 0.5 \\ 0.55 & 0.65 & 1 & 0.6 \\ 0.65 & 0.5 & 0.6 & 1 \end{pmatrix}; \qquad \mathbf{c}_2 = \begin{pmatrix} 1 & 0.2 & -0.5 & 0.3 \\ 0.2 & 1 & 0.2 & -0.5 \\ -0.5 & 0.2 & 1 & 0.2 \\ 0.3 & -0.5 & 0.2 & 1 \end{pmatrix}$$

Each observation generated is treated as an individual potentially aberrant observation presented to a multivariate control chart (we do not examine treatment of sequences of observations or attempt any control actions or chart modification). Figure 4.3 below represent the flow of this comparative study.

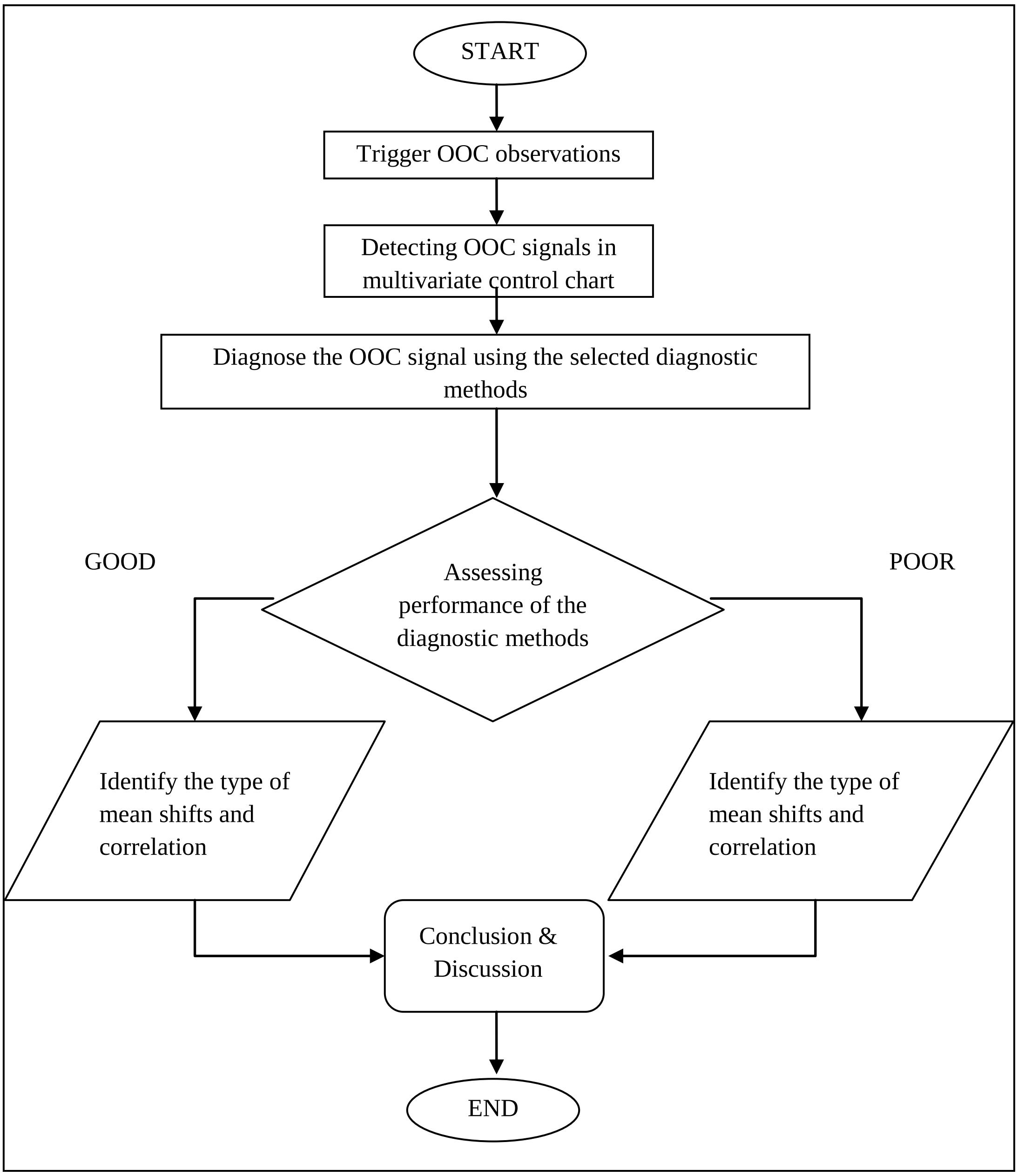


Figure 4.3: Flowchart of the programming work

Out-of-control observations are generated from multivariate normal distribution with a contaminated mean vector. A contaminant is introduced in the mean vector to enable contaminated observations to be generated and so trigger OOC signals. A contaminant refers to a mean shift in one or more variables in the mean vector. We consider three types of mean deviation or shift(s) in mean; *Type I* refers to shift of mean of one variable; *Type II* to shifts in mean vector of two process variables in the same direction, and *Type III* to shifts in mean on two process variables in opposite directions (this is of interest since we suspect a dependence of performance on whether shifts 'agree with' or 'contradict' the correlation structure. The contaminated process variables are process variable 1 and, in *Type II* and *III,* process variable 3. The diagnostic methods are performed on 5000 out-of-control individual multivariate observations from each combination of correlation matrix and type of shift(s) in mean vector. We subject each generated observation to a standard Hotelling's T^2 test to establish whether it can be deemed OOC or not. The simulation proceeds until 5000 OOC observations have been detected. In some cases we needed to generate substantial more than 5000 observation to achieve to achieve this as such random shifts in mean will not be detected. However, it makes performance comparison much easier if the number of times the diagnostic method is employed is holds constant. The simulation is done in **R**-programming.

In this preliminary comparison study, most of the simulation format of Doganaksoy, Faltin and Tucker (1991) is followed, as well as the size of the shift used, which is fixed at 2 or -2. One thing that differs is the value for *n* in equation [2.8]. In this preliminary comparison study, the number of OOC observations is fixed at 5000 whereas in Doganaksoy, Faltin and Tucker (1991), the value is varies with respect to the shift(s) in the mean vector and correlation matrices. This study believes that by making the number of OOC observations similar throughout different shifts in mean vector, the comparison between the correlation matrices can be made easily and clearly.

Performance of the diagnostic method explained previously is assessed by calculating the percentage of the correct identification as proposed by Das and Prakash (2008) and shown in [2.8]. In this study, we would like to give a clearer definition for the performance of the diagnostic methods in which hereafter called as Power (P) and define as

$$Power\ (P) = \frac{n(detect)}{n(OOC)} 100 \qquad [4.4]$$

where

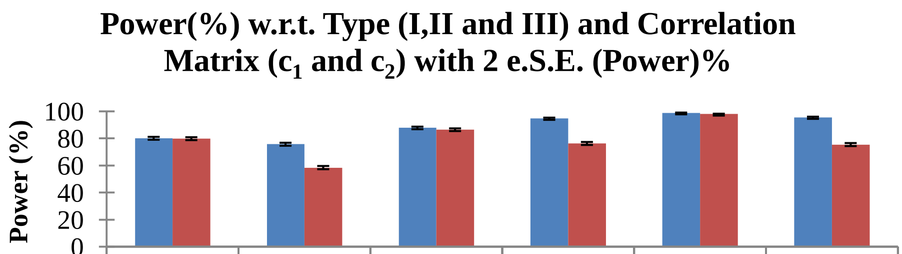
n(OOC) = the number of times T^2 detects the shift or out-of-control observations

n(detect) = the number of times the aberrant variable(s) is detected as aberrant by a diagnostic method

A diagnostic method is considered to be successful in detecting the right aberrant variable(s) when the aberrant variable(s) has the highest "value" among all the variables. The "value" here is referring to the ratio [2.4] or *K_ind* [2.5]. For shift in mean *Type I*, a diagnostic method is considered successful in detecting the correct aberrant variable when variable 1 has the highest ratio or *K_ind* among all variables. Whereas for shifts in mean *Type II* and *Type III*, a diagnostic method is considered successful in detecting the correct aberrant variable when variable 1 or variable 2 has the highest ratio or *K_ind* value among all variables. The percentage results are studied to identify any pattern of sensitivity towards the type of shift and correlation among variables.

4.5 Results

The results of the preliminary investigation are divided into separate sections to maintain a clear and easy understanding of the results. Basically, all the results are compared between two types of correlation matrices, and three types of mean shifts. The results from both diagnostic methods are compared with respect to these two factors. The following results will show that something unsatisfactory for one of the methods with one particular correlation matrix. Figure 4.4 illustrates the percentage of correct identification for both methods with three types of shift, single deviated variable (*Type I*) and two deviated variables (*Type II* and *Type III*) with respect to correlation matrices.



	c1	c2	c1	c2	c1	c2
	Type I		Type II		Type III	
DFT	80.10	75.68	87.84	94.78	98.84	95.52
Ratio	79.82	58.34	86.50	76.24	97.92	75.42

Figure 4.4: The distribution of Power (%) for method DFT and Ratio with respect to three different Types (I, II, III) and 2 correlation matrices ($\mathbf{c_1}$, $\mathbf{c_2}$).

In general, Figure 4.4 showed that the power is increasing throughout the type of the shifts in mean. The power for each method for different correlation matrix is lower for *Type I* and increasing a bit in *Type II* and so forth. The DFT and Ratio methods have quite similar power and increasing from *Type I* to *Type III* when the correlation matrix is $\mathbf{c_1}$. However, the power of the Ratio method is appeared to be considerably lower than the DFT method when the correlation matrix is $\mathbf{c_2}$ for each type of shifts in mean. The power distribution of both methods across different types of shifts in mean and correlation matrices is studied closely. The difference in power with its corresponding 2 estimated standard error with respect to the diagnostic methods, correlation matrices and the types of shifts in mean is given in Tables 4.1 and 4.2.

Power Comparison between the Diagnostic Methods

Table 4.1 shows the difference in Power (%) between the DFT and the Ratio methods with respect to the correlation matrices and the types of shifts in mean. For correlation matrix $\mathbf{c_1}$, there is small and not significant difference in power (%) between the two diagnostic methods for shifts in mean *Type I* and *Type II*. For *Type III* shifts in mean, there is also a small difference in power (%) between the two methods but the difference is found to be significant.

Table 4.1: The distribution of the Power Difference (%) between the diagnostic methods

	$\mathbf{c_1}$		$\mathbf{c_2}$	
	$(p_{DFT} - p_{Ratio})$%	2(S.E.)%	$(p_{DFT} - p_{Ratio})$%	2(S.E.)%
Type I	0.28	1.60	17.34	1.85
Type II	1.34	1.34	18.54	1.36
Type III	0.92	0.50	20.10	1.35

However for the correlation matrix $\mathbf{c_2}$, Table 4.1 shows that the difference in power between the two diagnostic methods is considerable. The DFT method consistently shows a significantly higher power than the Ratio method throughout the different types of shifts in mean. The largest difference is for shifts in mean Type III where the mean of the aberrant variables shifted in opposite directions.

Power Comparison between the Correlation Matrices

Table 4.2 shows the power difference of the diagnostic methods with respect to the different correlation matrices and the types of shifts in mean. Table 4.2 (a) provides the difference of power (%) for DFT method with correlation matrix $\mathbf{c_1}$ and the same diagnostic method with correlation matrix, $\mathbf{c_2}$.

Table 4.2: The distribution of the Power Difference (%) between correlation matrices

DFT	$(p_{c1}-p_{c2})$%	2(S.E)%
Type I	4.42	1.66
Type II	-6.94	1.12
Type III	3.32	0.66

(a)

Ratio	$(p_{c1}-p_{c2})$%	2(S.E)%
Type I	21.48	1.80
Type II	10.26	1.54
Type III	22.5	1.28

(b)

Table 4.2 (b) shows the power difference for the Ratio method between the two correlation matrices for each type of shifts in mean. The differences in power (%)

for DFT method between the two correlation matrices are considered small. The differences in power for each type of shifts in mean are within ± 7%. Even so, the difference in power (%) is found to be significant with respect to 2 standard errors in power (%) difference. Table 4.2 (a) showed that the power of the DFT method with correlation matrix $\mathbf{c_1}$ is found to be significantly higher compared to the power of DFT method with correlation matrix $\mathbf{c_2}$ for shifts in mean Type I and Type III. On the other hand, the power of the DFT method with correlation matrix $\mathbf{c_1}$ is significantly lower than the power of the diagnostic method with the correlation matrix $\mathbf{c_2}$ for shifts in mean Type II.

The power difference (%) for each type of shifts in mean shown in Table 4.2 (b) is much larger than the one shown in Table 4.2 (a). It indicates that the power of the Ratio method is very much affected by a correlation matrix. The power of the Ratio method with correlation matrix $\mathbf{c_1}$ is significantly higher than the power of the diagnostic method with correlation matrix $\mathbf{c_2}$. Unlike the DFT method, the power of the Ratio method with correlation matrix c1 is consistently higher than the power of the diagnostic method with correlation matrix $\mathbf{c_2}$ throughout all the types of shifts in mean. The highest power difference (%) is shown by the shifts in mean Type III and closely followed by the shifts in mean Type I.

4.6 Further Investigation of the Ratio Method

The results shown in Figure 4.4 have illustrated that in certain situations the Ratio method fails and hence poor power results are produced. The peculiar results only happen when correlation matrix $\mathbf{c_2}$ is used. Table 4.2 (b) has shown a large difference in power (%) for the Ratio method when different correlation matrices is used. A further study of the effect of the correlation matrices is given below.

4.6.1 The Effect of Correlation Matrix Structure

The theoretical eigenvalues and eigenvectors were obtained for both correlation matrices and are shown below in Table 4.3. First of all, focus on the theoretical eigenvalues between correlation matrix $\mathbf{c_1}$ and $\mathbf{c_2}$. It is clear from Table 4.3 that the first and the second eigenvalues of correlation matrix $\mathbf{c_1}$ are far apart whereas for correlation matrix $\mathbf{c_2}$, the first and the second eigenvalues are very close together. We suspect that the cause of the peculiar results is due to this closeness between the first two eigenvalues.

Table 4.3: Theoretical eigenvalues and eigenvectors for $\mathbf{c_1}$ and $\mathbf{c_2}$.

		Correlation Matrix							
		$\mathbf{c_1}$				$\mathbf{c_2}$			
Eigenvector		1	2	3	4	1	2	3	4
Eigenvalue		2.855	0.555	0.432	0.158	1.554	1.452	0.946	0.048
Loadings	var. 1	-0.521	0.407	0.403	0.633	0.540	0.482	-0.457	0.518
	var. 2	-0.522	0.510	-0.136	0.671	-0.457	0.518	-0.540	-0.482
	var. 3	-0.488	-0.323	-0.758	0.288	-0.457	-0.518	-0.540	0.482
	var. 4	-0.467	-0.686	0.495	-0.258	0.540	-0.482	-0.457	-0.518

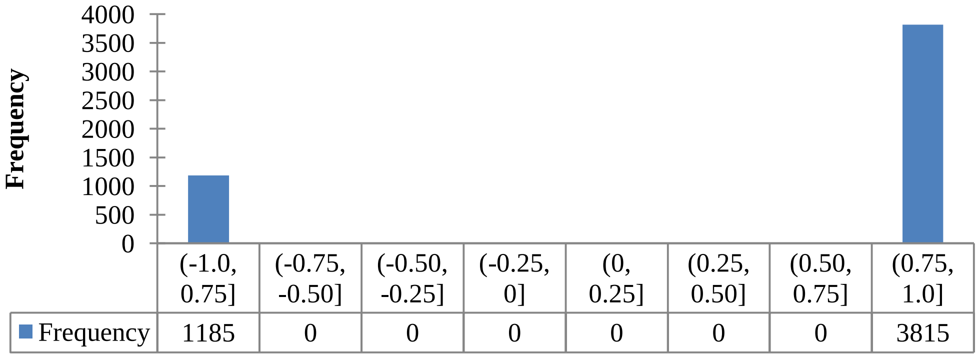
Since the sample eigenvalues have appreciable standard errors then there is a possibility that if the population values are close together the sample values could be in the wrong order. Unfortunately, the standard error of eigenvalues from correlation matrices seems to be unavailable (Anderson, 1993) but for eigenvalues from covariance matrices Anderson gives the standard error as $l_i\sqrt{\frac{2}{n}}$ where l_i is the *i*th sample eigenvalue based on a sample of size *n*. The standard error of correlation matrix eigenvalues will be of a similar order of magnitude and it is clear that if two eigenvalues are close together then for moderate sample sizes there is a possibility that the sample values are not in the same order as the population ones. This would lead to the 'wrong' eigenvector being selected.

4.6.2 The Inner Product of the First Theoretical and Sample Eigenvectors

This study tries to find the risk or the probability that a random sample from a multivariate normal distribution with a specified theoretical covariance matrix has a sample covariance whose eigenvalues have swapped around. The computation of the inner products is done to investigate whether the two eigenvectors have the same direction. The inner products values are always between -1 to +1. If the two vectors are in the same direction, the inner product will be close to +1 or -1. Inner product -1 means the eigenvectors going in the same line, but one in the reverse direction (out by 180^0).

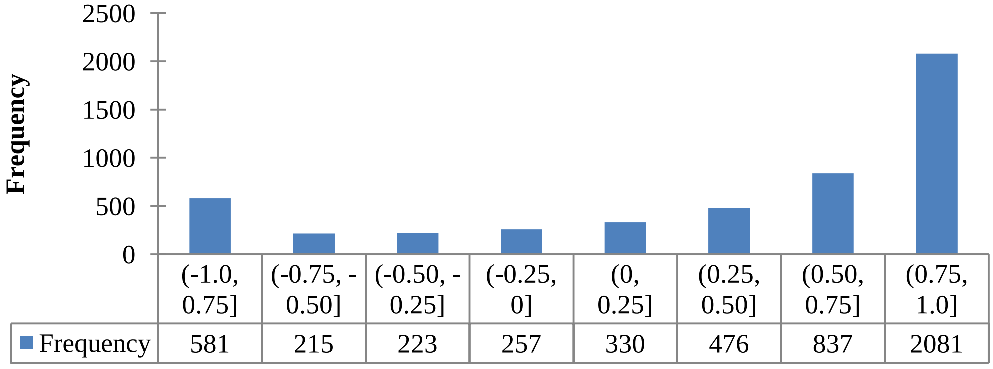
More interestingly, if the inner product is 0, this means the two eigenvectors are orthogonal. Thus, values for the inner product between the first sample principal component (PC_S) and the first theoretical principal component (PC_T) which are close to 0 suggest that it is a sample value of the 'wrong' principal component because all principal components are, by construction, orthogonal. The issue is investigated by simulation. A random sample size is fixed at 50 and the two theoretical scaled covariance matrices, $\mathbf{c_1}$ and $\mathbf{c_2}$, are tested. The simulation is done for 5000 times, the inner products of the first eigenvector of the sample covariance matrix with the actual first eigenvector of the theoretical covariance matrix are recorded.

Figure 4.5 (a) shows the frequency distribution of the inner products between the first eigenvectors of the sample covariance matrix with the actual first eigenvector of the theoretical covariance matrix $\mathbf{c_1}$.Whereas Figure 4.5 (b) shows the frequency distribution of the inner products between the first eigenvectors of the sample covariance matrix with the actual first eigenvector of the theoretical covariance matrix $\mathbf{c_2}$.



	(-1.0, 0.75]	(-0.75, -0.50]	(-0.50, -0.25]	(-0.25, 0]	(0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1.0]
Frequency	1185	0	0	0	0	0	0	3815

(a)



	(-1.0, 0.75]	(-0.75, - 0.50]	(-0.50, - 0.25]	(-0.25, 0]	(0, 0.25]	(0.25, 0.50]	(0.50, 0.75]	(0.75, 1.0]
Frequency	581	215	223	257	330	476	837	2081

(b)

Figure 4.5: The frequency distribution of inner products between the first theoretical eigenvector from correlation matrices $\mathbf{c_1}$ and $\mathbf{c_2}$ with the first sample eigenvector.

Figure 4.5 indeed shows an interesting and useful result. First let us assumed that for inner products within the interval (-0.75. 0.75) indicate swapping and values in the range (-1, -0.75) and (0.75, 1) indicate no swapping. Figure 4.5 (a) showed clearly that all the inner products are close to ± 1. Based on the assumption, shows no swapping. On the other hand, Figure 4.5 (b) showed the inner products are distributed throughout all the intervals. Even though nearly half (41.6%) of the inner products within the (0.75, 1) interval and some (11.6%) are within the (-1, -0.75) interval, there is also a considerable percentage (46.8%) outside the two intervals. This indicates a high risk or possibility of swapping between the eigenvectors.

4.7 Discussions and Conclusions

Many multivariate methods rely on identifying the eigenvalues of a covariance matrix and then the associated eigenvalues. Some rely on picking out particular, ordered vectors. The Ratio method is such a method. For some covariance matrices, two or three eigenvalues can be similar. Thus, in sample versions of them, we may see ‘swapping’. A result is that the associated eigenvectors are not in orientations close to those of the corresponding eigenvectors of the true (theoretical) covariance matrix. Since eigenvectors are, by construction, orthogonal, this can lead to quite different directions for what is labelled ‘the *k*-th eigenvector’. Matrix $\mathbf{c_2}$ is a matrix with this property.

The effect on the power of the Ratio method will also depend on the particular aberrant value. If the potential aberrant value was in a direction lying almost equally in between the two eigenvectors of $\mathbf{c_2}$ which are liable to swap over, then the power would not be so badly affected as when the aberrant value was in a direction that was parallel to one eigenvector but orthogonal to the other that it is swapping with.

The inconsistent performance of Ratio method, depending upon the factors discussed previously, indicates the instability of the method itself. The inconsistent results which have been illustrates by two different structures of correlation matrix, shows that the Ratio method has defects and sometimes during the identification of the aberrant variable, the “wrong one” is selected. When the two largest population eigenvalues are very close together, it makes the identification of aberrant variables almost impossible.

This particular situation with regards to eigenvalue and eigenvector in principal component analysis has been discussed by a few researchers. Zhang et al. (1997) has mentioned that when two eigenvalues are too close together, the corresponding eigenvector will be hardly distinguishable. Quadrelli et al. (2005) also stated that “the magnitude of the sampling error with a single eigenvector is depends on how much it tends to mix with each of the other eigenvectors of the sample”.

One might test the eigenvectors first as to whether they are informative enough before performing the ratio calculation by looking at their 95% confidence limits (Jackson, 1993 & Mehlman et al., 1995). On the other hand, one might first determine the typical mixing between pairs of eigenvectors by looking at the ratio between corresponding eigenvalues (Quadrelli et al., 2005). In practice, the user needs to check that the eigenvalues are sufficiently distinct for them to be estimated in the correct order with possibly very small sample sizes (i.e less than 10). However, it is impractical to limit applicability of the Ratio method further in this way and we must conclude that the usage of ratio calculation as suggested by Maravelakis et al. (2002) to diagnose aberrant variable from multivariate Hotelling's T^2 control chart signal is not advisable and should not be taken.

CHAPTER 5

New Approach with Union Intersection Test

5.1 Introduction

In this chapter, we discuss a technique in Union Intersection testing and how it relates to a new approach to our key diagnosis problem. The diagnostic task starts when a multivariate Hotelling's Control chart triggers an out-of-control (OOC) signal. An OOC signal indicates a process is no longer statistically stable and it happens when an observation plotted on a multivariate control chart falls beyond the upper control limit (UCL). It indicates that the observation has deviated sufficiently far from the in-control mean value to trigger the chart alarm. Since the multivariate observation is a composite measure based on all p variables, a multivariate Hotelling's control chart can only give an indication that a process is no longer statistically stable and is unable to pinpoint which variable(s) caused the alarm.

In this chapter we shall discuss to applying a new method which embraces the Union Intersection principle to diagnosing the OOC observation, i.e. identifying the variables(s) responsible for the OOC signal. Throughout the discussion in this chapter, the in-control mean vector and covariance matrix of the process is assumed known or very well estimated. Section 5.2 gives a brief introduction to the Union Intersection Test itself following the arguments of Mardia, Kent & Bibby (1994). Section 5.3 briefly discusses the application of the Union Intersection Test in a diagnostic problem. A preliminary simulation study is outlined in 5.4 with the results presented and discussed in 5.5. Section 5.6 investigates further on the results obtained in section 5.5 by looking at the identification by variables. The chapter concludes with the discussion of the strengths and the weaknesses of the proposed approach.

5.2 Union Intersection Test

The Union Intersection principle was first introduced by Roy (1953, 1957). A test involving the breaking down of a complicated hypothesis into the intersection of simpler hypotheses is identified by Casella and Berger (1990) as union intersection testing. Union intersection testing projects data into a particular single direction and tests the hypothesis in that particular direction. The direction chosen is that which shows the greatest deviation from the null hypothesis. The validity of this procedure relies on the Cramer-Wold Theorem. Consider a random vector, **x**, which has the $N_p(\boldsymbol{\mu}, \mathbf{I})$ distribution and a non-random *p*-vector $\boldsymbol{\beta}$.

Theorem 5.1 (Cramer-Wold)

The distribution of a random *p*-vector **x** is completely determined by the set of all one-dimensional distributions of linear combinations $\boldsymbol{\beta}'\mathbf{x}$, where $\boldsymbol{\beta} \in \Re^p$ ranges through **all** fixed *p*-vectors (see, for example, Mardia, Kent & Bibby, 1994).

The theorem establishes the connection between the set of all one-dimensional projections and the multivariate distribution. In other words, it implies that a multivariate probability distribution can be defined completely by specifying the distribution of **all** its linear combinations, though not just the marginal distributions.

5.2.1 Composite and Component Hypotheses

Consider again a random vector, **x**, which has the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$ distribution and a non-random *p*-vector **β**. The mean vector, $\boldsymbol{\mu}_0$, is called the target value or the in-control mean vector and the variance covariance matrix, $\boldsymbol{\Sigma}_0$, is assumed known. The *p*-vector **x** can be written as

$$\mathbf{x} = (x_1, x_2 ..., x_p)' \qquad [5.1]$$

Then with **β** for any non-random *p*-vector:

$$\begin{aligned} y_\beta &= \boldsymbol{\beta}'\mathbf{x} \\ &= (\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)' \text{; a scalar} \qquad [5.2] \end{aligned}$$

Suppose that we wish to test the mean of a process, the hypothesis statements for the test are given as

$$\mathrm{H}_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad \mathrm{H}_0: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0. \qquad [5.3]$$

Under this composite null hypothesis H_0, by the Cramer-Wald theorem,

$$\mathbf{y}_{\boldsymbol{\beta}} \sim \mathrm{N}(\boldsymbol{\beta}'\boldsymbol{\mu}_0, \boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta}) \qquad [5.4]$$

Expression [5.4] is true for all *p*-vectors $\boldsymbol{\beta}$. The null hypothesis for every $\boldsymbol{\beta}$ is a univariate null hypothesis. It is called a component of H_0 or a component hypothesis,$\mathrm{H}_{0\boldsymbol{\beta}}$. Thus, in union intersection tests, the multivariate null hypothesis H_0 can be written as the intersection of the set of all univariate hypotheses, $\mathrm{H}_{0\boldsymbol{\beta}}$, as shown below. The H_0 represents the intersection of all null hypotheses over $\boldsymbol{\beta}$

$$\mathrm{H}_0 = \bigcap_{\boldsymbol{\beta} \in \mathrm{R}^p} \mathrm{H}_{0\boldsymbol{\beta}} \qquad [5.5]$$

5.2.2 Acceptance and Rejection Regions

The intersection sign in [5.5] indicates that all of the component hypotheses must be true in order for the composite null hypothesis H_0 to be true. Let a test on any component hypothesis $\mathrm{H}_{0\boldsymbol{\beta}}$ be carried out. Suppose that we wish to test the hypothesis $\mathrm{H}_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ so, the component null hypothesis is $\mathrm{H}_{0\boldsymbol{\beta}}: \boldsymbol{\mu}_0 = \boldsymbol{\mu}_{0\boldsymbol{\beta}}$. The common test statistic for this problem, for observation given in [5.2] and with mean vector and covariance matrix given in [5.4], is

$$\mathrm{z}_{\boldsymbol{\beta}} = \frac{\mathrm{y}_{\beta} - \boldsymbol{\beta}'\boldsymbol{\mu}_0}{\sqrt{(\boldsymbol{\beta}'\Sigma_0\boldsymbol{\beta})}} \qquad [5.6]$$

This test statistic has the standard normal distribution. The test statistic z_β needs to be sufficiently extreme to enable the hypothesis test to reject the component null hypothesis, $\mathrm{H}_{0\boldsymbol{\beta}}$. The rejection region for $\mathrm{H}_{0\boldsymbol{\beta}}$ based on z_β would be of the form

$$\mathrm{R}_{\boldsymbol{\beta}} = \{\mathrm{z}_{\boldsymbol{\beta}}: |\mathrm{z}_{\boldsymbol{\beta}}| > \mathrm{c}_0\}$$

where c_0 is some suitably chosen critical value. To simplify by removing the modulus, the rejection region for every component hypothesis can also be written as

$$R_{\boldsymbol{\beta}} = \{z_{\boldsymbol{\beta}}: z_{\boldsymbol{\beta}}^2 > c_0^2\}.$$

Since the composite null hypothesis H_0 would be true if and only if all the component hypotheses $H_{0\boldsymbol{\beta}}$ are true, it is appropriate to say that H_0 would be rejected if any of the component hypotheses is rejected. So, the rejection region for the composite null hypothesis, R can be written as

$$R = \cup_{\boldsymbol{\beta} \in R^p} R_{0\boldsymbol{\beta}} \qquad [5.7]$$

The relationship between the acceptance regions and the rejection regions which are presented by the intersection of the component hypotheses given by [5.5] and the union of the rejection regions given by [5.7] respectively, provide the basis of the union intersection strategy (Mardia, Kent & Bibby, 1994) which has been introduced earlier by Roy (1957) and will now be applied in our new approach to the diagnosis problem of the out-of-control signal triggered from a multivariate control chart.

5.2.3 Union Intersection Test Statistic

We note that [5.6] can also be expressed as,

$$z_{\boldsymbol{\beta}}^2 = \frac{\boldsymbol{\beta}'(\mathbf{x}-\boldsymbol{\mu}_0)(\mathbf{x}-\boldsymbol{\mu}_0)'\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta}} \qquad [5.8]$$

Expression [5.8] has an exact χ_1^2 distribution under $H_{0\boldsymbol{\beta}}$ and so $z_{\boldsymbol{\beta}}^2$ is the chi-squared statistic for the union intersection test (UIT).

