STATISTICAL PROCESS CONTROL USING MODIFIED ROBUST HOTELLING'S T^2 CONTROL CHARTS

FIRAS SALEEM FARES HADDAD

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2013

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to :

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Carta Hotelling's T^2 adalah alat yang popular bagi memantau kawalan proses berstatistik. Walau bagaimanapun, carta ini sensitif pada titik terpencil. Bagi mengatasi masalah ini, tiga pendekatan terhadap carta Hotelling's T^2 teguh telah dicadangkan iaitu pendekatan pemangkasan, peWinsoran dan berasaskan median. Kesemua pendekatan ini menggunakan penganggar lokasi teguh dan penganggar skala teguh yang masing-masing menggantikan min biasa dan matriks kovarians. Bagi setiap pendekatan, tiga penganggar skala teguh: MAD_n, S_n dan T_n diperkenalkan, dan penganggar ini berfungsi sewajarnya mengikut pendekatan. Pendekatan pertama, ditandai sebagai T_t^2, menggunakan konsep pemangkasan melalui jarak Mahalanobis. Penganggar skala teguh digunakan untuk mengganti matriks kovarians dalam jarak Mahalanobis. Min terpangkas dan matriks kovarians terpangkas merupakan penganggar lokasi dan skala bagi carta T_t^2. Pendekatan kedua, T_w^2, menggunakan setiap penganggar skala sebagai kriteria Winsor. Pendekatan ini mengaplikasikan penganggar M-satu langkah terubahsuai terWinsor dan kovarians terWinsor yang sepadan, masing-masing sebagai penganggar lokasi dan matrik skala bagi carta T_w^2. Manakala dalam pendekatan ketiga, T_H^2, penganggar skala teguh berperanan sebagai matriks skala dengan Hodges-Lehman sebagai penganggar lokasi. Pendekatan ini menggunakan data asal tanpa sebarang pemangkasan atau peWinsoran. Secara keseluruhannya, sembilan carta kawalan teguh telah dicadangkan. Prestasi setiap carta kawalan teguh dinilai berdasarkan kadar penggera palsu dan kebarangkalian mengesan. Bagi mengkaji kekuatan dan kelemahan carta yang dicadangkan, pelbagai keadaan diwujudkan dengan memanipulasi empat pembolehubah iaitu bilangan ciri-ciri kualiti, kadar data terpencil, tahap anjakan min dan sifat ciri-ciri kualiti (bebas dan bersandar). Secara umumnya, carta yang dicadangkan menunjukkan prestasi yang baik dari segi kadar penggera palsu. Dari sudut kebarangkalian mengesan, prestasi kesemua carta yang dicadangkan mengatasi carta Hotelling's T^2 tradisional. Keseluruhannya, kajian mendapati carta Hotelling's T^2 teguh yang dicadangkan boleh dijadikan alternatif yang baik kepada carta tradisional yang dipertikaikan.

Katakunci: Hotelling's T^2, Carta kawalan, Penganggar teguh

Abstract

Hotelling's T^2 chart is a popular tool for monitoring statistical process control. However, this chart is sensitive to outliers. To alleviate the problem, three approaches to the robust Hotelling's T^2 chart namely trimming, Winsorizing and median based were proposed. These approaches used robust location and scale estimators to substitute for the usual mean and covariance matrix, respectively. For each approach, three robust scale estimators: MAD_n, S_n and T_n were introduced, and these estimators functioned accordingly to the approach. The first approach, denoted as T_t^2, applied the concept of trimming via Mahalanobis distance. The robust scale estimator was used to replace the covariance matrix in Mahalanobis distance. The trimmed mean and trimmed covariance matrix were the location and scale estimators for the T_t^2chart. The second approach, T_w^2, employed each scale estimator as the Winsorized criterion. This approach applied Winsorized modified one step M-estimator and its corresponding Winsorized covariance as the location and the scale matrix for T_w^2 chart, respectively. Meanwhile, in the third approach, T_H^2, the robust scale estimator took the role of the scale matrix with Hodges-Lehman as the location estimator. This approach worked with the original data without any trimming or Winsorizing. Altogether, nine robust control charts were proposed. The performance of each robust control chart was assessed based on false alarm rates and probability of detection. To investigate on the strengths and weaknesses of the proposed charts, various conditions were created by manipulating four variables, namely number of quality characteristics, proportion of outliers, degree of mean shifts, and nature of quality characteristics (independent and dependent). In general, the proposed charts performed well in terms of false alarm rates. With respect to probability of detection, all the proposed charts outperformed the traditional Hotelling's T^2 charts. The overall findings showed that, the proposed robust Hotelling's T^2 control charts are viable alternatives to the disputed traditional charts.

Keywords: Hotelling T^2, Control chart, Robust estimator

Acknowledgement

I am grateful to the Almighty Allah for giving me the opportunity to complete my PhD thesis.

In completing this thesis, I owe a debt of gratitude and thanks to many persons and institutions that have supported me throughout this difficult yet challenging journey. While being thankful to all of them, I must register my gratitude to some in particular. Primarily, I would like to express my deepest appreciation to my supervisor Associate Professor Dr. Sharipah Soaad Syed Yahaya who has been very patient in guiding me and supporting from the first day arrival here in Malaysia and throughout this thesis. She assisted me immensely in focusing my thinking and ideas towards the right direction and gave me her valuable ideas, insights, comments and suggestions towards understanding the empirical predicaments I have encountered. I would like to also thank my co-supervisor Dr. Nor Idayu who supported me and help me in all stages of the writing of the thesis. To all academic and administrative staff in College of Art and Science, my sincere gratitude goes to you.

I would like to express my never ending appreciation and gratitude to Prof. Jose Luis Alfaro and the people in Jordan. First and foremost, I would like to thank my father who had been a great and wise teacher in my life, my lovely mother for her infinite patience especially during my absence, and her sincere flow of love has accompanied me all the way in my long struggle and has pushed me to pursue my dreams. My lovely wife Maiss for her love and supported me, infinite patience especially during my absence and for her pushed me to pursue my dream. Big thanks and appreciate for my two sons Yazeed and Saleem and my brothers and sisters for their patients until finish this journey of the studying.

I would like to thank all my friends, Dr. Mustafa Abu- Shaweish, Dr. Ossama Badawi, Dr. Moatasem Smadi, Dr. Raed Khasawnweh, Hussam Haddad, Dr. Malek Kasasbeh, Mr. Mohammed Kasasbeh, Dr. Hamzeh Smadi, Dr. Belal Al-Wadi, Dr. Abed Alftah Al-Azam, Dr. Haidar Al-Dreybe, Dr. Mahmoud Al-Eqab, Dr. Tareg

Abusaa, Mr. Ramzi Al-Tarazi, Dr. Aymen Abu Alhija, Dr. Salem Al-Shra’a, Dr. Abdallah Alshamari, Dr. Ali Naimat, Dr. Abed Al-hameed Al-Eneze, Dr. Obaideh Alhazimeh, Mr Amer Alhazimeh Dr. Alla’a Alsiad, Dr. Eid Hassan, Dr. Adnan Almulhem, Dr. Saleh Al-Rasheed, Mr. Basem Ayoub, Dr. Younis Megdad, Dr. Amer Abu- Rashed, Dr. Mahmoud Megdadi, Dr. Hatim Megdadi. To all of you, I have this to say: I love you, respect you, pray for you, and may Allah bless you.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

MOM	Modified One-step *M*-estimator
HL	Hodges and Lehmann estimator
Med	Median
$\boldsymbol{MAD_n}$	Median absolute deviation
$\boldsymbol{S_n}$	A scale estimator
$\boldsymbol{T_n}$	A scale estimator
FA	False Alarms
POD	Probability of Detection
ARE	*A*symptotic Relative Efficiency
MD	Mahalonobis Distance

CHAPTER ONE
MULTIVARIATE QUALITY CONTROL CHARTS

1.1 Introduction

The invention of Statistical Process Control (SPC) chart was pioneered by Dr. Walter Shewhart while he was working for Bell Labs in 1920. He aimed to monitor the quality of a process mathematically. Since then, this tool has received tremendous attention and interest from many researchers and practitioners from various fields including statistics, engineering and education to name just a few. There are some definitions of SPC charts tool. We refer to Montgomery (2005), who defined the SPC charts as tool for optimizing the amount of information needed for decision-making purposes. In addition, Nedumaran and Pignatiello (2000) defined the charts as tools to monitor performance or state of the process.

In general, SPC charts are graphical presentations that display the stability of a process. Unlike other common charts, such as bar chart, line chart or pie charts, SPC charts have some main features such as the following:

(i) The upper limit and lower limit's lines that create a range to where a process output is considered "in control"

(ii) A center line which located in the middle of the lower and upper limits that reflects the average state of the process.

(iii) Data points of the collected output of the process.

In practice, the SPC charts display the fluctuation of data points, i. e. output of a process, across times in a way that allows a user to determine easily whether such variations fall within the process limits. Later, immediate actions can be taken accordingly in order to adjust the process. This is to accommodate to some erratic behavior if in variations or they fall outside the range of limits.

The earliest SPC charts found in the literature are such as the famous Shewhart control charts ($\bar{x}$, S and R control charts). They were designed to monitor a process with a single quality characteristic, e.g. room temperature and diameter of a ring. Hence, they are coined as univariate control charts. The demand of these tools have influenced various modifications of the control charts to suit with different type of characteristics of a process, C-chart, U-chart, $n\times p$-chart and p-chart are some examples to which further discussions, can be found in major statistical quality control textbooks.

However, in real application of a production process always involve with variety of processes. Hence, the quality of the product is usually determined by more than one quality's characteristics. Weight, degree of hardness, thickness, width, and its length for example, may determine the quality of a certain type of tablets. In such a case, one may monitor each quality separately by creating different individual control chart. Then, determination of the overall quality can now be conducted by summarizing all the variations obtained from the charts. Alternatively, one may combine all the variations from the measured quality's characteristics into a single control chart and make decision based on the display obtained from the chart.

Monitoring each process separately by using univariate control chart is easy and one can identify changes in the process much faster. However, the strategy demands extra works as many control charts need to be built. Therefore, the conclusion of the overall state of the processes will be difficult. This is resulted from many control charts that need to be examined. At worst, this strategy assumes that each quality characteristic is independent to each other, which may not represent the actual variation in the process. Therefore, all quality variables must be treated as a set of data and must be used simultaneously in the monitoring process. This can be done by constructing multivariate control charts, which work similar to the univariate control chart. Therefore, these charts have been developed and become more sophisticated by the availability of the numerous data collection methods. Typically, Oktay and Aricigil (2001) referred to three popular types of multivariate control charts namely, Hotelling's T^2control charts, *MEWMA* (Multivariate exponentially weighted moving average) and *MCUSUM* (Multivariate Cumulative Sum).

Subsequently, this study addresses such issue on handling multivariate variables in designing a control chart.

1.2 Terminology

In general, the process control is divided into two phases:

(i) Phase *I* the construction of control chart is made based on the analysis of historical data.

(ii) Phase *II* involves a real time process monitoring using the constructed control chart in Phase *I*.

Historical data set with p quality characteristics are used to construct a multivariate control chart in Phase *I*. Here, the unknown mean of the processes and variance-covariance structure are estimated from the samples in order to compute the upper control limit (*UCL*). Since the T^2 statistic is non-negative the lower control limit (*LCL*) is set to zero, (Thompson & Koronacki, 1993). A good control chart should imply that all points fall within the lower and upper limits at random. This condition is termed as in control state. If the samples are unable to produce supposedly as in suggested control chart, it is termed as out of control chart. If this occurs, some refinements have to be made until the in control state is finally obtained. Usually, the refinements to the computation of mean of the process and the limits should be made.

Once an “in control” multivariate control chart is obtained, then it is ready to be used for monitoring performance of future processes in Phase *II*. Hotelling's T^2 control chart is known as a tool for monitoring multi variables simultaneously (Mason &Young, 2001). Thus, this study focuses on such tool and detailed discussion is given in the next sections.

1.3 The Hotelling's T^2 control chart

The Hotelling's T^2 control chart is a multivariate extension version of a famous Shewhart's control chart ($\bar{X}$ -control chart). The Hotelling's T^2 control chart is suitable to monitor the average state of p variables in a single display. In practice, there are two types of the Hotelling's T^2 control charts to monitor for subgroups that contain individual observations and to monitor for subgroups of at least two individuals observations (Cheng, Away & Hasan, 2006). Both types of charts offer better perspective for controlling multi variables as its mathematical nature takes into consideration the correlation among the variables. Moreover, this aspect of the Hotelling's T^2 chart implemented to determine the covariance matrix. These charts enable us to detect the outliers, the shifted mean vector, and other deviations from the control distributions (Woodall & Ncube, 1985; Williams, Woodall, Birch & Sullivan, 2006). Therefore, Alt (1985) and Montgomery (2005) reported that the Hotelling's T^2 chart is one of the most common methods in the multivariate statistical control charts.

Let us assume that a $p \times 1$ multivariate vector of variables $\boldsymbol{X}=\{\boldsymbol{x}_{1,\ldots,}\boldsymbol{x}_p\}$which is observed from group sizes m be an independent multivariate normal $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. Then, the T^2 statistic is defined as

$$T^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \mu_0)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \mu_0). \quad (1.1)$$

such that $\boldsymbol{\mu}_0$ is the overall means and $\boldsymbol{\Sigma}^{-1}$ is the inversed matrix of pooled variance-covariance matrix with p-dimension. The higher value of T^2 indicates that the observation is at distant from the overall mean, μ_0 (Mason & Young, 2002). Thus, a significant distant of T^2 indicates that some processes are out of control.

Following Equation(1.1), it is easy to show that T^2 follows the chi-square distribution, χ_p^2 if the variables are jointly distributed in a *p*-multivariate normal with known parametersμ, Σ. However, the parameters involved in this distribution are usually unknown and normally been replaced by maximum likelihood estimators$\overline{\boldsymbol{x}}$ and$\boldsymbol{S}$, thatrepresent the arithmetic mean vector and covariance matrix, respectively. If the vector $\boldsymbol{x}$ is independent, the replacement of these estimators into Equation (1.1) gives

$$T^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \sim \left[\frac{p\,(m+1)(m-1)}{m(m-p)}\right] F_{\alpha,p,m-p}. \qquad (1.2)$$

where$F_{(p,m-p)}$ is *F*-distribution with its parameter *p* and *m-p*.

For cases, when vector $\boldsymbol{x}$ is dependent of the estimators $\overline{\boldsymbol{x}}$ and$\boldsymbol{S}$ then the T^2 statistic follows Beta distribution with parameters $p/2$ and $(n-p-1)/2$ such that

$$T^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \sim \left[\frac{(m-1)^2}{m}\right] \beta_{\left(\frac{p}{2},\frac{m-p-1}{2}\right)}. \qquad (1.3)$$

Similarly, the upper control limit (UCL) is determined by the dependencies of the vector $\boldsymbol{x}$ to $\overline{\boldsymbol{x}}$ and$\boldsymbol{S}$. If both are independent, then the computed T^2 in Equation (1.2) will be compared to

$$UCL = \left[\frac{p\,(m+1)(m-1)}{m(m-p)}\right] F_{\alpha,p,m-p}. \qquad (1.4)$$

Otherwise, T^2 in Equation (1.3) will be compared to

$$UCL = \left[\frac{\frac{p}{2}}{\frac{m-p-1}{2}}\right]\beta_{(\alpha,\frac{p}{2},\frac{m-p-1}{2})}. \tag{1.5}$$

1.4 Some issues about the Hotelling's T^2

The traditional Hotelling's T^2 control chart introduced in Section 1.2 may face serious problems if the underlying populations deviate from normal distribution. The construction of this traditional chart in Phase I uses the computed $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$, which depend heavily on the assumption of a population that is normally distributed. To ignore such assumption and the deviation may lead to the construction of a biased control chart and may put us at risk by using it to monitor future processes in Phase II.

The existence of outliers may lead to non normality. This phenomenon is common in huge data gathering and multi processes (Jensen, Birch & Woodall, 2007). Outliers may easily affect the reliability of both of $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ (Alfaro & Ortega, 2009). Hence, they have to be addressed with much care. The simplest strategy to cater this problem is to remove outliers from sample. However, this step is impractical, especially in multivariate case, as one may lose too much important information although even an individual is omitted.

Alternatively, one may opt to use robust estimators in construction of a control chart. This strategy demands on vigilant methodology to ensure that a reliable and a valid control chart is produced. Many works have been conducted to produce a control chart that is robust towards the occurrence of outliers. However, much extensive work on robust estimators is in univariate cases. Hodges and Lehmann (1963) for example proposed Hodges-Lehmann estimator in improvising robust location of a variable. Only in recent years, work on multivariate control chart has come to the limelight. Alloway and Raghavachari (1991) initiated a multivariate control chart based on a robust location called Hodges-Lehmann estimator. However, from previous study, yet a robust control chart is deemed the best option. The choice is given to the users to correctly match the data in hand with available options of control charts. This study is aspiring to provide wider options for end user to use suitable robust control charts.

1.5 Problem Statement

From previous studies, Hotelling's T^2 control chart is proved to be efficient and at its optimum performance if the data considered follows a multivariate normal distribution (Montgomery, 2005; Alfaro & Ortega, 2009). Therefore, any deviations from normality may cause bias in obtaining the range of acceptable processes. Hence, its capability to monitor future processes is questionable(Mason & young, 2001). Moreover, Tiku and Singh (1982) reported that the resistance of the traditional Hotelling's T^2chart against the non-normality is weak and ineffective. Johnson (1987) validated this finding and found that the robustness of the traditional Hotelling's T^2chart does not exist against the non-normal data. Everitt (1979) reported the non-normality of the distribution data could severely affect the traditional Hotelling's T^2statistic. This is due to the fact that the classical estimators in the traditional chart are sensitive to non normal data. Later; Brooks (1985) has shown that the occurrence of the errors will be increased as manufacturing systems become more sophisticated and are able to collect huge data. Besides; the assumptions of normality are not easily satisfied in multivariate case in contrast to univariate case. Moreover, Peña and Prieto (2001) stated that the existence of outliers makes the values of the location and the scale estimators meaningless. Hence, the robustness is necessary in the case of multivariate quality control approach (Crosier, 1988).

Previous studies showed the use of robust location estimators such as trimmed mean and Hodges-Lehmann, in which they have been proven effective (Abu Shawiesh & Abdullah, 2001; Alfaro & Ortega, 2009). However, the trimmed mean depends heavily on the choice of percentage of the objects to be omitted in the computation (Alfaro & Ortega, 2009). Trimming need to be done meticulously so that no information will go to waste. The recommended trimming percentage by most researchers is between 15% to 25% (Rocke, Downs & Rocke, 1982; Wilcox & Keselman, 2003; Othman, Keselman, Padmanabhan, Wilcox & Fradette, 2004). While Hodges-Lehmann is a good choice if the distribution possesses slightly heavier tails, but otherwise when the distribution is heavy tailed. (Abu Shawiesh & Abdullah, 2001). Wilcox and Keselman (2003) later proposed one-step M-estimator (*MOM*) which trims or deletes proportion of extreme values according to the types of distribution considered. Unlike trimmed mean, this estimator empirically determines whether an observation should be trimmed, or the possibility of no trimming as well as different amount of trimming in the left versus the right tail. Meaning, the *MOM* estimators offer better default trimming to improve the Type I error opposed to other potential robust scale estimators proposed by Rousseeuw and Croux (1993).

Although many studies have discussed the process of constructing a robust Hotelling's T^2 chart, none has yet investigated on the use of *MOM* estimator. Prior work on this estimator by Syed Yahaya *et al.* (2006) showed the feasibility of this estimator in univariate case but its capability for multivariate case has yet to be tested. Therefore, this study contemplates on possibilities to design a Hotelling's T^2 control chart with modified one-step *M*-estimator.

The idea to replace a maximum likelihood estimator with robust estimators seemed straightforward. However, the methodology for integrating these robust estimators into the control chart needs to be critically planned and executed to ensure the outcome of the control chart neither over perform nor underperform the process. Besides, the control chart should be valid for real application.

To address this issue, this study opts to use three robust location estimators - the trimmed mean, winsorized *MOM* and Hodges-Lehmann estimators. The robust scale estimators -the median absolute deviation Mad_n, S_n and T_n are also considered to replace the common *S* in the traditional chart. Therefore, the combination of these three robust location estimators and three robust scale estimators will produce nine robust charts. Trimmed mean, winsorized *MOM* and Hodges Lehmann represent three different approaches of estimating location measures. The first approach is by trimming the data, while the second is by winsorizing i.e. trimming and then replacing the trimmed data with certain values, and the last is median based approach where no trimming or replacing of data is needed.

This study then continued with the measurement of the control charts' performance which involved two phases namely Phase I and Phase II. In Phase I, in control parameters (mean and covariance) were calculated. While in Phase II, these parameters were used to construct the control charts and the performance of the charts based on false alarm and probability of detection were assessed.

1.6 Research Objectives

The goal of this study is to propose alternatives to the traditional Hotelling's T^2 charts. These alternatives should be very effective to detect the occurrence of assignable causes of process shifts. They must also simultaneously control on the false alarms rates and detect the occurrence of outliers. Thus, the following objectives need to be accomplished.

(a) Integrate some highly efficient robust location and scale estimators in place of the common location and scale estimators in the traditional Hotelling's T^2 statistic·

(b) Assess the performance of the modified Hotelling's T^2 control charts.

(c) Compare the performance of the modified Hotelling's T^2 control charts with the traditional charts.

(d) Apply the investigated control charts on real industrial data

1.7 Significance of the Study

This study intents to contribute towards the development of the knowledge concerning statistical process control (SPC) charts, in particular the manufacturing sectors. Because the data are not always normally distributed, statisticians believe that the use of Hotelling's T^2 statistic that depending on the assumption of normality may caused significant risk towards the producing of control charts in real application. Previous studies showed that the robust univariate alternative control charts proved to achieve better results than the traditional ones in the presence of the contaminated data or multiple outliers. These promising results have given us confidence that the implementation of the robust estimators in the traditional multivariate control chart may provide better results than their counterparts.

This proposed methods offer practitioners a smart way to control multi processes. They can work with the original data without having to worry about the distributions of the data or the existence of the data outliers. Besides, it will help researchers to minimize concern on having to eliminate outliers in Phase *I* prior to the construction of the control chart. Subsequently, the process of analyzing and making conclusion about the processes will be much faster.

1.8 The scope of this thesis

This study focuses upon the problem of monitoring statistical process control through traditional Hotelling T^2 control chart when the measured multi-quality characteristics are contaminated with outliers. Many robust estimators have been investigated to improve the traditional control chart, such as the new estimator, i.e. modified one-step *M*-estimator (*MOM*) has not yet been tested. Earlier work on this estimator by Syed Yahaya *et al*., (2004) has shown the feasibility of this estimator, however, it limits to the general use in univariate case. Therefore, this study will extend the use of *MOM* to the multivariate case and to avoid from trimming the data, we employed winsorization approach in the computation of the estimator. Another alternative estimator is trimmed mean, where the trimming executed by using the Mahalonobis distance as we will explained the method of trimming later (section 3.3). Alloway and Raghavachari (1990) used this method but they used it for the bivariate case for subgroup data. This study adopted this method for the multivariate case with individual data. However, instead of using the classical mean and covariance to calculate the Mahalanobis Distance to determine the data to be trimmed, this study respectively used median and the robust scale estimators mentioned earlier. In addition, the trimming percentage used followed the amount mostly recommended by researchers i.e. 20%.

The last robust location estimator suggested for this study is Hodges - Lehmann where this estimator and robust scale estimator *MAD* (median absolute deviation) were used by Abu- Shaweish and Abdullah, 2001 in constructing Hotelling T^2 control chart for bivariate characteristic variables. Nonetheless, for this study, we employed the three robust scale estimators (Mad_n, S_n and T_n) as the covariance matrix.

A major concern in constructing quality control chart is to use suitable estimation method so that an unbiased control charts are produced. The traditional practice based on maximum likelihood estimators (*MLE*) $\bar{x}$ and s may be a good choice but it tends to be unreliable when outliers exist in the measured characteristics. The use of robust estimators in the construction of Hotelling T^2 control chart could help to alleviate the problem.

Proposals for estimating parameters in Phase *I* of the construction of T^2 control chart and evaluating the chart in Phase *II* are presented in Chapter 3. The investigations and numerical examples are presented through the chapters 4, 5 and 6. Results show that the proposed methods are competitively much better than the traditional methods. Conclusions and suggestions for further studies for improvement are given in Chapter 7.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This study inspires to propose new alternatives to robust Hotelling's T^2 charts via the use of high breakdown robust location and scale estimators. This chapter is divided into five sections. First section aims to introduce some previous studies on multivariate control charts. It also concerns on the statistical methods that are used simultaneously to monitor the multiple characteristics variables. Second section discusses reports that appear in the literature on robust Hotelling'sT^2 control charts. In addition, the third section describes the robust location estimators utilized in the current work as well as in previous works. In particular, it describes the types of estimators that will be used in this work such as α -trimmed mean and winsorized modified one-step *M*-estimator (w*MOM*). This section also includes discussion on the Hodges-Lehmann estimator. The forth section focuses on the four types of the robust scale estimators, *i. e*. the Median Absolute Deviation $(Mad_n), S_n$and $T_n$and α-Trimmed variance covariance matrix, and the final section is on the quality control charts.

1.9 Multivariate Control Charts

The multivariate statistical process (*MSP*) is improving rapidly in the areas of statistical process control (Woodal & Montgomery, 1999). Historically, Hotelling introduced *MSP* via Hotelling's T^2 control chart in his important worth published in 1947. Since then, the Hotelling's T^2 control chart has been frequently used as a control tool for multiple settings which monitors multiple quality characteristics simultaneously. In practice, the implementation of multivariate Hotelling's T^2 control chart passes two distinct phases: Phase I checks the status of the processes either they are in control or out of control, based on the first m subgroups which were drawn from historical data (Williams, Woodall, Birch & Sullivan, 2006), while Phase II checks the status of the ongoing processes of future subgroups.

Jackson (1985) supported the use of *MSP* control charts under the following three conditions:

(i) The interested processes are in control, thus considering multiple characteristics simultaneously.

(ii) The existence of interdependency among processes.

(iii) When false alarms are well-controlled.

However, some practical drawbacks may occur to *MSP* control charts especially the possibility of the occurrence of certain bias, which may distort the signal. In such a case, these charts are not able to determine exactly which characteristic among all characteristics is responsible for such particular signal (Jackson, 1985). Besides, the simplest idea for estimating the parameters of Hotelling's T^2 via maximum likelihood, i.e. sample mean vector $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{S}$ may lead to bias estimates as these estimators are known to be very sensitive to outliers and any shifts in mean and covariance (Vargas, 2003).

A study by Mason and Young (1999) showed that the Hotelling's T^2 statistics can be used to identify outliers in Phase *I* and the statistics may detect process shifts based on new observations in Phase *II*. However, several common causes that occur in the processes due to trends, cycles and autocorrelation, may lead to biased statistics, which may cause the estimation to be deviated further from the real value. This will lead to extreme or out of control values of the T^2 statistics. (Mason, Chou, Sullivan, Stoumbos &Young, 2003). Therefore, in recent years, many investigations have been done to propose several alternatives such as the robust Hotelling's T^2 charts in order to analyze and to monitor the multivariate quality characteristics.

The choice of estimators is important for the successiveness of detecting and handling outliers in multivariate historical data. Sullivan and Woodall (1996) tackled the issue of biased estimators by proposing estimators based on the covariance matrix of the historical data. Their study aimed to rectify the poor properties in detecting the shifts in the mean vector when the individual observations were used. Williams *et al.*(2006) proposed an estimator, which depends on the multivariate successive differences for the individual observations. Such estimator increased the false alarm rate when the size of step shifts increases. Since the distribution of the Hotelling's T^2 statistic has not yet able to determine when the successive estimator is practically in use, hence they demonstrated many properties for Hotelling's T^2 statistic, based on successive differences of the estimator. Their findings demonstrated: (i) an approximate distribution to calculate the upper control limit (UCL) of the individual observations and (ii) the precision of false alarm probability for the control charts would be increased if the successive difference covariance matrix estimator in case of in-control observations is independent and identically distributed.

Recently, many researchers have started to endeavor robust statistics in multivariate Hotelling's T^2 chart. The following section revises on some extensive studies that have dealt with robust statistics which lead to robust multivariate control charts.

1.10 Robust Multivariate Control Charts

The traditional sample multivariate Hotelling's T^2 control chart is solely based on the sample mean vector $\bar{\boldsymbol{x}}$ that represents the center of the quality characteristics and the sample covariance matrix $\boldsymbol{S}$ that represents the dispersion of the data from $\bar{\boldsymbol{x}}$. These two statistics are known to be very sensitive against outliers and will be greatly influenced by their presence (Vargas, 2003). The multiple outliers considered are observable and they are the ones that occur when a large value appeared in Hotelling's T^2 statistic (Croiser, 1988). Similarly, Brooks (1985) noticed that outliers increased because of the of huge occurrence data collection. Earlier, Johnson (1987) discovered that the traditional Hotelling's T^2 statistics could not resist the departure from the normal distribution. For this reason, the need for robust multivariate control charts is vital and practical. The robust estimators may become good alternative to eliminate the influence of outliers in the discussed control chart.

Previous researchers proposed modification of Hotelling's T^2 charts by using insensitive robust location and scale estimators. As a result, various types of robust multivariate control charts have been proposed. Alloway and Raghavachari (1990) proposed a robust multivariate control chart of the Hotelling's T^2 based on trimmed mean and trimmed covariance matrix. They proved the proposed statistics are robust and resistant to the contamination observations in the case of symmetrical distribution. They trimmed the outliers by trimming paired vectors of the data containing the largest two values of the Mahalonobis distances. Then, they replaced them with another two vectors of data that have the third and fourth ordered values of Mahalonobis distance.

This thesis discusses the robust Hotelling's T^2 statistic when trimming is symmetric. Besides, the trimming method of Alloway and Raghavachari (1990) is considered suitable to apply and the behavior of the control chart based on trimmed outliers is part of our concern. However, we modified Mahalonobis distance formula in order to ensure that it permits with the data of interest. This is done by replacing the arithmetic mean with the median and the covariance matrix with the three trimmed covariance matrices of the robust scale estimators, the $MAD_n, S_n and\ T_n$. The main concern of this study is the permitted percentage of trimming. Whereas, it is equal to 40% of the outlier's data, which represent the data that have the largest 40% of the values of modified Mahalonobis distance and then replacing them by the next 40% of the data that have the next 40% of the largest values of modified Mahalonobis distance. Moreover, this study used the multivariate case with the number of variables 2, 5, and 10, an extension to the study by Alloway and Raghavachari (1990) which only discussed the bivariate case.

Besides, Alloway and Raghavachari (1991) proposed a control chart based on the robust location Hodges-Lehmann estimator with the Wilcoxon signed-rank statistic. Their proposed control chart is a nonparametric and maintains the nominal specifications of the specified Type I error. The data used were generated from the normal distribution approach that is then compared with Shewhart`s control chart. With regard to the moderate sample size and the long-tailed symmetric distribution, it has been shown clearly that their performance is stronger than the performance of the traditional approach. These properties of the distribution make these charts appropriate for early production. Nonetheless, it functions in limited size sample whenever the distribution of the process statistic is unknown. The study concluded that there is a little difference between the trimmed and the untrimmed method in case of heavier tails existence.

In other similar interest on robust control charts, Abu-Shawiesh and Abdullah (2001) proposed a new robust Hotelling'sT^2 chart for a bivariate data. The proposed chart based on the Hodges-Lehmann and Shamos-Bickel-Lehmann estimators as alternatives to the robust location and scale estimators, respectively. In order to judge for the performance of the new robust chart, they used the contaminated normal distribution with outliers' percentages of 10% and 20% in the data. Then, they applied the robust chart on 20 subgroups data, which comprised several sample sizes for each subgroup. They confirmed that the new robust Hotelling's T^2chart has stronger performance in case of symmetrical contamination application. They also concluded that there is small variation between the traditional Hotelling's T^2 chart and the proposed robust charts when the distribution has a slightly heavier tail. They also concluded that the robust method is more dominant in its performance than the traditional one when the size of tail increases. Such evidence has led us to consider the robust Hodges-Lehmann estimator as an alternative to the arithmetic mean, $\bar{\boldsymbol{x}}$ and the three covariance matrices of the three scale estimators, Mad_n, S_n and T_n as alternatives to the covariance matrix in the traditional Hotelling's T^2 chart.

It is worth to highlight some extensive studies that have been done on an attempt to identify possible robust Hotelling's T^2 charts. Alfaro and Ortega (2008) suggested new robust Hotelling's T^2charts where they replaced the sample mean vector and covariance matrix in the traditional Hotelling's T^2 chart with the trimmed mean vector and the trimmed covariance matrix, respectively. Their investigation considered *p* variables with *n* size of individual observations. They generated the data mixture normal density:

$$(1\text{-}\varepsilon)MVN_p(\mu_0, \boldsymbol{\Sigma}_0) + \varepsilon MVN_p(\mu_{1}, \boldsymbol{I}_p). \qquad (2.1)$$

where ε denotes the proportion of outliers, $\mu_0$and Σ_0 represent the in control parameters, $\mu_1$is out of control mean vector and I_p is the identity matrix for *p*-variables. The percentage of detection outliers and the rate of false alarms were computed in order to evalute the performance of the charts. Their results revealed that the proposed robust Hotelling's T^2chart is more effective than the traditional Hotelling's T^2chart especially in detecting outliers.

Later, Alfaro and Ortega (2009) developed four robust Hotelling's T^2charts; each used the minimum volume ellipsoid (*MVE*), the minimum covariance determinant (*MCD*), the reweighted *MCD* and the trimmed mean estimators. In order to evaluate the performances of these robust Hotelling's T^2charts, they used various types of upper control limits. These involved the quantile Snedecor F- distribution, the quantile of the beta distribution, and the chi-square distribution. This simulation study was applied by taking into account that the distribution of Hotelling's T^2 statistic is unknown and the sizes of data sets are small. Their study depended on simulation, and went through two phases, namely, phase I and II. In phase I, they calculated the traditional and robust estimators whereas in phase II they generated a new observation. They ended the study by calculating the value of Hotelling's T^2 statistic for each new observation. The results concluded that the alternative robust Hotelling's T^2 charts behaved better than the traditional Hotelling's T^2 chart in terms of performance with the presence of outliers. In addition, they recommended the use of the robust Hotelling's T^2charts that depend on the trimmed mean and the modified *MCD* estimators when the amount of outliers is small. While as, they recommended to use the two robust Hotelling's T^2 charts that are used the robust estimators *MVE* and *MCD* when the detection of outliers is more interested.

Chenouri *et al*. (2009) proposed multivariate robust Hotelling'sT^2 charts for multiple individual observations by using the location and scale of reweighted minimum covariance determinant (*RMCD*) estimators. These estimators are highly robust and more efficient if compared to the ordinary location and scale *MCD* estimators. To get the control limit formulas, they calculated the empirical quantile by using Monte Carlo simulations. Then, they evaluated the performance of the robust Hotelling'sT^2chart by means of monitoring the values of false alarms rates and the probability of detection of outliers. They concluded that the proposed robust control charts perform as par when compared with the traditional control charts in terms of false alarms. However, they proved that these charts perform better and more efficient than the traditional chart in terms of detecting process shifts and outliers.

Even though Hotelling's T^2 control charts using the distance based approach such as *MVE* and *MCD* were proven better than the traditional Hotelling's T^2 control charts, but the approach is not time efficient (Syed Yahaya, Ali & Omar, 2011). The computation which requires reiteration process consume a lot of computational time.

Another approach in computing robust Hotelling's T^2 statistic is via coordinate wise, which adopts the concept of univariate robust estimators in multivariate setting. This approach provide faster computation due to its direct substation without any reiteration as in the distance based approach.

This thesis proposed the use of coordinate wise approach using robust location estimators namely trimmed mean, winsorized modified one-step M-estimator, and Hodges-Lehman with three robust scale estimators, Mad_n, S_n and T_n. These estimators were chosen based on their high breakdown point and efficiency, which will be discussed in the next section.

1.11 Robust Location Estimators

By far, the most common measurement of location estimator is the arithmetic mean $\bar{\boldsymbol{x}}$that represents the center of the quality characteristics in Hotelling's T^2statistic. However, $\bar{\boldsymbol{x}}$ is known to be very sensitive to outliers. Their presence will influence greatly outcomes of the statistic. To overcome such problem, researchers search for alternatives via the use of the traditional Hotelling's T^2 statistic that implements the robust location estimators in place of $\bar{\boldsymbol{x}}$ which are regarded insensitive to outliers. The main goal of these estimators is to give a reasonably high efficiency for a range of distributional shapes, such as the normal and longer tail distributions, symmetrical and asymmetrical distribution. Initially, the robust estimators of locations have been considered as symmetrical distributions. On the other hand, asymmetrical distributions do not imply general arguments to indicate what aspects of the distribution should be studied. Therefore, Ansell and Margaret (2009) provided three main properties of combined distribution of the potential interest (i) mean or median of the original uncontaminated distribution, (ii) mean of the contaminated distribution and (iii) median of the contaminated distributions.

The following subsections present some important types of robust location estimators, which are considered in this study. These estimators are chosen as each of them will represent various possible shapes of the data distribution. One of these estimators is the trimmed mean that is efficient when used with the symmetrical distribution (Pei-Chen, 2007). Another robust location estimator is the modified one-step *M*-estimator (*MOM*), which is suitable for asymmetrical distribution (Syed Yahaya, Othman & Keselman, 2006). While the third location estimator is the Hodges-Lehmann which is an efficient median based robust estimator and need no trimming to deal with the problem of outliers (Brown & Kildea, 1978). In this study, these three different types of robust location estimators will replace the $\bar{x}$ in the Hotelling's T^2 statistic.

1.11.1 α -Trimmed Mean

Trimming is a process that aims to remove the extreme values from each tail of the ordered statistics. According to Rock *et al*. (1982), this process resulted in providing the best percentage of the amount of trimming from each side of any ordered statistics in range of 20%-25% in symmetric distribution. Rosenberger and Gasko (1983) and Wilcox (1995) suggested a trim of 20% from each tail of the ordered statistics. Othman *et al*. (2004) confirmed that the best achievement of false alarms control appears when the percentage of trimming is moderate, *i.e*. 10%-15%.

According to Pei-Chen (2007) there is a practical concern regarding trimmed data, that is the appropriate amount of trimming needed. He indicated that many statisticians suggested that trimming as much as 20% in the simulation process is unsatisfactory in case of asymmetric distribution with heavy tails. He added that this percentage would be satisfied when we assume that the data are symmetric.

The α-trimmed mean,$\bar{x}_\alpha$is determined by removing α100% from each end of the ordered statistics and thereafter, calculating the mean of the remaining observations. The trimmed mean is simple, flexible, and easy to compute and understandable. Hence, if the assumption of symmetry is violated, the trimmed mean will estimate the quantity depending on the size of trimming on the two sides (Hogg, 1974 and Mehrotra *et al*., 1991). According to Bickel (1965) α-trimmed mean estimator will have approximate normal distribution when the sample size is large and this estimator is robust in the presence of outliers.

Another important advantage of the trimmed mean is it can accommodate the dissatisfaction arises from two choices of the situation, the sensitivity of the sample mean and insensitive of the sample median. Therefore, the trimmed mean has been enhanced greatly so that it could be considered as a compromise between these two estimators (Siegel, 1988). The trimmed mean estimator is also considered better than the traditional sample mean upon the presence of data outliers. Despite this advantage, the implementation of the trimmed mean may cause loss of efficiency when there are no extreme outliers in the data. In addition, the efficiency and the robustness make the trimmed mean more important than the median since the efficient estimators make the control charts more capable of detecting outliers and allow for another departure from the stable process. Besides, the breakdown of the trimmed mean is equal to the percentage of the trimmed values of the two sides, for example if the percentage of the trimming is α % from each side of the ordered statistics, then the breakdown point is equal to 2α %.

Mathematically, the formula of the α - trimmed mean $\bar{\boldsymbol{x}}_{\alpha}$ of *n* observations $\mathrm{x}_1, \ldots, \mathrm{x}_n$ after *k* smallest and *k* largest observations are eliminated from each tail of order statistics is defined as follows:

$$\bar{\mathrm{x}}_{\alpha} = \frac{1}{(m-2k)} \sum_{i=k+1}^{m-k} \mathrm{x}_{(i)}. \tag{2.2}$$

where

$x_{(i)}$: is i-th ordered statistic of m individual observations.

$k = [\alpha\, m]$ denotes to greatest integer less than or equal to $\alpha\, m$, where $0 \le \alpha < 0.50$.

In multivariate case, Alloway and Raghavachavari (1990) examined three methods of trimming. First, the multiple outliers trimming method that is capable of trimming the largest and the smallest values of many variables individually. It, then, calculates the trimmed mean and the trimmed variance covariance matrix of the remaining data. The problem in this method does not lie in the calculation of the trimmed mean but in the calculation of the variance covariance matrix. The covariance depends on its calculation on the pair values of each one of the two random variables. Therefore, in this case, we cannot satisfy this condition. Second the trimming method based on the idea that some variables are more important than the other. These variables contain more outliers. The trimming is done according to these outliers observations. Third, one depends on all sample information like the covariance, variance, and the distance of the data away from the center. This method is called the Mahalonobis squared distance. It selects the data pairs in the individual observations that can be trimmed and then winsorized. The formula of Mahalonobis distance is written as follows:

$$MD\,(\boldsymbol{x}) = (\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathbf{T}} \boldsymbol{S}^{-\mathbf{1}} (\boldsymbol{x} - \overline{\boldsymbol{x}}). \qquad (2.3)$$

where $\overline{\boldsymbol{x}}$ and $\boldsymbol{S}$ depend on the original data. The procedures of this method are as follows:

1. Trim the pairs of data that have the largest values of Mahalonobis squared distance.

2. Construct the winsorized sample by replacing the trimmed two pairs of data with the two pairs of data of the third and the fourth largest values of $MD\ (\boldsymbol{x})$.