5.2.4 Hypothesis Testing

The null hypothesisH_0 is not rejected if and only if $z_{\boldsymbol{\beta}}^2 \le c_0^2$ for all $z_{\boldsymbol{\beta}}$ or equivalently,

$$\max_{\boldsymbol{\beta}\in\mathbf{R}^p} z_{\boldsymbol{\beta}}^2 = \frac{\boldsymbol{\beta}'(\mathbf{x}-\boldsymbol{\mu}_0)(\mathbf{x}-\boldsymbol{\mu}_0)'\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta}} \le c_0^2 \qquad [5.9]$$

Thus, the component hypothesis $\mathrm{H}_{0\boldsymbol{\beta}}$ is rejected if and only if the statistic given by [5.9] is sufficiently large and exceeds c_0^2.

$$\max_{\boldsymbol{\beta}\in\mathbf{R}^p} z_{\boldsymbol{\beta}}^2 > c_0^2$$

In fact, in most cases we have pre-standardized our variance so that $\boldsymbol{\mu} = \mathbf{0}$, in which case considering a case of individual observations in a multivariate process monitoring and by assuming that $\boldsymbol{\mu}_{0\boldsymbol{\beta}} = 0$, the test statistic in [5.9] becomes;

$$\max_{\boldsymbol{\beta}\in\mathbf{R}^p} z_{\boldsymbol{\beta}}^2 = \frac{\boldsymbol{\beta}'\mathbf{x}\mathbf{x}'\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta}} \qquad [5.10]$$

The Union Intersection strategy based on [5.5] and [5.7] leads to a rejection of the composite hypothesis, H_0. Therefore, this case is clearly a maximization problem and we are able to solve it directly.

5.2.5 Maximization Problem

A Lagrange function, $\Omega(x)$ with a Lagrange multiplier, λ is introduced for this maximization problem subject to a constraint, $\boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta} = 1$. The maximization problem therefore can be set up with one Lagrange multiplier as follows

$$\text{maximize } \Omega(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}\mathbf{x}'\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta} - 1) \quad \text{with respect to } \boldsymbol{\beta}. \qquad [5.11]$$

By differentiating $\Omega(\mathbf{x})$ with respect to $\boldsymbol{\beta}$

$$\frac{\partial\{\Omega(\mathbf{x})\}}{\partial\boldsymbol{\beta}} = \frac{\partial}{\partial\boldsymbol{\beta}}\{\boldsymbol{\beta}'\mathbf{x}\mathbf{x}'\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta} - 1)\} \qquad [5.12]$$

$$= 2\mathbf{x}\mathbf{x}'\boldsymbol{\beta} - 2\lambda\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta}$$

By setting $\frac{\partial\{\Omega(\mathbf{x})\}}{\partial\boldsymbol{\beta}}$ = 0, the maximising value of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^*$ where

$$2.\mathbf{x}\mathbf{x}'\boldsymbol{\beta}^* - 2\lambda\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta}^* = 0$$

$$\therefore \quad \mathbf{x}\mathbf{x}'\boldsymbol{\beta}^* - \lambda\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta}^* = 0$$

or

$$\boldsymbol{\Sigma}_0^{-1}\mathbf{x}\mathbf{x}'\boldsymbol{\beta}^* - \lambda\boldsymbol{\beta}^* = 0 \qquad [5.13]$$

Expression [5.13] shows that $\boldsymbol{\beta}^*$ is the right eigenvector of the positive definite matrix $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'$. The maximum of $\boldsymbol{\beta}\mathbf{x}\mathbf{x}'\boldsymbol{\beta}$ given $\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta} = 1$ is attained when $\boldsymbol{\beta}^*$ is the eigenvector of $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'$ corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'$. Thus, if λ is the largest eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'$ then subject to the constraint $\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{0}}\boldsymbol{\beta} = 1$, $\max_{\beta} \boldsymbol{\beta}'\mathbf{x}\mathbf{x}'\boldsymbol{\beta} = \lambda$ (Mardia, Kent & Bibby, 1994). In other words, the maximum of objective function [5.11] is the largest eigenvalue of matrix $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'$ and is achieved at the corresponding eigenvector $\boldsymbol{\beta}^*$.

5.2.6 UIT Direction

Differentiating $\Omega(\mathbf{x})$ with respect to $\boldsymbol{\beta}$ in [5.11] shows that $\boldsymbol{\beta}$ satisfies [5.13] when

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x} \qquad [5.14]$$

By substituting [5.14] into [5.13], it can be shown that

$$\left(\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\mathbf{x}'\right).\left(\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\right) - \boldsymbol{\lambda}\left(\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\mathbf{x}\right) = 0$$

hence that

$$\lambda = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{0}})'\,\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{0}}) \qquad [5.15]$$

or in the case with $\boldsymbol{\mu}_{\mathbf{0}\boldsymbol{\beta}} = 0$ then $\lambda = (\mathbf{x})'\,\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}(\mathbf{x})$.

The UIT has an advantage where it can provide details about the **reasons** for rejection of H_0 (Mardia, Kent & Bibby, 1994). So, if H_0 is rejected, when the statistic in [5.10] is sufficiently large, then one can tell the direction of deviation is along the vector $\mathbf{\Sigma_0^{-1}x}$. We can interpret this direction by looking at the magnitude of the loadings on the individual components of the vector $\mathbf{\Sigma_0^{-1}x}$. In other words, we are not just looking in the direction of **x** (or $(\mathbf{x}-\boldsymbol{\mu}_\mathbf{0})$ for $\boldsymbol{\mu}_\mathbf{0} \neq \mathbf{0}$), but we are also taking into account the differing variances of the components of $\mathbf{\Sigma_0^{-1}}$.

5.3 Application to the Diagnosis Problem

In the diagnosis problem in statistical process monitoring for individual observations, we study the multivariate observation that triggered the signal in a multivariate control chart. The multivariate observation has to be sufficiently large to trigger the signal and this has been explained in sections 5.2.3, 5.2.4 and 5.2.5. Based on the sections 5.2.1 and 5.2.2, we can ask and find out which variables actually caused the signal. Instead of going back from a multivariate test to a univariate approach in order to identify the responsible component(s) or variable(s) that caused the out of control signal, we can go further with the multivariate test statistic itself in [5.10]. What we need to do is look at the eigenvector of the maximization problem in [5.13].

The identification of the largest deviated component(s) is proposed by looking at the loadings of the vector. This new method is known hereafter as the "largest deviation", method or in short LD, method. The identification of the aberrant variable in the LD method is based on the loadings of the components in vector **u** which is defined as

$$\mathbf{u} = \mathbf{\Sigma_0^{-1}x} \qquad [5.16]$$

5.4 Simulation Background

A few data sets of *n* individual *p*-dimensional observations, $\mathbf{x}_{p\times 1}$, were generated from the multivariate normal distribution using **R** (programming language) routine mvrnorm( ) preceded by set.seed ( ) with seed 2014. The datasets are distinguished from each other by their mean vector, later defined as contaminated mean vector, used in the mvrnorm( ) procedure.

The mean vector, $\boldsymbol{\mu}_0$, and the variance covariance matrix, $\boldsymbol{\Sigma}_0$, of the multivariate process are assumed known. A contaminant, $\boldsymbol{c}$ is introduced to selected process variable(s), which represented in a contaminated mean vector, $\boldsymbol{\mu}_c$. Typically most of the elements of $\boldsymbol{c}$ are zero with only a few elements non-zero, indicating that the out-of-control state is attributable to just those few variables.

$$\boldsymbol{\mu}_0 = \begin{pmatrix} \mu_{01} \\ \mu_{02} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{0p} \end{pmatrix}; \quad \boldsymbol{\mu}_c = \boldsymbol{\mu}_0 \pm \mathbf{c} = \boldsymbol{\mu}_0 + \begin{pmatrix} 0 \\ \cdot \\ c_k \\ \cdot \\ \cdot \\ 0 \end{pmatrix}; \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

The n individual p-dimensional observations, $\mathbf{x}_{p\times 1}$, were generated with a specific contaminated mean vector until 5000 out-of-control observations obtained. The proposed method, LD method, along with the other two diagnostic methods, DFT and Ratio, is carried out on the 5000 out-of-control observations.

5.5 Preliminary Study

The main purpose of this investigation is to study the performance of the proposed method given in Section 5.3 compared to the other two selected methods, DFT's method and the Ratio method, as given in [2.7] and [4.3] respectively. The performance of the diagnostic methods in this section is solely based on one criterion which is the percentage of correct identification given by [2.8]. The correct identification is defined as the identification of the contaminated process variables as the responsible variable for the out-of-control (OOC) signals.

This preliminary investigation can be divided into two parts. The first part employed correlation matrices used by Doganaksoy, Faltin and Tucker (1991) and the second part utilised four correlation matrices used by Das and Prakash (2008). Each part has four separate cases as listed below:

i) **Case I** - is the case when only one process variable is contaminated which is process variable 1.
ii) **Case II** – variable 1 and variable 2 are both contaminated with negative contaminants in different sizes.
iii) **Case III** – the same variables, variable 1 and variable 2 are contaminated and this time with positive contaminants in different sizes.
iv) **Case IV** - the same variables, variable 1 and variable 2 are contaminated but this time in opposite directions. One of the variables has a positive contaminant and vice versa.

In **Case I**, the performance is determined by calculating the number of times of variable 1 been identified as the most likely been the aberrant variable. The highest value of the univariate *t*-statistic in the method proposed by Doganaksoy, Faltin and Tucker (1991) will indicating that a particular process variable is most probably the responsible variable causing the out-of-control signal in the multivariate Hotelling's control chart. Whereas, in the Ratio's method and the proposed method, the highest ratio and the highest coefficient of the largest deviation vector will indicates the aberrant variable respectively. The performance of the other three cases is calculated by the total number of variable 1 or variable 2 having the highest value of ratio or coefficient of the largest deviation vector.

The first part is conducted to see whether all the tested diagnostic methods show a similar and consistent performance under equi-correlation matrices (**Part 1**) and the second part is conducted to see the performance of the diagnostic methods under two types of correlation matrix (**Part 2**) which are considered as the more realistic correlation matrices.

Part 1 - Datasets with Equi-Correlation Matrices

Four equi-correlation matrices are used in this first part of the investigation, with a unit variance and correlation values of -0.2, 0.2, 0.5 and 0.8. The combinations of the mean shifts of the variables are also taken from the study done by Das and Prakash(2008). This is to maintain the structure and the consistency of the simulation work.

Part 2- Datasets with Non-Equi Correlation Matrices

Datasets were generated with contaminants introduced into mean vector, $\boldsymbol{\mu}_0$. A similar combination of the shifts in mean as in Das and Prakash (2008) and the covariance matrices used in Doganaksoy, Faltin and Tucker (1991) and shown in Section 4.4 were used. Doganaksoy, Faltin and Tucker (1991) stated that "*the covariance matrices were carefully chosen so as to cover a wide range of possible situations*" but no further explanation provided on the covariance matrices selection. This study only chose one of the three proposed covariance matrices with positive correlations between variables. The other two covariance matrices are not chosen because a pair of the variables has correlation value equal to 0.

5.5.1 Power with Equi-Correlation Matrix

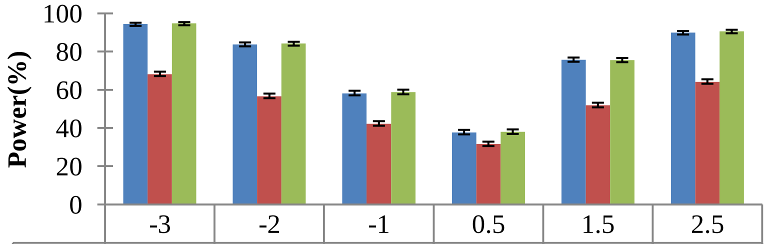
As stated before, this study used the correlation matrices used by Das and Prakash (2008) in this second part of the simulation study. The difference is Das and Prakash (2008) used 5 difference correlation matrices with ρ = -0.45, -0.2, 0.2, 0.5 and 0.8. This study only used four out of five correlation matrices used by Das and Prakash with the ρ = -0.2, 0.2, 0.5 and 0.8. It is not possible to use correlation matrix with ρ = -0.45 in this study since it gives a non positive definite matrix for p = 4. It is easy to show that the determinant of the $p \times p$ equi-correlation matrix is $[1 + (p - 1)\rho][1 - \rho]^{(p-1)}$ so we must have $\rho > \frac{-1}{(p-1)}$ otherwise the matrix is not positive definite.

Case I

There are four correlation matrices considered under this case and Figure 5.1 depicts the distribution of power of the three diagnostic methods for low correlation matrices.

The general pattern that can be seen from Figure 5.1(a) and Figure 5.1(b) is that the power for all the diagnostic methods is reduced when the shift in mean vector for variable 1 decreases in magnitude regardless of the direction of the shift. The proposed method LD has showed a good and similar result to the DFT method. Both methods have power much higher than the Ratio method. The Ratio method performed lowest throughout all the shifts in mean for both types of correlation matrix.

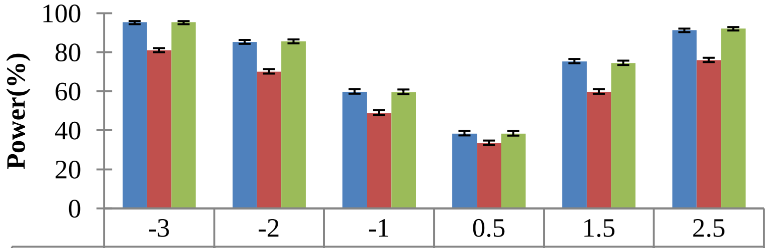
Power(%) and 2S.E.(%) of the diagnostic methods with ρ= - 0.2



	-3	-2	-1	0.5	1.5	2.5
DFT	94.5	83.8	58.1	37.7	75.7	90.0
Ratio	68.2	56.6	42.2	31.6	51.9	64.2
LD	94.8	84.1	58.7	37.9	75.5	90.6

(a)

Power(%) and 2S.E.(%) of the diagnostic methods with ρ = 0.2



	-3	-2	-1	0.5	1.5	2.5
DFT	95.3	85.2	59.7	38.3	75.3	91.2
Ratio	80.9	70.0	48.8	33.4	59.7	75.9
LD	95.3	85.5	59.5	38.2	74.4	92.1

(b)

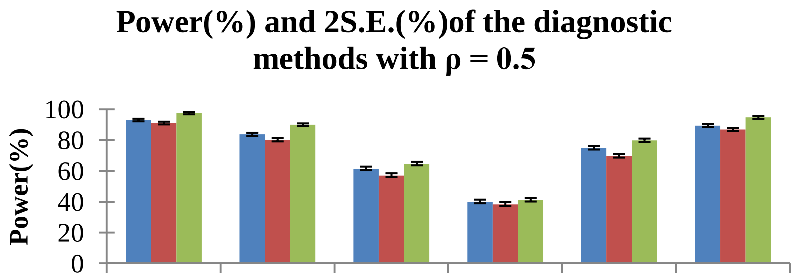
Figure 5.1: The distribution of the estimated power between the diagnostic methods when the correlation is low with respect to various contaminant values in variable 1.

The power for the LD and DFT methods in Figure 5.1(a) is similar to the power shown in Figure 5.1(b). This indicates that both methods are not affected by the sign of the low correlation. On the other hand, the Ratio method showed an appreciable difference in power between these two correlation matrices. The power of the Ratio method shown in Figure 5.1(b) is much higher than the power shown in Figure 5.1(a). This indicates that the Ratio method did not performed equally under the two correlation matrices. The power is higher when the values of the low correlation matrix are positive. Figure 5.2 shows the distribution of power between the three diagnostic methods throughout various shifts in mean vector for variable 1.

For higher correlations, Figure 5.2(a) depicts the power when the correlations between the variables are moderate positive with $\rho = 0.5$ and Figure 5.2(b) when the correlations between the variables are strong positive with $\rho = 0.8$. The general pattern that can be seen from Figure 5.2(a) and Figure 5.2(b) is the same as shown in Figure 5.1. The power of all the diagnostic methods decreased when the shift in mean vector for variable 1 is reduced in magnitude.

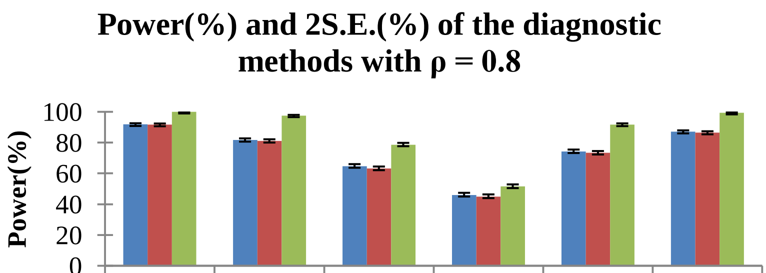
The proposed method LD has showed the highest power compared to the other two diagnostic methods in both Figure 5.2(a) and Figure 5.2(b) for every shift in mean vector for variable 1. The LD power is consistently higher in Figure 5.2(b) than in Figure 5.2(a) for every shift in mean. This indicates that the performance of the LD method improves when the correlation between the variables is stronger.

For the other two methods, both of them performed quite similarly especially when the correlation between the variables is equal to 0.8. In Figure 5.2(a), the DFT method showed slightly higher power than the Ratio method for a few shifts in mean vector such as $(1.5,0,0,0)'$ and $(2.5,0,0,0)'$. However, in Figure 5.2(b) the power of the two diagnostic methods are very close. The DFT and the Ratio methods do not share the same consistent improvement in performance with higher correlation as shown by the proposed LD method. For smaller shifts, their power is higher when the correlation between variables is strong positive, but for the intermediate and large shifts in mean, the DFT method showed a higher power when the correlation is moderate positive. However, the Ratio method does not showed any specific pattern in its performance for the intermediate and large shifts in mean.



	-3	-2	-1	0.5	1.5	2.5
DFT	93.1	83.7	61.4	40.0	74.9	89.4
Ratio	91.1	80.1	57.0	38.4	69.7	86.8
LD	97.7	90.0	64.6	41.1	79.8	94.8

(a)



	-3	-2	-1	0.5	1.5	2.5
DFT	91.8	81.6	64.7	46.0	74.2	87.0
Ratio	91.6	81.0	63.1	45.0	73.3	86.4
LD	99.9	97.5	78.6	51.5	91.7	99.3

(b)

Figure 5.2: The distribution of the estimated power between the diagnostic methods when the correlation is medium and high with respect to various contaminant values in variable 1.

Case II

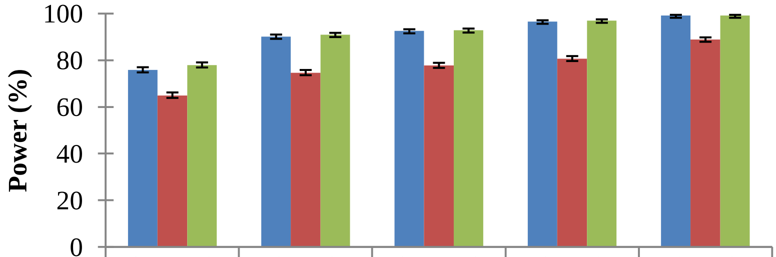
Figure 5.3 shows the distribution of power between the three diagnostic methods with respect to various combinations of shifts in mean vector with negative contaminants introduced in variables 1 and 2. Figure 5.3(a) showed the distribution of power throughout 5 combinations of shifts in mean when the correlation between variables is low negative (ρ = -0.2) and Figure 5.3(b) when it is low positive (ρ = 0.2).

In general, both figures, Figure 5.3(a) and Figure 5.3(b) show as expected, a lower power for the combination of small shifts in mean for variables 1 and 2. The power increases when at least one of the shifts increases in magnitude. Figure 5.3(a) showed that the power of the Ratio method is considerably lower than the other two methods. Figure 5.3(b) also showed the same performance pattern for the Ratio method but with smaller difference in power between it and the other two methods.

The proposed LD method has the highest power when the correlation is low negative (Figure 5.3(a)) except for the combination of shifts in mean (-3,-3). The power is the same with the power shown by the DFT method for that combination of shifts in mean. The power of the DFT is very close to the LD method for the other combinations of shifts in mean.

Figure 5.3(b) showed that the DFT performed best among the three methods when the correlation between the variables is low positive. However, the power of the LD method is quite close behind the DFT method especially for the combination of big and small shifts in mean. The power shown by both methods are higher in Figure 5.3(a) but the difference is considered small except for the combination of small shifts for the LD method. On the other hand, the Ratio method showed a higher power in Figure 5.3(b) and the power difference between the two figures for the Ratio method is appreciably large except for the combination of small shifts in mean.

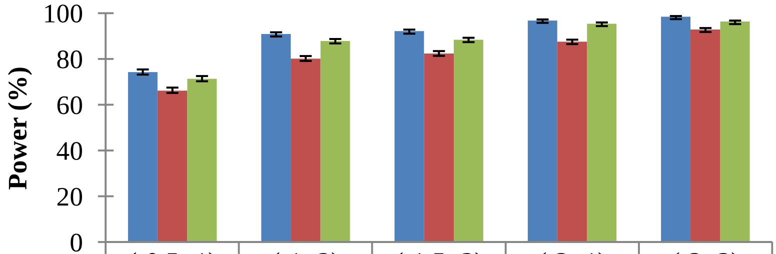
Power(%) and 2S.E.(%) of diagnostic methods with ρ = - 0.2



	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	75.8	90.1	92.5	96.6	99.2
Ratio	64.8	74.6	77.7	80.6	88.9
LD	77.9	90.9	92.8	97.0	99.2

(a)

Power(%) and 2S.E.(%) of diagnostic methods with ρ = 0.2



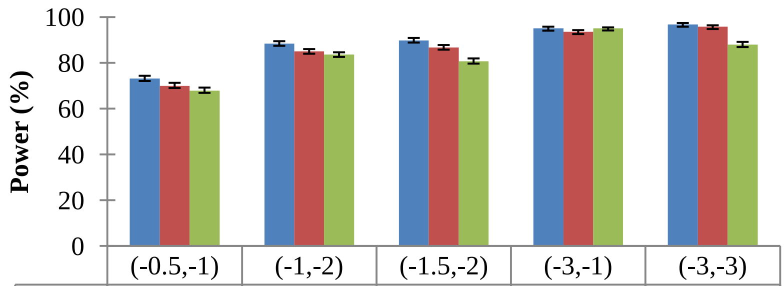
	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	74.2	90.9	92.1	96.8	98.4
Ratio	66.2	80.2	82.4	87.5	92.8
LD	71.3	87.8	88.4	95.4	96.3

(b)

Figure 5.3: The distribution of the estimated power between the diagnostic methods for low correlation matrices with respect to various shifts in mean with negative contaminant values in variable 1 and variable 2.

Figure 5.4(a) and (b) shows a power distribution between the three diagnostic methods when the correlation between the variables is moderate or strongly positive. The proposed method LD has shown the highest power for the combination of shifts in mean (-3,-1) but weaker performance than both other methods for other shifts combinations.

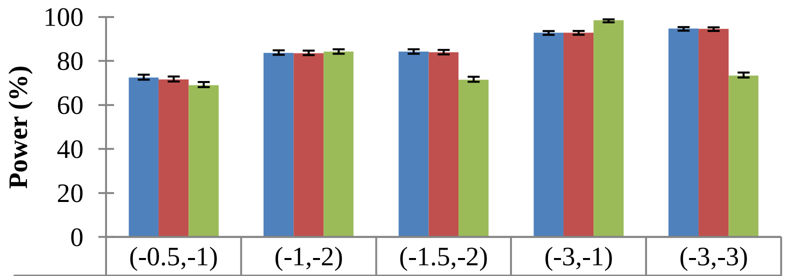
Power (%) and 2S.E.(%) of diagnostic methods with ρ = 0.5



	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	73.1	88.4	89.8	95.0	96.7
Ratio	70.0	84.9	86.7	93.5	95.7
LD	67.9	83.6	80.6	95.1	87.9

(a)

Power (%) and 2S.E. (%) of diagnostic methods with ρ = 0.8



	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	72.4	83.7	84.2	92.8	94.7
Ratio	71.6	83.6	83.9	92.8	94.6
LD	69.0	84.2	71.5	98.5	73.4

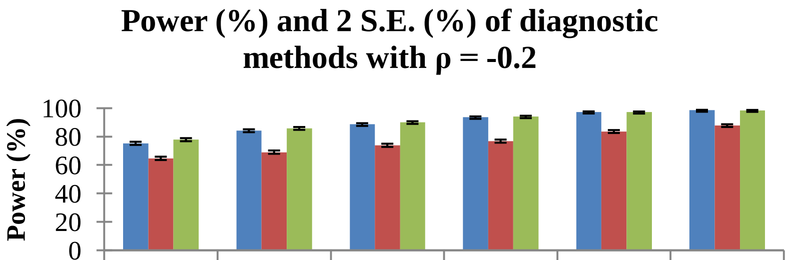
(b)

Figure 5.4: The distribution of the estimated power between the diagnostic methods for medium and high correlation matrix with respect to various combinations of shifts in mean with negative contaminant values in variable 1 and variable 2.

This indicates that the LD method performed better when the correlation between the variables is low and the power is more or less similar regardless whether the shifts are in accordance or not in accordance with the correlation structure. When the correlation between variables increased positively, the LD method showed a very good performance when the shifts are combination of large and small values. The power dropped considerably from moderate to strong positive correlation when the shifts are very close to each other in magnitude.

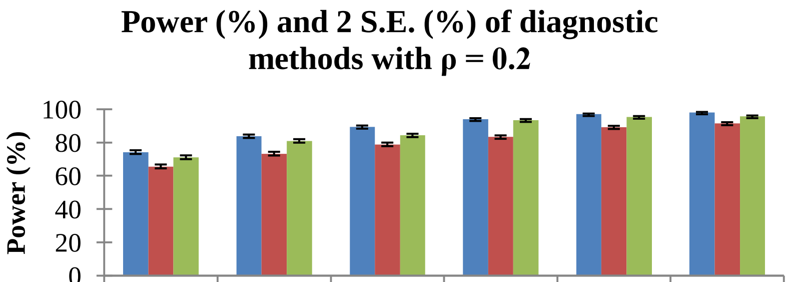
Case III

Figure 5.5 shows the distribution of power between the three diagnostic methods with respect to various combinations of shifts in mean vector with negative contaminants introduced in variable 1 and variable 2. Figure 5.5(a) showed the distribution of power throughout 5 combinations of shifts in mean when the correlation between variables is low negative (ρ = -0.2), Figure 5.5(b) when the correlation between variables is low positive (ρ = 0.2).



	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	75.2	84.1	88.6	93.5	97.3	98.5
Ratio	64.6	69.0	73.8	76.8	83.6	87.8
LD	77.8	85.8	90.0	94.0	97.2	98.4

(a)



	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	74.2	83.8	89.4	94.0	97.0	98.0
Ratio	65.5	73.2	78.8	83.3	89.2	91.5
LD	71.0	81.0	84.3	93.4	95.3	95.7

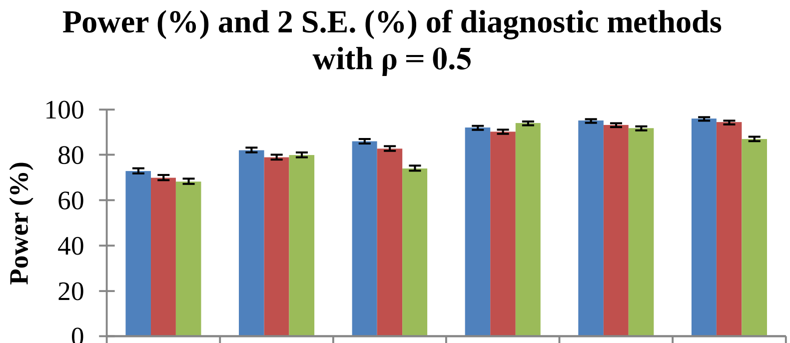
(b)

Figure 5.5: The distribution of the estimated power between the diagnostic methods for low correlation matrix with respect to various shifts in mean with positive contaminant values in variable 1 and variable 2.

In general, Figure 5.5 again shows that the power of all diagnostic methods is lower when the shifts are smaller. The Ratio method showed the lowest power throughout all combinations of shifts in mean. The power of the Ratio method is much lower than the other two methods when the correlation between the variables is low negative as shown in Figure 5.5(a). In Figure 5.5(a), the proposed LD method has a very good performance throughout almost all combinations of shifts in mean. The power is slightly better or at least almost equal to the power showed by the DFT method. Whereas in Figure 5.5(b), the power of the LD method is slightly lower than the power showed by the DFT method except for combination of shifts ((1.5, 1.5). The power of the LD method is considered appreciably lower than the power showed by the DFT method for that combination of shifts in mean.

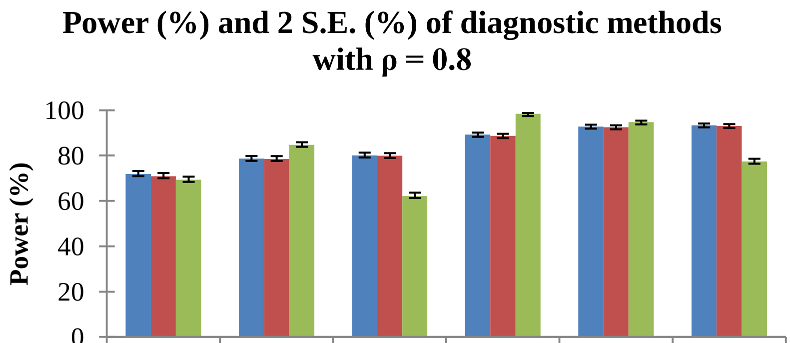
Figure 5.6(a) shows that the LD method has the highest power for combination of shifts in mean (2.5, 0.5) and the power is increased and much higher for the same shifts when the correlation between the variables is strong positive. For the combination of shifts (1.5, 0.5) in Figure 5.6(a), the power of the LD method is quite close to the highest power given by the DFT method. However, the power of the LD method increased higher and surpassed the power showed by the DFT method when the correlation between the variables is strong positive as shown in Figure 5.6(b). It looks as though the LD method performs well when the shifts are combinations of big and small shifts or intermediate and small shifts in mean.

However, when the shifts are very close together in magnitude such as combinations of shifts (1.5, 1.5) and (3, 2.5), the power of the LD method dropped as shown in Figure 5.6(a) and it dropped more when the correlation between the variables is strong positive which is shown by Figure 5.6(b).



	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	72.8	82.1	86.0	92.0	95.1	96.1
Ratio	69.8	78.9	82.8	90.3	93.2	94.4
LD	68.2	79.9	74.0	94.0	91.8	87.0

(a)



	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	71.9	78.6	80.1	89.2	92.8	93.4
Ratio	70.9	78.5	79.9	88.6	92.5	93.1
LD	69.3	84.8	62.2	98.3	94.7	77.3

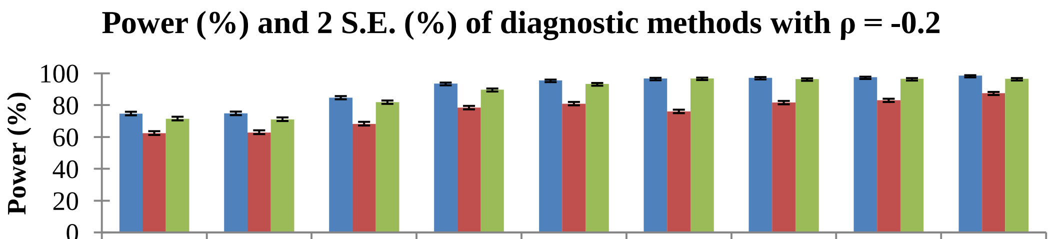
(b)

Figure 5.6: The distribution of the estimated power between the diagnostic methods for medium and high correlation matrix with respect to various shifts in mean with positive contaminant values in variable 1 and variable 2.