Previous works employed the aforementioned methods on subgroup data while in this study, the methods will be used on individual data. However, in this study, the percentage of trimming was based on the amount commonly suggested i.e. 20% on each tail or a total of 40% on both tails (Rock et al., 1982; Rosenberger & Gasko, 1983; Wilcox, 1995; Pei-Chen, 2007).

1.11.2 Modified one step *M*-estimator (*MOM*)

It is notable from earlier studies that the trimmed mean method suffers from two practical concerns. First, the amount of the trimming that is fixed priori. This implies that if the percentage of the trimming is 20% (Rosenberger & Gasko, 1983; Wilcox, 1995), the efficiency is good compared to the mean under normality condition. However, sampling from sufficiently heavy tailed distribution may lead to poor efficiency in case of applying more trimming. While the second practical concern is the suitability of trimmed mean when the distribution is asymmetric. When using the usual symmetric trimmed mean, same amount of trimming is employed on both tails even when the data are skewed which by right, more data should be trimmed on the skewed tail as compared to the other. This problem could be alleviated by using *MOM* and furthermore *MOM* has high breakdown point whereas the usual trimmed mean has breakdown point of 2 α %, which depends on the percentage of trimming.

If the above two concerns are taken care of, the next question is how to determine the best trimming amount that would satisfy the need of having a good false alarm rates control? One solution to this problem lies in the use of the modified one-step *M*-estimators (*MOM*). Empirically *MOM* determines whether the observations should be trimmed. It also determines the possibility of non-trimming as well as the different amount of trimming in the left versus the right side (Wilcox & Keselman, 2003). The breakdown point here is not like the breakdown points of the trimmed mean, where the amount of trimming is fixed in *MOM* and equal to 0.5 (Pei-Chen, 2002).

Mathematically, Wilcox and Keselman (2003) defined the *MOM* estimator as follows: when a random sample from any distribution that represents the number of the random variables accordingly, then the *MOM* estimator is defined as the follows:

$$\hat{\theta}_j = \frac{\sum_{i=i_1+1}^{m-i_2} \mathrm{x}_{(i)j}}{m-i_1-i_2}. \tag{2.4}$$

where $\mathrm{x}_{(\mathrm{i})\mathrm{j}} = i-th$ order statistic in $j-th$ characteristic variable.

i_1: Number of x_{ij} that satisfies the criteria $\left(\mathrm{x}_{ij} - \widehat{M}_{\mathrm{j}}\right) < -K *$ (scale estimator)

i_2: Number of x_{ij} that satisfies the criteria $\left(\mathrm{x}_{ij} - \widehat{M}_{\mathrm{j}}\right) > K *$ (scale estimator)

m: The size of the data set for each variable.

$\widehat{M}_{\mathrm{j}}$: The median of the data in each j-th variable.

and the scale estimator is the median absolute deviation ($\mathrm{Mad_n}$).

The constant *K*= *2.24* is motivated to give a good efficiency for the robust scale estimators, Mad_n when the sample is taken from a normal distribution (Othman *et al.*, 2004). This study also uses *MOM* estimator to trim the outliers when the data are asymmetric. This can be done by modifying the criteria of the trimming by using alternative two robust scale estimators in addition to the use of the Mad_n estimator. Othman *et al*. (2004) and Wilcox and Keselman (2003) found that when correction factor *K*= *2.24* was used, they achieved high efficiency for *MOM* estimator. Therefore, in order to modify the formula of *MOM* estimator using the same correction factor *K*= *2.24* together with the new robust scale estimators, S_n and T_n and calculate the efficiency of the *MOM* again. Syed Yahaya *et al*. (2006) proved that these estimators are able to control on Type I error even under extreme violation of the assumptions. It also modified the criteria to choose the sample of values in order to modify the *MOM*.

1.11.3 Winsorized Modified one step *M*-estimator (*MOM*)

The Winsorized *MOM* is one of the central tendency measurements. It is the arithmetic mean, which is resulted by replacing outliers from each end of the data with the next largest and smallest values of the continuity of the consistent data after performing the trimming by using the *MOM* criteria (Mahir & Al-Khazaleh, 2009). According to Wilcox (1997), the mean is one of the most popular location estimators. Nevertheless, there is a problem concerns with this measurement in the tail of the distribution that may affects its value. This problem becomes clearer through the unbounded influence function of the breakdown point of zero, where the influence function measures how the estimator reacts to the small proportion of outliers and the lack of robustness against outliers Hampel (1974). In order to solve this problem, more attention should be given to the value of means whenever they are nearer to the center. Therefore, the advantage of using the robust measures of the center, winsorized *MOM* is that it can be used instead of the usual mean, which is thought to be another easier solution for tackling the sensitivity of the mean. As winsorized mean is known to be less sensitive than the mean but still give a reasonable estimate of central measure (Wilcox & Keselman, 2003), thus, we assume that winsorized *MOM* should also perform better than the usual mean. Furthermore, an initial investigation on the performance of Hotelling T^2 statistic using MOM and winsorized MOM showed that the latter performed better.

The construction of the Winsorized sample is proceeding as follows: (Wilcox, 1997).

For each random variable, $X_l = \{\mathrm{x}_{1l,} \ldots, \mathrm{x}_{ml}\}, l = 1, \ldots, p,$ the winsorized sample is constructed as:

$$\mathrm{w}_{kl} = \left\{ \begin{array}{ll} \mathrm{x}_{(k_1+1)l} & \text{if } \mathrm{x}_{kl} \leq \mathrm{x}_{(k_1+1)l} \\ \mathrm{x}_{kl} & \text{if } \mathrm{x}_{(k_1+1)l<\mathrm{x}_{kl}} < \mathrm{x}_{(m-k_2)l} \\ \mathrm{x}_{(m_l-k_2)l} & \text{if } \mathrm{x}_{kl} \geq \mathrm{x}_{(m-k_2)l} \end{array} \right\} \tag{2.5}$$

where

k_1: Number of the smallest outliers data.

k_2: Number of the largest outliers data.

Therefore, the estimated winsorized *MOM* for *l-th* variable as follows:

$$\overline{\mathrm{w}}_l = \frac{\mathbf{1}}{m_l} \left[\sum_{\boldsymbol{k=1}}^{m_l} \mathrm{w}_{\boldsymbol{kl}} \right] \tag{2.6}$$

1.11.4 Median

Median is the most popular robust estimator among the robust measures of center. Its application could be observed early in the statistical process control charts. According to Bluman and Allan (2009), the median is the midpoint of the data array.

$$med\,\{\mathrm{x}_1, \dots, \mathrm{x}_m\} = \begin{Bmatrix} \mathrm{x}_{\left(\frac{m+1}{n}\right)} & if\ m\ even \\ \frac{X_{\left(\frac{m}{2}\right)} + X_{\left(\frac{m}{2}+1\right)}}{2} & if\ m\ even \end{Bmatrix} \qquad (2.7)$$

where $\mathrm{x}_{(1)}, \dots, \mathrm{x}_{(m)}$ are the values of the sample order statistics (Chernobai & Rachev, 2006).

We illustrate main properties of the sample median, summarized as follows:

1. The maximum value of the breakdown point is equal to 0.5 (Geyer, 2006; Hampel, 2000).

2. This value is easy to calculate since it is the value of the middle point among data array (Bluman, 2009).

3. The efficiency of the median in the normal distribution will be decreased as the sample size increased. Accordingly, its efficiency reaches to 64%.

4. Since the calculation of the median depends only on the middle value of the data array, then, the data may cause confusion in the distribution of the long tails because it will leave out a lot of information (Janacek & Meikle, 1997).

5. Finally, the median estimator is used when we deal with open-ended distribution (Blum, 2009).

1.11.5 Hodges–Lehmann estimator

Hodges Lehmann estimator is a robust location measure. In a one sample estimation, it is the median of a set of *n*(*n*+1)/2 Walsh averages. According to Brown and Kildea (1978), the Hodges-Lehmann estimator is defined as the following:

"A simple Hodges-Lehmann estimator for that $x_j = \theta + y_j$ *, j = 1, 2, ...,n where* y_j *are i.i.d rv's symmetric about zero. The Hodges-Lehmann estimator of θ is the median of* $\{\frac{x_{i+}x_j}{2}, 1 < i,j < n\}$*, and an asymptotically equivalent estimator* $\widehat{\theta_n}$ *is the median of* $\{\frac{x_{i+}x_j}{2}, 1 \le i < j \le n\}$*."*

The main properties of the sample Hodges-Lehmann are:

1. The breakdown point is 29% (Hampel, 2000).
2. There is a symmetric distribution about the parameter θ (Hodges, 1967).
3. It is robust against the gross error, and that its asymptotic relative efficiency (*ARE*) is the same as the Wilcoxon signed rank test (Hodges, 1967).
4. When a random sample $x_1, \ldots, x_1$ come from continuous distribution and is symmetric about the parameter θ, then the distribution of Hodges-Lehmann is also symmetric about θ. Therefore, the statistic HL is unbiased estimator of θ where the unbiased of θ is equal to X_i (Hodges & Lehmann, 1963; Randles & Wolf, 1979).
5. It has some "robust" properties (Bickel, 1965).

1.12 Robust Scale Estimators

For the robust scale estimators, the sample standard deviation is the most commonly used scale estimator. It has the capability to eliminate data outliers resulting from the normal distribution and according to Alfaro and Ortega 2009 this estimator is sensitive to outliers. It represents the dispersion of quality and is considered as an important part in the Hotelling's T^2 statistic. To overcome the sensitivity problem, researchers substitute this estimator with robust scale estimators in the Hotelling's T^2 statistic. We used the median absolute deviation (Mad_n), S_n, T_n and α-Trimmed Covariance Matrix, in this thesis.

Robust scale estimators, S_n and $T_n$were proposed by Rousseeuw and Croux (1993). Some of their characteristics are highest breakdown point, and the bounded influence function. These estimators are also easy to compute, have reasonable efficiency when the observations are from normal distribution and positively definite.

The following subsections give more information about the construction and the main properties of these estimators:

1.12.1 Median Absolute Deviation (MAD_{n})

The median absolute deviation (MAD_n) is regarded as a robust alternative to the variance and was promoted first by Hampel (1974) according to the following formula:

$$Mad_n = 1.4826 * med\{|\mathrm{x}_i - med\{\mathrm{x}_1, \dots, \mathrm{x}_m\}|\}, i = 1, \dots, m. \qquad (2.8)$$

1.4826 is chosen in case of the usual parameter σ for a normal distribution and is made consistent with the parameter σ, therefore, Mad_n is unbiased to σ. The main properties of estimator, investigated by Rousseeuw and Croux (1993) are:

1. The breakdown point of it is 50% (Hampel, 2000).
2. In case of standard normal distribution, the influence function is bounded.
3. The gross-error sensitivity of the estimator is equal to 1.167, which is the smallest value in case of normal distribution for scale estimator.
4. It applies symmetric dispersion concerning the sample median. This means that the symmetric interval around the median contains 50% of the data.
5. It is easy to compute.
6. The efficiency at normal distribution is equal to 37%.
7. It needs the location estimator of the data, which is called the median.
8. By simulation, we proved that this estimator (scale matrix) is positive definite in multivariate form, refer to section 3.6 for further discussion.

However, MAD_n takes a symmetric view on dispersion, because one first estimates a central value (the median) and then attaches equal importance to positive and negative deviations from it and this does not seem to be a natural approach at asymmetric distributions.

1.12.2 T_n

Another scale estimator that proposed by Rousseeuw & Croux (1992) is the robust scale estimator the T_n. Its formula and the main properties are as follows:

$$T_n = 1.38\ \frac{1}{h}\sum_{k=1}^{h}\left\{med\left|\mathrm{x}_i - \mathrm{x}_j\right|\right\}_{(k)}, i \neq j, i, j = 1, \ldots, m. \tag{2.9}$$

where $h = \left[\frac{m}{2}\right] + 1$. This estimator has the highest breakdown point- approximately 50%, has continuous bounded influence function, and its efficiency is approximately, 52%. It is a positive definite, simple and explicit formula. This is a useful estimator in case of asymmetric distributions.

2.5.3 S_n

Let $x_1, \ldots, x_m$ be a sample of data set, and then the robust scale estimator S_n is defined as the following:

$$S_n = c * med_i\{med_j|x_i - x_j|\}, \quad i, j = 1, \ldots, m, i \neq j. \tag{2.10}$$

where $c = 1.1926$ is a correction factor which make S_n unbiased for finite samples (Rouesseuw & Croux, 1993).

The estimator S_n is very similar to Mad_n. It is used as an alternative to Mad_n. The only difference is that the med_i operation is moved outside the absolute value. The estimator S_n is a simple mixture of medians having absolute values. It measures the distance between the values. The median absolute deviation (Mad_n), on the other hand, measures the distance between the observations and the median.

The main properties of the S_n estimator, investigated by Rousseeuw and Croux (1993) are:

1. The estimator has a maximal breakdown of 50%.
2. In case of the standard normal distribution, the influence function of the S_n estimator is bounded.
3. The efficiency at the normal distribution is 58%, which is very high.

4. S_n estimator is measured by how far the observations at a typical distance between each two observations. Thus, this estimator is valid to be used at symmetric as well as asymmetric distributions since it is a location free estimator unlike MAD_n.
5. It is positive definite.

Due to these promising properties, it may be suitable estimator for the symmetric and asymmetric distributions. It is able to measure the distance between values, whereas the standard deviation measures the values from the central location.

2.5.4 α-Trimmed Covariance Matrix

From previous sections, the covariance matrix is known to be sensitive to the extreme data. Therefore, it is necessary to find alternative estimator for data when outliers are present. Consequently, the trimmed covariance matrix can be used instead of the usual covariance matrix.

The calculation of this estimator depends on the winsorized covariance matrix and its formula is as follows:

$$\boldsymbol{S}_t = \frac{m-1}{m_t-1} * \boldsymbol{S}_w \tag{2.11}$$

where m is the number of all data and m_t is the number of the data after the trimming. The estimator, $\mathrm{S_w}$ is the winsorized covariance matrix. $\boldsymbol{S}_w$is calculated similarly as the calculation of the usual covariance matrix to the winsorized sample. Accordingly, the trimmed mean with the trimmed covariance matrix were used to construct the modified Hotelling's T^2 statistic.

The trimming of the outliers is done through Mahalonobis distance by replacing sample mean vector with median vector, and the covariance matrix with one of the three robust scale covariance matrices of$Mad_n S_n$ and T_n. The trimming and the replacement of the data are dependent on the percentage, which is equal to 40%. Since α-trimmed covariance matrix depends on the winsorized covariance matrix, then, the properties of α-trimmed variance covariance matrix also depend on the properties of the winsorized covariance matrix. Consequently, according to Mingxin (2006) the properties of winsorized covariance matrix are as follows:

1. It provides highest breakdown points and is bounded for different distributions.
2. It is efficient at mild tailed symmetric models.
3. It results in high efficiency on the use of the heavy tailed or skewed distribution.
4. When the contaminated points are concentrated around the center, then, it provides the best performance among other high breakdown estimators.

1.13 The Quality Control Charts

The quality is defined as excellent achievement manifested in the products and services, means to exceed expectations. According to ISO 9000 (2000), quality is defined as the degree of a set of original characteristics that meet the needs. This can be considered as poor, good, excellent or original. Naturally, the characteristics may be qualitative or quantitative (Besterfield, 2004).

The shifted mean and the deviation from the control distribution affect the control charts, which are used to monitor the performance and the capability of the process. The control charts consist of two kinds- univariate and multivariate. Since this study is interested in the multivariate control charts, the most important tasks of the multivariate control charts (Sepulveda & Nicholas, 1997) are as follows:

1. Type I error must be fixed without any effect regarding the changes of the number of random variables.
2. It is available for the judgment criterion to explain the signals in the control charts.
3. The computational effort should be modest enough to analyze each sample when dealing with non-automated processes.

In general, control charts are one of the techniques and the activities that are necessary to improve the quality of the products. According to Besterfield (2004), these techniques and activities consist of the following relations:

1. Specifying certain specifications of what is required.
2. Propose the product or the service that suit the specifications.
3. Production has to suit the goal specified in the specifications.
4. Check the compatibility of the products and specifications.
5. Review the usage of the product to give information about the required specifications.

Thus, through these activities, customers can possess the best products or services at the cheapest prices or costs.

CHAPTER THREE
RESEARCH METHODLOGY

3.1 Introduction

This study proposed three alternative approaches to Hotelling's T^2 control chart subscripted by the location estimators used in the calculation namely the trimmed mean (T_t^2), winsorized modified one-step *M*-estimator (*MOM*) (T_w^2), and Hodges Lehman (T_H^2). Each of these approaches demonstrates a different integration of robust scale estimator in the computation of the Hotelling's T^2 chart, be it directly or indirectly. The scale estimators are the median absolute deviation (Mad_n), S_n and T_n. In general, these approaches can be categorized under trimming and non-trimming data technique. Trimmed mean and winsorized *MOM* approaches belong to the trimming data technique while Hodges Lehman approach belongs to the non-trimming technique. Under trimming technique, each approach trims data differently. When using the trimmed mean technique, the data are trimmed symmetrically, while for winsorized *MOM* technique, data are trimmed asymmetrically. The integration of the three scale estimators into the three approaches produced 9 alternative procedures for Hotelling's T^2 control charts as shown in Figure 3.1.

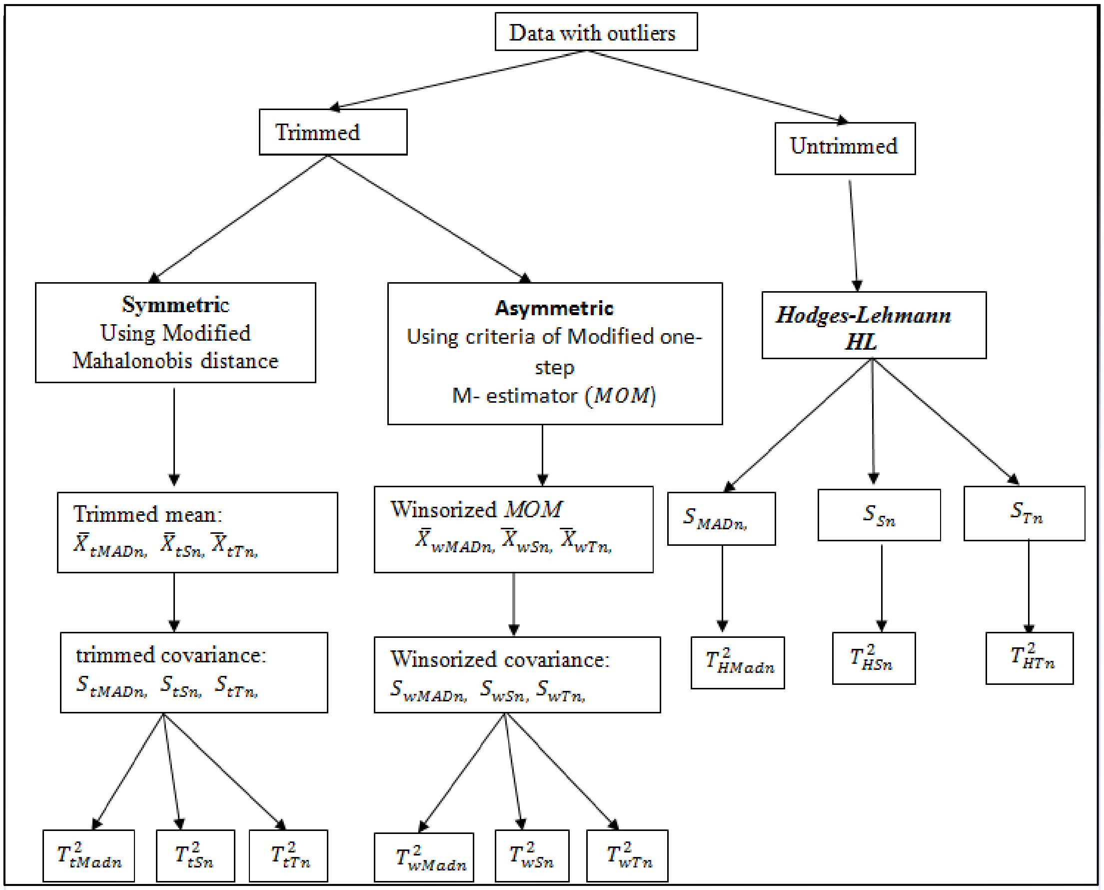


Figure 3.1: The whole procedures to construct the nine modified Hotelling's T^2*charts*

The list of the new procedures and their representations is as follows:

i. *Hotelling* T^2 *with trimmed mean and* MAD_n -T^2_{tMadn} .

ii. *Hotelling* T^2 *with trimmed mean and* S_n- T^2_{tSn}.

iii. *Hotelling* T^2 *with trimmed mean and* T_n - T^2_{tTn}.

iv. *Hotelling* T^2 *with winsorized mean and* MAD_n- T^2_{wMadn}.

v. *Hotelling* T^2 *with winsorized mean and* S_n- T^2_{wSn}.

vi. *Hotelling* T^2 *with winsorized mean and* T_n- T^2_{wTn}.

vii. *Hotelling* T^2 *with Hodges Lehman and* Mad_n- T^2_{HMadn}.

viii. *Hotelling* T^2 *with trimmed mean and* S_n- T^2_{HSn}.

ix. *Hotelling* T^2 *with trimmed mean and* T_n- T^2_{HTn}.