Case IV

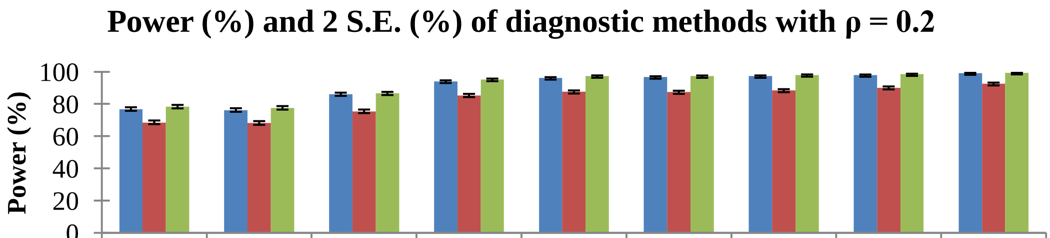
Figure 5.7 showed the power distribution for the three diagnostic methods when the shifts are in opposite directions. In general, Figure 5.7 showed that all the diagnostic methods have a lower power when the shifts are small or when the shifts are combination of intermediate and small shifts. The Ratio method performed worst among the three diagnostic methods for both low correlation matrices.

The proposed method LD and the DFT method showed a consistently very high power for the rest of the combinations of shifts in mean in Figure 5.7. The DFT method showed slightly higher power than the LD method in Figure 5.7(a) except for combination (-3, 0.5). On the other hand, the LD method showed slightly higher power for all the combinations of shifts in mean in Figure 5.7(b). However, the power difference between the two methods is considered small (< 5%) in Figure 5.7(a) and very small (< 2%) in Figure 5.7(b). So it can be said that both diagnostic methods showed a similar good performance with respect to the correlation matrices. In terms of the power difference between the low correlation matrices, the DFT method showed a very similar power for both correlation matrices. But for the LD method, the power is appreciably higher with low positive correlation matrix when the shifts small.



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3.-1)	(-3,1.5)	(-3,2.5)
DFT	74.7	74.8	84.7	93.5	95.5	96.7	97.2	97.4	98.5
Ratio	62.3	62.9	68.2	78.4	80.9	76.0	81.6	83.0	87.4
LD	71.5	71.1	81.8	89.6	93.2	96.8	96.3	96.5	96.5

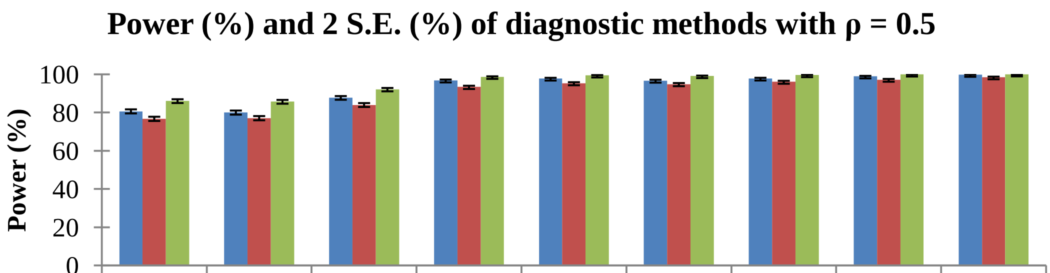
(a)



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3,-1)	(-3,1.5)	(-3,2.5)
DFT	76.8	76.2	86.1	94.0	96.2	96.8	97.3	98.0	99.2
Ratio	68.5	68.1	75.4	85.3	87.5	87.3	88.4	90.1	92.6
LD	78.3	77.6	86.6	95.2	97.3	97.3	98.0	98.5	99.4

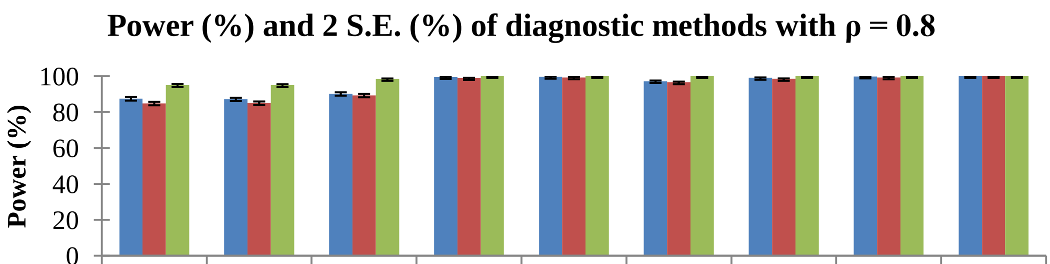
(b)

Figure 5.7: The distribution of the estimated power between the diagnostic methods for low correlation matrix with respect to various shifts in mean with positive and negative contaminant values in variable 1 and variable 2.



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3,-1)	(-3,1.5)	(-3,2.5)
DFT	80.6	80.0	87.7	96.8	97.8	96.7	97.8	98.9	99.7
Ratio	76.7	77.0	84.0	93.4	95.3	94.8	96.1	97.2	98.5
LD	86.0	85.7	92.1	98.7	99.4	99.1	99.6	99.9	100.0

(a)



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3,-1)	(-3,1.5)	(-3,2.5)
DFT	87.4	87.1	90.2	99.4	99.7	97.2	99.1	99.8	100.0
Ratio	84.8	84.9	89.3	98.8	99.3	96.5	98.4	99.3	100.0
LD	95.0	94.9	98.4	100.0	100.0	100.0	100.0	100.0	100.0

(b)

Figure 5.8: The distribution of the estimated power between the diagnostic methods for medium and high correlation matrix with respect to various shifts in mean with positive and negative contaminant values in variable 1 and variable 2.

Figure 5.8 shows that all the diagnostic methods improved considerably high in power compared to the power showed in Figure 5.7 for the combination of small shifts and the combination of small and intermediate shifts. The Ratio method is also improved and shows a very high power for the rest of the combinations of shifts in mean and is quite close to the power shown by the DFT method.

The proposed method LD has the highest power among the three methods throughout all the combinations of shifts in mean. The power is much better when the correlation between the variables is strong positive. This indicates that the proposed method LD is very good in detecting the correct aberrant variable(s) when the shifts are not in accordance to the correlation structure of the variables. However, the LD method is also able to detect the correct aberrant variable(s) when the shifts are in accordance to the correlation structure which is low negative, with a high percentage in power as shown in Figure 5.7(a).

5.5.2 Power with Non-equi Correlation Matrix

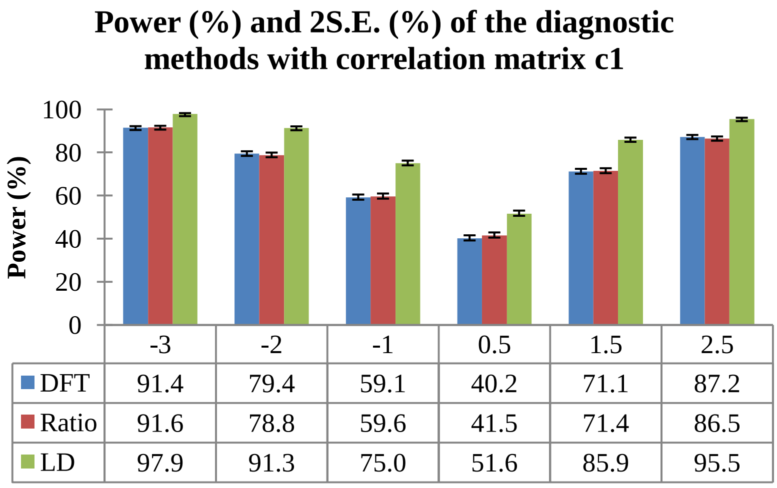
Performance as defined previously is the total percentage of the correct identification of the aberrant variable(s). Figures 5.9 to 5.12 represent the power of the three diagnostic methods when the correlations between the variables are either all positive or when the correlations are mixed in sign.

Case I

The proposed method, LD, showed the highest power among the three diagnostic methods throughout all the contaminant values in variable 1 regardless of whether the correlation matrix is all positive in values (Figure 5.9(a)) or when the correlation matrix consist of mixed sign values (Figure 5.9(b)).

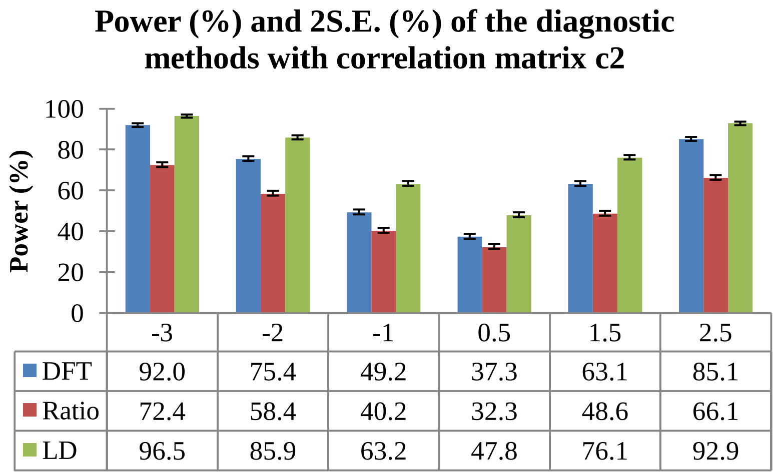
A general pattern shown in Figure 5.9(a) and Figure 5.9(b) is that the power is lower when the shift in mean is small. The power is increasing when the shift in mean becomes larger regardless of the direction of the shift in mean. This is true for all the diagnostic methods.

The power in Figure 5.9(a) is noticeably higher than the power in Figure 5.9(b) throughout all the contaminant values in variable 1 for all the diagnostic methods. It indicates that the diagnostic methods can detect variable 1 as the aberrant variable better when the correlation matrix is all positive in values.



	-3	-2	-1	0.5	1.5	2.5
DFT	91.4	79.4	59.1	40.2	71.1	87.2
Ratio	91.6	78.8	59.6	41.5	71.4	86.5
LD	97.9	91.3	75.0	51.6	85.9	95.5

(a)



	-3	-2	-1	0.5	1.5	2.5
DFT	92.0	75.4	49.2	37.3	63.1	85.1
Ratio	72.4	58.4	40.2	32.3	48.6	66.1
LD	96.5	85.9	63.2	47.8	76.1	92.9

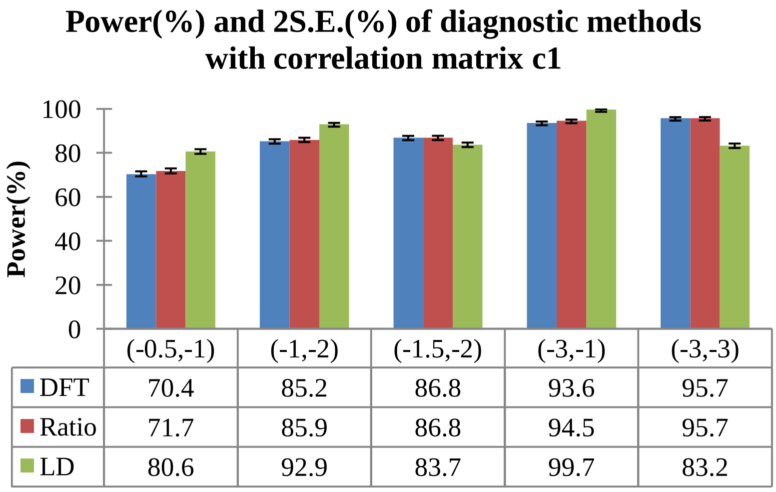
(b)

Figure 5.9: The distribution of the estimated power between diagnostic methods DFT, Ratio and LD with respect to the contaminant values in variable 1.

In Figure 5.9(a), the DFT and the Ratio method show similar power, but rather lower than the power shown by the LD method. In Figure 5.9(b), an appreciable drop in power for the Ratio method is shown. This is not surprising for correlation matrix c2 as the drawback of the Ratio method has already been discussed in the previous chapter.

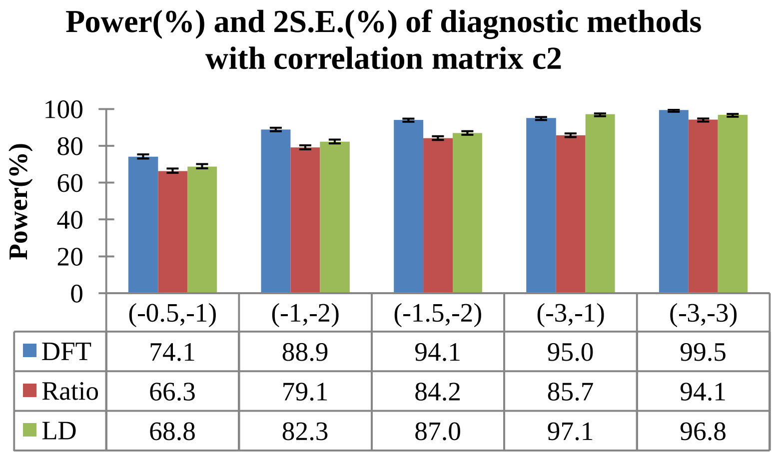
Case II

Figure 5.10(a) shows that the proposed method LD has the highest power for three combination of shifts in mean which are (-0.5, -1), (-1,-2) and (-3,-1). There is an obvious dropped in power when the shifts in mean for variable 1 and variable 2 are equal. The reduction in power is not large, and results are comparable to the other two methods when the shifts are approximately equal as shown in combination of shifts in mean (-1.5, -2).



	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	70.4	85.2	86.8	93.6	95.7
Ratio	71.7	85.9	86.8	94.5	95.7
LD	80.6	92.9	83.7	99.7	83.2

(a)



	(-0.5,-1)	(-1,-2)	(-1.5,-2)	(-3,-1)	(-3,-3)
DFT	74.1	88.9	94.1	95.0	99.5
Ratio	66.3	79.1	84.2	85.7	94.1
LD	68.8	82.3	87.0	97.1	96.8

(b)

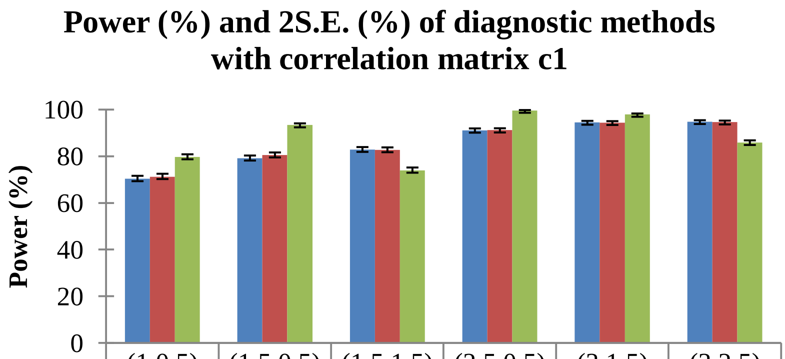
Figure 5.10: The distribution of the estimated power between diagnostic methods DFT, Ratio and LD with respect to various shifts in means with negative contaminant values.

In Figure 5.10(b), LD has the highest power for the shifts in mean (-3,-1) only. The proposed method LD, performed better when one of the shifts in mean is much larger in value than the other shift. Although the LD method does not perform best for the combination (-3,-3), the power is much higher compared to the power obtained under correlation matrix c1. The proposed method LD performed well with correlation matrix c1 as long as the shifts in mean are not equal or large and approximately shifts. It also performed best for a combination of large and small shifts in mean vector for both correlation matrices.

Case III

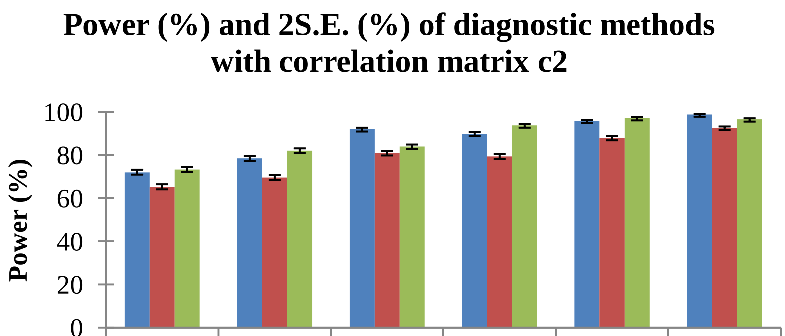
Figure 5.11(a) shows that the proposed method LD has the highest power in all of the combinations of shifts in mean vector except for the combinations (1.5, 1.5) and (3, 2.5). The power of the LD method is very high (> 90%) when the larger shift is at least double in magnitude to the smaller shift except for the combination of small shifts. There is an obvious drop in power compared to the other two methods when the shifts in mean for variable 1 and variable 2 are equal or approximately equal except for the combination of small shifts in mean vector. However, even the lower power is still at an acceptable level. The LD method in Figure 5.11(b) showed similar patterns to Figure 5.11(a) but with higher power for the combinations of shifts in mean (1.5, 1.5) and (3, 2.5).

The DFT and Ratio method showed very similar power in Figure 5.11(a). However, Figure 5.11(b) showed the Ratio method as having the lowest power among the three diagnostic methods. This is also not surprising for the reason discussed in the previous chapter.



	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	70.3	79.1	82.8	91.1	94.5	94.8
Ratio	71.2	80.5	82.7	91.2	94.4	94.6
LD	79.7	93.4	73.9	99.6	97.9	85.8

(a)



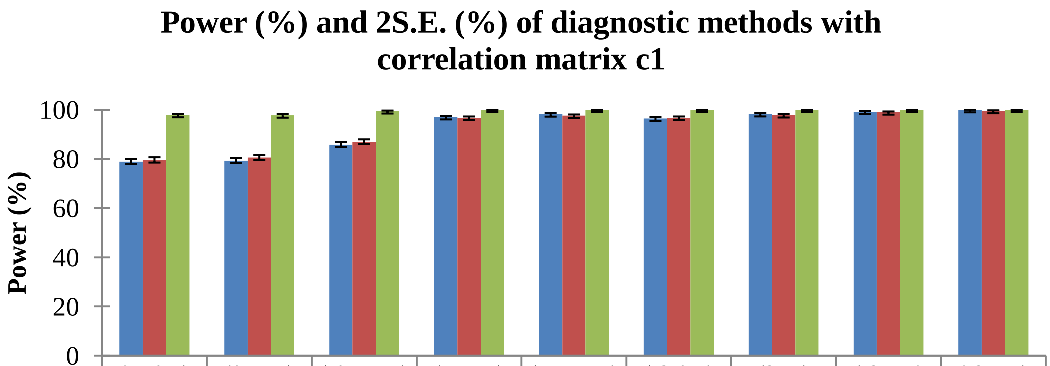
	(1,0.5)	(1.5,0.5)	(1.5,1.5)	(2.5,0.5)	(3,1.5)	(3,2.5)
DFT	72.0	78.4	91.9	89.8	95.8	98.8
Ratio	65.1	69.5	80.8	79.3	87.9	92.5
LD	73.3	82.0	83.9	93.7	97.1	96.5

(b)

Figure 5.11: The distribution of the estimated power between diagnostic methods DFT, Ratio and LD with respect to various shifts in means with positive contaminant values.

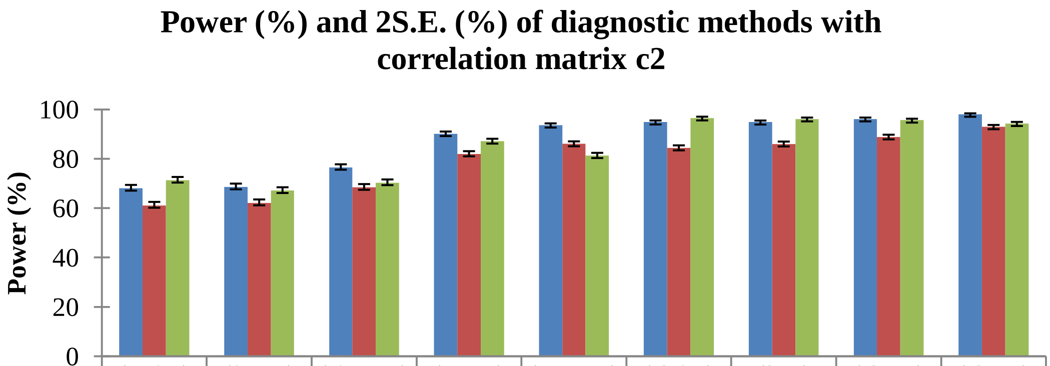
Case IV

The proposed method LD has consistently showed the highest power among the three diagnostic methods throughout all the combinations of the shifts in mean vector for variable 1 and variable 2 in Figure 5.12(a). The obvious difference in power is shown for the smaller shifts where the LD method power is at least 12% higher than the other two methods.



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3,-1)	(-3,1.5)	(-3,2.5)
DFT	78.8	79.3	85.8	97.1	98.2	96.5	98.3	99.2	99.9
Ratio	79.5	80.6	87.0	96.8	97.6	96.8	97.8	99.0	99.6
LD	97.9	97.8	99.5	100.0	100.0	100.0	100.0	100.0	100.0

(a)



	(-1.0.5)	(0.5,-1)	(-0.5,1.5)	(-2,1.5)	(-1.5,2.5)	(-3,0.5)	(3,-1)	(-3,1.5)	(-3,2.5)
DFT	68.1	68.6	76.5	90.2	93.6	94.9	94.8	96.1	98.0
Ratio	61.1	62.1	68.4	82.0	86.0	84.4	86.0	88.8	92.9
LD	71.3	67.1	70.3	87.1	81.3	96.5	96.1	95.6	94.2

(b)

Figure 5.12: The distribution of the estimated power between diagnostic methods DFT, Ratio and LD with respect to various shifts in means with positive and negative contaminant values.

The LD method does not show similar high power in Figure 5.12(b) but again shows the highest power for combinations of large shift with small shift. The Ratio method consistently shows the lowest power among the three methods except for the shifts (-1.5, 2.5), and again much lower power compared to that in Figure 5.12(a).

5.5.3 Discussions

Tables 5.1 to 5.4 show the estimated power difference (%) between the LD method and the other two diagnostic methods. The standard error (%) of each estimated power difference is also given to determine the significance of the power difference between the diagnostic methods. The tabulated result of the power comparison is presented with respect to the correlation between the variables. The first two sub-tables in each table present the estimated power difference (%) between the diagnostic methods for the non equi-correlation matrices, c1 and c2. The two sub-tables in the middle present the estimated power difference for low equi-correlation matrices with ρ = -0.2 and iρ = 0.2. Whereas, the two sub-tables at the bottom present the comparison between the diagnostic methods for equi-correlation matrices too, but with moderate and strong positive correlation between the variables (ρ = 0.5 and ρ = 0.8).

Table 5.1 shows the estimated power difference (%) for Case I when only one variable mean is shifted in the mean vector. By looking at the estimated power difference (%), in general the power of the LD method is significantly higher than the power shown by the DFT method except for the low correlation matrices and for one combination with small shifts in moderate positive correlation. The proposed LD method has similar power with the DFT method for the low correlation matrices regardless of the sign of the correlation coefficient. Even here there is usually a difference in power in favour of LD though the magnitude is small and not significant. As we move down Table 5.1, consider increasing positive correlation, we see that method LD begins to outperform the method DFT with increasing difference in their estimated power.

The LD method also showed a significantly higher power than the Ratio method and in fact the LD power is higher than the Ratio method for all the correlation matrices used. Their power difference (%) is much higher in magnitude compared to the power difference (%) between the LD and the DFT methods in Table 5.1 (b), Table 5.1 (c) and Table 5.1 (d).

Table 5.1: The estimated power difference (%) and the corresponding estimated two standard errors (%) between the proposed method LD with the other two methods for various shifts in mean of variable 1.

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	6.4 (0.9)(+)	6.3 (0.9) (+)
-2	11.9 (1.4)(+)	12.5 (1.4) (+)
-1	15.9 (1.9)(+)	15.4 (1.9) (+)
0.5	11.4 (2.0)(+)	10.1 (2.0) (+)
1.5	14.8 (1.6)(+)	14.5 (1.6) (+)
2.5	8.3 (1.1)(+)	9.1 (1.1) (+)

(a) c1

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	4.5 (0.9) (+)	24.1 (1.4) (+)
-2	10.5 (1.6) (+)	27.5 (1.7) (+)
-1	14.0 (2.0) (+)	23.0 (1.9) (+)
0.5	10.5 (2.0) (+)	15.5 (1.9) (+)
1.5	12.9 (1.8) (+)	27.5 (1.9) (+)
2.5	7.7 (1.2) (+)	26.7 (1.5) (+)

(b) c2

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	0.4 (0.9)	26.6 (1.5) (+)
-2	0.3 (1.5)	27.5 (1.7) (+)
-1	0.6 (2.0)	16.5 (2.0) (+)
0.5	0.2 (1.9)	6.4 (1.9) (+)
1.5	-0.2 (1.7)	23.6 (1.9) (+)
2.5	0.7 (1.2)	26.4 (1.6) (+)

(c) $\rho = -0.2$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	0.0 (0.8)	14.4 (1.3) (+)
-2	0.3 (1.4)	15.5 (1.6) (+)
-1	-0.2 (2.0)	10.7 (2.0) (+)
0.5	-0.1 (1.9)	4.9 (1.9) (+)
1.5	-0.9 (1.7)	14.7 (1.9) (+)
2.5	0.8 (1.1)	16.2 (1.4) (+)

(d) $\rho = 0.2$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	4.6 (0.8) (+)	6.6 (0.9) (+)
-2	6.2 (1.3) (+)	9.8 (1.4) (+)
-1	3.2 (1.9) (+)	7.6 (1.9) (+)
0.5	1.2 (2.0)	2.8 (2.0) (+)
1.5	5.0 (1.7) (+)	10.2 (1.7) (+)
2.5	5.4 (1.1) (+)	8.1 (1.1) (+)

(e) $\rho = 0.5$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
-3	8.1 (0.8) (+)	8.3 (0.8) (+)
-2	15.9 (1.2) (+)	16.5 (1.2) (+)
-1	14.0 (1.8) (+)	15.5 (1.8) (+)
0.5	5.5 (2.0) (+)	6.5 (2.0) (+)
1.5	17.5 (1.5) (+)	18.4 (1.5) (+)
2.5	12.3 (1.0) (+)	12.9 (1.0) (+)

(f) $\rho = 0.8$

(+) the estimated power of the LD method is significantly different than the other method, the LD power is higher

Table 5.2: The estimated power difference (%) and the corresponding estimated two standard errors (%) between the proposed method LD with the other two methods for various shifts in mean in variable 1 and variable 2 with positive contaminants.

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	9.4 (1.7) (+)	8.5 (1.7) (+)
(1.5,0.5)	14.2 (1.3) (+)	12.9 (1.3) (+)
(1.5,1.5)	-8.9 (1.6) (-)	-8.8 (1.6) (-)
(2.5,0.5)	8.5 (0.8) (+)	8.4 (0.8) (+)
(3,1.5)	3.4 (0.8) (+)	3.5 (0.8) (+)
(3,2.5)	-9.0 (1.2) (-)	-8.8 (1.2) (-)

(a) c1

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	1.3 (1.8)	8.1 (1.8) (+)
(1.5,0.5)	3.7 (1.6) (+)	12.6 (1.7) (+)
(1.5,1.5)	-8.1 (1.3) (-)	3.0 (1.5) (+)
(2.5,0.5)	3.9 (1.1) (+)	14.4 (1.3) (+)
(3,1.5)	1.3 (0.7) (+)	9.2 (1.0) (+)
(3,2.5)	-2.3 (0.6) (-)	4.0 (0.9) (+)

(b) c2

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	2.6 (1.7) (+)	13.2 (1.8) (+)
(1.5,0.5)	1.7 (1.4) (+)	16.8 (1.6) (+)
(1.5,1.5)	1.4 (1.2) (+)	16.2 (1.5) (+)
(2.5,0.5)	0.5 (1.0)	17.3 (1.4) (+)
(3,1.5)	-0.1 (0.7)	13.7 (1.1) (+)
(3,2.5)	-0.1 (0.5)	10.6 (1.0) (+)

(c) $\rho = -0.2$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	-3.1 (1.8) (-)	5.5 (1.9) (+)
(1.5,0.5)	-2.8 (1.5) (-)	7.7 (1.7) (+)
(1.5,1.5)	-5.1 (1.3) (-)	5.4 (1.5) (+)
(2.5,0.5)	-0.6 (1.0)	10.1 (1.3) (+)
(3,1.5)	-1.6 (0.8) (-)	6.2 (1.1) (+)
(3,2.5)	-2.3 (0.7) (-)	4.2 (1.0) (+)

(d) $\rho = 0.2$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	-4.6 (1.8) (-)	-1.6 (1.8)
(1.5,0.5)	-2.2 (1.6) (-)	1.0 (1.6)
(1.5,1.5)	-12.0 (1.6) (-)	-8.7 (1.6) (-)
(2.5,0.5)	2.0 (1.0) (+)	3.8 (1.1) (+)
(3,1.5)	-3.3 (1.0) (-)	-1.5 (1.1) (-)
(3,2.5)	-9.0 (1.1) (-)	-7.4 (1.2) (-)

(e) $\rho = 0.5$

Shift in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(1,0.5)	-2.5 (1.8) (-)	-1.6 (1.8)
(1.5,0.5)	6.2 (1.5) (+)	6.3 (1.5) (+)
(1.5,1.5)	-17.9 (1.8) (-)	-17.7 (1.8) (-)
(2.5,0.5)	9.2 (1.0) (+)	9.7 (1.0) (+)
(3,1.5)	1.9 (1.0) (+)	2.2 (1.0) (+)
(3,2.5)	-16.0 (1.4) (-)	-15.7 (1.4) (-)

(f) $\rho = 0.8$

(+) the estimated power of the LD method is significantly different than the other method, the LD power is higher

(-) the estimated power of the LD method is significantly different than the other method, the LD power is lower

Table 5.2 show that the power of the LD method is significantly higher than the DFT methods when the shifts are a combination of large and small shifts such as the combination (2.5, 0.5), for c1, c2, and moderate and strong positive correlation matrices. The power of the LD method also significantly higher than the DFT method for some cases with respect to the correlation matrix such as small shifts for c1 and combination of small shifts and intermediate and small shift for negative low equi-correlation matrix. The power is similar between the LD and DFT method when the correlations are low for that particular combination of shifts in mean. The power of the LD method is significantly higher than the Ratio method throughout all combinations of shifts in mean in Table 5.2(b), Table 5.2(c) and Table 5.2(d). For the other correlation matrices, the power is significantly lower than the Ratio method for combination of large shifts and intermediate shifts.

Table 5.3: The estimated power difference (%) and the corresponding estimated two standard errors (%) between the proposed method LD with the other two methods for various shifts in mean in variable 1 and variable 2 with negative contaminants.

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	10.2 (1.7) (+)	8.9 (1.7) (+)
(-1,-2)	7.7 (1.2) (+)	7.0 (1.2) (+)
(-1.5,-2)	-3.1 (1.4) (-)	-3.2 (1.4) (-)
(-3,-1)	6.1 (0.7) (+)	5.2 (0.7) (+)
(-3,-3)	-12.5 (1.2) (-)	-12.5 (1.2) (-)

(a) c1

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	-5.3 (1.8) (-)	2.4 (1.9) (+)
(-1,-2)	-6.6 (1.4) (-)	3.2 (1.6) (+)
(-1.5,-2)	-7.1 (1.2) (-)	2.8 (1.4) (+)
(-3,-1)	2.1 (0.8) (+)	11.4 (1.1) (+)
(-3,-3)	-2.7 (0.5) (-)	2.6 (0.8) (+)

(b) c2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	2.1 (1.7) (+)	13.0 (1.8) (+)
(-1,-2)	0.8 (1.2)	16.3 (1.5) (+)
(-1.5,-2)	0.3 (1.0)	15.1 (1.4) (+)
(-3,-1)	0.4 (0.7)	16.4 (1.2) (+)
(-3,-3)	0.0 (0.4)	10.3 (0.9) (+)

(c) ρ = -0.2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	-2.9 (1.8) (-)	5.1 (1.9) (+)
(-1,-2)	-3.0 (1.2) (-)	7.6 (1.5) (+)
(-1.5,-2)	-3.7 (1.2) (-)	6.0 (1.4) (+)
(-3,-1)	-1.4 (0.8) (-)	7.9 (1.1) (+)
(-3,-3)	-2.2 (0.6) (-)	3.5 (0.9) (+)

(d) ρ = 0.2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	-5.2 (1.8) (-)	-2.1 (1.9) (-)
(-1,-2)	-4.8 (1.4) (-)	-1.4 (1.5)
(-1.5,-2)	-9.2 (1.4) (-)	-6.1 (1.5) (-)
(-3,-1)	0.1 (0.9)	1.6 (0.9) (+)
(-3,-3)	-8.9 (1.1) (-)	-7.9 (1.1) (-)

(e) ρ = 0.5

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-0.5,-1)	-3.4 (1.8) (-)	-2.5 (1.8) (-)
(-1,-2)	0.5 (1.5)	0.6 (1.5)
(-1.5,-2)	-12.7 (1.6) (-)	-12.4 (1.6) (-)
(-3,-1)	5.7 (0.8) (+)	5.7 (0.8) (+)
(-3,-3)	-21.3 (1.4) (-)	-21.2 (1.4) (-)

(f) ρ = 0.8

(+) the estimated power of the LD method is significantly different than the other method, the LD power is higher

(-) the estimated power of the LD method is significantly different than the other method, the LD power is lower

Table 5.3 shows similar results to Table 5.2. Both shows shifts in the same directions for variable 1 and variable 2. The power of the LD method is still significantly higher than the Ratio method throughout all the combinations of shifts in mean for the same correlation matrices (Table 5.3(b), Table 5.3(c) and Table 5.3(d)). For the combination of large shifts and intermediate shifts in the mean vector, the LD power is significantly lower than the Ratio method for the same correlation matrices too (Table 5.3(a), Table 5.3(e) and Table 5.3(f)).

The LD method has a significantly higher power (%) than the DFT method for the combination of large and small shifts (Table 5.3 (a), Table 5.3(b) and Table 5.3(f)) and combination of small shifts and intermediate and small shifts for correlation matrix c1. For the other combinations of shifts in mean, the LD method has a similar or significantly lower power than the DFT method.

Table 5.4 showed the power of the diagnostic methods for the shifts in the opposite directions. The LD method consistently has showed a significantly higher power than the Ratio method throughout almost all combinations of shifts in mean for all the correlation matrices used. There is only one combination of shifts in Table 5.4 (b) that showed the power of the LD method is significantly lower than the Ratio method. In a few cases of large shifts, the power of the two methods is similar.

The LD method has significantly higher power than the DFT method in almost all combinations of shifts in mean in Table 5.4(a), Table 5.4(e) and Table 5.4(f). For the large shifts in mean vector, the two methods performed equally. For the negative low correlation (Table 5.4(c)), the LD method is significantly lower than the DFT method in all combinations of shifts in mean except for combination (-3, 0.5) in which the power of the two methods did not differ significantly. The LD method always has a significantly higher power than the other two methods for the combination of large and small shifts regardless of the correlation between the variables.

Table 5.4: The estimated power difference (%) and the corresponding estimated two standard errors (%) between the proposed method LD with the other two methods for various shifts in mean in variable 1 and variable 2 with negative and positive contaminants.

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	19.1 (1.2) $^{(+)}$	18.4 (1.2) $^{(+)}$
(0.5,-1)	18.5 (1.2) $^{(+)}$	17.2 (1.2) $^{(+)}$
(-0.5,1.5)	13.6 (1.0) $^{(+)}$	12.5 (1.0) $^{(+)}$
(-2,1.5)	2.9 (0.5) $^{(+)}$	3.2 (0.5) $^{(+)}$
(-1.5,2.5)	1.8 (0.4) $^{(+)}$	2.4 (0.4) $^{(+)}$
(-3,0.5)	3.5 (0.5) $^{(+)}$	3.2 (0.5) $^{(+)}$
(3,-1)	1.7 (0.4) $^{(+)}$	2.2 (0.4) $^{(+)}$
(-3,1.5)	0.8 (0.2) $^{(+)}$	1.0 (0.3) $^{(+)}$
(-3,2.5)	0.1 (0.1)	0.4 (0.2) $^{(+)}$

(a) c1

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	3.3 (1.8) $^{(+)}$	10.2 (1.9) $^{(+)}$
(0.5,-1)	-1.5 (1.9)	5.0 (1.9) $^{(+)}$
(-0.5,1.5)	-6.2 (1.8) $^{(-)}$	1.9 (1.8) $^{(+)}$
(-2,1.5)	-3.0 (1.3) $^{(-)}$	5.2 (1.4) $^{(+)}$
(-1.5,2.5)	-12.4 (1.3) $^{(-)}$	-4.8 (1.5) $^{(-)}$
(-3,0.5)	1.6 (0.8) $^{(+)}$	12.1 (1.2) $^{(+)}$
(3,-1)	1.3 (0.8) $^{(+)}$	10.2 (1.1) $^{(+)}$
(-3,1.5)	-0.5 (0.8)	6.8 (1.1) $^{(+)}$
(-3,2.5)	-3.7 (0.8) $^{(-)}$	1.3(1.0) $^{(+)}$

(b) c2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	-3.2 (1.8) $^{(-)}$	9.2 (1.9) $^{(+)}$
(0.5,-1)	-3.7 (1.8) $^{(-)}$	8.2 (1.9) $^{(+)}$
(-0.5,1.5)	-2.9 (1.5) $^{(-)}$	13.6 (1.7) $^{(+)}$
(-2,1.5)	-3.9 (1.1) $^{(-)}$	11.2 (1.4) $^{(+)}$
(-1.5,2.5)	-2.2 (0.9) $^{(-)}$	12.3 (1.3) $^{(+)}$
(-3,0.5)	0.1 (0.7)	17.8 (1.3) $^{(+)}$
(3,-1)	-0.9 (0.7) $^{(-)}$	14.7 (1.2) $^{(+)}$
(-3,1.5)	-1.0 (0.7) $^{(-)}$	13.5 (1.2) $^{(+)}$
(-3,2.5)	-2.0 (0.6) $^{(-)}$	9.1 (1.1) $^{(+)}$

(c) ρ = -0.2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	1.5 (1.7)	9.8 (1.8) $^{(+)}$
(0.5,-1)	1.4 (1.7)	9.5 (1.8) $^{(+)}$
(-0.5,1.5)	0.5 (1.4)	11.2 (1.6) $^{(+)}$
(-2,1.5)	1.2 (0.9) $^{(+)}$	9.9 (1.2) $^{(+)}$
(-1.5,2.5)	1.2 (0.7) $^{(+)}$	9.8 (1.0) $^{(+)}$
(-3,0.5)	0.5 (0.7)	9.9 (1.0) $^{(+)}$
(3,-1)	0.7 (0.6) $^{(+)}$	9.6 (1.0) $^{(+)}$
(-3,1.5)	0.5 (0.5)	8.4 (0.9) $^{(+)}$
(-3,2.5)	0.3 (0.3)	6.8 (0.8) $^{(+)}$

(d) ρ = 0.2

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	5.4 (1.5) $^{(+)}$	9.4 (1.5) $^{(+)}$
(0.5,-1)	5.7 (1.5) $^{(+)}$	8.6 (1.5) $^{(+)}$
(-0.5,1.5)	4.4 (1.2) $^{(+)}$	8.2 (1.3) $^{(+)}$
(-2,1.5)	1.9 (0.6) $^{(+)}$	5.3 (0.8) $^{(+)}$
(-1.5,2.5)	1.6 (0.5) $^{(+)}$	4.1 (0.6) $^{(+)}$
(-3,0.5)	2.4 (0.6) $^{(+)}$	4.2 (0.7) $^{(+)}$
(3,-1)	1.8 (0.5) $^{(+)}$	3.5 (0.6) $^{(+)}$
(-3,1.5)	1.0 (0.3) $^{(+)}$	2.7 (0.5) $^{(+)}$
(-3,2.5)	0.2 (0.2)	1.5 (0.4) $^{(+)}$

(e) ρ = 0.5

Shifts in mean	$p_{LD}-p_{DFT}$ (%) (2S.E.)%	$p_{LD}-p_{Ratio}$ (%) (2S.E.)%
(-1.0.5)	7.5 (1.1) $^{(+)}$	10.2 (1.2) $^{(+)}$
(0.5,-1)	7.8 (1.1) $^{(+)}$	10.0 (1.2) $^{(+)}$
(-0.5,1.5)	8.2 (0.9) $^{(+)}$	9.1 (0.9) $^{(+)}$
(-2,1.5)	0.6 (0.2) $^{(+)}$	1.2 (0.3) $^{(+)}$
(-1.5,2.5)	0.3 (0.2) $^{(+)}$	0.7 (0.2) $^{(+)}$
(-3,0.5)	2.8 (0.5) $^{(+)}$	3.5 (0.5) $^{(+)}$
(3,-1)	0.9 (0.3) $^{(+)}$	1.6 (0.4) $^{(+)}$
(-3,1.5)	0.2 (0.1) $^{(+)}$	0.7 (0.2) $^{(+)}$
(-3,2.5)	0.0 (0.0)	0.0 (0.0)

(f) ρ = 0.8

(+) the estimated power of the LD method is significantly different than the other method, the LD power is higher

(-) the estimated power of the LD method is significantly different than the other method, the LD power is lower

5.6 Identification by Variables

The results shown in section 5.5 provide the overall performance of the diagnostic methods. As given before, the definition of Power is the percentage of identifying variable 1 (in Case I) or contaminated variables which are variable 1 and variable 2 (in Cases II, III and IV) as the aberrant variable(s). For Case I, the identification of the aberrant variable is very straight forward since there is only one contaminated process variable. But for the other cases, when there are two process variables deviated from the in-control mean vector, it is presumed that the percentage of identification of each variables would be depending on how far the process variable mean deviates from it. In other words, if the contaminants introduced to both process variables are equal in value, the percentage of identifying each variable as the aberrant variable should be more or less similar. If one of the contaminated process variables has higher contaminant value, it is presumed that the process variable will be identified more in frequency compared to the other process variable. In later sections 5.6.1 and 5.6.2, the identification of the aberrant variables will be presented separately across all types of correlations. A diagnostic method that can identify the highest contaminated process variable more than a lower contaminated process variable would be considered as a consistent diagnostic method and therefore will be regarded as the best diagnostic method among the three methods. The definition of power in this section is slightly different than the one defined in 5.5. The power is defined as the percentage of the number of times a diagnostic method identified variable k, $k = 1, 2$ as the correct aberrant variable.

5.6.1 Equi-correlation Matrix

Identification by variables is necessary when there is more than one aberrant variables is suspected to caused the OOC signal in a multivariate control chart. Hence, only three cases presented here which are cases II, III and IV. A general pattern that can be seen from Figures 5.13 to 5.20 is that the power is higher for variable k when the shift in mean for variable k is larger.

Case II

Figure 5.13 shows the power of identifying the aberrant variables separately between variable 1 and variable 2. Figure 5.13 shows that the proposed method can identify a bigger shift more frequent and share similar performance with the DFT method. The Ratio method shows a considerably lower percentage in identifying the variable with a bigger shift in mean when the correlation between the variables is -0.2.

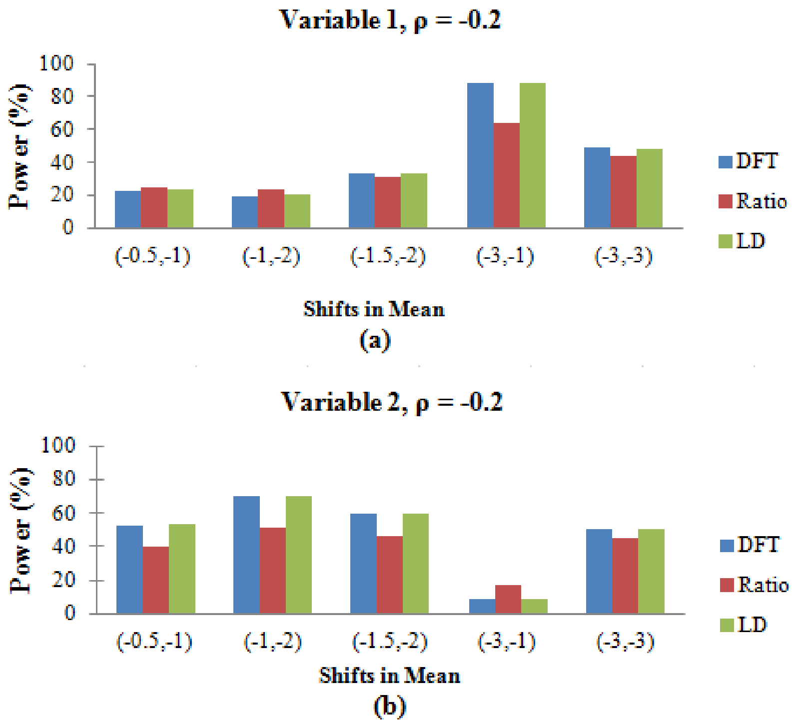
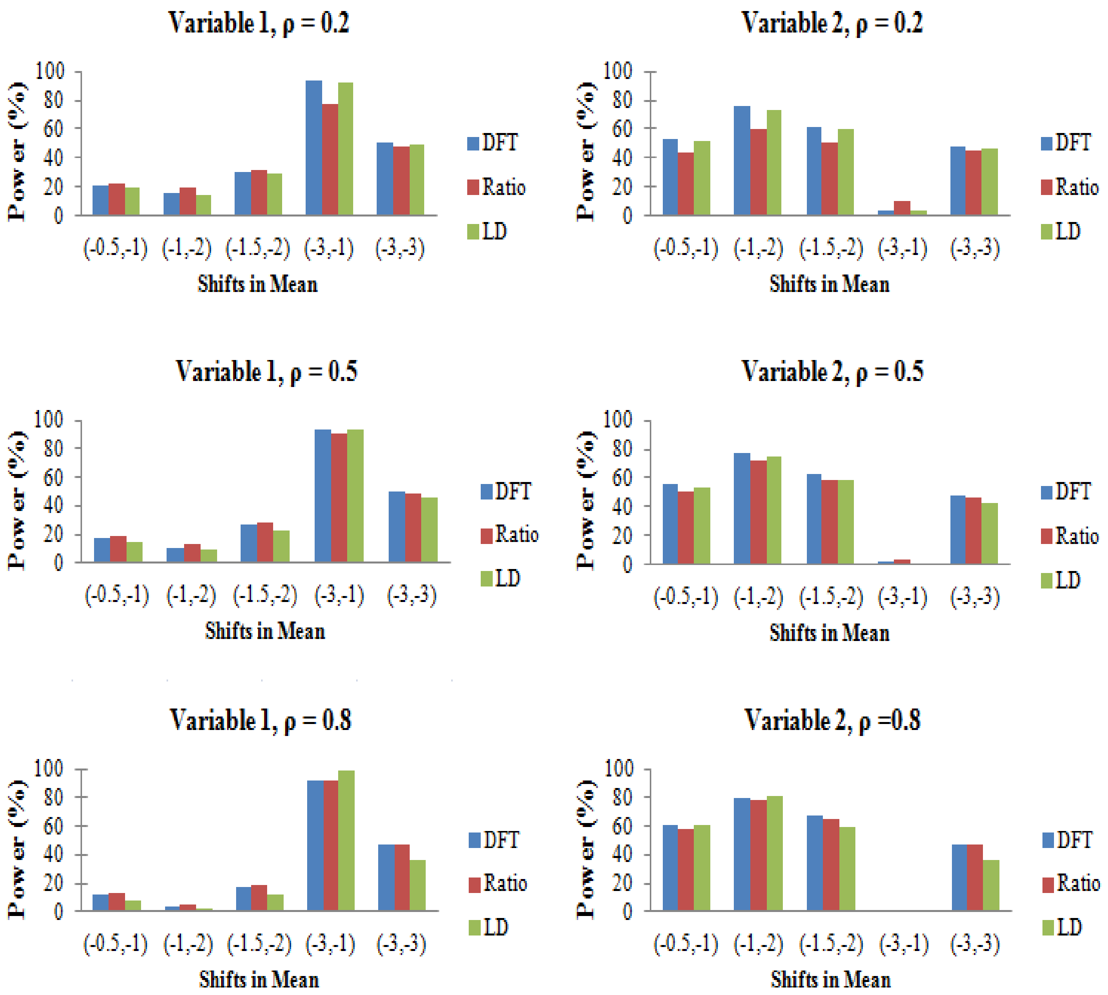


Figure 5.13: The power distribution with respect to contaminated variables and diagnostic methods for low negative correlation, ρ = - 0.2, between variables.

The power of identifying the smaller shift in mean is quite similar for all the diagnostic methods except for combination (-3,-1) in Figure 5.13(b). The Ratio method has higher power than the other two methods.

Figure 5.14 shows the distribution of the percentage of correct identification with respect to variables 1 and 2 for positive correlation between variables. The proposed method LD shows a high percentage of correct identification for a bigger shift in mean. Figure 5.14(a) shows that the LD method has the highest power for combination of shifts (-3,-1) with strong positive correlation between the variables. The LD method maintains high power for moderate and low positive correlation but slightly lower than the DFT method for the low correlation. For combination (-3,-3), the LD method has the lowest power for ρ = 0.8 but the power increases when the correlation decreases.

The same pattern is also shown by the LD method in Figure 5.14 for the same combination of shifts in mean. Figure 5.14(b) shows that the LD method can identify the larger shift quite good and slightly lower than the DFT method but higher that the Ratio method. Interestingly, for combination of shifts in mean (-3,-1) in Figure 5.14(b), the identification of the aberrant variable with a smaller shift is nearly null for ρ = 0.8 and only detected very poorly for the moderate and low positive correlation.



(a) (b)

Figure 5.14: The power distribution with respect to contaminated variables and diagnostic methods for positive correlation.

Case III

Figure 5.15 shows the distribution of the correct identification with respect to the contaminated variables when the correlation is between the variables is low negative. Figure 5.15(a) shows that the proposed method maintains a high power in identifying the contaminated variable with a bigger shift in mean. The power is similar or slightly higher than the DFT method. Figure 5.15(b) shows that the LD method is also quite good compared to the other methods in picking up the aberrant variable with smaller shift for combinations (1, 0.5), (1.5, 1.5) and (3, 2.5).

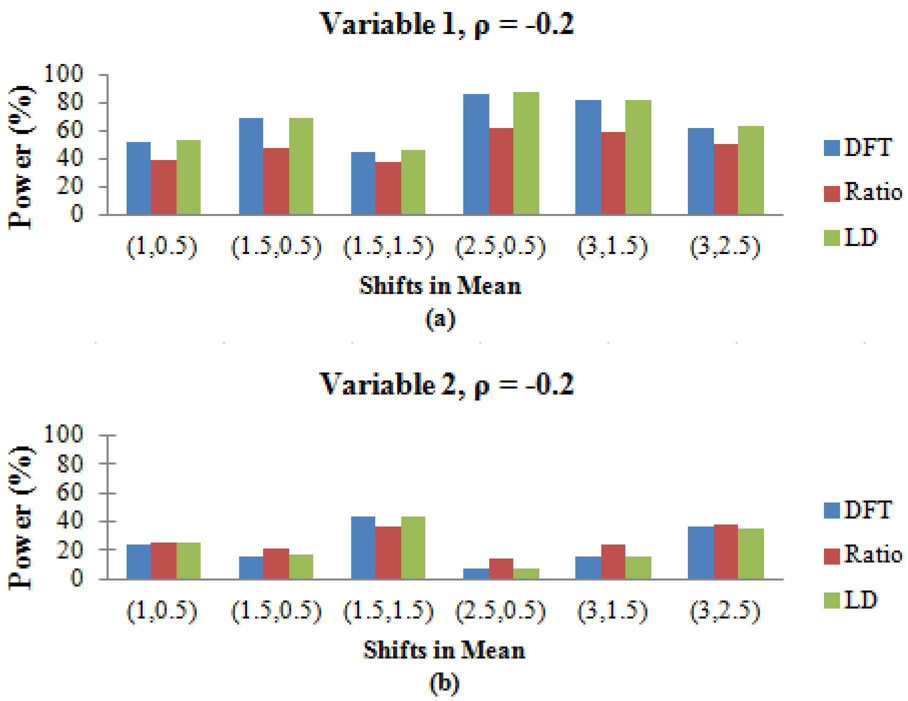
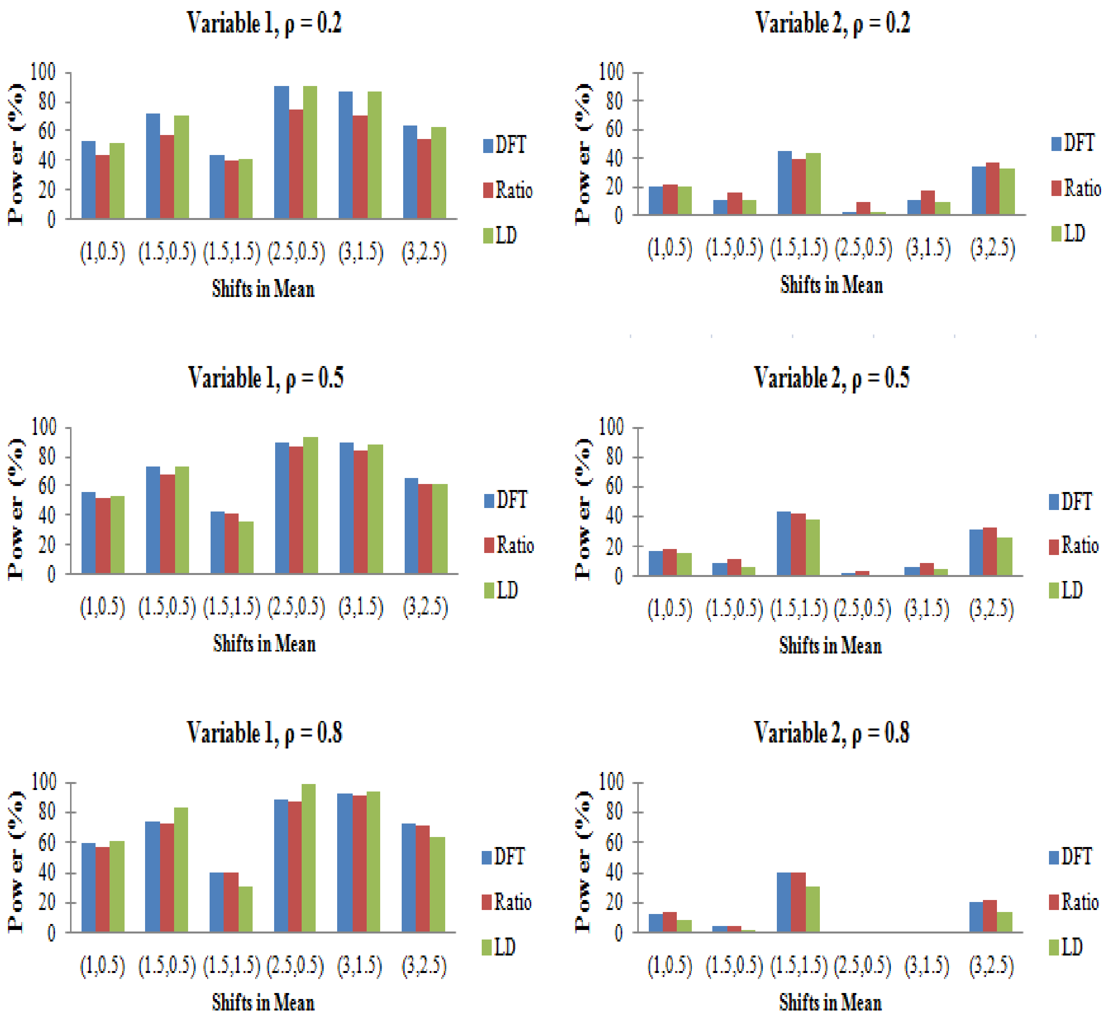


Figure 5.15: The power distribution with respect to contaminated variables and diagnostic methods for low negative correlation, ρ = -0.2, between variables.

Figure 5.16 shows the distribution of the power for positive correlations between the variables. The LD method consistently shows a high power in detecting the aberrant variable with a bigger shift in Figure 5.16(a) except for combinations (3, 2.5) and (1.5, 1.5). The power shown by the LD method is the highest for the strong positive correlation, similar or slightly lower than the DFT method for some combinations when the correlation is moderate or low positive.

Figure 5.16(b) shows a higher percentage in identifying the second aberrant variable if the shifts in mean are approximately equal in magnitude. For this type of combination, the power of the LD method for a variable with a smaller shift increases when the correlation is decreases. The power is higher for variable with a smaller shift when the correlation between the variables is low positive.



(a) (b)

Figure 5.16: Performances across three diagnostic methods when two process variables contaminated by positive contaminants and correlations between variables are moderate and high.

Case IV

Figure 5.17 shows the distribution of power for shifts in mean in opposite direction when the correlation between the variables is low negative. The LD method consistently shows the highest power in identifying the aberrant variable with a bigger magnitude of shift whenever a large shift is one of the shifts in mean. The power is similar or slightly lower than the DFT method whenever the large shift is absent in the combination of shifts in mean. The LD method is able to identify the aberrant variable with a smaller shift (slightly lower than the other methods) except for combinations (-3, 0.5), (3, -1) and (-3, 1.5).

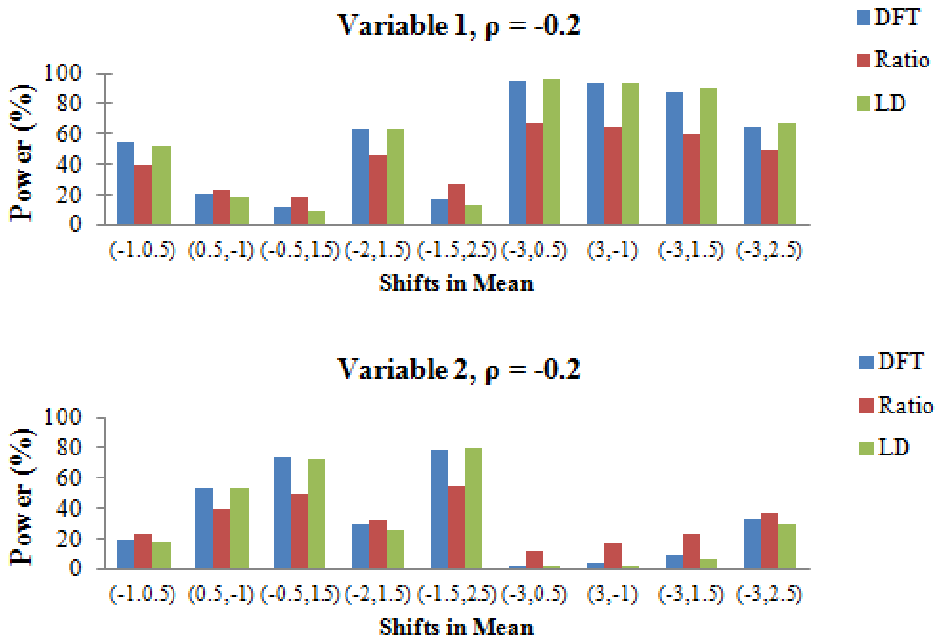


Figure 5.17: The power distribution with respect to contaminated variables and diagnostic methods for moderate positive correlation, ρ = -0.2, between variables.

For low positive correlation between the variables, the pattern of the power distribution is quite similar. The LD method is more responsive for the aberrant variable with smaller shift in mean as shown in Figure 5.18. Figure 5.19 shows the power distribution when the correlation between the variables is moderate positive. In most of the combinations, the LD method is able to identify the aberrant variable with a larger shift better than the other two methods but slightly higher than the power shown by the DFT method. The LD method is also able to identify the aberrant variable with a smaller shift with approximately equal power or slightly lower compared to the other methods.

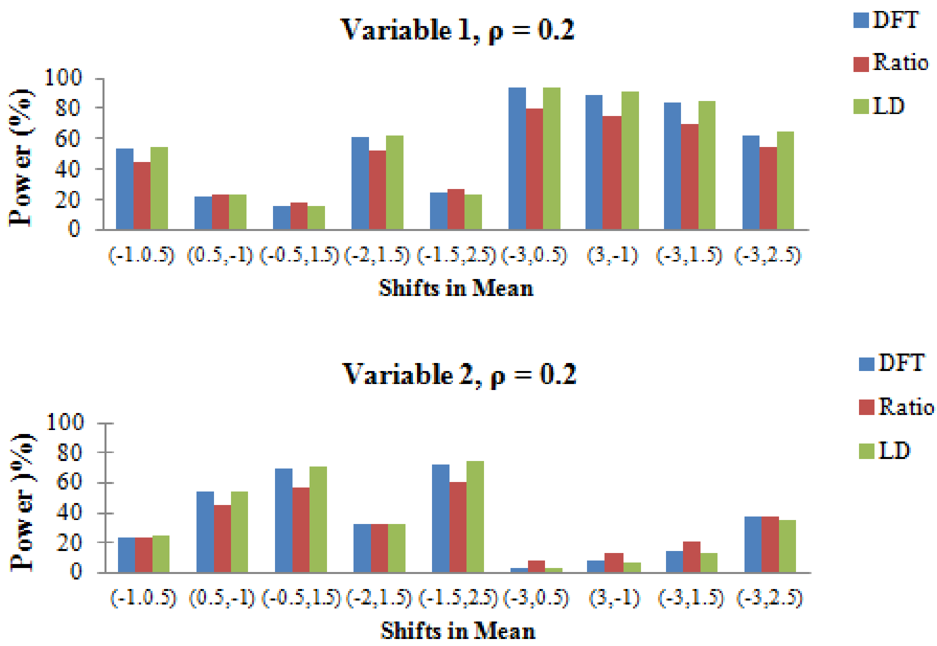


Figure 5.18: The power distribution with respect to contaminated variables and diagnostic methods for moderate positive correlation, ρ = 0.2, between variables.