All these procedures are compared with the traditional Hotelling's T^2 chart in terms of the false alarms rates and probability of detecting outliers. In the case of traditional Hotelling's T^2, this study presents two different procedures, with and without data cleaning in Phase I as listed below,

x. *Traditional Hotelling* T^2 without data cleaning- T^2_O

xi. *Traditional Hotelling* $T^2$with data cleaning - T^2_1

In the above list, procedures *i* to *iii* belong to the Hotelling T^2 using trimmed mean as the location estimator. Procedures *iv* to *vi* used winsorized mean as the location estimators and procedure *vii* to *ix* are the Hotelling's T^2 with Hodges-Lehman as the location estimators. The role of each of the three scale estimators in every approach will be discussed in depth in the next sections.

In the following sections, we present the traditional Hotelling's T^2statistics followed by the statistics used in the construction of the robust Hotelling's T^2 statistics.

1.14 Traditional Hotelling's T^2 statistic

Let $\boldsymbol{x}_j = (\mathrm{x}_{1j}, \dots, \mathrm{x}_{mj})$, $j = 1, \dots, p$ be a sample from multivariate normal distribution with mean zero and identity covariance matrix $\boldsymbol{I}_p$, where p is the number of quality characteristics.The Hotelling's T^2 statistic is as follows:

$$T_0^2(\boldsymbol{x_i}) = (\boldsymbol{x_i} - \mu)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x_i} - \mu). \quad (3.1)$$

However, since the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, the parameters have to be estimated by using $\overline{\boldsymbol{x}}$ vector and $\boldsymbol{S}$ covariance matrix, respectively as follows:

$$T_0^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}). \quad (3.2)$$

such that $\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\mathrm{x}}_1 \\ \overline{\mathrm{x}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \overline{\mathrm{x}}_p \end{bmatrix}$

where the arithmetic mean for j-th vector calculated using the following formula:

$$\overline{\mathrm{x}}_j = \frac{1}{n_j} \sum_{i=1}^{m_j} \mathrm{x}_{ij} \text{ for } j = 1, \dots, p. \quad (3.3)$$

and the covariance matrix as follows:

$$S = \begin{bmatrix} \mathrm{s}_1^2 & \cdots & \mathrm{s}_{1p} \\ \vdots & \ddots & \vdots \\ \mathrm{s}_{p1} & \cdots & \mathrm{s}_p^2 \end{bmatrix}$$

where the variance and covariance of variable $\boldsymbol{x}_j$ are as follows respectively :

$$s_j^2 = \frac{1}{m_j}\sum_{i=1}^{m_j}(\mathrm{x}_{ij} - \bar{\mathrm{x}}_j)^2. \tag{3.4}$$

$$\boldsymbol{s} = cov(\boldsymbol{x}_j, \boldsymbol{x}_g) = \frac{1}{m-1}\sum_{i=1}^{m}(\mathrm{x}_{ij} - \bar{\mathrm{x}}_j)(\mathrm{x}_{ig} - \bar{\mathrm{x}}_g). \tag{3.5}$$

where $j = 1, \dots, p$; $g = 1, \dots, p$; $j \neq g$.

This study focuses on two types of traditional statistics. The first type, denoted as T_0^2, is using the original data without any outliers data cleaning,while the other type, denoted as T_1^2, is used on cleaned outliers data.

The process of the cleaning outliers data (for the T_1^2 type) is performed as follows:

Step 1: Generate data set from $MVN_p\left(\mathbf{0}, \boldsymbol{I}_p\right)$.

Step 2: Put the outliers in the data set according to the two proportions 0.1 and 0.2.

Step 3: Calculate the location and scale estimators for the data set.

Step 4: Calculate the Hotelling's T^2 statistic for each observation in the data set.

Step 5: Compare each value of the Hotelling's T^2 statistic with *UCL*. If any value of the Hotelling's T^2 statistic is greater than the *UCL* then this observation is considered outlier and then eliminated.

Step 6: The remaining observations after eliminating the outliers is considered the new data set.

Step 7: Repeate steps 1-6 until all outliers in the data set are eleminated (all values of Hotelling's T^2 statistic are less than the *UCL*).

There is no difference between these two types in formulation. The difference is only their types of samples.

As mentioned in the previous chapters, Hotelling's T^2 statistic is one of the most frequently used statistics under the multivariate aspect. However, this statistic is not free from weaknesses. To overcome these weaknesses, researchers in this area are focusing into some robust procedures, in particular trying to improve the existing procedures to be more robust. Due to the limitations of the non robust usual mean as the location measure, in this study, the usual mean vector in Hotelling T^2 is replaced with the robust location measures vector i. e. trimmed means, winsorized means, and Hodges Lehman.

Even though each location measure offers different approach in Hotelling's T^2 statistic computation, all procedures proposed in this study have the same goal that is to improve the Hotelling's T^2 statistic performance in terms of controlling the false alarm rates as well as the percentage of detecting outliers. The following section will discuss on the proposed modified Hotelling's T^2 statistic with different location measures and the different roles of each scale estimators in each approach.

1.15 Robust Hotelling's $\boldsymbol{T^2}$ statistic using trimmed mean

The first approach replaces the usual mean vector with the trimmed means as the location measures. The process involves two stages. The first stage aims to determine the observations to be trimmed before the calculation of the trimmed mean could be performed. Trimming is conducted via Mahalonobis distance method (*MD*) and 20% trimming is employed on each end of the data. The choice of 20% trimming was based on Rock*et al*. (1982), Rosenberger and Gasko (1983), Wilcox (1995) and Pei-Chen (2007). The formula of Mahalonobis distance method ($MD(\boldsymbol{x}_i)$) is as follows:

$$MD(\boldsymbol{x}_i)=(\boldsymbol{x}_i-\overline{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}). \tag{3.6}$$

To eliminate the usual problems due to masking and swamping effect, we replace the traditional location and scale estimators with the median and the robust scale estimators Mad_n respectively. This modified Mahalonobis distance is used to select the observations to be trimmed. Once trimming is done in this stage, the second stage that is to calculate the mean and covariance matrix using the trimmed sample can be conducted. The process is described as follows:

Let $x_{i1,\dots,}x_{ip}$, where $i=1,\dots,m$ be a matrix of rank mxp where m is the group size and p is the number of quality characteristics. We construct the modified Mahalonobis distance values, $MD(\boldsymbol{x}_i)$ for each observation $x_{i1,\dots,}x_{ip}$ in the data set as follows:

$$MD(\boldsymbol{x}_i)=\left(\boldsymbol{x}_i-\widehat{\boldsymbol{M}}\right)'\boldsymbol{S}_{Mad_n}^{-1}\left(\boldsymbol{x}_i-\widehat{\boldsymbol{M}}\right) \tag{3.7}$$

The steps in constructing Hotelling's T^2 with trimmed mean and Mad_n which is denoted by $T^2_{\mathrm{t}Mad_\mathrm{n}}$ is as follows,

1) Calculate Mahalonobis distance for each observation in the data set using median and MAD_n as the location and scale estimators, respectively:

   i) Median vector for the observations of each variable:

$$\widehat{\boldsymbol{M}}=\begin{bmatrix}\widehat{M}_1\\ \vdots\\ \widehat{M}_p\end{bmatrix} \tag{3.8}$$

ii) p x p covariance matrix for Mad_n i.e. $\boldsymbol{S}_{Mad_n}$:

First, compute the diagonal elements of the $p \times p$ covariance matrix which are the sample variances of each variable which are represented by

$$Mad_{n_j}^2 = Mad_n\,(\boldsymbol{x}_j) Mad_n\,(\boldsymbol{x}_j) \quad \text{where } j = 1, \dots, p. \tag{3.9}$$

Based on Abu-Shaweish and Abdullah (2001), the remaining elements of $p \times p$ covariance matrix are calculated as follows:

(a) Compute the Mad_n for the vectors $\boldsymbol{x}_j$ and $\boldsymbol{x}_g$ which are denoted by $Mad_n\,(\boldsymbol{x}_j)$ and $Mad_n\,(\boldsymbol{x}_g)$ where $j = 1, \dots, p$ and $g = 1, \dots, p, j \neq g$.

(b) Compute the spearman correlation for ranks between the variables$\boldsymbol{x}_j$ and $\boldsymbol{x}_g$, which is denoted by *corr* $(\boldsymbol{x}_j, \boldsymbol{x}_g)$.

(c) The sample covariance between the variables X_j and X_g is

$$Mad_n(\boldsymbol{x}_{j,} \boldsymbol{x}_g) = Mad_n\,(\boldsymbol{x}_j) Mad_n\,(\boldsymbol{x}_g)\; corr\,(\boldsymbol{x}_j, \boldsymbol{x}_g)\,. \tag{3.10}$$

(d) Thus, the p x p covariance matrix is

$$\boldsymbol{S}_{MADn} = \begin{bmatrix} Mad^2{}_1 & \cdots & Mad_{1p} \\ \vdots & \ddots & \vdots \\ Mad_{p1} & \cdots & Mad^2{}_p \end{bmatrix} \tag{3.11}$$

iii) Next, for each observation, calculate the modified Mahalonobis distances using formula (3.7).

iv) Arrange all values of modified Mahalonobis distances in ascending form.

2) For all variables, delete simultaneously 40% of the observations which have the largest values of mahalonobis distances.

3) Calculate the trimmed mean for the remaining data by dividing the total of the remaining data by$m_t = 0.6\, m$, where *m* denotes to the whole sample size.

4) Construct the winsorized samples by replacing these observation that are trimmed by the next observation that represents the next largest 40% of the values of mahalonobis distance.

5) Calculate the winsorized covariance matrix in the same way as untrimmed covariance matrix using the previous winsorized sample. The winsorized covariance matrix are denoted by $\boldsymbol{S}_{wMadn,}$.

6) Determine the trimmed covariance matrix $\boldsymbol{S}_{tMad_n}$by multiplying the winsorized covariance matrix $\boldsymbol{S}_{wMad_n}$ by $\frac{m-1}{m_t-1}$ (Alloway & Raghavachari, 1990).

7) Calculate the inverse of the trimmed covariance matrix $\boldsymbol{S}_{tMadn,}$ which is denoted by $\boldsymbol{S}_{tMad_n}^{-1}$.

8) The alternative robust Hotelling T^2 statistic is constructed by replacing the sample mean vector by the robust location estimator $\overline{\boldsymbol{x}}_{tMad_n}$ and also by replacing the inverse of the usual covariance matrix with the inverse of covariance matrix of the robust scale estimator $\boldsymbol{S}_{tMadn}$, denoted by $\boldsymbol{S}_{tMad_n}^{-1}$. Then, the resulting Hotelling's T^2 statistic is constructed as follows:

$$T^2_{tMAD_n}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tMadn})^{\boldsymbol{T}} \boldsymbol{S}^{-\boldsymbol{1}}_{tMadn} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tMadn}). \tag{3.12}$$

The steps for constructing another two modified Hotelling's T^2 statistics using trimmed mean using the robust scale estimators S_n and T_n which are denoted by $\overline{\boldsymbol{x}}_{\boldsymbol{tSn}}$ and $\overline{\boldsymbol{x}}_{\boldsymbol{tTn}}$and their inverses $\boldsymbol{S}^{-1}_{tS_n}$ and $\boldsymbol{S}^{-1}_{tT_n}$are similar to the previous steps of the robust scale estimator $Mad_{\boldsymbol{n}}$, except now we replace Mad_{n} by S_{n} and T_n. Thus, the two new modified Hotelling's T^2statistics are as follows:

$$T^2_{tS_n}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tSn})^T \boldsymbol{S}^{-\boldsymbol{1}}_{tS_n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tSn}). \tag{3.13}$$

$$T^2_{tT_n}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tTn})^T \boldsymbol{S}^{-\boldsymbol{1}}_{tT_n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{tTn}) . \tag{3.14}$$

1.16 Robust Hotelling's T^2 statistics using winsorized modified one step *M*-estimator (*MOM*)

The second approach uses the winsorized *MOM* as the location measures to replace the usual mean vector in the traditional Hotelling's T^2 statistics. The process involves two stages. The first stage determines the observations to be trimmed before the calculation of the mean could be performed. Trimming is done via the trimming criterion used for MOM estimator. The *MOM* estimator (Wilcox & Keselman, 2003) is defined as follows:

$$\hat{\theta}_j = \sum_{i=i_1+1}^{m_j - i_2} \frac{\mathrm{x}_{(i)j}}{m_j - i_1 - i_2}. \tag{3.15}$$

where $\mathrm{x}_{(i)j} =$ ith order statistic in jth characteristic variable.

i_1: Numberof x_{ij}that satisfies the criterion$\left(x_{ij} - \widehat{M}_j\right) < -K * \left(Mad_{nj}\right)$. (3.16)

i_2: Number of x_{ij} th at satisfies the criterion $\left(x_{ij} - \widehat{M}_j\right) > K * (Mad_{nj})$. (3.17)

m_j: Denote to the group size for jth variable.

$\widehat{M}_j = med\left\{x_{1j,\dots,}x_{nj}\right\},\ j = 1, \dots, p$.

$$Mad_{nj} = 1.4826 * med_i\left\{\left|x_{ij} - \widehat{M}_j\right|\right\}. \tag{3.18}$$

Constant *K* = 2.24 was adjusted so that efficiency is good under normality especially for small sample sizes (Wilcox & keselman, 2003; Othman *et al*., 2004; Syed Yahaya *et al*., 2006). Wilcox and Keselman (2003) found that the efficiency is equal to 0.9 for *m* = 20 when *K* = 2.24. Our investigation on *MOM* estimator using S_n and T_n reveals that its efficiency is high, that is 0.92 and 0.91, respectively.

Since the efficiency of the $\hat{\theta}$ is also high when the robust scale estimators S_n and T_n are used, it can also be used in the criterion of the modified one step *M*-estimator. This is done by replacing Mad_n by the two robust scale estimators S_n and T_n, and leave the median as the robust location estimator.

The construction of the winsorized sample using MOM follows the winsorization process suggested by Wilcox (1997)

$$\boldsymbol{w}_j = \begin{Bmatrix} \mathrm{x}_{(i_1+1)\mathrm{j}} & \text{if } \mathrm{x}_{\mathrm{ij}} \leq \mathrm{x}_{(i_1+1)\mathrm{j}} \\ \mathrm{x}_{\mathrm{ij}} & \text{if } \mathrm{x}_{(i_1+1)\mathrm{j}} < \mathrm{x}_{\mathrm{ij}} < \mathrm{x}_{(m-i_2)\mathrm{j}} \\ \mathrm{x}_{(m-i_2)\mathrm{j}} & \text{if } \mathrm{x}_{\mathrm{ij}} \geq \mathrm{x}_{(m-\mathrm{i}_2)\mathrm{j}} \end{Bmatrix}. \tag{3.19}$$

where

i_1: Number of smallest outliers in the data.

i_2: Number of largest outliers in the data.

The winsorized sample is constructed from original observations following these steps:

1. Delete i_1,
2. Replace i_1 by $\mathrm{x}_{(i_1+1)j}$, then
3. Delete i_2, then
4. Replace i_2 by $\mathrm{x}_{(m-i_2)j}$.

For each random variable $\boldsymbol{x}_j = \{\mathrm{x}_{1j,}\mathrm{x}_{2j,\dots,}\mathrm{x}_{mj}\}$, $j = 1, \dots, p$, the estimated winsorized *MOM* and the corresponding winsorized covariance matrix are calculated as:

$$\overline{\mathrm{w}_j} = \frac{\sum_{i=1}^{m_j} \mathrm{w}_{ij}}{m_j}. \tag{3.20}$$

$$\mathrm{s}_w(\boldsymbol{w}_i, \boldsymbol{w}_j) = \frac{1}{m-1} \sum_{k=1}^{m} (\mathrm{w}_{ki} - \overline{\mathrm{w}}_i)(\mathrm{w}_{kj} - \overline{\mathrm{w}}_j). \tag{3.21}$$

To construct new robust Hotelling's T^2statistic, the winsorized *MOM*,$\overline{\boldsymbol{w}}$and winsorized covariance matrix,$\boldsymbol{S}_{\boldsymbol{w}}$ replace the usual mean vector and the covariance matrix in the traditional Hotelling's T^2statistic, respectively, and we get

$$T_w^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{w}})^T \boldsymbol{S}_w^{-\mathbf{1}} (\boldsymbol{x}_i - \overline{\boldsymbol{w}}) \ . \tag{3.22}$$

Hotelling T^2 statistics in equation 3.22 represent the statistic using winsorized *MOM* based on the default criterion withMad_nas the scale estimator. Since this study also extends to other scale estimators namely $S_n$andT_n therefore, for the ease of reference, the statistics corresponding to each scale estimator is subscripted with that particular scale estimator such that $\overline{\boldsymbol{w}}_{\boldsymbol{Madn}}$, $\overline{\boldsymbol{w}}_{\boldsymbol{Sn}}$ and $\overline{\boldsymbol{w}}_{\boldsymbol{wTn}}$ representing the winsorized *MOM* for the criterion using Mad_n, S_n and T_n, respectively. Thus, the three new robust Hotelling's T^2statistics are as follows:

$$T_{wMADn}^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{Madn})^T \boldsymbol{S}_{wMadn}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{wMadn}). \tag{3.23}$$

$$T_{wSn}^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{wSn})^T \boldsymbol{S}_{wSn}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{wSn}). \tag{3.24}$$

$$T_{wTn}^2(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{wTn})^T \boldsymbol{S}_{wTn}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{w}}_{wTn}). \tag{3.25}$$

1.17 Robust Hotelling's T^2 using Hodges-Lehmann estimator

The third approach uses the Hodges-Lehmann estimator and one covariance of MAD_n, S_n or T_n to replace the usual mean and covariance matrix, $\boldsymbol{S}$, respectively. The three new robust Hotelling's T^2 control charts are denoted as T_{HMadn}^2, T_{HSn}^2 andT_{HTn}^2 depending on the scale estimators used to calculate the covariance matrix. The calculation of Hodges-Lehmann estimator is demonstrated below:(Abu- Shawiesh & Abdullah, 2001; Wei, 2007; Majid, Haron & Midi, 2010):

1. Calculate Walsh averages, $\mathbf{w}_r$by using $\mathrm{w}_{rj}= \frac{\mathrm{x}_{ij+}\mathrm{x}_{kj}}{2}$, where$r = 1, \ldots, M$, such as $M = \frac{m\,(m+1)}{2}$, $i \leq k, andi, k = 1, \ldots, m, j = 1, \ldots, p$, m is the group size.

2. Calculate Hodges-Lehmann estimator

$$HL_j\{\mathrm{x}_{1j,} \ldots, \mathrm{x}_{mj,}\} = \begin{cases} \mathrm{w}_{(k+1)j} & ifM\ is\ odd \\ \frac{\mathrm{w}_{(k)j}+\mathrm{w}_{(k+1)j}}{2} & ifM\ is\ even \end{cases} \tag{3.26}$$

where

$$k = \begin{Bmatrix} \frac{M-1}{2} ifMisodd \\ \frac{M}{2} ifMiseven \end{Bmatrix}. \tag{3.27}$$

Now, we can compute the Hotelling's T^2 statistic:

i. Let$\mathrm{x}_{i1,\ldots\ldots,}\mathrm{x}_{ip}$, a matrix of mxp where$i = 1, \ldots, m$, with m the number of observations and p is the number of quality characteristics.

ii. Calculate the mean vector using Hodges-Lehmann estimator for matrix of mxp as follows:

$$\boldsymbol{HL} = \begin{bmatrix} HL_1 \\ \vdots \\ HL_p \end{bmatrix}. \tag{3.28}$$

iii. For p-variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$, the robust covariance matrix using scale estimator Mad_n is constructed as below:

$$\boldsymbol{S}_{Madn} = \begin{bmatrix} Mad^2{}_1 & \cdots & Mad_{1p} \\ \vdots & \ddots & \vdots \\ Mad_{p1} & \cdots & Mad^2{}_p \end{bmatrix}. \tag{3.29}$$

iv. The formula for the new Hotelling's T^2control chart using Hodges Lehman as the location estimator and the covariance of *MADn* as the scale estimator is

$$T^2_{HMadn}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \boldsymbol{HL})^T \boldsymbol{S}^{-1}_{Madn}(\boldsymbol{x}_i - \boldsymbol{HL}). \tag{3.30}$$

Similar steps are needed to construct the other two new Hotelling's T^2 statistics using Hodges-Lehmann estimatorwith robust scale estimators S_n andT_n. The formulas for the respective Hotelling's T^2 statistics are given as

$$T^2_{HSn}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \boldsymbol{HL})^T \boldsymbol{S}^{-1}_{Sn} (\boldsymbol{x}_i - \boldsymbol{HL}). \tag{3.31}$$

$$T^2_{HTn}(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \boldsymbol{HL})^T \boldsymbol{S}^{-1}_{Tn} (\boldsymbol{x}_i - \boldsymbol{HL}). \tag{3.32}$$

1.18 Positive definite

The covariance matrix is a positive definite matrix if it satisfies the Cholesky decomposition. The Cholesky decomposition is defined as the product of a lower-triangular matrix and its transpose. A simulation method is employed in order to prove that the covariance matrix of robust scale estimators$Mad_{n,}S_n$ and T_n are positive definite and satisfies the Cholesky decomposition. The algorithm is:

i. Generate data set from the standard normal distribution $MVN_p\,(\mathbf{0}, \boldsymbol{I}_p)$ of size m = 25.

ii. Put outliers in the data according to the two cases, case A and case B which will be discussed later.

iii. Calculate the robust covariance matrices.

iv. Apply Cholesky decomposition these covariance matrices.

As a result, the covariance matrices now are positive definite.

v. Repeat steps (*i*-*iv*) 10,000 times and recheck whether covariance matrices are positive definite in each repetition. The results proved that the covariance matrices satisfied Cholesky decomposition in all 10,000 repetitions.

1.19 Variables Manipulated

Six variables are manipulated to investigate the strengths and the weaknesses of the traditional and new robust Hotelling's T^2 charts. The variables are, number of quality characteristics (p), proportion of contamination (ε), mean shifts ($\boldsymbol{\mu}$), group size (m), significance level (α) and nature of quality characteristics (dependent or independent). The selections of the variables were based from previous studies. Alloway and Raghavachari (1990), Wilcox (1995), Abu-Shawiesh and Abdullah (2001), Johnson (1987; 2007), Vargas (2003), Jensen, Birch and Woodall, (2006), Alfaro and Ortega (2008; 2009), Chenouri, Variyath and Steiner (2009), Midi, Shabbak, Al-Talib and Hassan (2009) and Mohammadi, Midi, Arasan and Al- Talib (2011) extensively used these variables in their works.

This study deals with the multivariate Hotelling's T^2 statistics**,** which are sensitive to the contamination of data. Therefore, the data are generated from the standard normal distribution, contaminate it with different proportions of outliers, and mean shifts. To judge the performance and capability of these new robust Hotelling's T^2 charts, false alarms rates and the percentage of detection of outliers are calculated. A good control chart should be able to control its false alarm rate close to the nominal level, α, and simultaneously have strong ability in detecting outliers. Bradley`s (1978) inequality of $0.5\alpha < \acute{\alpha} < 1.5\alpha$ is used as a criterion to evaluate whether the chart is regarded as in control of its false alarm rate. If the empirical false alarm rate, $\acute{\alpha}$, is within the Bradley's inequality, then the Hotelling's T^2 chart is regarded as robust. The closer the false alarm rate to the nominal value α, the better is the chart in terms of controlling false alarm rate. As for the percentage of detecting outliers, the higher the percentage, the better is the chart in detecting outliers. The following subsections will discuss in detail about the six quality characteristics that are considered in this study.

1.19.1 Quality Characteristics (*p*) and Group Sizes (*m*)

According to Vargas (2003), Alfaro and Ortega (2008; 2009), Rousseeuw and Zomeren (1990) and Jensen *et al*., (2006), they reported that for the traditional charts, if the group sizes m is fixed and the number of the quality characteristic, p, increased, then the effect will cause the probability of detection to decrease. Correspondingly, if one fixed the number of p and increased the group sizes, m, then the Hotelling's T^2 charts will have different effect on their performance. Vargas (2003) showed that when the value of $\frac{m}{p}$ is relatively small, the values of the false alarm rate and the percentages of detecting of outliers are approximately the same. In addition, Ruosseeuw and Zomeren (1990) suggested that p is selected by using the formula $\frac{m}{p} > 5$ since small $\frac{\mathrm{m}}{p}$ value may increase complication in detecting outliers. Other works considered different number of quality characteristics depending on the group sizes (*m)* such that Jensen *et al*. (2006)

used $p = 2,3,4, \ldots, 10$ with m = 30, 50, 75,100 and 125, while Alfaro and Ortega (2009) chose $p = 2, 3, 5,$ with m = 25, 50, 100, 1000 and Chenouri *et al*. (2009) used $p = 2, 6, 10$ with m values of 50, 100, 150. To check whether our proposed methods work regardless of the above quandary, we set the values of p at 2, 5, and 10 with m = 25, 50, 100, and 150. However, under dependence case, the Hotelling's T^2 charts, using trimmed mean and Hodges Lehman, failed to perform when m = 25 and p = 10. This is due to the existence of singular matrix in the calculation of the robust scatter matrix. Thus, for the two robust Hotelling's T^2 charts, we considered m = 50, 100 and 150 while for Hotelling's T^2 chart using winsorized *MOM*, the initial values of m= 25, 50 and 100 are adopted.

1.19.2 Proportion of Outliers (ε) and mean shifts (μ).

Outliers as defined by Hellerstein (2008) are the inconsistent observations, which are located far away from the bulk of the data. Penna and Parieto (2001) reported that the enhancement of technology helped in collecting huge numbers of data for multivariate studies, which consequently will increase the number of outliers in the collected data. According to Rocke and Woodruff (1998) most method fail to detect outliers if the proportion of the outliers greater than the fraction $1 / (p + 1)$, where p denote to the number of quality characteristics. This means that if p is large then there is small proportion of outliers present in the data which cause difficulty for the traditional charts to detect outliers. To alleviate the problem due to outliers, many statisticians like Williams *et al*. (2006); Alfaro and Ortega (2008;2009); Vargas (2003); Chenouri *et al*. (2009); Midi *et al*. (2009) and Mohammadi *et al*. (2011) used robust control charts to improve the probability of detection. The robust charts encouraged the practitioners to collect huge quantity of data without worrying about the presence of outliers. To check whether our proposed robust charts will be able to handle the problem of outliers, thus, in this study, we shifted the mean (centrality) to a certain value ($\boldsymbol{\mu}$) for a certain proportion (ε). The shifts in mean indicate the extent of the outliers. The larger the shift, the more extreme is the values of the outliers. In this study, we used 3 levels of mean shifts ($\boldsymbol{\mu}$) i.e. 0 representing no outliers, 3 and 5 for moderate and extreme values respectively. The proportion of outliers (ε) to be included in our study, we adopt the values that most researchers used i.e. 0.1 and 0.2 (Alfaro & Ortega, 2009; Chenouri *et al*., 2009; Midi *et al*., 2009; Mohammadi *et al*., 2011) which these values regarded suitable to measure the performance of proposed charts.

1.19.3 Level of Significance

Type I error in the hypothesis testing is defined as the rejection of null hypothesis H_0 while H_0 is true. The probability of making type I error is denoted by α i. e level of the significance for the hypothesis test. In the domain of the statistical control charts, the level of significance, α is called false alarm rate (Steiner, 1994). Bersimis *et al*. (2006)defined the false alarm rate as the probability of obtaining at least one out of control signal while the statistical process is in control under assumed probability distribution. A study by Chenouri *et al*. (2009) which considered a few values of α revealed that the performance of their robust charts in terms of detecting outliers dwindled when α is small. To check on the performance of our proposed robust charts, we chose α value that is commonly used i.e. 5% and 1%.

1.19.4 Nature of Quality Characteristics

This study investigated on the two natures of quality characteristics i.e. dependent and independent. Case (A) represents the independent nature while Case (B) represents the dependent nature.

Case (A) was needed to compare behaviours of modified Hotelling's T^2 charts when there exist different-sized changes for the mean of all independent variables, i. e. without correlations. Case (B) was needed when there is high correlation among variables considered. As mentioned in chapter one, the traditional control chart works well when the data is normal and the quality characterestics are independent but not in the case of dependent.

Therefore in this study we also considered the case of dependent quality characteristics with high correlation to measure the robustness and the capability of new robust charts in detecting outliers if the condition of independent is violated. However, for dependent case, only two values of the shifted mean were considered such that when no outliers exists (μ = 0) and when there exist good leverage points (μ = 5). When there are leverage points, there is a good chance to create masking effect points. So as to measure the performance of the new proposed charts in detecting these masking points. Masking effect occurs when one outlier mask the second outlier such as if the second outlier is considered as an outlier with itself only without the first outliers. Then if we delete the first outlier from the data then the second outliers will merge with outliers. Therefore, the masking effect exist when the outlying data skew the mean and covariance matrix to it (Ben-Gal, 2005).

The generation of data for Case A and Case B follow different models as explained in the next subsections.

1.19.4.1 Case A: Independent Characteristics

The following model represents the data generation for Case A:

$$(1-\varepsilon)\, MVN_p\big(0, \boldsymbol{I}_p\big) + \varepsilon MVN_p\, (\boldsymbol{\mu}_{\mathbf{1}},\ \boldsymbol{I}_P). \tag{3.33}$$

Without loss of generality, the in control mean vector is zero, while out of control mean vector of size p is $\boldsymbol{\mu}_{\mathbf{1}}$and the covariance matrix with no correlation is the identity matrix $\boldsymbol{I}_P$. Johnson (2007) reported that the covariance matrix is homogenous when there is no correlation among the variables, which indicates identity matrix $\boldsymbol{I}_p$.

1.19.4.2 Case B: Dependent Characteristics

In this case the data were generated from the following distribution:

$$(1\text{-}\varepsilon)\, MVN_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{0}}) + \varepsilon MVN_p\, (\boldsymbol{\mu}_{\mathbf{1}},\ \boldsymbol{\Sigma}_{\mathbf{0}}). \tag{3.34}$$

Without loss of generality the in control mean vector of size p, $\mu_{\mathbf{0}}$is equal to zero while out of control mean vector of size p is $\mu_{\mathbf{1}}$. $\boldsymbol{\Sigma}_{\mathbf{0}}$ represents simulteanously the in control and out of control $p \times p$ covariance matrices. For simplicity, we consider covariance matrix $\boldsymbol{\Sigma}_{\mathbf{0}}$ with high correlation among variables. For $\boldsymbol{\Sigma}_{\mathbf{0}}$, the elements of the main diagonal equal to 1's and the remaining elements equal to 0.9 (Jensen *et al.*, 2006; Johnson, 2007). The out of control parameter $\mu_{\mathbf{1}}$onlytakes the values of 0 (no change) and 5 (good leverage points) but not 3 (moderate change) because in this case the outliers are too small to be detected in case of dependent variables. Case (B) was considered in this study in order to judge the performance of proposed robust Hotelling's T^2 charts when there is high correlation among the characteristics. According to Holmes and Mergen (1996) Hotelling's T^2 charts becomes out of control when the correlation among the variables (characteristics) has changed. The effect could be observed when we changed from Case A (no correlation) to Case B (high correlation). If the chart works for the highest, then it will hopefully works for the moderate and low correlation.

1.20 Construction of control charts

The construction of control charts involves many steps as shown in the following subsections. This study used simulated data for the construction of the various control charts.

1.20.1 Data Generation

The generation of data was done using models of mixture normal for independent and dependent variables. To investigate on the performance of the charts in terms of false alarms rates and the rates of detection of outliers, four levels of contaminations were considered for the independent case namely no contamination (ideal condition), mild contamination, moderate contamination and extreme contamination. These levels of contaminations were created through manipulation of the variables discussed earlier. The manipulation generated five different types of distributions representing the levels of contamination which are categorized as below,

1) $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$ - Ideal (no contamination)

2) $(0.9)MVN_p(\mathbf{0}, \boldsymbol{I}_p) + (0.1)MVN_p(\mathbf{3}, \boldsymbol{I}_p)$ - Mild contamination

3) $(0.8)\ MVN_p(\mathbf{0}, \boldsymbol{I}_p) + (0.2)MVN_p(\mathbf{3}, \boldsymbol{I}_p)$- Moderate contamination

4) $(0.9)\ MVN_p(\mathbf{0}, \boldsymbol{I}_p) + (0.1)MVN_p(\mathbf{5}, \boldsymbol{I}_p)$- Moderate contamination

5) $(0.8)\ MVN_p(\mathbf{0}, \boldsymbol{I}_p) + (0.2)MVN_p(\mathbf{5}, \boldsymbol{I}_p)$ - Extreme contamination

All of these distributions were tested for each combination of group size *m*, and number of dimensions, *p*.

Nevertheless, for the dependent case, there is a slight changes in the number of shifted means and the covariance. Instead of using 3 values of shifted means i.e. 0, 3, and 5 as in the case of independent, we only used 2 values i.e. 0 and 5 for the dependent case . Another changes is in the covariance matrix. In the independent case, the covariance matrix is I_p since there is no correlation among the quality characteristics. However, in the dependent case, some amount of correlation should be considered in the covariance matrix. Thus, in this study, we consider a correlation as high as 0.9 as suggested by Alfaro and Ortega (2009). The contamination levels and the distributions representing the levels are as follows,

1) $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$ - Ideal (no contamination)

2) $(0.9)\ MVN_p(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{0}) + (0.1)MVN_\mathrm{p}(\mathbf{5}, \boldsymbol{\Sigma}_\mathbf{0})$ Moderate contamination

3) $(0.8)MVN_p(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{0}) + (0.2)\ MVN_\mathrm{p}(\mathbf{5}, \boldsymbol{\Sigma}_\mathbf{0})$ Extreme contamination

1.20.2 Estimation of control limit

In the construction of a control chart, the most important step is to calculate the control limits. In a multivariate control chart only the upper control limit is required since the lower control limit is always set at 0. This section will discuss on the two methods used in estimating control limits (*CL*). First is the exact method used on the traditional Hotelling's T^2 control chart. Next is the simulation method employed on robust Hotelling's T^2 charts. The simulation method is employed in this study since the underlying distribution is unknown for the robust statistics. According to Alfaro and Ortega (2009)and Chenouri *et al*. (2009), the simulation method is used to calculate upper control limit (*UCL*) when the sample size is small and the distributions of the Hotelling's T^2 statistics are unknown. In addition, Abu- Shaweish and Abdullah (2001) stated that the results of *UCL* from the simulation on the traditional Hotelling's T^2 control chart indicated close agreement with the axact method used on the traditional Hotelling's T^2 control chart. Therefor, the simulation method is used in determining of *UCL* for the robust charts.

For the traditional Hotelling's T^2 chart, we adopted the *UCL* given by Tracy, *et al*. (1992). The calculation of *UCL* depends on the how the estimators in the Hotelling's T^2 statistic i.e. $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$ are determined. If the they are not determined from the sample of observations being monitored, then the computed T^2 will be compared to

$$UCL = \left[\frac{p(m+1)(m-1)}{m(m-p)} \right] F_{\alpha,p,m-p} . \qquad (3.35)$$

Otherwise, T^2 will be compared to

$$UCL = \left[\frac{p/2}{(m-p-1)/2} \right] B_{\alpha, p/2, (m-p-1)/2} \,. \qquad (3.36)$$

For the robust Hotelling T^2 charts, the simulation method used to calculate the *UCL* followed Alfaro and Ortega (2009) as below,

Phase I:

1. Generate a data set of group size m such that $\boldsymbol{x}_j = (\mathrm{x}_{1j}, \ldots, \mathrm{x}_{mj}),\ j = 1, \ldots, p$ from standard normal distribution $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$.
2. For each data set $\boldsymbol{x}_j$, the robust estimators (location and scale estimators) are calculated.

Phase II:

3. Generate a new additional observation vector $\boldsymbol{x}_i = (\mathrm{x}_{i1,\ldots,}, \mathrm{x}_{ip})$ from the multivariate standard normal distribution $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$ for each data set.
4. Calculate the corresponding robust Hotelling's T^2 statistics for each new additional observation vector using robust estimators in phase I.
5. Repeat phase I and II to 5000 replications. According to Alfaro and Ortega (2009) and Mohamadi (2011), it is enough to estimate *UCL* if we repeat phase I and II for 5000 replications. .
6. Calculate the 95th and 99th percentile of the 5,000 values of the robust Hotelling's T^2statistics. These two percentiles are commonly used in most previous studies. Therefore, it helps us to make comparison with these previous studies.

7. The *UCL* represents the 95th and 99th percentile since the values of α were set at 5% and 1%, respectively.

The study used the same values of *UCL* for cases A and B, in order to make the comparison of the performance for all robust and traditional charts.

1.20.3 The Construction and Evaluation of Control Charts

The construction of the control charts follows two phases as explained below,

Phase I:

i. Generate data set of group size $m\boldsymbol{x}_j = (\mathrm{x}_{1j}, \dots, \mathrm{x}_{mj}), j = 1, \dots, p$ from the multivariate standard normal distribution $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$.

ii. Put outliers in data set in step *i* according to two cases (A) and (B).

iii. Calculate the traditional and the robust estimators for each data set.

iv. Repeat step i to step ii for 1000 replications. According to Alfaro and Ortega (2009), it is enoough to repeat step *i* to step *ii* for1000 replications to determine the estimators for evaluation of the control chart.

Phase II:

This phase calculates the Hotelling's T^2 statistics and evaluates the performance of the corresponding charts in terms of false alarm and probability of detection. The generation of the observations differs based on the types of performance i.e. the false alarm or the probability of detection of outliers.

i. False alarms:

1. Generate a new additional observation vector for each data set from the "in control" distribution based on the nature of characteristics. For Case A, the distribution is $MVN_p(\mathbf{0}, \boldsymbol{I}_p)$ while for Case B, it is $MVN_p(\mathbf{0}, \boldsymbol{\Sigma}_0)$.
2. Calculate the traditionals and the robust Hotelling's T^2 statistics for each new observation vector by using the estimators calculate in phase I.
3. Compare each value of the Hotelling's T^2statistics with their corresponding UCL.
4. Repeat steps 1 to 3 for 1,000 replications.
5. The false alarm is equal to the proportion of the number Hotelling's T^2 statistics that are greater than *UCL* divided by the number of replications (i.e. 1000)

ii. Probability of detection of outliers:

1. Generate a new observation vector for each data set from "out of control" distribution for the two cases. For Case A, the distribution is MVN_p ($\mu_{\mathbf{1}}$, $\boldsymbol{I}_P$) while for Case B, it isMVN_p ($\mu_{\mathbf{1}}$, $\boldsymbol{\Sigma}_{\mathbf{0}}$)
2. Calculate the traditionals and the robust Hotelling's T^2 statistics for each new additional observation vector by using the robust estimators in phase I.
3. Compare each value of the Hotelling's T^2 statistics with their corresponding UCL.
4. Repeat steps 1 to 3 for 1000 replications.
5. The probability of detection of outliers is equal to the proportion of the number

Hotelling's T^2 statistics that are greater than *UCL* divided by the number of replications (i.e. 1000).

1.20.4 Performances of the Robust Hotelling's T^2 Control Charts

The performance of the robust Hotelling's T^2 chartsare judged based on the values of false alarms rates and the percentage of detection of outliers. To evaluate the performance of the competing methods in phase I, the probability of signal must be calculated. If the data comes from in control process, the false alarms rates must be close to the nominal rate, α. In addition, if the data comes from an out of control process, then the probability of detection of outliers should be large enough to ensure that the chart is able to monitor on-line data and quickly detect shifts in the process of Phase II. Consequantly, Bradley's inequality (Bradley,1978) were used to measure the robustness of the Hotelling's T^2 charts as follows:

i) For α = 5%, the Hotelling's T^2 charts are considered robust and in control of their false alarms rates if the empirical false alarm rates are within the 2.5% and 7.5%interval.

ii) For α = 1%, the Hotelling's T^2 charts are considered robust and in control of their false alarm rates, if the empirical false alarm rates are within the 0.5% and 1.5%interval.

In case of probability of detection of outliers, the closer the values to the 100% level, the better is the performance of the charts (Chenouri *et al*., 2009). The probability of detection of outliers gives indication to the ability of the charts to detect the outliers in the data. For example, if the level of the detection outliers reaches 100%, this would be an indication that the charts detected all the outliers in the data. This means that, if we put 20% outliers in the data, the charts will be able to detect all these outliers. In our study, the benchmark for a good detection was set at 80% based on the rule of thumb for evaluating power (Linden, Adams & Robert, 2004)

Finally, computer programs in MATLAB version 7.8 (2009a) were developed to calculate and evaluate the robustness and probability of detection of the proposed robust Hotelling's T^2 charts.

1.21 Flowchart

Diagram 3.2 represents the flowchart for calculating the false alarm rates and the percentages of detection of outliers for the traditional and robust Hotelling's T^2 charts.