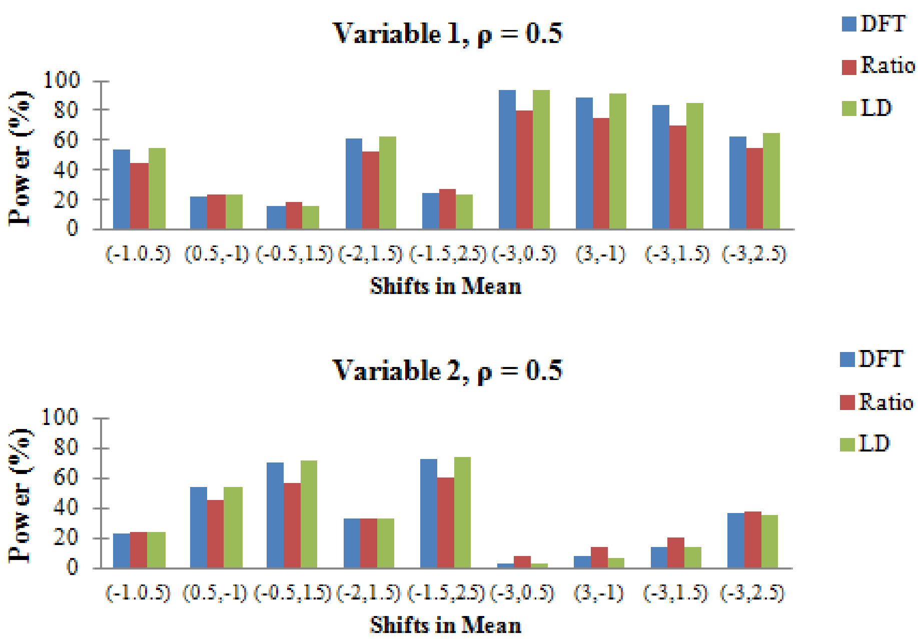


Figure 5.19: The power distribution with respect to contaminated variables and diagnostic methods for moderate positive correlation, ρ = 0.5, between variables.

Figure 5.20 shows the same pattern of power distribution throughout the combinations of shifts in mean but with better and higher power. The LD method consistently shows the highest power for the aberrant variable with a larger shift when the correlation between the variables is strong positive. The identification of the aberrant variable with a smaller shift is approximately equal with the other methods in power when both shifts are small. The power of the LD method is lower than the other methods whenever the bigger shift in the combination is a large shift.

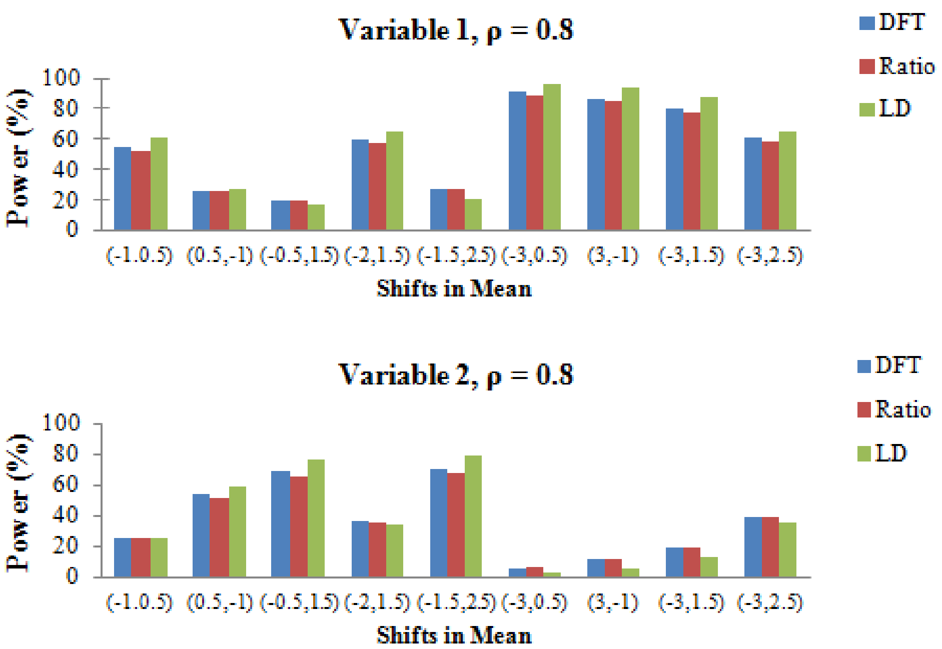


Figure 5.20: The power distribution with respect to contaminated variables and diagnostic methods for moderate positive correlation, ρ = 0.8, between variables.

5.6.2 Non Equi-Correlation Matrix

The combinations of the shifts in mean in these results are also similar with the previous sub-section 5.6.1. The difference is that the correlations between the variables are considered more realistic, instead of assuming that the process variables are all equi-correlated. The same general pattern for the power distribution shown in all the figures in section 5.6.1 is also seen in this section. Figures 5.21 to 5.24 show that the power is higher for the aberrant variable with a bigger shift.

Case II

Figure 5.21 shows the distribution of power in identifying the aberrant variables separately. The two graphs on the top show the power distribution for correlation matrix with all positive values. The LD method shows the highest power for the aberrant variable with a larger shift except for the combination (-3,-3).

The two graphs at the bottom show the power distribution for correlation matrix with mixed signs values. The LD method still maintains having the highest power for the aberrant variable with a larger shift. The shifts in mean with combination (-3, -3) shows an interesting power pattern between variable 1 and variable 2. The power is higher for variable 2 for correlation matrix c1 whereas the power is higher for variable1 when the correlation matrix is c2.

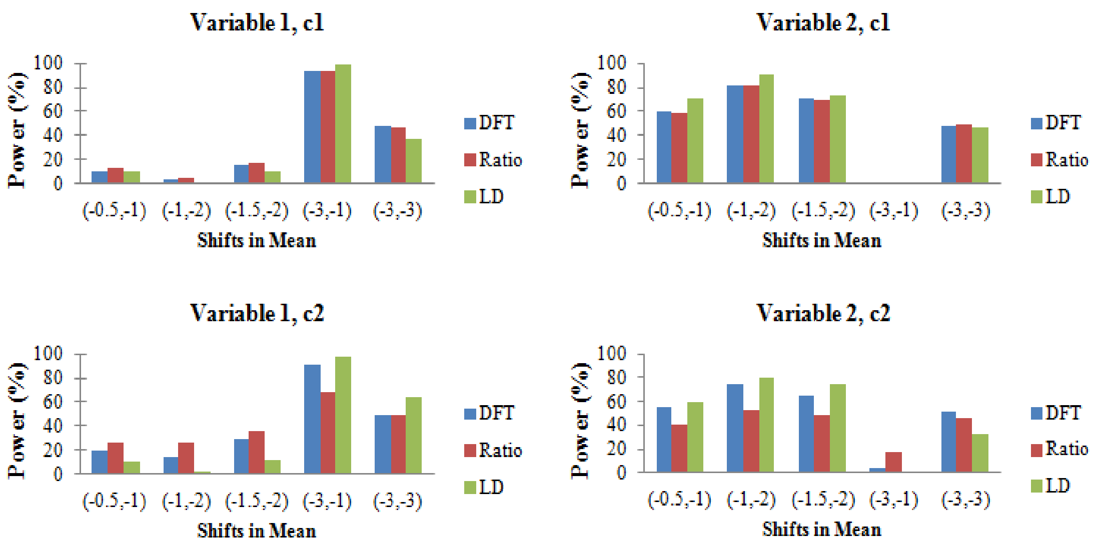


Figure 5.21: Performances across three diagnostic methods when two variables contaminated by negative contaminants.

Case III

Figure 5.22 shows that the LD method performed better than the other two methods in identifying the aberrant variable with a larger shift in mean for correlation matrix c1 except for combination (1.5, 1.5) and (3, 2.5) where the power is slightly lower.

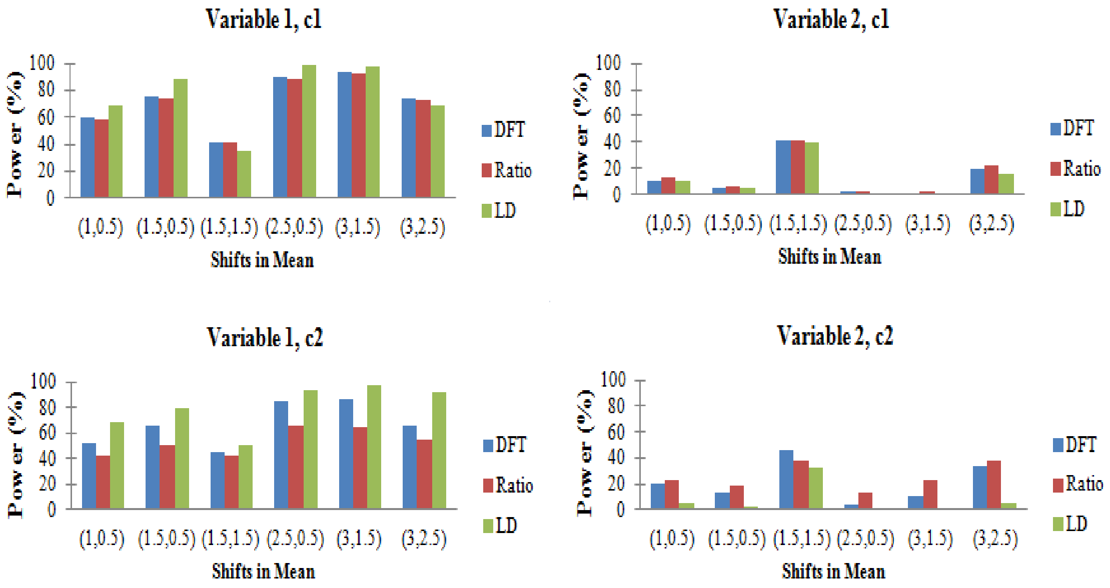


Figure 5.22: Performances across three diagnostic methods when two variables contaminated by positive contaminants.

Case IV

Figure 5.23 shows that the LD method consistently shows the highest power in identifying the aberrant variable with a larger shift except for combination (-2, 1.5) where the power is approximately similar or slightly lower than the other methods. For aberrant variable with a smaller shift, the power is highest for combinations (0.5, -1), (-2, 1.5) and (-3, 2.5). The rest of the combinations show the power of the LD method is slightly lower than the other methods.

For correlation matrix with mixed sign values, the LD method shows better and higher power in identifying the aberrant variable with a larger shift except for combinations (0.5, -1), (-0.5, 1.5) and (-1.5, 2.5). For these combinations, the LD method surprisingly gives a higher power for a smaller shift in mean.

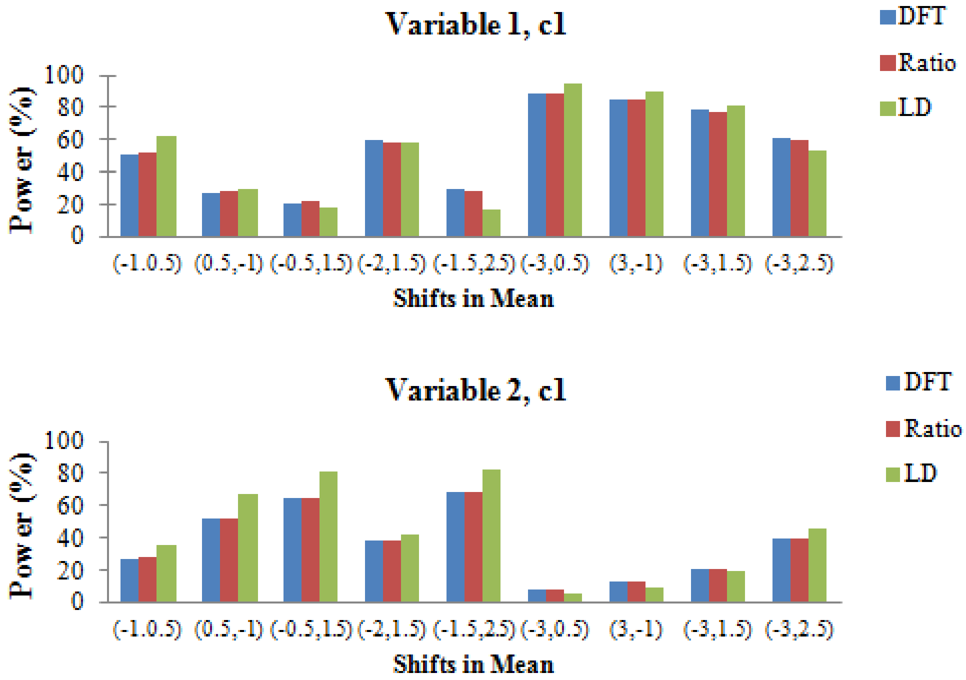


Figure 5.23: Performances across three diagnostic methods when two process variables deviated from the in-control mean in the opposite directions.

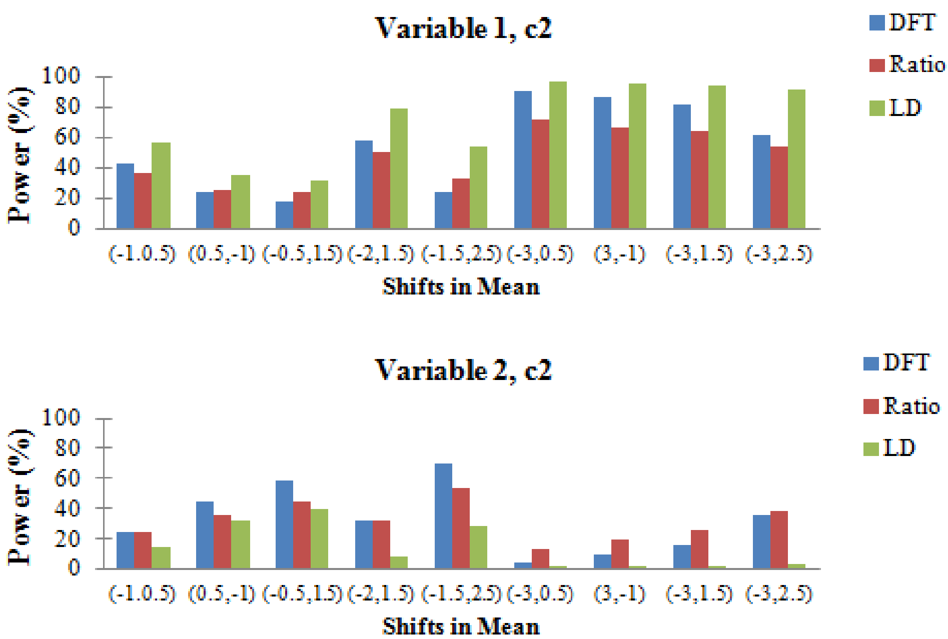


Figure 5.24: Performances across three diagnostic methods when two process variables deviated from the in-control mean in the opposite directions.

5.6.3 Summary and Conclusion

The proposed method, LD, shows a very good performance in identifying an aberrant variable with a bigger shift when one of the shifts is large and the other is small. The power is higher compared to the other methods regardless whether the shifts are in counter-correlational or in accordance with the correlation structure. The proposed method LD showed a very good performance, better than the other two methods, when only one variable has deviated from the in-control mean vector. The result shown by the LD method in Case I was consistently the best among the three methods regardless of the structure of the correlations between the variables in four dimensional data sets. The power of the LD method is significantly higher than the Ratio method throughout all shifts in mean for every correlation matrices. The LD method also has power significantly higher than the DFT method for both non equi-correlation matrices and for the strong positive equi-correlation matrix. For the moderate positive equi-correlation matrix, the LD method has a significantly higher power than the DFT method except for the smallest shift. For the low equi-correlation matrices, both methods are on a par. The proposed method is thus very much recommended when there is a priori knowledge that one of the tested variables has a high tendency to deviate easily from the in-control mean vector compared to the other variables. It may also prove valuable in other situation.

For the combinations of shifts in mean in the same direction, the LD method showed a good performance when one of the shifts is large and the other is small specifically when the deviated variables are strongly correlated. So, the proposed method is very much recommended under this condition. However, the estimated power of the LD method dropped considerably when the shifts are equal and large or intermediate, specifically when the variables are strongly correlated too. Hence, it is not recommended for this condition.

For the combinations of shifts in mean in the opposite directions, the LD method has shown a very good performance. The power is significantly higher than the other two methods when the shifts are not in accordance with the correlation structure especially when the correlation between the variables is moderate and strong positive. Even though the estimated power of the LD method is not substantially higher than the other two methods for any combination of at least one large shift, it is for the combination of small shifts and combination of small and intermediate shifts. This feature makes the proposed method recommended over the DFT method for strong positive correlations between the variables.

The results given above showed that the proposed method is a promising method in indentifying aberrant variable in multivariate processes. Furthermore, the performance of this new approach can be improved in a way in which cannot be done to the other two methods. The performance of this approach can be increased by selecting formal threshold values which allow it to identify multiple variables as aberrant simultaneously and hence can give a higher correct identification. A discussion of the extension of the proposed method is given in Chapter 7.

5.7 Random Correlation Matrices

The correlation matrices used in the previous sections are taken from two studies by Doganaksoy, Faltin and Tucker (1991) and Das and Prakash (2008). There are four types of correlation structure proposed by the first study and two of them were used in this chapter. The study claimed that the correlation matrices "*were carefully chosen to typify a mathematically broad range of situations, are believed to provide an adequate basis for judgement as to the usefulness in practice...*". No further explanation of the method of selection was given. The second study, unlike the first, took a different approach by considering equi-correlation matrices. The selection of the correlation values was again unexplained.

In this section, we tested the performance of the diagnostic methods under randomly selected correlation matrices. A total of 20 random positive definite correlation matrices are simulated using R-programming. Five combinations of shifts in mean vector are. The selected combinations are shown in Table 5.5.

Table 5.5: Selected mean vectors with three types of shifts

Type of shifts in mean vector	Shifted mean vector
One aberrant variable	$(2.5 \quad 0 \quad 0 \quad 0)'$
Two aberrant variables (same directions)	$(3.0 \quad 1.5 \quad 0 \quad 0)'$ $(1.5 \quad 0.5 \quad 0 \quad 0)'$
Two aberrant variables (opposite direction)	$(-3.0 \quad 2.5 \quad 0 \quad 0)'$ $(3.0 \quad -1.0 \quad 0 \quad 0)'$

Samples of 1000 OOC observations are then simulated using the mvrnorm procedure for each random positive definite correlation matrix and shifted mean vector. The three diagnostic methods are performed on the sample and the power of each diagnostic method is estimated. The estimated Power is defined as the percentage of the number of times the aberrant variable(s) has the highest coefficient value, out of 1000 OOC observations. The list of the estimated powers for DFT, Ratio and LD diagnostic methods are given in Tables 5.6, 5.7, 5.8, 5.9 and 5.10. The corresponding random correlation matrices are given in Appendix.

The estimated power for LD method is excellent in many cases and even reached 100% correct identification of the aberrant variable(s) in some simulations and sometimes from the same correlation matrices for all types of shifts such as matrix 4 and matrix 15. Unfortunately, for several other random correlation matrices, the estimated power is only good or fair and sometimes very low or even complete failure to detect any aberrant variable. DFT method produced more consistent estimated power compared to the Ratio and LD methods even though in quite a number of cases, LD method outperformed it. Ratio method also showed inconsistencies in performance, like the LD method. The estimated power dropped approximately equal in magnitude or sometimes lower than the LD method. For a particular matrix such as matrix 12, LD method is totally unresponsive and the Ratio method performs only slightly better.

Table 5.6: Estimated Power (%) for every random correlation matrix with respect to diagnostic methods for shifted mean vector = $(2.5 \quad 0 \quad 0 \quad 0)'$.

Random correlation matrix	Estimated Power (S.E.)%		
	DFT	Ratio	LD
1	89.4 (1.0)	89.7 (1.0)	93.7 (0.8)
2	85.8 (1.1)	88.2 (1.0)	96.8 (0.6)
3	84.9 (1.1)	88.8 (1.0)	78.6 (1.3)
4	84.0 (1.2)	91.9 (0.9)	100 (0)
5	84.9 (1.1)	40.1 (1.5)	67.0 (1.5)
6	84.3 (1.2)	84.3 (1.2)	85.8 (1.1)
7	86.1 (1.1)	34.0 (1.5)	35.3 (1.5)
8	85.6 (1.1)	87.8 (1.0)	0.2 (0.1)
9	85.0 (1.1)	92.0 (0.9)	97.3 (0.5)
10	84.0 (1.2)	92.1 (0.9)	99.5 (0.2)
11	89.6 (1.0)	90.8 (0.8)	88.0 (1.0)
12	84.1 (1.2)	0 (0)	0.3 (0.2)
13	88.7 (1.0)	59.9 (1.5)	83.5 (1.2)
14	91.5 (0.9)	88.5 (1.0)	70.1 (1.4)
15	88.1 (1.0)	80.0 (1.3)	2.6 (0.5)
16	93.3 (0.8)	2.2 (0.5)	55.7 (1.6)
17	85.0 (1.1)	9.1 (0.9)	79.5 (1.3)
18	85.9 (1.1)	90.6 (0.9)	90.4 (0.9)
19	85.1 (1.1)	87.8 (1.0)	83.8 (1.2)
20	89.9 (1.0)	64.0 (1.5)	74.1 (1.4)

Table 5.7: Estimated Power (%) for every random correlation matrix with respect to diagnostic methods for shifted mean vector = $(3.0 \quad 1.5 \quad 0 \quad 0)'$.

Random correlation matrix	Estimated Power (S.E.)%		
	DFT	Ratio	LD
1	97.8 (0.5)	95.0 (0.7)	96.6 (0.6)
2	96.9 (0.5)	98.5 (0.4)	99.9 (0.1)
3	95.4 (0.7)	93.2 (0.8)	83.8 (1.2)
4	94.4 (0.7)	95.6 (0.6)	100 (0)
5	95.0 (0.7)	74 (1.4)	66.0 (1.5)
6	95.7 (0.6)	91.3 (0.9)	86.1 (1.1)
7	97.4 (0.5)	83.3 (1.2)	72.8 (1.4)
8	92.8 (0.8)	94.6 (0.7)	0 (0)
9	94.5 (0.7)	96.7 (0.6)	97.5 (0.5)
10	93.7 (0.8)	96.3 (0.6)	98.6 (0.4)
11	97.0 (0.5)	98.8 (0.3)	93.8 (0.8)
12	95.8 (0.6)	47.3 (1.6)	0.3 (0.2)
13	96.6 (0.6)	90.0 (0.9)	92.9 (0.8)
14	94.7 (0.7)	94.6 (0.7)	16.3 (1.2)
15	97.4 (0.5)	97.9 (0.5)	86.6 (1.1)
16	96.7 (0.6)	56.3 (1.6)	8.6 (0.9)
17	97.8 (0.5)	72.2 (1.4)	100 (0)
18	95.2 (0.7)	93.5 (0.8)	92.7 (0.8)
19	98.0 (0.4)	99.5 (0.2)	100 (0)
20	97.4 (0.5)	93.6 (0.8)	96.5 (0.6)

Table 5.8: Estimated Power (%) for every random correlation matrix with respect to diagnostic methods for shifted mean vector = $(1.5 \quad 0.5 \quad 0 \quad 0)'$.

Random correlation matrix	Estimated Power (S.E.)%		
	DFT	Ratio	LD
1	84.9 (1.1)	79.4 (1.3)	81.5 (1.2)
2	82.4 (1.2)	88.5 (1.0)	92.4 (0.8)
3	78.2 (1.3)	75.1 (1.4)	68.8 (1.5)
4	74.8 (1.4)	78.0 (1.3)	99.0 (0.3)
5	81.6 (1.2)	45.6 (1.6)	56.1 (1.6)
6	78.0 (1.3)	71.4 (1.4)	72.1 (1.4)
7	81.1 (1.2)	54.4 (1.6)	63.4 (1.5)
8	72.2 (1.4)	80.4 (1.3)	0 (0)
9	77.9 (1.3)	83.3 (1.2)	85.2 (1.1)
10	75.0 (1.4)	83.9 (1.2)	93.7 (0.8)
11	83.1 (1.2)	89.7 (1.0)	83.9 (1.2)
12	75.1 (1.4)	19.9 (1.3)	6.1 (0.8)
13	83.9 (1.2)	74.9 (1.4)	82.2 (1.2)
14	80.1 (1.3)	83.7 (1.2)	39.3 (1.5)
15	80.9 (1.2)	87.4 (1.0)	87.6 (1.0)
16	83.0 (1.2)	26.9 (1.4)	28.3 (1.4)
17	80.5 (1.3)	41.0 (1.6)	97.7 (0.5)
18	82.3 (1.2)	82.7 (1.2)	84.1 (1.2)
19	82.8 (1.2)	93.4 (0.8)	98.5 (0.4)
20	84.2 (1.2)	79.0 (1.3)	86.5 (1.1)

Table 5.9: Estimated Power (%) for every random correlation matrix with respect to diagnostic methods for shifted mean vector = $(-3.0 \quad 2.5 \quad 0 \quad 0)'$.

Random correlation matrix	Estimated Power (S.E.)%		
	DFT	Ratio	LD
1	96.4 (0.6)	94.2 (0.7)	92.9 (0.8)
2	96.1 (0.6)	96.6 (0.6)	65.4 (1.5)
3	98.6 (0.4)	95.2 (0.7)	80.6 (1.3)
4	96.9 (0.5)	94.9 (0.7)	99.9 (0.1)
5	99.2 (0.3)	91.0 (0.9)	78.3 (1.3)
6	97.6 (0.5)	92.9 (0.8)	93.4 (0.8)
7	98.0 (0.4)	91.2 (0.9)	88.4 (1.0)
8	99.8 (0.1)	99.9 (0.1)	100 (0)
9	99.0 (0.3)	98.9 (0.3)	99.8 (0.1)
10	98.5 (0.4)	98.4 (0.4)	100 (0)
11	96.9 (0.5)	98.6 (0.4)	81.1 (1.2)
12	97.3 (0.5)	73.3 (1.4)	0 (0)
13	98.3 (0.4)	95.3 (0.7)	96.5 (0.6)
14	99.7 (0.2)	99.6 (0.2)	99.7 (0.2)
15	99.0 (0.3)	99.6 (0.2)	100 (0)
16	98.7 (0.4)	78.9 (1.3)	42.3 (1.6)
17	98.5 (0.4)	91.1 (0.9)	92.9 (0.8)
18	97.3 (0.5)	95.1 (0.7)	95.3 (0.7)
19	96.1 (0.6)	98.4 (0.4)	86.7 (1.1)
20	97.8 (0.5)	94.7 (0.7)	87.5 (1.0)

Table 5.10: Estimated Power (%) for every random correlation matrix with respect to diagnostic methods for shifted mean vector = $(3.0 \quad -1.0 \quad 0 \quad 0)'$.

Random correlation matrix	Estimated Power (S.E.)%		
	DFT	Ratio	LD
1	95.0 (0.7)	93.6 (0.8)	95.2 (0.7)
2	94.2 (0.7)	96.0 (0.6)	94.6 (0.7)
3	95.5 (0.7)	93.2 (0.8)	80.9 (1.2)
4	93.1 (0.8)	95.6 (0.6)	100 (0)
5	97.1 (0.5)	73.6 (1.4)	75.0 (1.4)
6	93.1 (0.8)	89.4 (1.0)	90.2 (0.9)
7	94.5 (0.7)	74.4 (1.4)	79.7 (1.3)
8	98.1 (0.4)	99.0 (0.3)	20.8 (1.3)
9	95.7 (0.6)	97.4 (0.5)	99.2 (0.3)
10	94.9 (0.7)	96.5 (0.6)	100 (0)
11	96.1 (0.6)	97.1 (0.5)	87.7 (1.0)
12	93.1 (0.8)	33.5 (1.5)	0 (0)
13	96.6 (0.6)	89.7 (1.0)	91.9 (0.9)
14	98.4 (0.4)	99.2 (0.3)	98.0 (0.4)
15	97.2 (0.5)	98.6 (0.4)	100 (0)
16	97.9 (0.5)	41.2 (1.6)	71.9 (1.4)
17	95.5 (0.7)	54.4 (1.6)	99.2 (0.3)
18	93.6 (0.8)	93.7 (0.8)	93.9 (0.8)
19	96.3 (0.6)	98.5 (0.4)	98.9 (0.3)
20	97.1 (0.5)	92.6 (0.8)	92.2 (0.8)

This scenario can be investigated by studying the estimated power in relation to the correlation structures. For instance, in correlation matrix 8, there are mixed signed correlations between variables. Variables 1 and 2 are positively strong correlated but there is another variable, variable 4, which is more strongly correlated to variable 1 in the same direction. Table 5.11 showed that the LD method most of the time has incorrectly diagnosed variable 4 as aberrant whenever the shift(s) is not counter to the correlation structure and totally failed to identified variable 2.

This result is totally different from the result shown in Figures 5.8 and 5.11 where the estimated powers for the same shifts are excellent. There are a couple of differences in the correlation structure compared to $\mathbf{c_1}$ and equi-correlation matrix with ρ = 0.8. Unlike matrix 8, both matrices have only positive correlation values and there is no other correlation value higher than the one between the aberrant variables. So, it is suspected that these two conditions caused the dramatic drop in performance.

Table 5.11: Identification by variables with correlation matrix 8 for LD method

Shifted mean vector	Random correlation matrix 8			
	var. 1	var. 2	var. 3	var. 4
(2.5, 0, 0, 0)'	2	0	0	998
(3.0, 1.5, 0, 0)'	0	0	0	1000
(1.5, 0.5, 0, 0)'	0	0	0	1000
(-3.0, 2.5, 0, 0)'	1000	0	0	0
(3, -1, 0, 0)'	208	0	0	792

The method works tremendously well when the shifts are in counter correlation given that the magnitudes of the shift are sufficiently large. This particular result is comparable to the result showed by Figure 5.8 for combination (-3.0, 2.5, 0, 0)'. The LD method did shows 100% correct identification for this combination in Figure 5.8. The same performance was not repeated for combination (3, -1, 0, 0)'. Variable 4 is seems very dominant and been incorrectly identified in most of the time.

The estimated power of the LD method is much worse for correlation matrix 12. Regardless of whether the shifts are counter correlational or not, the LD method showed either very poor performance or was totally unresponsive. Matrix 12 gives negative low correlation value for the aberrant variables, 1 and 2. Correlation values with other variables are either moderate or low with mixed signs. Table 5.12 shows that most of the time, for all the shifts, variable 3 was incorrectly identified as the aberrant variable followed by variable 4 (less frequently than variable 3). Unfortunately, this result is not similar to any of the correlation structures from the previous sections. Even though we tested the method on an equi-correlation matrix with ρ = -0.2, but correlation between other variables (non-aberrant) in the matrix are all low negative in value. Unlike correlation matrix 12, there are two variables with moderate and moderate to strongly correlate with opposite signs. This condition might have different kind of impact on the estimated power of the LD method.

Table 5.12: Identification by variables with correlation matrix 12 for LD method

Shifted mean vector	Random correlation matrix 12			
	var. 1	var. 2	var. 3	var. 4
(2.5, 0, 0, 0)'	3	0	947	50
(3.0, 1.5, 0, 0)'	3	0	987	10
(1.5, 0.5, 0, 0)'	61	0	824	115
(-3.0, 2.5, 0, 0)'	0	0	939	61
(3, -1, 0, 0)'	0	0	966	34

5.7.1 Summary and Conclusion

The investigation of the performance of the diagnostic methods under several random correlation matrices has revealed a few findings. Generally, the DFT method has shown a consistent performance across different combinations of shift in mean vector. The Ratio method as well as the LD method has shown less consistency than the DFT method. In many cases, the LD method outperformed the other two methods with estimated power 100% or nearly 100%. Unfortunately, under a few random correlation matrices, LD method failed to perform. A future study on this problem is necessary to investigate how much the correlation structure affects the estimated power of LD method.

CHAPTER 6

Simplifying UIT Assessment by Application of Spectral Decomposition

6.1 Introduction

Most studies of the interpretation of multivariate control charts use simulated data sets to assess performance of proposed methods. Data sets are usually presumed to follow a multivariate normal distribution with known variance covariance matrix. Many studies considered different types of covariance or correlation matrices. Doganaksoy, Faltin & Tucker (1991) considered four types of correlation matrices. Das & Prakash (2008) used equi-correlation matrices, whereas, Maravelakis et al. (2002) considered two types of covariance matrices: all positive values and mixed positive and negative values in the covariance matrix.

In this chapter, we will show that it is not necessary to consider multiple types of covariance matrices in looking for the UIT direction. We demonstrate that no generality is lost when dealing with data from a distribution with known covariance matrix, if we only consider the identity covariance matrix. This is because any covariance matrix may be transformed into a diagonal, and hence identity, covariance matrix by use of spectral decomposition.

Further discussion of this matter is covered in Section 6.2 where we will show how a transformation of a random vector, or a potential aberrant observation, $\mathbf{x}$, from a general variance data space to a data space with identity covariance matrix can save us the trouble of studying multiple types of covariance matrices. A brief discussion of singular value decomposition and its special case spectral decomposition is given in Section 6.3. The link between spectral decomposition procedures and the proposed UIT approach is given in Section 6.4. Section 6.5 explains critical features of the UIT direction and distance in both general data space and transformed variable data space. A numerical example is given in Section 6.6 to demonstrate the use of the UIT method in a data space with identity covariance matrix followed by some discussions in the final section of this chapter.

6.2 Assessing Aberrance in General and Standardized Data Space

The main aim of this study is to assess how aberrant **x** is in a general data space i.e. against data from a distribution with general variance covariance matrix, $\boldsymbol{\Sigma_0}$. Given that the $\boldsymbol{\Sigma_0}$ is known, our proposed UIT test means we need to look in the direction $(\boldsymbol{\Sigma}_0^{-1})\mathbf{x}$ to find the potential aberrant variable(s). This assessment needs to be performed in the original, general covariance data space, defined hereafter as **x**-space.

We find later on that this is not very helpful in proposing a general diagnosis method, based on threshold values (as discussed further in Chapter 7) since it seems that we need to find different threshold values for being aberrant for every possible covariance matrix. This is clearly impractical in real situations. Instead, we want to transform the whole problem to a new, standardized data space in which the transformed data has identity covariance matrix. This can always be achieved for data with a multivariate normal probability distribution with known parameters. Let **x** be defined as a random vector with a *p*-variate normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}$, i.e $\mathbf{x} \sim \mathrm{N_p}(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$. Hence **x** has probability distribution function as shown below,

$$f(\mathbf{x}) = \ |2\pi\boldsymbol{\Sigma}_0|^{-1} exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}-\boldsymbol{\mu}_0)\right\} \qquad [6.1]$$

Now consider

$$\mathbf{y} = \boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_0) \qquad [6.2]$$

where $\boldsymbol{\Sigma}_0^{-1/2}$ is the symmetric positive-definite square root of $\boldsymbol{\Sigma}_0^{-1}$ then the transformed variables $\mathbf{y}$ = $(y_1, y_2, \ldots y_p)'$ are independent and identically distributed (IID), specifically $y_i \sim N_1(0,1)$ where i = 1,2,...,p (Mardia, Kent & Bibbly ,1994).

When we transform **x** to **y** as shown in [6.2], we are no longer working in **x**-space but in **y**-space instead, i.e. in a data space with an identity covariance matrix. So, the assessment of the potentially aberrant variable now needs to be performed in a new data space. Once in this standardized **y**-space we can assess aberrance of each component individually and against a known ($N_1(0,1)$) distribution. This realization forms the basis of the formalization of our UIT procedure beyond 'examine the largest' in Chapter 7. Let us define $\mathbf{u_x}$ as the UIT vector in **x**-space and $\mathbf{u_y}$ as the UIT vector in **y**-space. Based on the relationship of vectors **x** and **y** given in [6.2] with $\boldsymbol{\mu_0}$ = 0 we have the inverse transformation

$$\mathbf{x} = \boldsymbol{\Sigma}_0^{1/2}\mathbf{y} \qquad [6.3]$$

By referring to equation [5.15], the UIT direction in **x** -space is

$$\mathbf{u_x} = \boldsymbol{\Sigma}_0^{-1}\mathbf{x} \qquad [6.4]$$

So, by substituting **x** in [6.3] into [6.4], the UIT direction in **y** -space would be

$$\begin{aligned}\mathbf{u_y} &= \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{\Sigma}_0^{1/2}\mathbf{y}\right)\\ &= \boldsymbol{\Sigma}_0^{-1/2}\mathbf{y} \qquad [6.5]\end{aligned}$$

With the variance covariance matrix assumed known, or well estimated, then we can always deal with observations from independent identically distributed N(0,1) variables because we can standardize all variables by pre-multiplying by the inverse of the square root of that matrix. Therefore, we need an easy way to find $\boldsymbol{\Sigma}_0^{-1/2}$. This can be done using spectral decomposition procedure in which the square root of a known positive definite matrix $\boldsymbol{\Sigma}_0$ can be defined via the spectral decomposition $\mathbf{UDU^T}$ as $\mathbf{UD^{1/2}U}$ where **U** is a matrix of eigenvectors of $\boldsymbol{\Sigma}_0$ and **D** is a diagonal matrix with diagonal elements the eigenvalues of $\boldsymbol{\Sigma}_0$ (which necessarily are strictly positive). Further discussion of spectral decomposition is preceded by a brief explanation of singular value decomposition, which is the more general factorisation of this kind, is given below.

6.3 Singular Value Decomposition and Spectral Decomposition

Spectral decomposition is a special case of singular value decomposition. Singular value decomposition is a factorization of a matrix into its canonical or normal form and can be applied to any $m \times n$ matrix. Let **M** be any $m \times n$ matrix whose entries can be either real or complex numbers. In our case, we only focus on real numbers. The singular value decomposition of a matrix **M** is the factorization of **M** into the product of three matrices as shown in equation [6.6], where the columns of matrices $\mathbf{U}_{\boldsymbol{m}\times\boldsymbol{m}}$ and $\mathbf{V}_{\boldsymbol{n}\times\boldsymbol{n}}$ are the singular vectors of matrix **M** and the matrix **D** is a diagonal matrix with positive real entries.

$$\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}} \qquad [6.6]$$

The diagonal entries of **D** are known as the singular values of **M**. The columns of **U** are known as the left singular vectors and the columns of **V** are the right singular vectors. If **M** is a symmetric positive definite matrix, where $\mathbf{M} = \mathbf{M}^{\boldsymbol{T}}$ and $\mathbf{x}^{\boldsymbol{T}}\mathbf{M}\mathbf{x} > 0$ for any nonzero vector **x**, then its eigenvectors are orthogonal and we can write

$$\mathbf{M} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathbf{T}} \qquad [6.7]$$

Matrix **P** is an orthogonal matrix, i.e. is such that

$$\mathbf{P}^{-1}\mathbf{M}\mathbf{P} = \mathbf{P}^{\mathbf{T}}\mathbf{M}\mathbf{P} = \mathbf{D}$$

where **D** is a diagonal matrix and the eigenvalues of **M** lie on the main diagonal of **D** (Kolman, 1991). The columns of **P** correspond to the eigenvectors of **M** and the diagonal entries of **D** correspond to the eigenvalues of the matrix **M**.

This kind of factorization is a special case of singular value decomposition where **U** = **V** = **P**. Equation [6.7] represents a special case of a singular value decomposition for a square and symmetric positive definite matrix and which is commonly known as spectral decomposition. Of course our covariance matrix $\mathbf{\Sigma}_0$ is of this form.

6.4 Spectral Decomposition and the Union Intersection Technique

Given that the known, square $p \times p$ and positive definite covariance matrix **M** has real entries, matrix **M** can be factorized under spectral decomposition as shown below,

$$\mathbf{M} = \mathbf{PDP}^{\mathbf{T}} \qquad [6.8]$$

where matrix **P** consist of p orthogonal vectors which makes **P** an orthogonal matrix and also invertible. Thus,

$$\mathbf{P}^{\mathbf{T}} = \mathbf{P}^{-\mathbf{1}} \qquad [6.9]$$

and

$$\mathbf{PP}^{\mathbf{T}} = \mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I} \qquad [6.10]$$

As stated in section 6.1, **D** is a diagonal matrix and the relation between **M** and **D** can be expressed as [6.11] due to the special properties of matrix **M** which is symmetric and positive definite.

$$\mathbf{M} = \mathbf{PDP}^{\mathbf{T}} \qquad [6.11]$$

We now demonstrate that the spectral decomposition procedures provide an easy way for us to obtain $\boldsymbol{\Sigma}_0^{-1/2}$ for the transformation of variables in **x**-space to **y**-space that we wish to use in our UIT approach. Spectral decomposition procedures factorized the positive definite covariance matrix **M** into its eigenvectors and eigenvalues in the form shown in [6.11]. This makes the task of finding the inverse square root of covariance matrix **M** easier since which can be done by finding the square root of the diagonal matrix **D** first as shown in [6.12] below (as shown and proven by Harville (1997)).

$$\begin{aligned} \mathbf{M}^{1/2} &= (\mathbf{PDP}^{\mathbf{T}})^{1/2} \\ &= \mathbf{P}(\mathbf{D})^{1/2}\,\mathbf{P}^{\mathbf{T}} \qquad [6.12] \end{aligned}$$

Hence,

$$\mathbf{M}^{-1/2} = \mathbf{P}(\mathbf{D})^{-1/2}\,\mathbf{P}^{\mathbf{T}} \qquad [6.13]$$

The inverse square root of **M** can be obtained in **R**-programming by following [6.13] and thus the transformation of the variables from **x**-space to **y**-space can be carried out easily.

6.5 Examples

Here we consider two examples, Example 1 is for dimension *p*=2, while Example 2 covers *p*=4. In Example 1, the correlations are equal whereas in Example 2 we use the correlation matrices previously used for illustration in Sections 4.4 and 5.5, taken from Doganaksoy, Faltin & Tucker (1991). Their matrices *c1* and *c2* look at cases of all correlations positive and correlations of mixed sign respectively.

We denoted all the correlation matrices as **M** and the matrix of the eigenvectors and eigenvalues as **P** and **D** respectively, to make it similar to the preceding section and hoping that it would make the explanation easier. One difference is that matrix **M** is a correlation matrix instead of a covariance but given that it is still has the special properties of symmetric and positive definiteness, this make no difference. Given that **d** is the vector of the eigenvalues of **M** then **D =** diag(**d**). The potentially aberrant observation (reflecting deviation from zero men) for Example 1 and Example 2 is fixed as **x** = $(1 \quad 1)'$ or $(1 \quad 1 \quad 1 \quad 1)'$ respectively. The calculation is done using the package **R**.

Example 1: Equi-correlation

Case a: $\mathbf{M} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$;

Spectral decomposition of **M** gives; $\mathbf{P} = \begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}$; $\mathbf{d} = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}$,

Using **R,** the square root and the inverse square root of matrix **M** based on [6.14] and [6.15] are

$$\mathbf{M}^{1/2} = \begin{pmatrix} 0.966 & 0.259 \\ 0.259 & 0.966 \end{pmatrix}; \qquad \mathbf{M}^{-1/2} = \begin{pmatrix} 1.115 & -0.299 \\ -0.299 & 1.115 \end{pmatrix}$$

The UIT direction on **x**-space based on [6.4] is

$$\mathbf{u_x} = (0.667 \quad 0.667)'$$

The transformed variable, **y**, on **y**-space based on [6.3] is;

$$\mathbf{y} = (0.816 \quad 0.816)'$$

So the potential aberrant variable $\mathbf{x} = (1 \quad 1)'$ with correlation matrix, **M,** is equivalent to an observation $\mathbf{M}^{-1/2}\mathbf{x} = (0.816 \quad 0.816)'$ with covariance or correlation matrix, $\mathbf{I_2}$.

Case b:

$$\mathbf{M} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix};$$

Spectral decomposition of **M** gives; $\boldsymbol{P} = \begin{pmatrix} -0.707 & -0.707 \\ 0.707 & -0.707 \end{pmatrix}$; $\mathbf{d} = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}$.

Using **R**, the square root and the inverse square root of matrix **M** are

$$\mathbf{M}^{1/2} = \begin{pmatrix} 0.966 & -0.259 \\ -0.259 & 0.966 \end{pmatrix}; \qquad \mathbf{M}^{-1/2} = \begin{pmatrix} 1.115 & 0.299 \\ 0.299 & 1.115 \end{pmatrix}$$

The UIT direction on **x**-space based on [6.4] is

$$\mathbf{u_x} = (2 \quad 2)'$$

The transformed variable, **y**, on **y**-space based on [6.2] is

$$\mathbf{y} = (1.41 \quad 1.41)'$$

So the potential aberrant variable $\mathbf{x} = (1 \quad 1)'$ with correlation matrix, **M,** is equivalent to an observation $\mathbf{M}^{-1/2}(1 \quad 1)' = (1.41 \quad 1.41)'$ in a space with covariance matrix, $\mathbf{I_2}$.

This clarifies a point which has been noted several times previously. The direction of the deviations in both cases is the same but it seems that it is easier to pick out aberrant observations if the shift goes against that suggested by the correlation structure instead of being in accordance with it (**y** in case b is 'greater' than in case a).

Example 2: Non equi-correlation

***Case a*:**

$$\mathbf{M} = \begin{pmatrix} 1 & 0.8 & 0.55 & 0.6 \\ 0.8 & 1 & 0.65 & 0.5 \\ 0.55 & 0.65 & 1 & 0.6 \\ 0.6 & 0.5 & 0.6 & 1 \end{pmatrix}$$

Spectral decomposition of **M** gives

$$\mathbf{P} = \begin{pmatrix} -0.521 & 0.407 & 0.403 & 0.633 \\ -0.522 & 0.510 & -0.136 & -0.671 \\ -0.488 & -0.323 & -0.758 & 0.288 \\ -0.467 & -0.686 & 0.495 & -0.258 \end{pmatrix}; \quad \mathbf{d} = \begin{pmatrix} 2.855 \\ 0.555 \\ 0.432 \\ 0.158 \end{pmatrix}.$$

Using **R**, the square root and the inverse square root of matrix **M** are

$$\mathbf{M}^{1/2} = \begin{pmatrix} 0.848 & 0.409 & 0.204 & 0.269 \\ 0.409 & 0.844 & 0.299 & 0.176 \\ 0.204 & 0.299 & 0.891 & 0.274 \\ 0.269 & 0.176 & 0.274 & 0.906 \end{pmatrix};$$

$$\mathbf{M}^{-1/2} = \begin{pmatrix} 1.6375 & -0.71138 & -0.0314 & -0.33879 \\ -0.7114 & 1.66894 & -0.3996 & 0.00808 \\ -0.0314 & -0.39959 & 1.3633 & -0.32533 \\ -0.3388 & 0.00808 & -0.3253 & 1.30077 \end{pmatrix}$$

The UIT direction in **x**-space based on [6.4] is

$$\mathbf{u_x} = (0.270 \quad 0.312 \quad 0.374 \quad 0.457)'$$

The transformed variable, **y**, in **y**-space based on [6.3] is

$$\mathbf{y} = (0.556 \quad 0.556 \quad 0.607 \quad 0.645)'$$

The potentially aberrant variable $\mathbf{x} = (1 \quad 1 \quad 1 \quad 1)'$ with correlation matrix $\mathbf{M}$ is equivalent to an observation $\mathbf{M}^{-1/2}(1 \quad 1 \quad 1 \quad 1)' = (0.556 \quad 0.556 \quad 0.607 \quad 0.645)'$ with an identity covariance matrix, $\mathbf{I}_4$.

***Case b*:**

$$\mathbf{M} = \begin{pmatrix} 1 & 0.2 & -0.5 & 0.3 \\ 0.2 & 1 & 0.2 & -0.5 \\ -0.5 & 0.2 & 1 & 0.2 \\ 0.3 & -0.5 & 0.2 & 1 \end{pmatrix}$$

Spectral decomposition of **M** gives

$$\mathbf{P} = \begin{pmatrix} 0.540 & 0.482 & -0.457 & 0.518 \\ -0.457 & 0.518 & -0.540 & -0.482 \\ -0.457 & -0.518 & -0.540 & 0.482 \\ 0.540 & -0.482 & -0.457 & -0.518 \end{pmatrix}; \quad \mathbf{d} = \begin{pmatrix} 1.554 \\ 1.452 \\ 0.946 \\ 0.048 \end{pmatrix}.$$

Using **R,** the square root and the inverse square root of matrix **M** are

$$\mathbf{M}^{1/2} = \begin{pmatrix} 0.905 & 0.178 & -0.313 & 0.227 \\ 0.178 & 0.917 & 0.170 & -0.313 \\ -0.313 & 0.170 & 0.917 & 0.178 \\ 0.227 & -0.313 & 0.178 & 0.905 \end{pmatrix};$$

$$\mathbf{M}^{-1/2} = \begin{pmatrix} 1.861 & -0.873 & 0.984 & -0.964 \\ -0.873 & 1.747 & -0.813 & 0.984 \\ 0.984 & -0.813 & 1.747 & -0.873 \\ -0.964 & 0.984 & -0.873 & 1.861 \end{pmatrix}$$

The UIT direction in **x**-space is

$$\mathbf{u_x} = (1.02 \quad 1.09 \quad 1.09 \quad 1.02)'$$

The transformed variable, in **y**-space is

$$\mathbf{y} = (1.01 \quad 1.05 \quad 1.05 \quad 1.01)'$$

Thus potentially aberrant variable $\mathbf{x} = (1 \quad 1 \quad 1 \quad 1)'$ with variance covariance matrix, **M,** is equivalent to an observation $\mathbf{M}^{-1/2}\mathbf{x} = (1.01 \quad 1.05 \quad 1.05 \quad 1.01)'$ with an identity covariance matrix $\mathbf{I}_4$.

Example 2 also shows that it is easier to pick out aberrant observations if the shift goes against that suggested by the correlation structure instead of being in accordance with that suggested by the covariance structure.

6.6 Conclusion and Discussion

In this chapter we have shown that to find the aberrant variable(s) causing a multivariate OOC signal, it is sufficient to consider only it suitably transformed independent and identically distributed variables usually against a unit variance a unit variance threshold. This standardization means we do not have to proceed differently for each original variance structure in testing the performance of the proposed method. This opens up the possibility of formalizing our LD procedure, a topic investigated in Chapter 7. In reality, $\boldsymbol{\Sigma}_0$ is not usually known, though perhaps it can be well-estimated from substantial in-control data. The standardization procedure is, of course, just as sensitive to the quality of estimation as any other and the effect of the estimation is discussed and illustrated further in Chapter 7.

CHAPTER 7

UIT Assessment with Threshold Value

7.1 Introduction

This chapter discusses another method to diagnose an out of control (OOC) signal triggered by a multivariate control chart. This method still applyies the same concept as the Largest Deviation method. Instead of looking at the highest coefficient value of the largest deviation (LD) vector **u,** a threshold value is used to separate the correct aberrant variables from the rest. The rationale behind this extended application of the proposed method becomes clear when we try to answer a couple of questions below.

1) How to determine all the aberrant variables at one time when there may be more than one variable shifted from its' in control mean value?

2) If there is more than one variable shifted from its mean value, for instance two, how could we know that the second largest coefficient from the largest deviation's vector is genuinely indicating the second aberrant variable?

Diagnostic methods based on the highest value of some statistic values such as univariate *t*-statistic, ratio, or coefficient value of a particular vector can only identify one aberrant variable at a time. The common practice adopted by other researchers is to remove the aberrant variable identified by a diagnostic method and then re-test the corresponding multivariate observation as to whether it still triggers an out of control signal. If it does then it will be diagnosed again in order to look for a second aberrant variable. The diagnostic process will stop when the corresponding multivariate observation no longer produces an out of control signal. This study aims to propose a simpler and faster way to identify all the genuinely aberrant variables simultaneously.

Simplifying components

We have been able to take advantage of spectral decomposition to move our diagnosis problem to a standardized space in which we can assess significance formally (subject to suitable handling of multiple testing, testing strategy etc.). However, when we examine these significant features back in the original data space, in general all p components could contribute to their size. Thus we still need a strategy for identifying the important components back now in a space for which we do not have a natural measure of scale.

There are numerous methods which have been suggested in a variety of multivariate settings for handling similar problems and attempting to find combinations of original variable. All the approaches here are aimed to seek simplicity in interpreting principal components, hence, a clear definition of "simplicity" is absolutely necessary (Rousson and Gasser, 2004). Hausman (1982) modify principal components loadings to -1, 0 and 1 to simplify the interpretation of principal components. Vines (2000) proposed the usage of approximate components with integer value to assist the interpretation whereas Jolliffe and Uddin (2000) shrink the loadings of principal components towards 0. Chipman and Gu (2005) introduced two types of constraint for the coefficients of principal components, homogeneity and sparsity constraints. These two constraints will either make the new components closer or more orthogonal to the original directions. Trendafilov and Vines (2009) adapted the approach proposed by Chipman and Gu (2005) where the loadings are classified into homogeneous, contrast and sparse. Vines (2000) proposed the use of drastic rounding (Jackson, 1991) in which can be applied with the approach proposed by Jeffers (1967) where the loadings of principal components which are less than 70% are set to zero.

In our case, we will also rely on a drastic rounding (Jackson, 1991) as described in Vines (2000) that is by setting the loadings to zero when a certain condition is violated. We will assess whether we can determine a suitable threshold for the application of rounding based on an examination of the root mean square error (RMSE) of discrepancy of generating aberrant mean and the back-transformed components in **y**-space in example according to a simple test of significance in the space with identity covariance matrix. The improved application of the proposed method is presented below. The choice of initial threshold values for the method based on the percentage points of N(0,1) probability distribution for vectors in **y**-space and the procedure to determine the threshold values in **x**-space are explained in Section 7.2. A few examples are given in Section 7.3 which illustrate the application of the threshold values obtained in Section 7.2. The results are discussed in Section 7.4.

7.2 Threshold Value in y-space

Firstly we determine the threshold values for the loadings of vector **y** of the proposed diagnostic approach. Vector **y** is the transformed vector [6.2] in a new data space which denoted in the previous chapter **y**-space. As a result of the transformation in [6.2], the transformed variables follow Normal distribution with zero mean and unit variance, $y_i \sim N_1(0,1)$ where $i = 1, 2, \ldots, p$ (Mardia, Kent & Bibly,1994).

7.2.1 The Quantiles of the Loadings Distribution

The threshold values are determine by the α/2 and (1-α/2) of the ordered loadings of vector **y**. For the illustration, α = 0.05 is used. This is a theoretical result and no simulation is needed. The theoretical cut point values for the vector in **y**-space is based on the 2.5% and 97.5% quantile of the standard normal distribution which are -1.96 and 1.96. These two values are the lower and the upper limit of the loadings. Any loading value falls beyond the interval of ± 1.96 is suspected of being the responsible variables(s) for the OOC signal triggered by a multivariate control chart. A 'simplified vector', $\mathbf{y}^*$, is introduced where any values in vector **y** as defined in [6.2],

$$\mathbf{y} = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{x}$$

within the range of the interval of ± 1.96 is set to 0.

7.2.2 The Threshold Values

The threshold values in the **x**-space are obtained by considering the distribution of the back-transformed significance indicator vector, $\mathbf{y}^*$, as follows

$$\mathbf{x}^* = \boldsymbol{\Sigma}_0^{1/2}\mathbf{y}^* \qquad [7.1]$$

where

$\mathbf{\Sigma}_0^{1/2}$ = the square root of known covariance matrix

$\mathbf{y}^*$ = the simplified vector $\mathbf{y}$ in $\mathbf{y}$-space

This can only be achieved if we have knowledge, or good estimation, of the covariance matrix, so cannot be given in general terms for all situations.

The values are determined by following a procedure given below. Say, the threshold values are for the investigation on multivariate data with p = 4

Step 1: Generate 4000 observations using rnorm(0,1).

Step 2: Form 1000 4-dimensional vectors from values obtained in Step 1.

Step 3: Simplify the vectors and then back-transform them according to covariance matrix $\mathbf{\Sigma}_0$.

Step 4: Sort the loadings of the vectors in an ascending order separately with respect to variables. So, in this case we have 4 variables and we should get 4 sets of threshold values for the **x**-space.

Step 5: Get the sample $\alpha/2 = 0.025$ and $(1-\alpha/2)=0.975$ quantiles from the distributions of the loadings obtained in **Step 4**.

In this study, we consider two covariance matrices, $\mathbf{c_1}$ and $\mathbf{c_2}$, for illustration. The threshold values obtained for each variable with respect to the covariance matrix in two decimal places are given below.

Table 7.1: The threshold values for **x**-space with respect to the covariance matrix

	$\mathbf{c_1}$		$\mathbf{c_2}$	
	lower limit	upper limit	lower limit	upper limit
var. 1	-1.78	0.87	-1.89	0.72
var. 2	-1.75	0.98	-1.90	0.79
var. 3	-0.80	1.78	-0.81	1.84
var. 4	-1.91	0.76	-1.85	1.27

The assessment of the extended LD method and the examples illustrated in later sections used the threshold values given in Section 7.2.1 (for **y**-space) and in Table 7.1 (for **x**-space).

7.2.3 Power Assessment

A dataset of 5000 out-of-control (OOC) multivariate observations is generated from the mvrnorm() procedure with contaminated mean vector and known covariance matrix. A selective combinations of shift(s) in mean (for the contaminated mean vector) is used in the assessment of the performance for the extended LD method. The procedures outlined in Sections 7.2.1 and 7.2.2 are followed and the result obtained is given below.

Table 7.2: The power (%) of the LD method with threshold values

	Shift(s) in mean	LD with Threshold Values var.1	var.2	var.3	var.4
$\mathbf{c}_1$	(2.5, 0, 0, 0)'	99.86	56.66	5.06	16.8
	(-3, -3, 0, 0)'	78.46	79	59	13.98
	(3, 2.5, 0, 0)'	91.96	80.72	20.3	57.02
	(-3, 2.5 ,0, 0)'	90.34	94.82	15.04	15.96
$\mathbf{c}_2$	(2.5, 0, 0, 0)'	96.32	45.6	38.9	39.74
	(-3, -3, 0, 0)'	83.7	73.42	8.64	13.92
	(3, 2.5, 0, 0)'	92.44	51.58	53.64	35.08
	(-3, 2.5, 0, 0)'	89.1	93.9	25.06	24.28

The performance of the LD method is assessed based on the power definition given in Section 5.6. The power (in section 5.6) is defined as the percentage of times a diagnostic method identified variable k, k = 1, 2 as the correct aberrant variable. The power of the LD method without the threshold values presented in Section 5.6 for the selected shift(s) in mean is given below.

Table 7.3: The power assessment of the LD method without threshold values

		LD			
	Shift(s) in mean	var.1	var.2	var.3	var.4
c_1	(2.5, 0, 0, 0)'	95.54	4.38	0	0.08
	(-3, -3, 0, 0)'	37.3	45.92	12.44	4.34
	(3, 2.5, 0, 0)'	69.42	16.38	9.28	4.92
	(-3, 2.5 ,0, 0)'	53.7	46.3	0	0
c_2	(2.5, 0, 0, 0)'	92.86	0.74	1.3	5.1
	(-3, -3, 0, 0)'	64.6	32.18	2	1.22
	(3, 2.5, 0, 0)'	91.92	4.62	2.42	1.04
	(-3, 2.5, 0, 0)'	90.98	3.26	0.04	5.72

We do not intend specifically to compare the performance of the LD method (as shown in Tables 7.2 and 7.3) with or without the threshold values since they are directed towards slightly different tasks. What we aim to do is to highlight the strengths as well the weakneses of the proposed extended LD method. For the case with one aberrant variable, the LD method with threshold values has shown a very high power in identifying variable 1 as the aberrant one. It is, of course, also better than the LD method in the sense that LD just chose the largest without any idea of whether it was significantly large. For the cases with two aberrant variables, the LD method with threshold values shows a high power in detecting both aberrant variables, variables 1 and 2. Thus it also has a fair chance of finding the second aberrant variable. The power in identifiying the second aberrant variable is much higher when the shifts in mean are counter-correlational.

However, the proposed extended method has one problem. The chance of finding a mistaken variable is very high for the LD method with threshold values. As shown in Table 7.2, a lot of cases for variables 3 and 4, the percentage of identification is 20% or more. Apparently, this problem needs further investigation in the future.

7.3 Examples

The application of threshold value in determining the responsible variable(s) for an OOC signal is examined in three separate sub-sections. The first section, 7.3.1, considers an observation with one aberrant variable whereas Sections 7.3.2 and 7.3.3 look at an observation with two aberrant variables in the same and opposite directions, respectively. The examples will illustrate how to single out variable x_1 (Section 7.3.1) or variables x_1 and x_3 (Sections 7.3.2 and 7.3.3) as the responsible variable(s) that caused the OOC signal in a multivariate control chart using the approach discussed in Section 7.2.1.