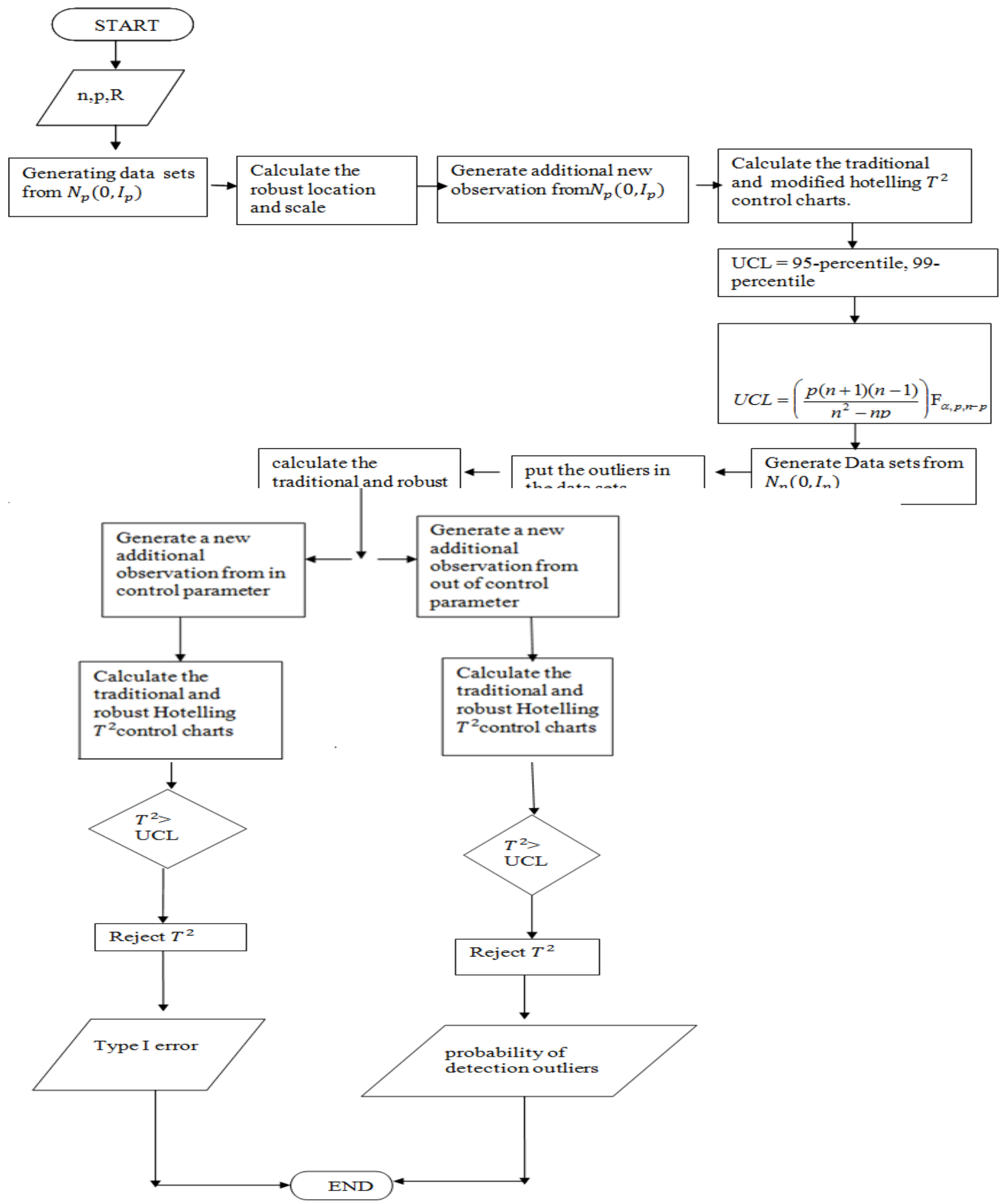


Figure 3.2: Represents the steps of calculating false alarm rates and probability detection of outliers

CHAPTER FOUR

ROBUST HOTELLING'S T^2 CONTROL CHARTS USING TRIMMED MEAN WITH TRIMMED VARIANCE COVARIANCE MATRIX

2.1 Introduction

The results of the analysis will be presented and thoroughly discussed in Chapters 4, 5 and 6. Each chapter presents a unique proposed control chart. Three different robust multivariate Hotelling's T^2charts were proposed as explained in Chapter 3. Each of the charts was assessed under different conditions to check on their strengths and weaknesses. These conditions were created by manipulating the group sizes, number of quality characteristics, proportion of contamination and the shifted mean. Each of these conditions was tested for independent (Case A) and dependent (Case B) cases of individual observations.

This chapter aims to present the performance of robust multivariate Hotelling's T^2 charts using trimmed mean as the location vector and trimmed covariance matrix as the scale matrix. However, the trimming approach via Mahalonobis distance formula was modified using three different scale estimators suggested by Rousseeuw and Croux (1993), namely $MAD_{n,}S_n$ and T_n. Hence, in this chapter the Hotelling's T^2 charts with the aforementioned different scale estimators are respectively denoted as T^2_{tMadn},T^2_{tSn}, and T^2_{tTn}. The process for computing robust Hotelling's T^2 statistic is as follows:

Step 1: Obtain original data (with outliers)

Step 2: Compute modified Mahalonobis distance such that

$$MD(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \widehat{\boldsymbol{M}})' \boldsymbol{S}_A^{-1} (\boldsymbol{x}_i - \widehat{\boldsymbol{M}}). \tag{4.1}$$

Where *A* represents the robust scale estimator, MAD_n, S_n and T_n.

Step 3: Trim the original data

Step 4: Compute trimmed mean

Step 5: Compute trimmed covariance

Step 6: Compute Hotelling's T^2 statistic

The performance of each chart is measured in terms of false alarm rates and percentage of detecting outliers. The tables, which record the false alarm rates, are arranged based on ascending number of quality characteristics (variables), namely p = 2, 5 and 10, with group sizes of$m$ = 50, 100 and 150. The first column in each table displays the group sizes, followed by the proportion of outliers (ε) and shifted means ($\boldsymbol{\mu}$), respectively in the second and third column. The proportion of outliers takes the values of 0 (no outliers), 0.1 and 0.2 while the shifted mean takes the values of 0 (no shift), 3 and 5 for case (A), and the values of 0 and 5 for case (B). The rest of the columns display the values of false alarm rates for the two traditional charts (i.e.$T_0^2$and T_1^2) and the three modified Hotelling's T^2charts. The control chart is considered to be in control of its false alarm if the empirical value is close to the nominal value α. Based on Bradley's (1978) liberal criterion of robustness, the chart is considered robust (in control) if its empirical false alarm rate, $\hat{\alpha}$, lies within $0.5\,\alpha \le \hat{\alpha} \le 1.5\,\alpha$interval. The nearer the rate to the nominal value, the better (in terms of robustness) would be the chart at that particular condition. For comparison purposes, the empirical false alarm rates within

Bradley's interval are shaded. This chapter is organized in two cases of variables, namely the independent, denoted as Case A, and dependent, denoted as Case B.

In terms of percentage of detection, the performance of the control chart is regarded as more effective in detecting outliers when the value of the percentage is closer to 100%. Based on the rule of thumb, a method is considered good in detecting outliers if the percentage of detection achieves the 80% level. For ease of comparison, the performance of the three control charts in terms of percentage of detection under each condition is graphically presented.

2.2 Independent Variables (Case A)

The performance of the control charts for independent variable is based on two measurements namely false alarm rates and percentage of detecting outliers.

2.2.1 False alarm rates and Percentage detecting outliers at $\alpha = 5\%$

The results of the analysis for the false alarm rates and percentage of detecting outliers when $\alpha = 5\%$are summarized in Table 4.1 – Table 4.2.

Table 4.1: False alarms rates (percent) under independent case for α = 5%

			P=2					P=5					P=10				
m	ε	μ	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	5.6	6	4.9	5.4	5.3	5.3	5.6	5.5	4.8	4.8	5.7	5.7	6.4	6	6.1
	0.1	3	2	5.7	3.1	3.4	3.3	2.7	3.6	3.6	3.8	3.6	4.1	3.5	4.3	4.2	3.9
		5	1.6	5.7	3	3.4	3.3	2.6	3.7	3.5	3.9	3.7	3.8	3.4	4.4	4.1	4
	0.2	3	2.1	3.1	1.9	1.9	1.9	2.6	3.1	1.9	1.9	2	4.2	4.8	2.9	3	2.9
		5	1.6	2.5	1.8	1.8	1.8	2.5	3.6	1.5	2.1	1.7	4.0	4.8	3.5	3.4	3.2
100	0	0	5.5	6.3	4.1	4.3	4.2	5.4	5	3.7	3.9	3.7	5.5	5.8	5.1	4.9	5.1
	0.1	3	2.1	5.7	2.7	2.8	3	2.9	4.3	2.5	2.3	2.3	3.3	4.3	3.6	3.8	3.9
		5	1.6	5.7	2.6	2.9	3.2	2.8	3.8	1.6	2.5	2.2	3.4	4.4	3.6	3.4	3.7
	0.2	3	2.1	3.2	1.1	1.2	1.1	3.0	4	1.4	1.7	1.7	3.5	4.9	1.8	1.8	1.8
		5	1.6	2.6	1.1	1.2	1	2.9	3.5	1.6	1.7	1.6	3.4	4.6	2.3	2.4	2.7
150	0	0	5.3	6.4	4.8	4.9	5.0	4.3	4.4	4.4	4.5	4.3	4.6	5.4	4.6	7	6.2
	0.1	3	2.3	5.7	3.6	3.9	3.7	2.1	4.4	2.8	3	2.8	3.7	3.8	3.8	3.5	3.5
		5	1.9	5.7	3.4	3.6	3.6	2.1	3.7	2.8	2.9	3	3.6	3.5	3.9	3.5	3.4
	0.2	3	2.2	3.1	1.6	1.9	1.9	2.1	3.6	1.2	1.3	1.5	3.6	3.6	2.5	2.6	2.7
		5	1.9	2.7	1.8	2	1.9	2.1	3.9	1.4	1.2	1.3	3.6	4.8	2.6	2.4	2.1
%			**20**	**100**	**60**	**60**	**60**	**73**	**93**	**53**	**47**	**47**	**100**	**100**	**87**	**80**	**87**

.

The false alarm rates in Table 4.1are presented in percentages for α = 5%. To measure the performance of the investigated charts, Bradley's robust interval is used such that the false alarm rates should be in between 2.5% and 7.5% for a chart to be considered robust at a certain condition. The last row of the table shows the percentage of the cells, which are considered as robust for each chart.

Table 4.1 shows three robust charts that perform as good as the traditional charts in controlling false alarm under ideal condition (ε = 0, $\boldsymbol{\mu}$ = 0), regardless of the group sizes$m$, ε and p. However, under non-ideal condition, under the influence of small proportion of outliers (ε = 0.1), the robust charts are still in control of the false alarm rates when the group sizes m are small. However, the control becomes conservative (smaller in values than the lower bound of robust interval) as the proportion of outliers increases (ε = 0.2) regardless of the shifted means $\boldsymbol{\mu}$. When p increases, the false alarm rates for the robust charts also increases but most of the rates are still within Bradley's interval. The robust charts perform better when p = 10 as compared to p = 2 and 5.

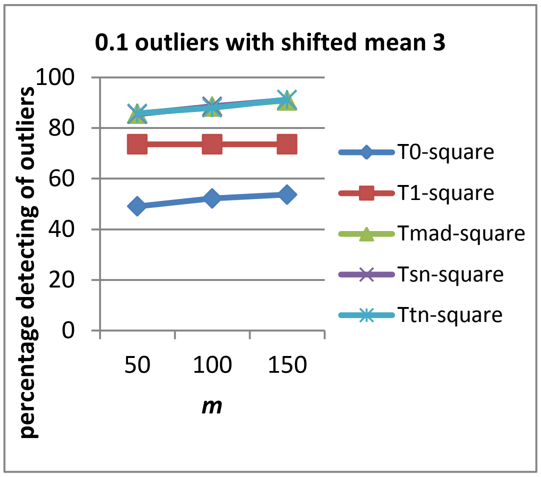
When the sample size increases to m = 100, the false alarm rates for the robust charts slightly drop, but the rates are still within the robust interval. As the sample size increases to m = 150, the false alarm rates improve. In regards to mean shift, we can say that the false alarm rates are not affected by the shifts since the changes in the false alarm rates are diminutive. However, with the increase of the proportion of outliers, ε, the false alarm rates for the robust and the traditional charts affected negatively causing the performance to drop.

In general, T^2_{tMadn} chart performs better than the other two robust charts at ε = 0.1, because it has the largest percentages of robust cells for all values of quality characteristics, p. When compared with the traditional charts, T^2_1 chart surpasses the performance of all charts.

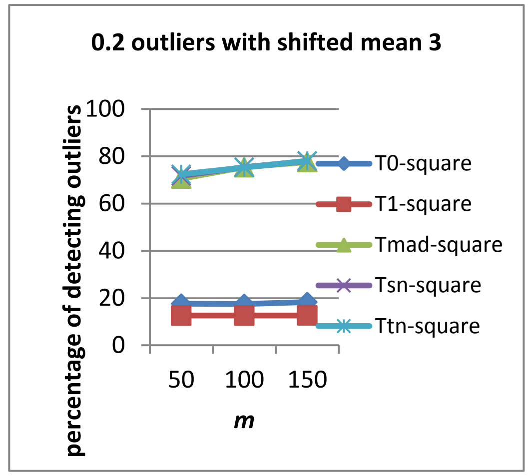
Table 4.2: Percentages of detecting outliers for independent case at α = 5%.

			p=2					*p*=5					*P*=10				
m	ε	μ	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_O^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	5.6	6	4.9	5.4	5.3	5.3	5.6	5.5	4.8	4.8	5.7	5.7	6.4	6	6.1
	0.1	3	49.1	73.6	85.7	85.5	85.8	36.4	15.2	98.2	98.1	98.3	24.7	23	100	100	100
		5	74.7	98.7	100	100	100	44.2	20.6	100	100	100	26.5	25.8	100	100	100
	0.2	3	17.6	12.6	70.6	71.7	72.4	11.9	7.1	89.7	88.8	90.1	10.6	9.9	79.8	76.9	79
		5	17.1	10.7	100	100	100	10.9	9.4	100	100	100	10.6	10.1	100	100	100
100	0	0	5.5	6.3	4.1	4.3	4.2	5.4	5	3.7	3.9	3.7	5.5	5.8	5.1	4.9	5.1
	0.1	3	52.2	73.6	88.5	88.6	88.5	41.7	20	99.5	99.5	99.5	29.4	27.6	100	100	100
		5	80.6	98.7	100	100	100	50.7	20.9	100	100	100	31.4	29.1	100	100	100
	0.2	3	17.5	12.6	75.4	75.4	75.3	11.8	11.2	93.7	94.3	94.7	10.7	10.1	84.1	83.3	84.7
		5	17.5	10.8	100	99.9	99.9	12.0	9.7	100	100	100	10.8	10.1	100	100	100
150	0	0	5.3	6.4	4.8	4.9	5	4.3	4.4	4.4	4.5	4.3	4.6	5.4	4.6	7	6.2
	0.1	3	53.7	73.6	90.9	91.2	91.3	45.5	20.5	100	99.9	99.9	32	15.2	100	100	100
		5	80.4	98.7	100	100	100	54.2	16.4	100	100	100	33.7	15	100	100	100
	0.2	3	18.3	12.7	77.6	78	78.1	12.2	10.9	96.6	96.7	96.9	11.8	8.1	85.9	86.2	86
		5	18.3	11	100	100	100	12.5	13.5	100	100	100	12.1	9.8	100	100	100

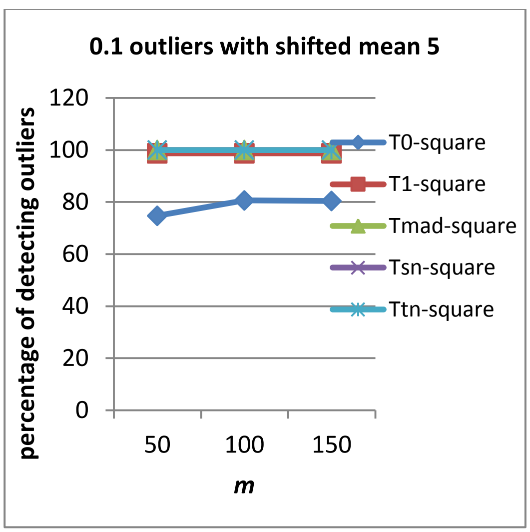
Next, we move to another measure of performance of the control charts i.e. the percentage of detecting outliers. Table 4.2 demonstrates that the robust charts improve concerning the rates of detection of outliers as the number of group sizes (*m*) increases. However, this is not the case for the traditional charts, where little or no improvement is detected for both charts. For a clearer and better comparison, we will refer to the graphs in the figures below.



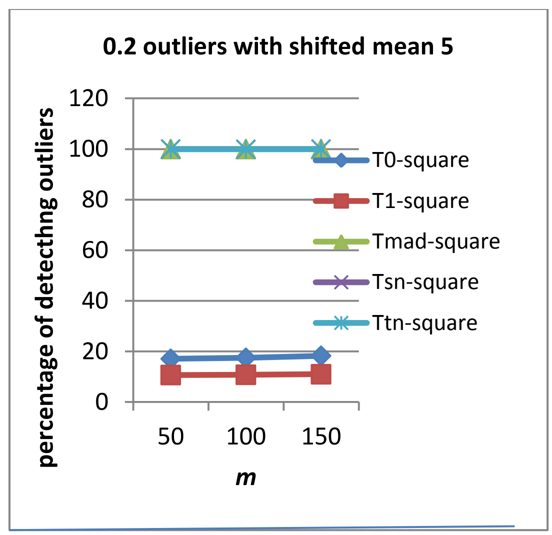
a



b



c



d

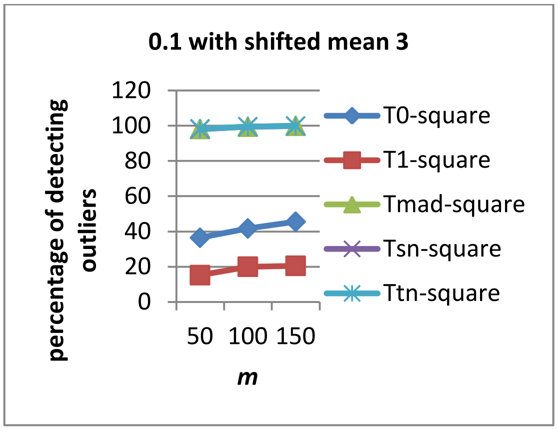
Figure 4.1: Percentages of detection of outliers when p = 2

Figure 4.1.a illustrates the case of mild contamination when the proportion of outliers $\varepsilon = 0.1$ and the shifted mean $\boldsymbol{\mu} = 3$. The three robust charts perform better than the traditional charts. There is an upward trend as the value of *m* increases for the robust

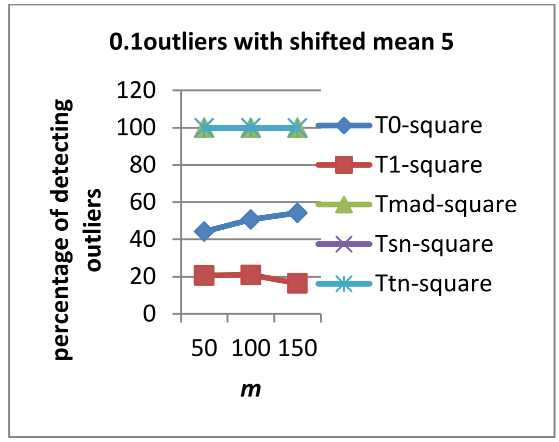
charts and T_0^2 chart, but not the one portrayed by T_1^2 chart. However, for this condition,T_1^2 chart perform better than T_0^2 chart.

Figure 4.1b shows when moderate contamination, the performance of all charts deteriorates as the proportion of outliers increases to $\varepsilon = 0.2$ while $\boldsymbol{\mu}$ stays constant. The upward trend for the robust charts maintains. We can observe a very wide gap between the lines representing robust charts and the traditional charts which indicates that under this condition, the performance of robust charts are far ahead of the traditional charts. Figure 4.1c shows the case of, the performance of the charts in detecting outliers when moderate contamination ($\boldsymbol{\mu} = 5$ and $\varepsilon = 0.1$). The performance of these three robust charts reaches the perfect level of 100% (refer to Table 4.2 for the values). In this case, T_1^2 chart also performs beautifully, may the reason in this case the data has not masking effects because this chart rebounded showed worsen performance in the next figure (4.1d), which proved that in this case the extreme contamination, the data has masking effect. The last figure (Figure 4.1d) illustrates the performance of the charts under extreme contamination ($\varepsilon = 0.2$ and $\boldsymbol{\mu} = 5$). It showed the same pattern as in Figure 4.1b, but with improvement in the performance of the robust charts and otherwise for the traditional charts especially T_0^2 chart.