7.3.1 One Aberrant Variable

Say $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$ is an out of control (OOC) observation or the multivariate observation of dimension *p* that falls beyond the upper control limit of a multivariate control chart with x_1 is the responsible variable. Basically, we have a contaminated mean vector μ = (3,0,0,0)', in which variable 1 has been contaminated (by contaminant = 3) and the OOC observation vector **x** is transformed to vector **y** under two non equi-correlation matrices used in Chapter 5 and given again below,

$$\mathbf{c_1} = \begin{pmatrix} 1 & 0.8 & 0.55 & 0.6 \\ 0.8 & 1 & 0.65 & 0.5 \\ 0.55 & 0.65 & 1 & 0.6 \\ 0.6 & 0.5 & 0.6 & 1 \end{pmatrix}; \quad \mathbf{c_2} = \begin{pmatrix} 1 & 0.2 & -0.5 & 0.3 \\ 0.2 & 1 & 0.2 & -0.5 \\ -0.5 & 0.2 & 1 & 0.2 \\ 0.3 & -0.5 & 0.2 & 1 \end{pmatrix}$$

where from equation [6.2]

$$\mathbf{y} = \mathbf{\Sigma}_0^{-1/2}\mathbf{x}; \quad \text{and } \mathbf{y} \sim \mathrm{N}(0, \mathrm{I})$$

Say we have two OOC observations, $\mathbf{x}_1$ and $\mathbf{x}_2$, from situation with correlation matrices $\mathbf{c_1}$ and $\mathbf{c_2}$ respectively

$$\mathbf{x}_1 = (1.84, -0.59, -0.30, -1.19)'$$

and

$$\mathbf{x}_2 = (2.80, -0.22, -0.33, -0.73)'$$

then, we obtain the transformed vectors in **y**-space as given in equation [7.1],

$$\mathbf{y}_{c1} = (3.84, -2.18, 0.16, -2.08)'$$

and

$$\mathbf{y}_{c2} = (5.78, -3.28, 2.99, -3.98)'.$$

Based on the "proposed threshold values" for the loadings of vector **y** (±1.96), the 'simplified vector' are

$$\mathbf{y}_{c1}^{*} = (6.05, -2.67, 0, -2.08)' \quad \text{and} \quad \mathbf{y}_{c2}^{*} = (5.78, -3.28, 2.99, -3.98)'.$$

These vectors both 'correctly pick out' variable x_1 as aberrant. The result of the back transformations of $\mathbf{y}^*$ give us vectors $\mathbf{x}^*$ with elements shown below,

$$\mathbf{x}^*{}_{c1} = (1.81, -0.63, -0.44, -1.23)' \quad \text{and} \quad \mathbf{x}^*{}_{c2} = (2.80, -0.22, -0.34, -0.73)'$$

The back-transformed vectors are more or less similar to the original vectors. Variable $x_1$still has the largest value. Variable x_1 has a value considerably larger than the other variables in both vectors, $\mathbf{x}^*{}_{c1}$ and $\mathbf{x}^*{}_{c2}$. What we just need now is a 'cut off rule' in **x**-space in order to determine that variable x_1 is the one and only and the most probable aberrant variable that caused the OOC signal. Using the threshold values given in Table 7.1, variable 1 has been successfully identified as the only variable with a loading outside the critical values, i.e. it is correctly identified as the sole aberrant variable.

7.3.2 Two Aberrant Variables (Same Direction)

Let say the same tests in Section 7.3.1 are repeated and the contaminated mean vector is μ = (3, 0, 3, 0)' . This time the OOC observations are

$$\mathbf{x}_1 = (1.84, -0.59, 2.70, -1.19)'$$

and

$$\mathbf{x}_2 = (4.01, -0.13, 1.86, -0.71)'$$

then, we obtain the transformed vectors,

$$\mathbf{y}_{c1} = (3.75, -3.38, 4.25, -3.06)'$$

and

$$\mathbf{y}_{c2} = (10.09, -5.94, 7.93, -6.94)'$$

After simplifying these vectors according to Normal quantiles and back transformation, we obtain

$$\mathbf{x}^*{}_{c1} = (1.84, -0.59, 2.70, -1.19)'$$

and

$$\mathbf{x}^*{}_{c2} = (4.01, -0.13, 1.86, -0.71)'$$

Variables x_1 and x_3 are still the two largest values. But again, what we need now is a ‘cut off rule’ in **x**-space. This time the aim is to separate both variables, x_1 and x_3, from the others and to determine that both of the variables are significantly aberrant and thus the probablecause of the OOC signal. Using the threshold values given in Table 7.1, both variables are successfully identified as the aberrant variables.

7.3.3 Two Aberrant Variables (Opposite Direction)

Say the OOC observations from contaminated mean vector, μ = (-3, 0, 3, 0)' are

$$\mathbf{x}_1 = (3.43, -0.64, -3.83, -1.05)'$$

and

$$\mathbf{x}_2 = (4.01, -0.13, -4.14, -0.71)'$$

then, we obtained the transformed vectors,

$$\mathbf{y}_{c1} = (6.55, -1.99, -4.73, -1.28)'$$

and

$$\mathbf{y}_{c2} = (4.19, -1.06, -2.55, -1.70)'$$

The vectors have a different simplified vectors which are

$$\mathbf{y}^*_{\mathbf{c1}} = (6.55, -1.99, -4.73, 0)' \quad \text{and} \quad \mathbf{y}^*_{\mathbf{c2}} = (4.19, 0, -2.55, 0)'$$

These are ‘good’’ too since both variables, x_1 and x_3, are ‘correctly picked out’ as aberrant variable. After back transformation on both vectors, we obtained

$$\mathbf{x}^*{}_{c1} = (3.78, -0.41, -3.45, 0.12)'$$

and

$$\mathbf{x}^*{}_{c2} = (4.59, 0.31, -3.65, 0.50)'$$

Variables x_1 and x_3 are still the two largest absolute values and the sign of the value of the aberrant variables are still the same, indicating the direction of the shifts is retained. Based on the threshold values given in Table 7.1, both variables, 1 and 2, are successfully identified as the aberrant variables.

7.4 Application to Real Data

The procedures shown in Section 7.2 and the examples given in Section 7.3 are under the assumption that the covariance matrix, $\mathbf{\Sigma}$ is known. In real applications, it is not always known and in most cases, the $\mathbf{\Sigma}$ needs to be estimated from the in-control dataset obtained from the Phase I process control monitoring. In this phase, the causes of out-of-signals on a multivariate control chart have been identified and dealt with. Thus, the out-of-control process is assumed to have been brought into in-control condition before the end of the Phase I process control monitoring. The remaining in-control observations from this phase are used to estimate the covariance matrix, $\mathbf{\Sigma}$, for later use in Phase II process control monitoring. It is also assumed that we can readily identify when we are back 'in-control' and so the estimation of $\mathbf{\Sigma}$ is based only on reliable observations. The procedure for the application of the LD method with new threshold values using the estimated covariance matrix is explained in the following sub-section.

7.4.1 Procedure

A dataset with 50 in-control multivariate observations are simulated using mvrnorm procedure with seed number 2015. The same correlation matrices, $\mathbf{c_1}$ and $\mathbf{c_2}$, are used in the mvrnorm() procedure with mean vector, $\boldsymbol{\mu}_0$. The covariance matrix, $\mathbf{\Sigma}$ is then estimated based on these in-control datasets. The estimated covariance matrices based on the in-control observations simulated using $\mathbf{c_1}$ and $\mathbf{c_2}$ are $\mathbf{S}_{c1}$ and $\mathbf{S}_{c2}$, respectively. Example of estimated covariance matrices are given below.

$$\mathbf{S}_{\mathbf{c}1}=\begin{pmatrix} 1.27 & 1.01 & 0.53 & 0.58 \\ 1.01 & 1.11 & 0.54 & 0.41 \\ 0.53 & 0.54 & 1.11 & 0.79 \\ 0.58 & 0.41 & 0.79 & 0.98 \end{pmatrix}$$

$$\mathbf{S}_{\mathbf{c}2}=\begin{pmatrix} 0.94 & 0.18 & -0.77 & 0.13 \\ 0.18 & 1.03 & -0.12 & -0.83 \\ -0.77 & -0.12 & 1.45 & 0.59 \\ 0.13 & -0.83 & 0.59 & 1.51 \end{pmatrix}$$

It should be noted that these estimates will not generally be correlation matrices (as are $\mathbf{c}_1$ and $\mathbf{c}_2$) but this does not affect the procedure which requires a covariance matrix is the fact that $\mathbf{c}_1$ and $\mathbf{c}_2$ are correlation matrices is a specified case and we revert here to the more general formulation given in the original theory (Section 5.3).

For computational simplicity, we will apply the LD method still using the same theoretical threshold values showed in Table 7.1. Further, distribution of the variables is still assumed to be Normal and not to follow the Student-t distribution, even though the covariance matrix is estimated.

Procedures in applying the LD method using an estimated covariance matrix

Suppose $\mathbf{x} = (x_1, x_2, x_3, x_4)'$ is a multivariate observation that triggers an out-of-control signal on a multivariate control chart. In order to diagnose the signal using the LD method with threshold values, $\mathbf{x}$ needs to undergo the steps explained below.

Step 1: Vector $\mathbf{x}$ is transformed to $\mathbf{y}$-space and becomes $\mathbf{y}$-vector as given in [6.2] where $\mathbf{\Sigma}$ is unknown and estimated from the in-control dataset in Phase I process control monitoring. In this section, the estimated covariance matrices used are $\mathbf{S}_{\mathbf{c}1}$ and $\mathbf{S}_{\mathbf{c}2}$. The inverse square root of the estimated covariance matrix is used in the transformation.

Step 2: The $\mathbf{y}$ -vector is then simplified according to the limits given in Section 7.2.1.

Step 3: The simplified vector $\mathbf{y}^*$ is back-transformed to vector $\mathbf{x}^*$ in the $\mathbf{x}$-space using the square root of the estimated covariance matrices. Aberrant variables are identified by comparing the relevant terms with the limits shown in Table 7.1.

7.4.2 Examples

The same observations in the illustration of the application of the LD method with the threshold values in Sections 7.3.1, 7.3.2 and 7.3.3 are used for the following examples. The observations are simulated from two covariance matrices, $\mathbf{c}_1$ and $\mathbf{c}_2$ but this time, the application of the LD method is based on the estimated covariance matrices, $\mathbf{S}_{c1}$ and $\mathbf{S}_{c2}$. Covariance matrix $\mathbf{S}_{c1}$ is estimated based on the 50 in-control multivariate observations simulated using mvrnorm() procedure with covariance matrix, $\mathbf{c}_1$. Whereas, covariance matrix $\mathbf{S}_{c2}$ is estimated based on the 50 in-control multivariate observations simulated using mvrnorm() procedure with covariance matrix, $\mathbf{c}_2$.

One Aberrant Variable

The shifted mean vector is (3, 0, 0, 0)' for this situation and the observation vectors simulated from covariance matrices $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively are

$$\mathbf{x}_1 = (1.84, -0.59, -0.30, -1.19)'$$

and

$$\mathbf{x}_2 = (2.80, -0.22, -0.33, -0.73)'$$

Then, we obtain the transformed vectors in $\mathbf{y}$-space as given in equation [7.1],

$$\mathbf{y}_{c_1} = (4.03, -2.91, 0.77, -2.55)'$$

and

$$\mathbf{y}_{c_2} = (7.21, -3.95, 4.11, -4.59)'.$$

The vectors are simplified by comparing the coefficients of the variables against the limits ± 1.96 as given in Section 7.2. The simplified vectors are

$\mathbf{y}^*_{c_1}$ = (4.03, -2.91, 0, -2.55)'

and

$\mathbf{y}^*_{c_2}$ = (7.21, -3.95, 4.11, -4.59)'.

The simplified vectors are then back-transformed to **x**-space and the final vectors are

$\mathbf{x}^*_{c_1}$ = (1.71, -0.76, -1.02, -1.50)'

and

$\mathbf{x}^*_{c_2}$ = (2.80, -0.22, -0.33, -0.73)'.

By applying the threshold values given in Table 7.1, variable 1 is identified as the only aberrant variable for both cases with $\mathbf{c_1}$ and $\mathbf{c_2}$. For interest, we can compare $\mathbf{x}^*_{c_1}$ and $\mathbf{x}^*_{c_2}$ with their equivalent from Section 7.3.1 which used $\mathbf{c}_1$ and $\mathbf{c}_2$ as known covariance matrices for $\mathbf{x}_1$ = (1.84, -0.59, -0.30, -1.19)' and $\mathbf{x}_2$ = (2.80, -0.22, -0.33, -0.73)' respectively. As we can see, there is very little difference and, unsurprisingly, we reach the same conclusion regarding aberrance.

Two Aberrant Variables (Same Direction)

The shifted mean vector is (3, 0, 3, 0)' for this situation and the observation vectors simulated from covariance matrices $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively are

$\mathbf{x}_1$ = (1.84, -0.59, 2.70, -1.19)'

and

$\mathbf{x}_2$ = (4.01, - 0.13, 1.86, - 0.71)'.

Then, we obtain the transformed vectors in **y**-space as given in equation [7.1],

$\mathbf{y}_{c_1}$ = (4.38, -3.97, 5.01, -4.50)'

and

$\mathbf{y}_{c_2}$ = (12.85, - 6.89, 9.62, - 8.24)'

The vectors are simplified by comparing the coefficients of the variables against the limits ± 1.96 as given in Section 7.2. The simplified vectors are

$$\mathbf{y}^*_{c_1} = (4.38, -3.97, 5.01, -4.50)'$$

and

$$\mathbf{y}^*_{c_2} = (12.85, - 6.89, 9.62, - 8.24)'.$$

The simplified vectors are then back-transformed to **x**-space and the final vectors are

$$\mathbf{x}^*_{c_1} = (1.84, -0.59, 2.70, -1.19)'$$

and

$$\mathbf{x}^*_{c_2} = (4.01, - 0.13, 1.86, - 0.71)'.$$

By applying the threshold values given in Table 7.1, variable 1 and variable 3 are correctly identified as aberrant variables for both cases. Again, we have reached similar conclusion to those in Section 7.3.2 where $\mathbf{c}_1$ and $\mathbf{c}_2$ are assumed known.

Two Aberrant Variables (Opposite Direction)

The shifted mean vector is (3, 0, -3, 0)' for this situation and the observation vectors simulated from covariance matrices $\mathbf{c}_1$ and $\mathbf{c}_2$ respectively are

$$\mathbf{x}_1 = (3.43, -0.64, -3.83, -1.05)'$$

and

$$\mathbf{x}_2 = (4.01, - 0.13, -4.14, - 0.71)'.$$

Then, we obtain the transformed vectors in **y**-space as given in equation [7.1],

$$\mathbf{y}_{c_1} = (6.22, -3.22, -4.11, -0.72)'$$

and

$$\mathbf{y}_{c_2} = (4.84, - 1.79, -1.30, -1.76)'$$

The simplified vectors are

$$\mathbf{y}^*_{c_1} = (6.22, -3.22, -4.11, 0)'$$

and

$$\mathbf{y}^*_{c_2} = (4.84, 0, 0, 0)'.$$

and the final vectors are

$$\mathbf{x}^*_{c_1} = (3.61, -0.56, -3.54, -0.43)'$$

and

$$\mathbf{x}^*_{c_2} = (4.03, \ 0.76, -2.10, \ 0.84)'.$$

Again, variable 1 and variable 3 are correctly identified as aberrant variables in both cases as was the cases in Section 7.3.3.

7.4.3 The Effect of Estimation

In the examples above we saw little effect from the estimation of $\mathbf{c}_1$ and $\mathbf{c}_2$. However, that was based on single examples. We investigate the effect of the estimated covariance matrices more fully by studying the change in the estimated power with respect to the corresponding Frobenius distance between the theoretical and the estimated covariance matrices over a larger simulation. Figures 7.1, 7.2 and 7.3 exhibit the relationship between the estimated powers and the corresponding Frobenius distance for the same combinations of shift in mean vector as used in the examples in Section 7.4.2.

We repeat here the procedure in Section 7.4.1 for 100 separate estimates of $\mathbf{c}_1$. The Frobenius distance is defined as the distance between $\mathbf{c}_1$ and it estimates, $\mathbf{S}_{\mathbf{c}1}$ or in other words, it measures how much the estimates of $\mathbf{c}_1$ differ from $\mathbf{c}_1$ itself. Note that although $\mathbf{c}_1$ is a correlation matrix, $\mathbf{S}_{\mathbf{c}1}$ is a more general, covariance matrix, being based on actual observations.

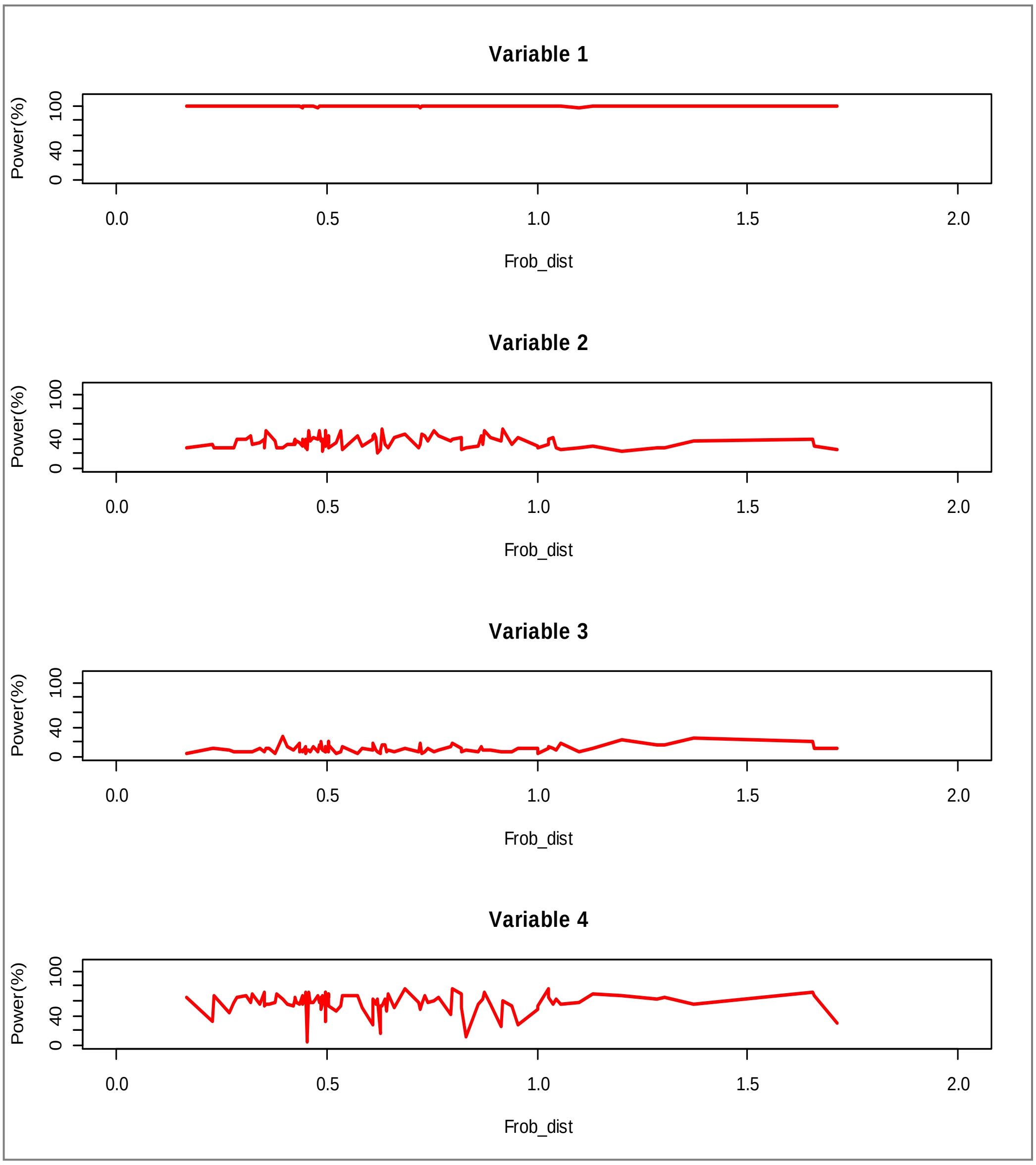


Figure 7.1: Identification with respect to Frobenius distance for case with one aberrant variable, (3, 0, 0, 0)'.

The plots in Figure 7.1 shows the estimated power for each variable with respect to the corresponding Frobenius distance. From the patterns shown by the plots, the true aberrant variable, in this case is variable 1, has consistent values of estimated power compared to the other variables. The plots of variables 2, 3 and 4 fluctuated more with much lower values for the percentage of identification though still disappointingly high being false alarm. There is no clear dependence of power on Frobenius distance. It is perhaps because all estimates are reasonably good.

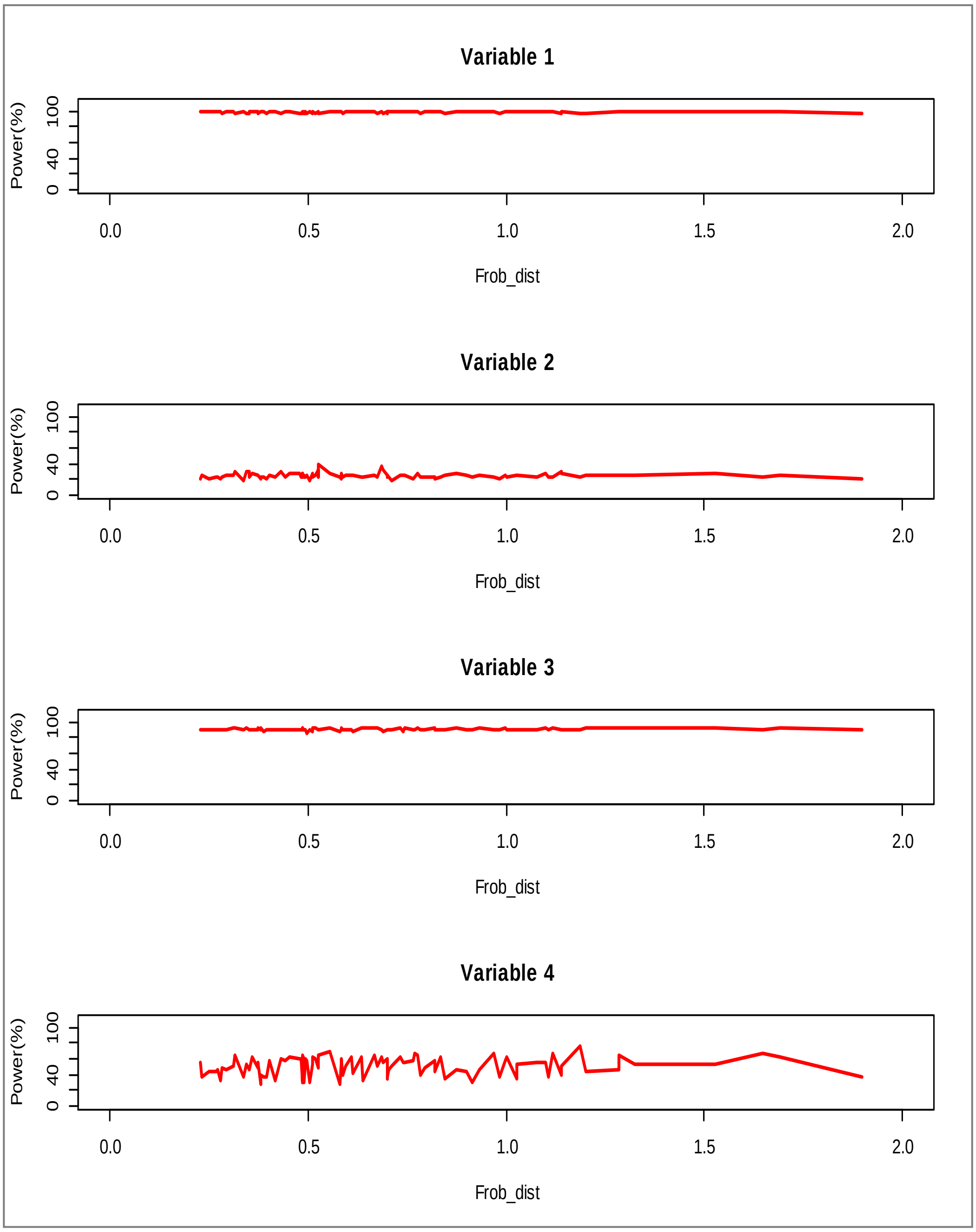


Figure 7.2: Identification with respect to Frobenius distance for cases with two aberrant variables in same direction, (3, 0, 0, 0)'.

Plots in Figures 7.2 and 7.3 show similar patterns to Figure 7.1. The plots of the aberrant variables, 1 and 3, are smoother with estimated power considerably higher than the other two variables. The estimated power for variable 3 in Figure 7.3 is slightly higher than the one shown by the same variable 3 in Figure 7.2. Again, little dependence on Frobenius distance is seen.

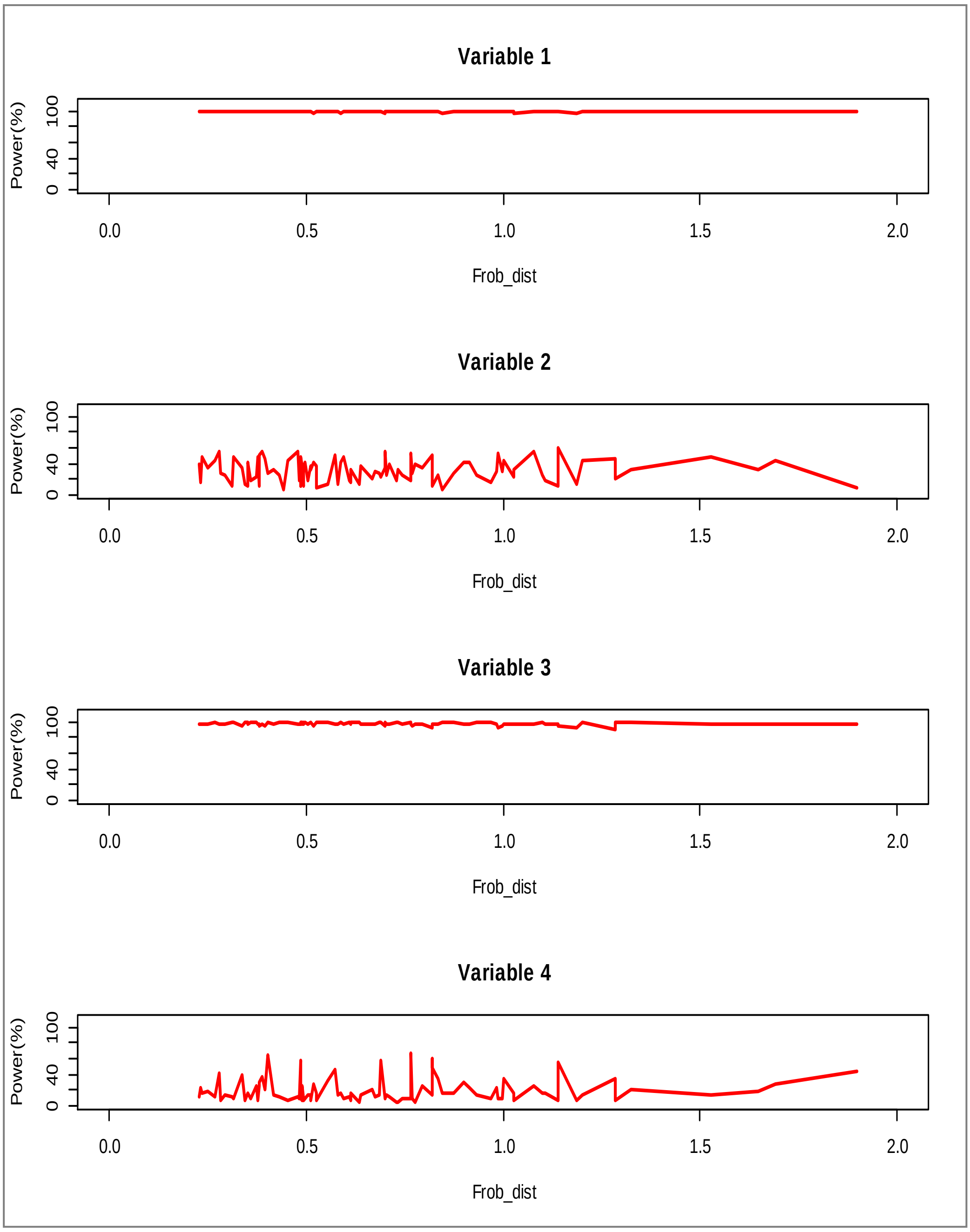


Figure 7.3: Identification with respect to Frobenius distance for cases with two aberrant variables in opposite direction, (3, 0, -3, 0)'.

7.5 Discussion

The LD method with threshold values has shown a good performance in identifying aberrant variables. The aberrant variables are identified correctly in many cases. The direction of the shifts are retained, in which is useful for follow-up action to get a process back to an in-control state. The LD method with threshold values also makes the task of identifying more than one aberrant variables at a time possible. This is truly a valuable advantage and better than the common practice where aberrant variables are identified one by one, the second being identified after the first has been removed.

The application of of the LD method on real data with estimated covariance matrix has shown promising results. The estimated power for the true aberrant variables consistently is very high throughout all Frobenius distances, whereas the non-aberrant variables have considerably lower estimated power (though further work is needed to address the level of these false alarm).

CHAPTER 8

SUMMARY AND FUTURE WORK

8.1 Introduction

This final chapter is divided into two parts; the first part summarizes the findings featured in Chapters 3, 4, 5, 6 and 7 and the second part outlines directions for future work. We will revisit these chapters, with outlining the aims and the objectives of the investigation in the beginning of the summary. The findings, conclusions and suggestions wherever available will be presented at the end to conclude the summary of each chapter.

8.2 Summary and Conclusion

The comparative study reported by Das and Prakash (2008) is the main topic discussed in Chapter 3. The aim is to fully utilise the findings reported by Das and Prakash to assist this study in conducting a new comparative study in Chapter 4 and an assessment of a new proposal in Chapter 5. Das and Prakash (2008) succeeded in conducting a detailed investigation of the performance of several diagnostic methods by performing the methods on various combinations of shifts in mean vector. A wider range of shifts in mean outlined in this simulation work is found to be the main strength of the study, exceeding the previous comparative studies by other researchers (Doganaksoy, Faltin and Tucker, 1991; Maravelakis et al., 2002) where the options presented are very few and limited. Because of this, the same range of shifts in mean vector has been adapted in the comparative study in Chapter 4. The simulation results revealed, with additional observations from this study, certain patterns of performances by the diagnostic methods which are believed related to not only to the size of the shift(s) in the mean vector but also to the combination of shifts in mean vector as well as the correlation between variables. The methods proposed by Hawkins (1991) and Murphy (1987) have shown some inconsistencies in their performance and are on many occasions unable to respond to the shifts in mean especially when the shifts are not in accordance with the correlation structure. The methods proposed by Mason, Tracy and Young (1995) and Doganaksoy, Faltin and Tucker (1991) have shown reliable performance and the pattern of their performance is more predictable. The method proposed by Doganaksoy, Faltin and Tucker (1991) (DFT) is chosen to be included in the comparison studies in Chapters 4 and 5 for this reason. It is chosen over the method proposed by Mason, Tracy and Young (1995) for its greater simplicity and practicality.

Apart from the strength, the investigation of Das and Prakash (2008) also carries several weaknesses such as some terms used in the investigation not being clearly defined. The most crucial is the definition of '*n(detect)'* in equation [4.4] for the computation of power for a diagnostic method for cases with more than one mean shifted. As a result, the meaning of "Power" for assessing the performance of a diagnostic method is left for the reader to assume. The performance criterion, which related to the power measurement, also lacks clarity in terms of its limits or ranges. As a result, ambiguities arise in describing the level of performance of the diagnostic methods studied. This study has taken necessary precautions to avoid doing the same. The "Power" of a diagnostic method has been clearly defined in the following chapters and suggestions for improving the assessment of diagnostic methods in comparative studies are given in Section 8.3 as one element of the future work.