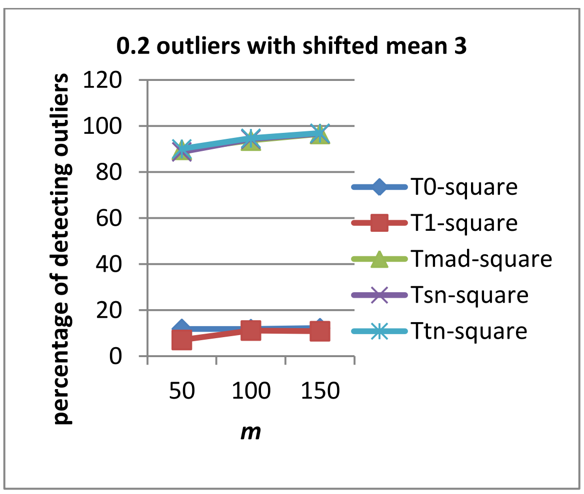
Overall, we observe that the performance of the robust charts in detecting outliers is always good regardless of conditions and their performance is on par with each other. The traditional charts, which are always below par when compared to the robust charts, worsen when $\varepsilon = 0.2$.



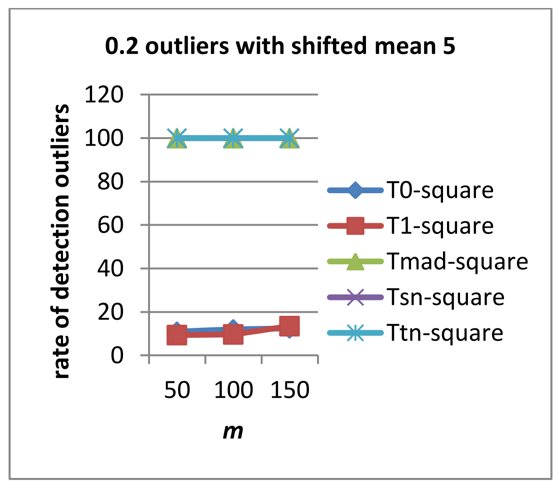
a



b



c



d

Figure 4.2: Percentages of detection of outliers when p = 5

Figure 4.2 graphically illustrates the performance of the charts in terms of detection of outliers for p = 5. Figure 4.2a shows when mild contamination, the proportion of outliers, ε = 0.1 and shifted mean, $\boldsymbol{\mu}$ = 3, the robust charts perform stronger than the traditional charts. The performance of the robust charts almost reached the highest level (100%), while the lines that represent traditional charts, hardly reach the 50% level. As group size increases, the performance of the robust and traditional charts is positively affected. When the mean shifts to $\boldsymbol{\mu}$ = 5 while the proportion of outliers maintains at ε = 0.1 (as shown in Figure 4.2b), the percentage values for the three robust charts improve to almost 100%, leaving the traditional charts far below the graph especially for T_0^2 chart. Even though there is some improvement in T_1^2 chart, the results are still far below the robust values. Increased in the values of *m* affect the robust and T_1^2 charts positively but the opposite occurs on T_0^2 chart. Figure 4.2c illustrates slight deterioration in the performance of the three robust charts when moderate contamination, ε = 0.2 and $\boldsymbol{\mu}$ = 3. The already under performed traditional charts, worsen to a level lesser than 20%. While under the extreme contamination, ε = 0.2 and $\boldsymbol{\mu}$ = 5 as shown in Figure 4.2d, the gap between the robust and the traditional charts gets wider with the three robust charts are at the peaked of their performance, while the traditional charts perform almost as bad as when ε = 0.2 and $\boldsymbol{\mu}$ = 3.

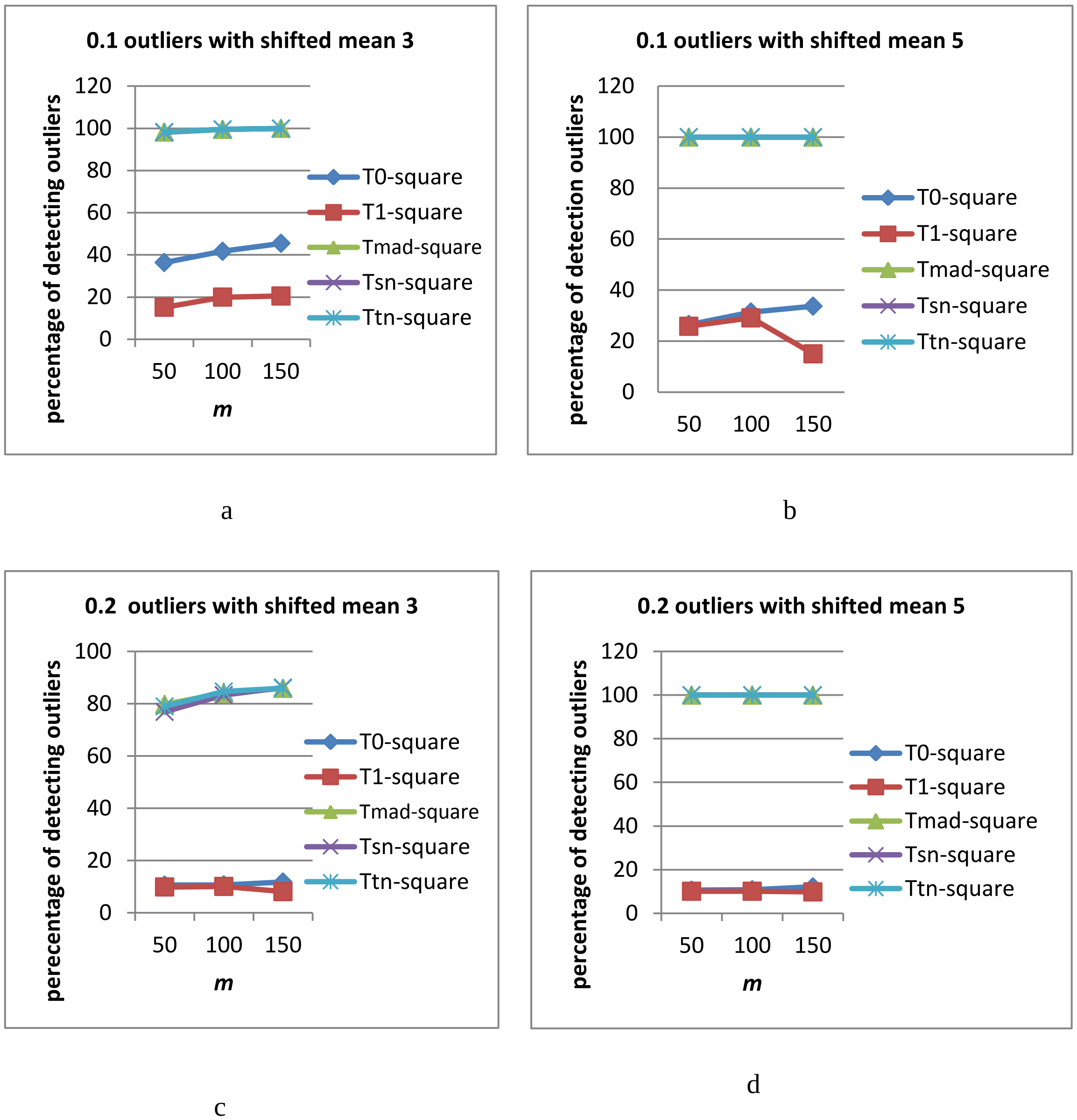


Figure 4.3: Percentages of detection of outliers when p = 10

The performance patterns of the investigated charts in terms of percentage of detection at p = 10 are shown in Figure 4.3. Figure 4.3a and Figure 4.3b represent the conditions when the percentage of outliers is 10% with mean shifts equal to 3 and 5 respectively. We could observe that both graphs display almost the same patterns. The robust charts achieve almost perfect detection leaving the traditional charts at the lower level of the percentage of detection ranging from 15% to 34%. Furthermore, T_1^2 chart worsens when p increases to 10.

However, in Figure 4.3c, as the proportion of outliers increases with shifted mean $\boldsymbol{\mu}$ = 3 some deterioration in the performance of robust charts occurs. The traditional charts deteriorated as well, resulting in the performance to be at their worst. Lastly, in Figure 4.3d in case of extreme contamination, the performance for the robust charts returns bounces back to 100%, while the performance of the traditional charts is still at the worst. Throughout the figures, we detect that as m increases, the changes in the performance of all charts show some positive effect except for T_1^2 chart.

2.2.2 False alarm rates and Percentage detecting outliers at α = 1%.

As mentioned earlier, this study adopted Bradley's robust interval to judge whether the false alarm is in control or not. Based on the interval, at α = 1% the false alarm rates must lie between 0.005 and 0.015 (0.5% to 1.5%). In the following subsections, we present tables 4.3 and 4.4, which display the results on the performance of the investigated charts according to the number of variables p in terms of false alarm and detection of outliers respectively. For the convenience of comparison, for each p in Table 4.4, there will be a corresponding figure graphically explains the results.

Table 4.3: False alarms rates (percent) under independent case for α = 1%.

			P=2					P=5					P=10				
m	ε	μ	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	1.1	1.1	1.2	1.1	1.3	0.8	0.4	0.6	0.6	0.5	0.8	1.2	1	0.9	0.9
	0.1	3	0.5	0.4	0.4	0.4	0.4	0.4	0.7	0.6	0.6	0.7	0.6	0.8	0.7	0.6	0.6
		5	0.5	0.5	0.3	0.4	0.4	0.4	1.1	0.5	0.6	0.7	0.6	0.8	0.9	0.5	0.7
	0.2	3	0.5	0.8	0.3	0.3	0.3	0.4	0.4	0.3	0.2	0.3	0.6	1	0.6	0.6	0.9
		5	0.5	0.4	0.3	0.3	0.3	0.4	0.5	0.3	0.3	0.3	0.6	1	0.8	0.5	0.6
100	0	0	1.4	1.6	1	1.2	1.1	0.8	0.4	0.5	0.4	0.6	0.8	1.4	1.1	1.3	0.9
	0.1	3	0.3	0.6	0.4	0.5	0.5	0.5	1	0.5	0.5	0.5	0.5	0.8	0.5	0.6	0.6
		5	0.3	0.5	0.4	0.5	0.5	0.6	0.8	0.4	0.5	0.5	0.5	1	0.4	0.6	0.6
	0.2	3	0.3	0.6	0.3	0.3	0.3	0.4	0.4	0.3	0.3	0.3	0.4	1.1	0.3	0.4	0.3
		5	0.3	0.6	0.3	0.3	0.3	0.4	0.	0.3	0.3	0.3	0.4	1	0.4	0.3	0.3
150	0	0	1	1.4	0.6	0 .7	0.7	0.7	1.4	1	1	1	0.8	1.1	0.8	0.7	0.8
	0.1	3	0.4	0.5	0.3	0.3	0.3	0.3	0.7	0.4	0.5	0.4	0.6	1.3	0.5	0.6	0.6
		5	0.4	0.6	0.3	0.3	0.3	0.3	1	0.4	0.5	0.4	0.6	0.9	0.6	0.5	0.6
	0.2	3	0.4	0.5	0.3	0.4	0.2	0.4	0.3	0.2	0.1	0.2	0.6	1.1	0.4	0.5	0.4
		5	0.4	0.8	0.3	0.3	0.3	0.4	0.3	0.3	0.1	0.2	0.6	1	0.7	0.5	0.6
			40	**80**	**20**	**33**	**33**	**33**	**53**	**40**	**53**	**47**	**87**	**100**	**73**	**87**	**80**

Table 4.3 displays the performance of the charts in terms of controlling false alarm rates at α = 1%. At p = 2, under ideal condition (ε = 0, $\boldsymbol{\mu}$ = 0), the false alarm rates are in control for T_0^2(1% to 1.4%) and the robust charts (0.6% to 1.3%) but not for T_1^2 chart (0.1% to 1.6%). When group size (m) equals to 100, T_1^2 chart produces a liberal value of false alarm where the value is above the upper interval limit. No specific pattern of performance could be observed across the values of m for other charts. However, for non-ideal condition, the results for the investigated charts somewhat drop to (0.2% - 0.5%) for the robust charts, (0.3% - 0.5%) for the T_0^2 chart and (0.4% – 0.8%) for the T_1^2 chart. There are smaller numbers of shaded cells for each robust chart as compared to

the traditional charts, especially T_1^2 chart. The change in the proportion of outliers (ε) from 0.1 to 0.2 negatively affects the performance of the robust charts but not the traditional charts. Nonetheless, the change in the mean shift ($\boldsymbol{\mu}$) seems to have no effect on the false alarm rates. Among the robust charts, the least to perform under p = 2 is T_{tMadn}^2 while T_{tSn}^2 chart is identified as the best in controlling false alarm rates. In general, for p = 2, T_1^2 chart performs the best.

For p = 5, under ideal condition, the false alarm rates for all robust charts and T_0^2 are within Bradley's robust criterion except for T_{tSn}^2 chart, which produces false alarm rate slightly below the interval limit (i.e. 0.4%) at $m = 100$. The T_1^2 chart also fails to control its false alarm rate at m = 50 and 100 producing false alarm rates equal to 0.4% for both conditions, and inflates to 1.4% as m changes to 150. For non-ideal condition (with proportion of outliers and shifted mean), the performance of robust charts are in control of the false alarm rates when the proportion of outliers ε = 0.1. However, it deteriorates when the proportion increases to ε = 0.2. No obvious effect could be detected when the mean shifts from 3 to 5. Under p = 5, T_{tSn}^2 chart still performs the best among the three robust charts producing the most numbers of robust conditions despite the conservative false alarm values at ε = 0.2. Although T_1^2 chart performs very well when ε = 0.1, the chart fails to perform when ε increases to 0.2, so does the T_0^2 chart.

When the number of quality characteristics, p, increases to 10, we could observe more shaded cells (robust conditions) as compared to smaller values of p. However, most of the non-shaded cells which concentrated at m = 100 with ε = 0.2 belong to the robust charts. Among the robust charts, T_{tSn}^2 still produces the most number of robust conditions while T_{tMadn}^2 chart produces the least. In general, T_1^2 performs the best with 100% robust conditions while the performance of T_0^2 is on par withT_{tSn}^2.

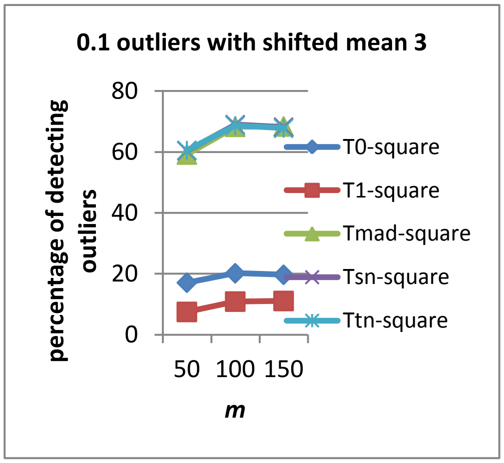
In general, no obvious pattern of changes could be detected on the performance of the charts in controlling false alarm rates as *m* increases but we could observe a slight glitch in the performance of the robust charts when *m* = 100 with $\varepsilon = 0.2$. When μ increases, a small increase in false alarm rates could be detected in the robust charts, but when ε increases, the rates slightly drop. However, the traditional charts are quite consistent in their performance across the conditions.

Table 4.4: Percentages of detecting outliers for independent case at α = 1%.

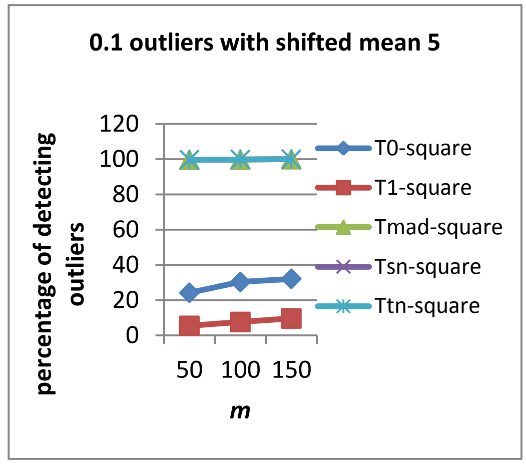
			P=2					P=5					P=10				
m	ε	μ	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	1.1	1.1	1.2	1.1	1.3	0.8	0.4	0.6	0.6	0.5	0.8	1.2	1	0.9	0.9
	0.1	3	17.1	7.5	59.2	60.4	60.6	9.7	10.9	89.8	90.6	89.6	6.6	5.3	98.6	97.8	98.4
		5	24.2	5.4	99.7	99.7	99.7	10.4	11.6	100	100	100	6.6	4.6	100	100	100
	0.2	3	3.4	2.8	36.0	37.8	38.2	1.8	2.7	71	72.2	72.2	2.2	2.8	69.6	65.3	69.2
		5	2.5	1.7	96.7	96.9	97.1	2	1.9	100	100	100	2.3	2.7	100	100	100
100	0	0	1.4	1.6	1	1.2	1.1	0.8	0.4	0.5	0.4	0.6	0.8	1.4	1.1	1.3	0.9
	0.1	3	20. 2	10.9	68.3	69	68.7	11.6	12.5	96.3	96.1	89.6	7.7	6.8	100	100	100
		5	30.3	7.6	99.8	99.9	99.9	13.0	12.7	100	100	100	8.2	7.5	100	100	100
	0.2	3	3.7	2.5	46.6	47.6	47.3	2.1	2.4	78.7	79	80.4	2	2.1	65.1	63.3	65
		5	2.8	2.3	99.3	99.4	99.4	2	2.6	100	100	100	2.2	2	100	100	100
150	0	0	1	1.4	0.6	0.7	0.7	0.7	1.4	1	1	1	0.8	1.1	0.8	0.7	0.8
	0.1	3	19.7	11.1	68.4	68.2	67.8	11.8	15.4	98.6	98.2	98.6	9	8.2	100	100	100
		5	32.1	9.6	100	100	100	14	11.4	100	100	100	9	9.6	100	100	100
	0.2	3	3.4	2.5	45.9	45.7	44.7	2.5	3	88.2	88.3	88.8	2.2	1.9	68	66	69
		5	2.9	2.1	99.3	99.3	99.3	2.1	2.9	100	100	100	2.2	1.7	100	100	100

Table 4.4 displays the results of the charts ability in detecting outliers. For the ease of comparison, the results in Table 4.4 are translated to graphical form in Figure4.4 to 4.6

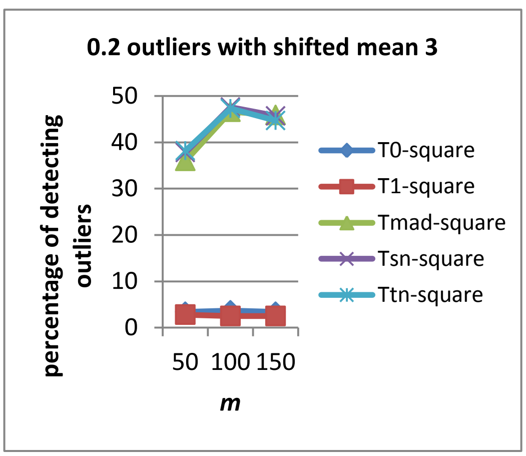
and our discussion will based on the graphs in the figures while referring back to Table 4.4 when necessary.