Chapter 4 is conducted as a preliminary comparison study for an extended comparison in Chapter 5. As a preliminary study, this chapter includes a selected diagnostic method from the study by Das and Prakash and compares its performance with another method called Ratio method. The method on the computation of ratio was taken from Maravelakis et al. (2002) but this study did not follow the exact approach proposed by them. A performance comparison between these two methods is based on a modified definition of "Power" in [4.4], and the assessment of successful diagnostics based on the largest value. The ranking of the K_ind [2.6] is used for the DFT method instead of the K_Bonf [2.7] in order to compute the power based on the number of out-of-control (OOC) signals produced by a multivariate control chart. For the Ratio method, the ratios are ranked and not plotted as proposed in the original procedure. This study has revealed a useful finding that is worthy of further investigation. Based on the simulation results in this chapter, a peculiar performance is shown by the Ratio method under a certain type of correlation matrix. Since the ratio's computation method proposed by Maravelakis et al. (2002) used the loadings in eigenvectors, it is suspected that the peculiar performance is due to a particular property of the correlation matrix itself. The problem arises in using the loading of eigenvectors from principal component analysis and has been highlighted before by several researchers (Jackson, 1993; Mehlman et al., 1995; Zhang et al., 1997 and Quadrelli et al., 2005). The problem is related to the situation where two eigenvalues are close together (Quadrelli et al., 2005 and Zhang et al., 1997). The computation of a ratio of the corresponding eigenvectors or the 95% confidence limits of the eigenvectors has been suggested (Jackson, 1993; Mehlman et al., 1995 and Quadrelli et al., 2005) as a "checking" step. This study investigated the risk of "swapped" or "mixed eigenvectors by studying the inner product of the first theoretical and sample eigenvectors and presented a graphical analysis of the frequency distribution of the inner products with fixed class intervals. Two correlation matrices used in this study, $\mathbf{c_1}$ and $\mathbf{c_2}$, were tested, and the frequency distribution of the inner products with respect to a number of fixed class intervals depicted in Figure 4.5 has uncovered two things. The first correlation matrix, $\mathbf{c_1}$, has most of its inner products close to +1 or -1 (Figure 4.5(a)). The Ratio method has shown a good performance on this correlation matrix. On the other hand, for correlation matrix $\mathbf{c_2}$, the inner products of the first theoretical and sample eigenvector are distributed throughout all the range [-1, 1] (Figure 4.5(b)). The Ratio method has shown a peculiar performance on this correlation matrix. The eigenvalues for the first two eigenvectors as shown in Table 4.3, for $\mathbf{c_1}$ these are far apart whereas for $\mathbf{c_2}$, they are very close together. The result of this investigation has reconfirmed the existence of the problem regarding the possibility of "mixed eigenvectors" when the eigenvalues between a pair of eigenvectors are close together and this study has successfully presented another way to identify the problem.

Chapter 5 has two objectives. The first objective is to propose a new method to assist in the interpretation of the multivariate control chart signals and the second objective is to carry out a new comparison study which includes the proposed method and the two methods from Chapter 4. The proposed method applies a union intersection technique (UIT), which procedure is validated by the Cramer-Wold theorem where the connection between the set of all one-dimensional projections and the multivariate distribution is established. The union of the rejection regions given in [5.7] provides the basis of the union intersection strategy which is being applied to the diagnosing of the OOC signal from a multivariate control chart. A test statistic is used to determine whether a multivariate observation is sufficiently extreme before a process is declared as ‘out-of-control’. In order to determine which raw variable (or which combination of them) is responsible for the signal, this chapter performs a union intersection test to obtain a UIT test-statistic and a UIT direction. Basically, a test statistic is referred to a set of critical values and a decision is made whether it is ‘significantly’ extreme or not. Generally these critical values will not depend upon that covariance of the observations (i.e it will be χ^2 if the covariance is known or Hotelling’s T^2 if the covariance is unknown because of the underlying assumption of normality) but in this chapter, the covariance is assumed known and the test statistic is referred to the χ^2 critical value. Then we look at the coefficients of the variables in the UIT direction and pick out the variable (or variables) whose coefficients are “sufficiently large”. The problem is we don’t know what “sufficiently large” is and typically this will depend upon the covariance. Because of this reason, the comparison study in this chapter has considers six different correlation matrices in which four of them are equi-correlation matrices with ρ = -0.2, 0.2, 0.5 and 0.8 (adapted from the correlation matrices used in Das and Prakash (2008)) and the other two are non equi-correlation matrices with the first one, $\mathbf{c_1}$, consisting of all positive values and the second $\mathbf{c_2}$, having mixed sign values (correlation matrices used by Doganaksoy, Faltin and Tucker (1991)). The comparison study in this chapter fully used the combinations of shifts in mean by Das and Prakash (2008). This study has taken one step ahead in re-defining the performance assessment proposed by Das and Prakash (2008) specifically for the cases with more than one aberrant variable for example, in the cases with two variable means are shifted in a mean vector, the “Power” is defined as the number of times a certain diagnostic method can correctly identify at least one of the aberrant variables.

This study also proposed other options in defining the “Power” for a comparison study of diagnostic methods which will be discussed further in Section 8.3. The results of the simulation study has been presented in such way that the effect of the combination of shifts in mean can be seen clearly with respect to whether the shifts are in accordance with the correlation structure or in counter-correlational. The newly proposed method has shown a very good performance and in many cases better than the other two methods especially for the cases with one aberrant variable or when one of the shifts in mean is much larger than the other, regardless of the structure of the correlations between the variables in four dimensional data sets. This new method is very much recommended when there is a priori knowledge that one of the variables has a much higher tendency than the other variables to deviate easily from the in-control mean. The proposed method also shows a very good performance in the situation where the shifts are not in accordance with the correlation structure. It has shown a considerably higher performance than the Ratio method in all of the combinations of shifts whenever the shifts are not in accordance with the correlation structure. For the shifts in mean with small magnitudes, or for combinations of small and intermediate shifts, the proposed method always showed a significantly higher power than the other two methods when the correlation between variables is moderate positive or strong positive. A good performance is also shown for the cases when the shifts are in accordance with the correlation structure but the power is not as high as the other methods when the shifts are counter-correlational. However, the power of the proposed method drops noticeably when the shifts in mean are close together in magnitude in the situation when the shifts are in accordance with the correlation structure. The drop in power is not that much when the correlation between the variables is low. The investigation in Section 5.6 has revealed that the proposed method has a higher tendency to detect a larger shift more frequent than a smaller shift compared to the other two methods. For the cases with non equi-correlation matrices, the proposed method shows a very good performance when the shifts are not in accordance with the correlation structure when all the correlation between variables are positive regardless of the combinations of shifts in mean.

The selection of the covariance matrices proposed by the other researchers is considered somewhat limited. In investigating the performance of the diagnostic methods, this study used additional randomly generated correlation matrices. The estimated power of the diagnostic methods is compared across this broader class. The proposed method has shown a very good performance in many cases, but sometimes the performance is only fair or even very poor under a few of the randomly generated correlation matrices. This underlines the fact that it is the relationship of a shift with the correlation structure which determines how easy it is to detect.

Chapter 6 aims to show how the the UIT assessment in Chapter 5 can be simplified by application of spectral decomposition. In this chapter, we again assume that we know the covariance matrix. In many situations in quality control, we deal with large datasets and we have a very good sample estimate of the covariance even if we don’t know it exactly. This means that we can transform all our observations (with a linear function) so that we have observations from iid N(0,1) variables. So, in this chapter, instead of working on the raw data, we will immediately transform the observations to iid N(0,1)s and perform the UIT on them. In this way, we will get a UIT test statistic (which will be identical to that from the raw data) and a UIT direction (which will be different). This time, we look at the coefficients of the (transformed) variables in the UIT direction obtained and pick those that are 'sufficiently large'. A few examples in 2 and 4 dimensional observations are presented in this chapter and the potentially aberrant variable is fixed as $\mathbf{x} = (1, 1)'$ (for 2 dimensional observation) or $\mathbf{x} = (1, 1, 1, 1)'$ (for 4 dimensional observation). The examples demonstrate that the pattern of the shifts are retained and similar to the shifts fixed in vector $\mathbf{x}$. This investigation makes formalizing the proposed method possible as is explained further in Chapter 7.

In Chapter 7, this study aims to present a formal procedure to determine threshold values for the coefficients of the transformed vectors discussed in Chapter 6. The transformed vectors are called $\mathbf{y}$ vectors in $\mathbf{y}$-space. The threshold values are obtained from iid N(0,1)s distributions. In general, this will yield a linear combination of (some of) the iid N(0,1)s as being the 'cause' of the aberrance. The simplified vector of this linear combination is called ‘significance indicator vector’. A reverse transform is applied to this linear combination to identify what linear combination of the original raw variables is the cause. Thresholds values dependent on the covariance matrix are required to decide which raw variables make significant contribution to this effect. These are determined by simulation of the quantities in the null (no OOC signals) case.

The practicality of this method is extended by demonstrates its use on real data where we must estimate the covariance matrix from in-control sample data. The estimated covariance matrix is used in the implementation of the proposed diagnostic method on out-of-control observations. The effect of the estimation upon the estimated power of the LD method is also studied by computing the Frobenius distance between the theoretical and the estimated correlation matrices. The findings showed that the estimated power of LD method is not much affected by the Frobenius distance.

8.3 Future Work

The performance of the proposed method, LD, is good and in many cases shows a better performance than the other two methods, DFT and the Ratio. The improved version of the LD method has a few advantages over the original version. An identified variable can be proven to be significantly aberrant with the application of the threshold values. Furthermore, more than one aberrant variable can be identified as aberrant at a time.

However, it is noticeable from the results shown in Chapter 7 (Table 7.2), that the LD method with threshold values has a considerable percentage of identification of the non-aberrant variables. This situation indicates a high possibility of making Type II error. A further investigation is obviously needed in order to improve the Type II error without worsening the Type I error.

A further investigation on the performance of the LD method with threshold values with respect to various combinations of shifts in mean is necessary to measure its potential as a good diagnostic method in interpreting a multivariate control chart signal. The same combinations of shifts in mean applied in Chapter 5, is proposed for the further investigation.

Some improvements in measuring the performance of diagnostic methods should be initiated and tested. The power measurement should be the same and applicable to all diagnostic methods. A clear definition is needed for it especially for the cases with more than one aberrant variable. The power of a diagnostic method can be measured in a few ways, i.e., the ability of a method to identify one of the aberrant variables or the ability of a method to identify both or all the deviated variables. A failure in defining the power of a method definitely affects the assessment of the power of diagnostic methods in any comparison study.

REFERENCES

Alt, F.B., (1977). Economic Design of Control Charts for Correlated, Multivariate Observations. Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA.

Alt, F.B., (1984). Multivariate Control Chart. The Encyclopaedia of Statistical Science. Edited by Johnson N.L. and Kots S. John Wiley and Sons, New York, pp. 110-122.

Alt, F.B., (1985). Multivariate Quality Control. Encyclopaedia of Statistical Science. Kotz and Johnson Eds, Vol. 6, John Wiley and Sons, N.Y, pp. 111-122.

Alt, F.B. and Smith, N.D., (1988). Multivariate Process Control. Handbook of Statistics, vol. 7, Krishnaiah PR, Rao CR eds.. Elsevier: Amsterdam, pp. 333–351.

Alwan L.C., (2000). Statistical Process Analysis. McGraw-Hill. New York.

Anderson, T.W., (1984). An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York.

Andrews, D.F., (1972). Plots of High-Dimensional Data. International Biometric Society, 28, pp. 125-136.

Barnard, G. A., (1959). Control Charts and Stochastic Processes. Journal of the Royal Statistical Society: Series B, 21, pp. 239–271.

Bissel A.F., (1969). CUSUM Techniques for Quality Control, Journal of Applied Statistic. 18, pp. 1-30.

Besterfield, D. H., (2004). Quality Control Seventh Edition. Pearson Prentice Hall. New Jersey.

Casella, G. and Berger, R.L., (1990). Statistical Inference Second Edition. Brooks & Cole Publishing Company.

Chipman, A. And Gu, H. (2005). Interpretable Dimension Reduction. Journal of Applied Statistics, Vol. 32, Issue 9, pp. 969-987.

Chua, M.K. and Montgomery, D.C., (1992). Investigation and Characterization of a Control Scheme for Multivariate Quality Control. Quality and Reliability Engineering International, 8, pp. 37–44.

Crossier, R. B., (1988). Multivariate Generalization of Cumulative Sum Quality-Control Schemes. Technometric, 30, pp. 291-303.

Das, N., (2006). Multivariate Diagnostic Tests: A Review of Different Methods. Economic Quality Control, 21, pp. 143-153.

Das, N. And Prakash, V., (2008). Interpreting the Out-of-Control Signal in Multivariate Control Chart- A Comparative Study. International Journal of Advanced Manufacturing Technology, 37, pp. 966-979.

Deming, W.E., (1982). Out of the Crisis Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study, pp. 507.

Deming, W.E., (1993). The New Economics: For Industry, Government, Education Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study, pp. 240.

Doganaksoy, N., Fatlin, F.W. and Tucker, W.T., (1991). Identification of Out-of-Control Quality Characteristic in a Multivariate Manufacturing Environment. Comunication in StatisticTheory and Methods, 209, pp. 2775-2790.

Duncan, A.J., (1974). Quality Control and Industrial Statistics Fourth Edition. Richard D. Irwin, Home Wood, IL.

Ewan, W.D., (1963). When and How to Use Cu-Sum Charts. Technometrics, 5: 1.

Fuchs, C. And Benjamini, Y., (1994). Multivariate Profile Chart for Statistical Process Control. Technometrics, 362, pp. 182-192.

Fuchs, C, Kenett, R.S., (1998). Multivariate Quality Control. Marcel Dekker INC, New York.

Hausman, R. E., (1982) Constrained Multivariate Analysis. In Optimization in Statistics (eds S.H. Zanakis and J.S. Rustagi), Amsterdam: North-Holland, pp. 137-151.

Hawkins, D.M., (1981). A CUSUM for a Scale Parameter. Journal of Quality Technology, 134, pp. 228-235.

Hawkins, D.M., (1991). Multivariate Quality Control Based on Regression-Adjusted Variables. Technometric, 331, pp. 61-75.

Hawkins, D.M., (1993). Regression Adjustment for Variables in Multivariate Quality Control, Journal of Quality Technology, 25, pp. 170–182.

Hayter, A. And Tsui, K., (1994). Identification and Qualification in Multivariate Quality Control Problems. Journal of Quality and Technology, 263, pp. 197-207.

Healy, J. D., (1987). A Note on Multivariate CUSUM Procedures. Technometric, 29, pp. 409-412.

Hinkley, D.V., (1969). Inference About the Intersection in Two-Phase Regression. Biometric, 563, pp. 495-504.

Hotelling, H., (1947). Multivariate Quality Control.Techniques of Statistical Analysis. In Eisenhart, C., Hastay M.W. and Wallis W.A. McGraw Hill, New York, pp. 111-184.

Hotelling, H., (1951). Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Canada, pp. 23-41.

Jackson, J.E., (1959). Quality Control Methods for Several Related Variables. Technometrics, 1, pp. 359-377.

Jackson, J. E., (1980). Principal Components and Factor Analysis Part II-Additional Topics Related to Principal Components. Journal of quality Technology, 131, pp. 46-58.

Jackson, J. E., (1981). Principal Components and Factor Analysis: Part I - Principal Components. Journal of Quality Technology, 14, pp. 201-213.

Jackson, J. E., (1985). Multivariate Quality Control. Communication in Statistics - Theory and Methods, 1411, pp. 2657-2688.

Jackson, J. E., (1991). Series in Probability and Mathematical Statistic. John Wiley and Sons.

Jackson, J. E., (1993). Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology, 74, pp. 2204-2214.

Jeffers, J. N. R., (1967). Two Case Studies in the Application of Principal Component Analysis. Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 16, No. 3, pp. 225-236.

Jolliffe, I. T. and Uddin, M., (2000). The Simplified Component Technique: An Alternative to Rotated Principal Components, Journal of Computational and Graphical Statistics, Vol. 9, No. 4, pp. 689-710.

Juran, J.M., (1989). Leadership for Quality: An Executive Handbook. Macmillan, The Free Press, New York, pp. 376.

Kresta, J.V., MacGregor, J.F. and Marlin, T.E., (1991). Multivariate Statistical Monitoring of Process Operating Performance. The Canadian Journal of Chemical Engineering, 69, pp. 35-47.

Kourti, T. And Mac Gregor, J.F., (1995). Process Analysis Monitoring and Diagnosis Using Multivariate Projection Methods. Chemometrics Intell. Lab. Syst., 28, pp. 3-21.

Kourti, T. And MacGregor, J.F., (1996). Multivariate SPC Methods for Process and Product Monitoring. Journal of Quality Technology, 284, pp. 409-427.

Krisnaiah, P. R., (1965). On The Simultaneous ANOVA and MANOVA Tests Annals of the Institute of Statistical Mathematics 17, pp. 35-53.

Krisnaiah, P. R., (1979). Some Developments on Simultaneous Test Procedures, In Developments in Statistics, edited by Krisnaiah P.R., North-Holland, New York.

Leavengood, S. and Reeb, J. E., (1999). Statistical Process Control: How and Why SPC Works. Oregon State University.

Lewis, R.G., (1994). Total Quality in Higher Education. St. Lucie Press, Delray Beach.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E., (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. Technometrics, 34, pp. 46-53.

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

Lucas, J. M., (1973). A Modified "V" Mask Control Scheme. Technometrics, 15, pp. 833-847.

Lucas, J. M., (1976). The Design and Use of Cumulative Sum Quality Control Schemes. Journal of Quality Technology, 8, pp. 1-12.

Lucas, J. M. and Crosier, R. B., (1982). Fast Initial Response for CUSUM Quality Control Scheme: Give Your CUSUM A Head Start. Technometrics, 24, pp. 199-205.

MacGregor, J. F. and Kourti, T., (1995). Statistical Process Control of Multivariate Process. Control Engineering Practice, 33, pp. 403-414.

Mandel, B. J., (1969). The Regression Control Chart. Journal of Quality Technology, 1, pp. 1-9.

Maravelakis, P.E., Bersimis S., Panaretos J. and Psarakis S., (2002). Identifying the Out-of Control Variable in a Multivariate Control Chart. Communication in Statistics Theory and Methods, 3112, pp. 2391-2408.

Maravelakis, P. E. and Bersimis, S., (2009). The Use of Andrews Curves for Detecting the Out-of-Control Variables when a Multivariate Control Chart Signals. Stat Papers.50, pp. 65. DOI 10. 1007/s00362-007-0060-9.

Mardia, K. V., Kent, J. T. and Bibby, J. M., (1994). Multivariate Analysis. Academic Press, New York.

Mason, R. L., Tracy, N. D. and Young, J. C., (1995). Decomposition of T^2 for Multivariate Control Chart Interpretation . Journal of Quality Technology, 272, pp. 99-108.

Mason, R. L., Tracy, N. D. and Young, J. C., (1996). Monitoring a Multivariate Step Process. Journal of Quality Technology, 281, pp. 39-50.

Mason, R. L., Tracy, N. D. and Young J. C., (1997). A Practical Approach for Interpretting Multivariate T^2 Control Chart Signals. Journal of Quality Technology, 974, pp. 396-406.

Mehlman, D. W., Shepherd, U. L. and Kelt, D. A., (1995). Bootstrapping Principal Component Analysis - A Comment. Ecology, 762, pp. 640-643.

Mitra, A., (1998). Fundamentals of Quality Control and Improvement. John Wiley and Sons, New York.

Mitra, A., (2012). Fundamentals of Quality Control and Improvement Second Edition, Pearson Prentice Hall, New Jersey.

Montgomery, D. C., (1995). Response Surface Methods and Designs., Invited Short Course at the ASQC/ASA Fall Technical Conference, St. Louis, MO, sponsored by the Statistic Division of ASQC.

Montgomery, D. C., (1996). Introduction to Statistical Quality Control., 3rd Edition, Wiley, pp. 9.

Montgomery, D. C., (1997). Introduction to Statistical Quality Control, 3rd Edition, New York: John Wiley & Sons, pp. 677.

Montgomery, D. C. (2005). Introduction to Statistical Quality Control, 5th Edition, John Wiley and Sons, New York.

Moran, M. A. and Murphy, B. J., (1979). A Closer Look at Two Alternative Methods of Discrimination. Applied Statistics, 28, pp. 223-232.

Murphy, B. J. (1987). Selecting Out of Control Variables with the T^2 Multivariate Quality Control Procedure. The Statistician, 36, pp. 571-583.

Nedumaran, G. & Pignatiello, J. J., (1998). Diagnosing Signals from T^2 and $\chi 2$ Multivariate Control Charts. Quality Engineering, 10, pp. 657–667.

Olkin, I. & Tmsky, J. L., (1981). A New Class of Multivariate Tests Based on the Union-Intersection Principle. The Annals of Statistics, 24, pp. 220-238.

Page, E. S., (1954). Continuous inspection schemes. Biometrika, 41, pp. 100–115.

Pignatiello, J. J. and Runger, G. C., (1990). Comparisons of Multivariate CUSUM Charts. Journal of Quality Technology, 22, pp. 173-186.

Qiu, P. H. and Hawkins D., (2001). A Rank-Based Multivariate CUSUM Procedure. Technometrics, 43, pp. 120-132.

Quadrelli, R., Bretherton, C. S. and Wallace J. M., (2005). On Sampling Errors in Empirical Orthogonal Functions. Journal of Climate, 18, pp. 3704-3710.

Richard, E. D., (1992). Statistical Quality Design and Control. Macmillan Publishing Company.

Roberts, S. W., (1959). Control Chart Tests Based on Geometric Moving Averages. Technometrics, 1, pp. 239-250.

Rousson, V. and Gasser, T., (2004). Simple Component Analysis. Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 53, No. 4, pp. 539-555.

Roy, S. N., (1953). On a Heuristic Method of Test Construction and its Use in Multivariate Analysis. The Annals of Statistics. 24, pp. 220-238.

Roy, S. N., (1957). Some Aspects of Multivariate Analysis. John Wiley and Sons, New York.

Runger, G. C., Alt, F. B. and Montgomery, D. C., (1996). Contributions to a Multivariate Statistical Process Control Chart Signal. Communications in Statistics Theory and Method, 25, pp. 2203-2213.

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

Ryan, T. P., (2000). Statistical Methods for Quality Improvement. John Wiley and Sons, New York.

Sparks, R. S., Adolphson, A. & Phatak, A., (1997). Multivariate Process Monitoring Using Dynamic Biplot. International Statistical Review, 65, pp. 325–349.

Sepulveda, A., (1996). The Minimax Control Chart for Multivariate Quality Control. Phd Dissertation, Virginia Polytechnic Institute and State University.

Sepulveda, A. And Nachlas, J. A., (1997). A Simulation Approach to Multivariate Quality Control. Computers Industrial Engineering, 331-2, pp. 113-116.

Shewhart, W. A., (1931). Economic Control of Quality of Manufactured Product Milwaukee, WI: Quality Press Van Nostrand. New York. pp. 501.

Stoumbos, et al., (2000). The State of Statistical Process Control as We Proceed into the 21st Century. Journal of American Statistician Association, 95451, pp. 992-998.

Timm, N. H., (1996). Multivariate Quality Control Using Finite Intersection Tests. Journal of Quality Technology, 282, pp. 233-243.

Tracy, N. D, Young J. C, and Mason, R. L., (1992). Multivariate Control Charts for Individual Observations. Journal of Quality Technology, 24, pp. 88-95.

Tracy, N. D., Young J. C, and Mason, R. L., (1995). A Bivariate Control Chart for Paired Measurements. Journal of Quality Technology, 27, pp. 370-376.

Trendafilov, N. T. and Vines, K., (2009). Simple and Interpretable Discrimination, Computational Statistics and Data Analysis, 53, pp. 979-989.

Vargas, V. C. C., Dias, L. F. D., & Souza, A. M., (2004). Comparative Study of the Performance of the CUSUM and EWMA Control Charts. Computer and Industrial Engineering, 464, pp. 707-724.

Vines, S. K., (2000). Simple Principal Components. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 49, No. 4, pp. 441-451.

Wade M.R. and Woodall W.H., (1993). A Review and Analysis of Cause Selecting Control Charts. Journal of Quality Technology, 25, pp. 161-170.

Wadsworth, H. M., Stephens, K. S. and Godfrey, A. B., (2002). Modern Methods for Quality Control and Improvement. John Wiley and Sons, New York.

Woodhall W. H. and Ncube M. M., (1985). Multivariate CUSUM Quality Control Procedures. Technometrics, 27, pp. 285-292.

Woodhall, W. H and Montgomery, D. C., (1999). Research Issues and Ideas in Statistical Process Control. Journal of Quality Technology, 32, pp. 376-378.

Woodhall, W. H, (2000). Controversies and Contratictions in Statistical Process Control. Journal of Quality Technology, 32, pp. 341-378.

Zhang G. X., (1985). Cause-Selecting Control Charts - A New Type of Quality Control Charts. Journal of Quality and Reliability, 12, pp. 221-225.

Zhang, J., Yan, Y. and Lades, M. (1997). Face Recognition: Eigenface, Elastic Matching, and Neural Nets. Proceedings of the IEEE, Vol. 85, No. 9.

APPENDIX

Table 1: A list of 20 random correlation matrices with values rounded to two decimal places.

Number	Random correlation matrix
1	$\begin{pmatrix} 1 & -0.26 & 0.69 & -0.60 \\ -0.26 & 1 & -0.31 & 0.49 \\ 0.69 & -0.31 & 1 & -0.71 \\ -0.60 & 0.49 & -0.71 & 1 \end{pmatrix}$
2	$\begin{pmatrix} 1 & -0.37 & 0.54 & -0.48 \\ -0.37 & 1 & -0.63 & -0.16 \\ 0.54 & -0.63 & 1 & 0.10 \\ -0.48 & -0.16 & 0.10 & 1 \end{pmatrix}$
3	$\begin{pmatrix} 1 & 0.23 & -0.12 & -0.71 \\ 0.23 & 1 & -0.30 & -0.01 \\ -0.12 & -0.30 & 1 & 0.44 \\ -0.71 & -0.01 & 0.44 & 1 \end{pmatrix}$
4	$\begin{pmatrix} 1 & -0.10 & 0.42 & 0.77 \\ -0.10 & 1 & 0.38 & -0.13 \\ 0.42 & 0.38 & 1 & -0.08 \\ 0.77 & -0.13 & -0.08 & 1 \end{pmatrix}$
5	$\begin{pmatrix} 1 & 0.43 & -0.29 & -0.24 \\ 0.43 & 1 & 0.28 & -0.34 \\ -0.29 & 0.28 & 1 & -0.68 \\ -0.24 & -0.34 & -0.68 & 1 \end{pmatrix}$

Cont.

Number	Random correlation matrix
6	$\begin{pmatrix} 1 & -0.17 & 0.71 & -0.19 \\ -0.17 & 1 & 0.29 & 0.45 \\ 0.71 & 0.29 & 1 & 0.28 \\ -0.19 & 0.45 & 0.28 & 1 \end{pmatrix}$
7	$\begin{pmatrix} 1 & -0.44 & -0.12 & -0.16 \\ -0.44 & 1 & 0.43 & -0.67 \\ -0.12 & 0.43 & 1 & -0.61 \\ -0.16 & -0.67 & -0.61 & 1 \end{pmatrix}$
8	$\begin{pmatrix} 1 & 0.73 & -0.37 & 0.92 \\ 0.73 & 1 & -0.71 & 0.65 \\ -0.37 & -0.71 & 1 & -0.10 \\ 0.92 & 0.65 & -0.10 & 1 \end{pmatrix}$
9	$\begin{pmatrix} 1 & 0.19 & -0.60 & -0.44 \\ 0.19 & 1 & -0.26 & -0.23 \\ -0.60 & -0.26 & 1 & -0.03 \\ -0.44 & -0.23 & -0.03 & 1 \end{pmatrix}$
10	$\begin{pmatrix} 1 & 0.46 & 0.69 & -0.14 \\ 0.46 & 1 & -0.01 & 0.17 \\ 0.69 & -0.01 & 1 & 0.03 \\ -0.14 & 0.17 & 0.03 & 1 \end{pmatrix}$
11	$\begin{pmatrix} 1 & -0.26 & -0.62 & -0.16 \\ -0.26 & 1 & 0.18 & -0.72 \\ -0.62 & 0.18 & 1 & 0.22 \\ -0.16 & -0.72 & 0.22 & 1 \end{pmatrix}$
12	$\begin{pmatrix} 1 & -0.20 & 0.50 & 0.32 \\ -0.20 & 1 & 0.09 & -0.37 \\ 0.50 & 0.09 & 1 & -0.62 \\ 0.32 & -0.37 & -0.62 & 1 \end{pmatrix}$

Cont.

Number	Random correlation matrix
13	$\begin{pmatrix} 1 & 0.20 & 0.21 & -0.47 \\ 0.20 & 1 & 0.60 & 0.37 \\ 0.21 & 0.60 & 1 & -0.04 \\ -0.47 & 0.37 & -0.04 & 1 \end{pmatrix}$
14	$\begin{pmatrix} 1 & 0.43 & -0.45 & -0.01 \\ 0.43 & 1 & -0.82 & -0.13 \\ -0.45 & -0.82 & 1 & -0.27 \\ -0.01 & -0.13 & -0.27 & 1 \end{pmatrix}$
15	$\begin{pmatrix} 1 & 0.54 & -0.08 & -0.15 \\ 0.54 & 1 & 0.28 & -0.75 \\ -0.08 & 0.28 & 1 & 0.10 \\ -0.15 & -0.75 & 0.10 & 1 \end{pmatrix}$
16	$\begin{pmatrix} 1 & 0.12 & 0.02 & 0.22 \\ 0.12 & 1 & 0.58 & 0.17 \\ 0.02 & 0.58 & 1 & -0.61 \\ 0.22 & 0.17 & -0.61 & 1 \end{pmatrix}$
17	$\begin{pmatrix} 1 & -0.58 & 0.26 & -0.14 \\ -0.58 & 1 & 0.44 & -0.29 \\ 0.26 & 0.44 & 1 & -0.54 \\ -0.14 & -0.29 & -0.54 & 1 \end{pmatrix}$
18	$\begin{pmatrix} 1 & -0.06 & 0.18 & 0.72 \\ -0.06 & 1 & -0.31 & -0.01 \\ 0.18 & -0.31 & 1 & 0.20 \\ 0.72 & -0.01 & 0.20 & 1 \end{pmatrix}$
19	$\begin{pmatrix} 1 & -0.70 & -0.17 & -0.12 \\ -0.70 & 1 & 0.50 & -0.15 \\ -0.17 & 0.50 & 1 & -0.32 \\ -0.12 & -0.15 & -0.32 & 1 \end{pmatrix}$

Cont.

Number	Random correlation matrix
20	$\begin{pmatrix} 1 & -0.32 & 0.06 & -0.01 \\ -0.32 & 1 & -0.14 & 0.68 \\ 0.06 & -0.14 & 1 & 0.17 \\ -0.01 & 0.68 & 0.17 & 1 \end{pmatrix}$