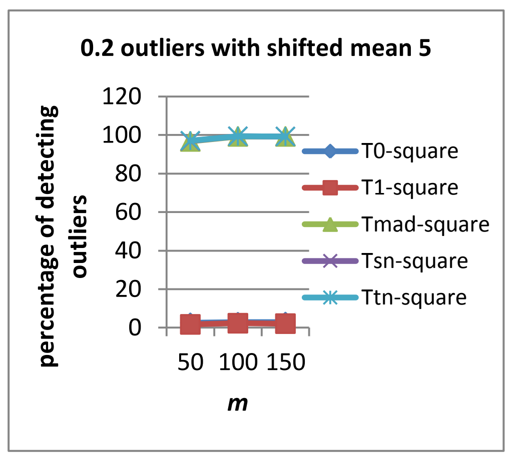
a



b



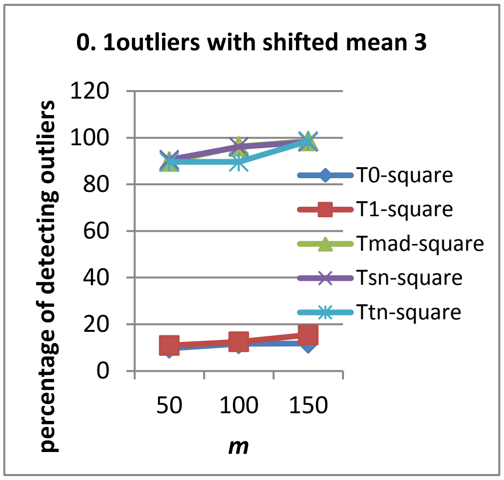
c

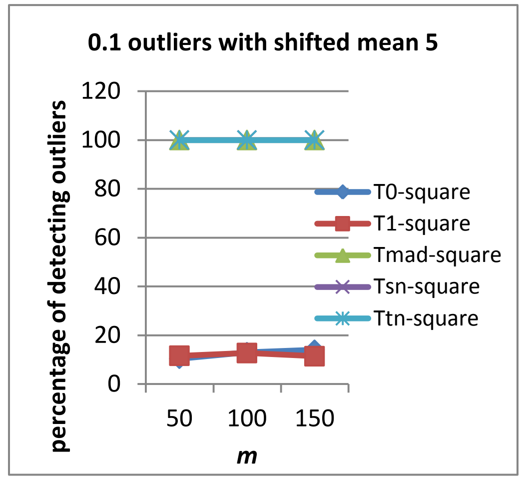


d

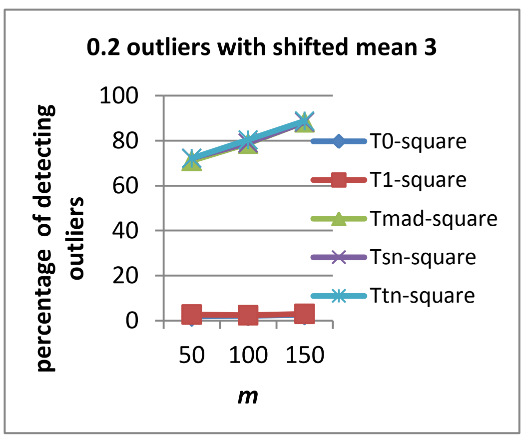
Figure 4.4: Percentages of detection of outliers when p=2

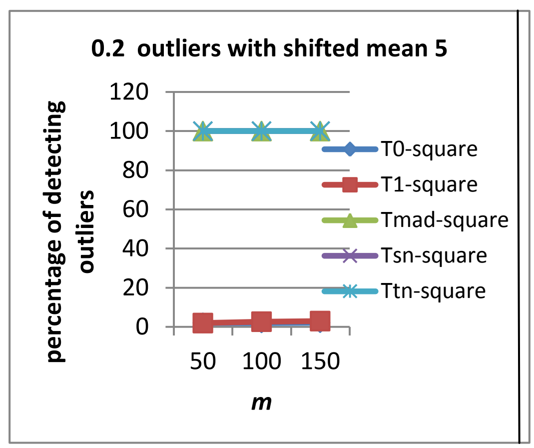
Figure 4.4 represents the performance for p = 2. As we glance through Figure 4.4a to 4.4d, we can observe that the disparity in performance between the robust and traditional charts is very obvious. Like in case of α = 5%, the robust charts perform excellently in detecting outliers as compared to the traditional charts. In Figure 4.4a, i.e. when the case of mild contamination (ε = 0.1 and $\boldsymbol{\mu}$ = 3), the performance of the robust charts is on par with each other with percentage of detection ranging from 59.2% to 69% leaving the traditional charts at 7.5% to 20.2%. We can see non straight lines across the m values which indicate that the changes in m do have some effect on the charts. When we increase $\boldsymbol{\mu}$ to 5, there is a sudden jump in the performance of the robust charts, achieving the 100% detection. The traditional charts still underperform as compared to the robust charts but there is also some improvement detected in T_0^2 chart but not in T_1^2 chart. The performance of the charts worsens when the moderate contamination (ε = 0.2 and $\boldsymbol{\mu}$ = 3). The highest recorded value is only at 47.6%, which belongs to T_{tSn}^2chart. As we shift the mean to 5, the performance for the robust charts bounces back to the almost perfect 100% level, while the performance of the traditional charts continue to worsen with the highest value of 2.9% only. At p = 2, the effect in the performance due to the changes in m could be detected in the robust charts when $\boldsymbol{\mu}$ = 3, however, no specific pattern (positive or negative) could be identified. We could not detect any obvious effect on the traditional charts except for the T_0^2 chart where there is a weak positive effect when ε = 0.1 and $\boldsymbol{\mu}$ = 5.





a **b**

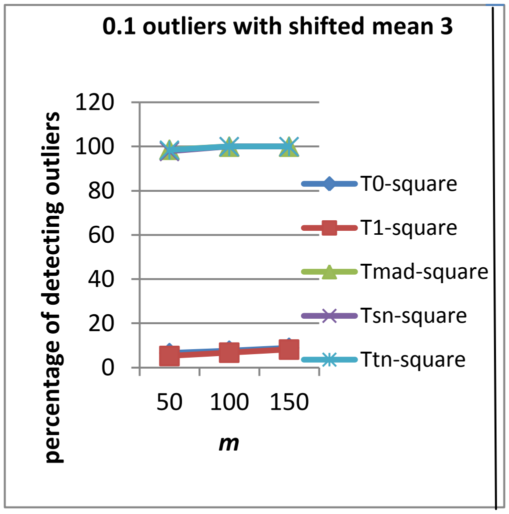




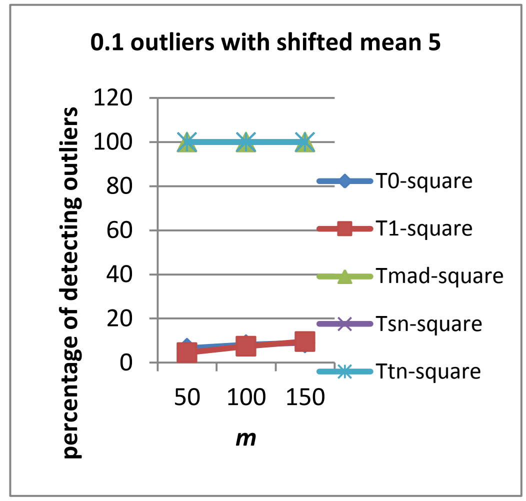
c d

Figure 4.5: Percentages of detection of outliers when p = 5

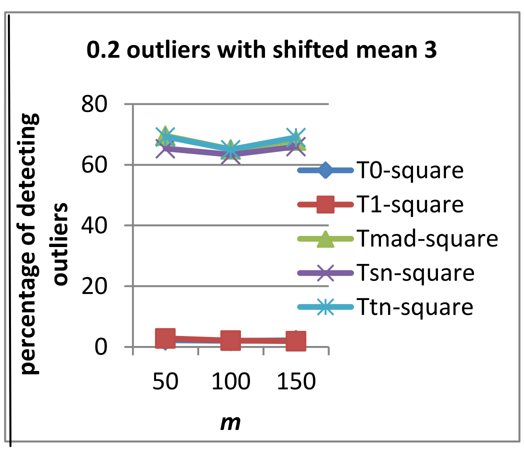
Figure 4.5 illustrates the performance of the charts when $p = 5$. Throughout the graphs in Figure 4.5, we notice that the gaps between the robust and traditional charts are wider as compared to $p = 2$. These situations occur due to the better performance of the robust charts while the performance of the traditional charts drop. Figures 4.5b and 4.5d which represent the conditions when $\boldsymbol{\mu} = 5$ with $\varepsilon = 0.1$ and 0.2 respectively, display perfect detection of outliers (100%) for all the robust charts regardless of the groups sizes (m). The already bad performance of the traditional charts worsens as ε increases. The percentage of detection drops to the value of less than 10%. When $\boldsymbol{\mu} = 3$, the performance of the robust charts increases towards 100% level as the group sizes increases and consistent at the 100% level when $\boldsymbol{\mu} = 5$. However, the change in mean only minutely affects the traditional charts. The change in the proportion of outliers from 0.1 to 0.2 does not show much effect on the robust charts, but worsens the performance of the traditional charts.



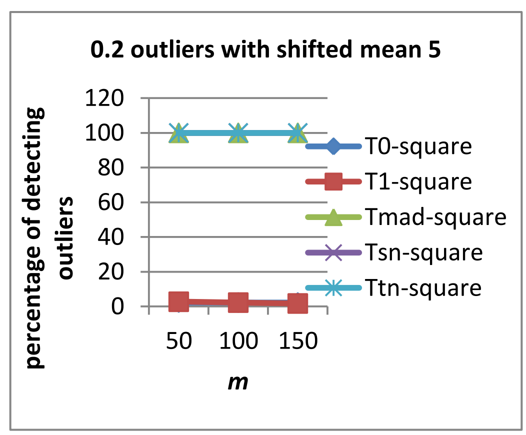
a



b



c



d

Figure 4.6: Percentages of detection of outliers when p =10

The graphical illustration on the percentage of detection for p = 10 is shown in Figure 4.6. Throughout the figure, we could observe three out of four graphs record almost perfect 100% detection for the robust charts except for the graph representing ε = 0.1 and $\boldsymbol{\mu}$ = 3. However, the performance of the robust charts (≈67%) is still far above the traditional charts. For p = 10, the performance of the traditional charts worsen to approximately 2.3% as compare to smaller p.

2.3 Dependent Variables (Case B)

For dependent case, this study only focuses on the shifted mean $\boldsymbol{\mu}$ = 5 since its main purpose is to analyze on the effect of correlation (versus no correlation) on the investigated charts. Alfaro and Ortega (2009) stated that it is sufficient to consider $\boldsymbol{\mu}$ = 0 which represents no change in mean and outliers free, and $\boldsymbol{\mu}$ = 5 which represents good leverage points. Similar to the independent case (Case A), the performance of the charts is measured based on the false alarm rates and percentage of detecting outliers. To reiterate, a control chart is considered robust under a particular condition when its false alarm rate satisfies Bradley's robust criterion. At a nominal level of α = 5% or α = 1%, the rate should lie within 2.5% to 7.5% or 0.5% to 1.5% respectively. A chart is considered to have good ability in detecting outliers when the percentage of detection is at least 80%.

2.3.1 False alarm rates and Percentage of detecting outliers at α = 5%

Tables 4.5 – 4.6 illustrate the performance of the control charts investigated for the case of dependent variables under high level of correlation (i.e. 0.9) with nominal false alarm rates of 5%.

Table 4.5: False alarms rates (percent) under dependent case for α = 5%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	5.6	6	6.2	5.4	6.1	5.3	5.6	6.4	6.6	5.7	4.7	5.7	4.9	5	3.2
	0.1	5	2.3	5.8	3.4	3.6	3.6	3.8	4.7	4	3.3	3.6	3.8	4.6	3.9	2.7	2.7
	0.2	5	2.3	3.1	1.8	2	2.4	3.6	3.8	3.2	2.4	2.9	4	3.8	2.9	3.1	3
100	0	0	5.5	6.3	4.1	4	4.6	5.4	5	4.8	4.9	4.2	4.5	5.8	4.5	4.2	3.3
	0.1	5	2.4	5.7	3	2.1	2.4	3.8	4.8	3.8	2.9	2.8	3.6	4.9	3.8	3.7	3.7
	0.2	5	2.2	3.1	1.4	1.3	1.4	3.7	4	3.4	1.8	2.5	3.4	4.7	3.4	3.4	3.3
150	0	0	5.3	6.4	5.4	4.9	5	4.3	4.4	4.8	4.3	3.6	4.6	5.4	4.8	4.4	4
	0.1	5	2.5	5.7	3.3	3.1	3	2.8	4.8	3.7	2.9	2.8	4.3	5	3.6	3.1	3.4
	0.2	5	2.4	3.2	1.6	1.7	2	2.7	3.6	2.5	2	2.1	4.3	4.8	2.5	3	3.3
%			44	100	67	56	56	100	100	100	67	89	100	100	100	100	100

Table 4.5 displays the performance of the charts in terms of their ability in controlling false alarm rates. First, let us look at p = 2. Under ideal condition (ε = 0, $\boldsymbol{\mu}$ = 0), all charts are robust regardless of the number of group sizes, m. Under non-ideal condition (with outliers), the performance of the three robust charts is in control when the proportion of outliers is low (ε = 0.1). Their performance dwindles to be out of control when the proportion of outliers increases to $\varepsilon = 0.2$, regardless of group sizes m. ForT_1^2 chart, its performance is in control regardless of the conditions investigated. However,T_0^2 chart fails to control the false alarm rates for all the investigated conditions except for when m is large (150) and ε is low (0.1). The false alarm rates for the three robust methods lie in between 1.3% and 3.6%, while T_0^2andT_1^2 charts produce values between 2.2% to 2.5% and 3.1% to 5.8%, respectively. From the results, we notice thatT_1^2 chart out performs its counterpart in controlling false alarm rates.

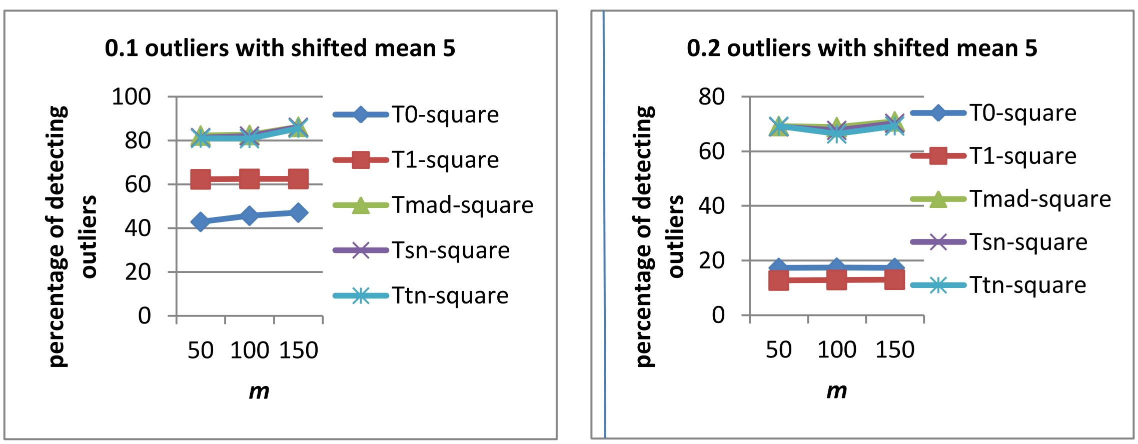
Next, for p = 5, under ideal condition, the false alarm rates for all robust and traditional charts are within Bradley's robust criterion. The rates for T_0^2, T_1^2 and T_{tTn}^2become closer to the nominal value when compared to p = 2, but the opposite for T_{tMadn}^2 and T_{tSn}^2. Under non-ideal conditions, we notice some improvement in the performance of the charts especially for T_0^2 andT_{tMadn}^2. The charts are robust regardless of m and ε. However, for T_{tSn}^2 charts we detect three non robust values while for T_{tTn}^2 chart, there is only one non robust value detected. All the non robust values are below the lower limit of robust interval and occur when $\varepsilon = 0.2$. Generally, we can see that as group size (m) increases, the rate of false alarm declines. This situation also occurs when the proportion of outliers increases. Among the three robust charts, T_{tMadn}^2 performs the best where 100% of its false alarm rates are within Bradley's criterion, followed by T_{tTn}^2 and T_{tSn}^2. The overall performance on false alarm rates shows that T_1^2 chart still outperforms all the other charts in terms of controlling false alarms rates.

As the number of quality characteristics, p increases to 10, we can observe great improvement in the performance of the charts in controlling false alarm rates. All the cells are shaded which indicates that the charts are robust. Overall, the robust charts perform the best when $m = 100$ under the influence of outliers. The change in the proportion of outliers (ε) from 0.1 to 0.2 has some inconsistent effect on the robust charts as well as the traditional charts depending on the group sizes. Even for the group sizes, the changes do not show any clear pattern in the performance of the charts.

Table 4.6: Percentages of detecting outliers for dependent case at α = 5%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	5.6	6	6.2	5.4	6.1	5.3	5.6	6.4	6.6	5.7	5.7	5.7	4.9	5	3.2
	0.1	5	42.9	62.3	82.3	81.4	81.0	15.3	12.6	38.1	32.9	31.5	10.6	7.5	19.1	12.9	12.1
	0.2	5	17.3	12.7	69.3	69.0	69.2	9.1	6.4	24.9	22.6	22.3	8	5.6	13.3	8.4	8.4
100	0	0	5.5	6.3	4.1	4	4.6	5.4	5	4.8	4.9	4.2	5.5	5.8	4.5	4.2	3.3
	0.1	5	45.7	62.5	82.6	82.1	80.9	18.2	12.7	54.3	32.8	29.5	9.9	8.4	21.1	14.9	12.6
	0.2	5	17.4	12.9	68.9	67.7	66.4	9.4	6.6	49.4	19.7	17.8	7.8	7.1	13.4	10.5	8.8
150	0	0	5.3	6.4	5.4	4.9	5	4.3	4.4	4.8	4.3	3.6	4.6	5.4	4.8	4.4	4
	0.1	5	47.1	62.5	86.3	86.1	85.6	17.5	13.1	36.3	32.9	29.5	10.6	8.6	17.2	14.4	11.2
	0.2	5	17.8	13	71	70.3	69.3	8.4	8	22.6	18	16.5	7.6	7.7	12	9.3	7.9

Table 4.6 depicts the results of the analysis on the ability of the investigated charts in detecting outliers. For ease of comparison, Figure 4.7 to Figure 4.9 graphically illustrate the performance of the charts in terms of the percentage of detecting outliers



a b

Figure 4.7: Percentages of detection of outliers when p = 2

For p = 2 as shown in Figure 4.7, we observe higher percentages among the three robust charts (66.4% to 86.3%) as compared to the traditional charts (12.7% to 62.5%). The differences in the percentages between the robust and traditional charts are large. When

ε = 0.1 the percentage for the three robust charts exceed 80%, but they decline to approximately 68% when ε = 0.2. In contrast the highest rates for T_0^2 and T_1^2 are 47.1% and 62.5% respectively at ε = 0.1 and furthermore respectively decline to 17% and 13% when ε = 0.2. No pattern could be detected among the charts when the group sizes (m) change.

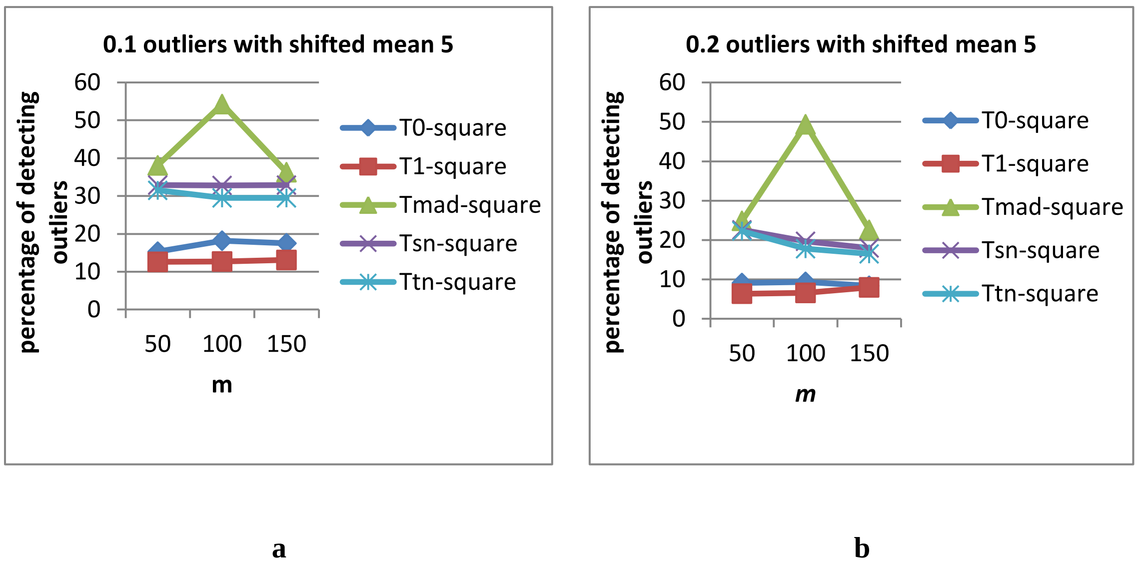
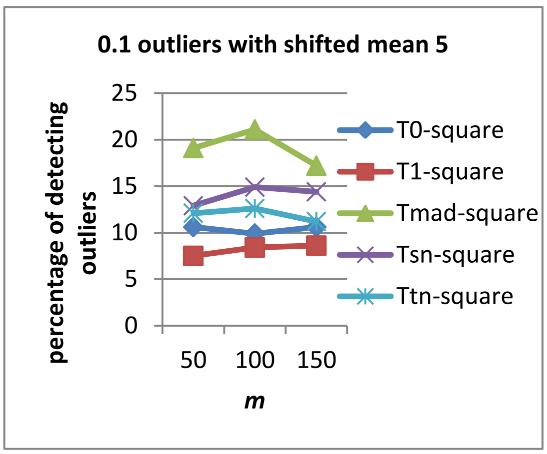
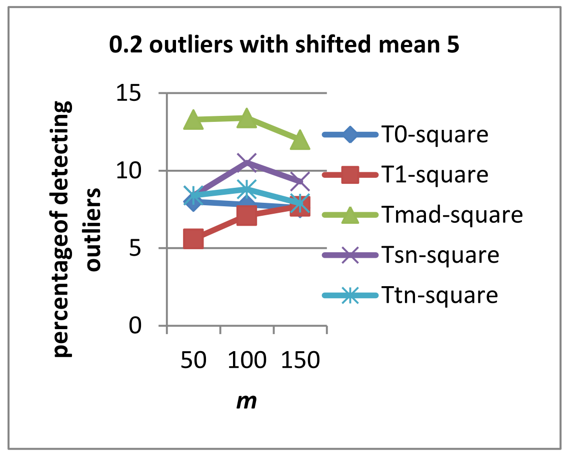


Figure 4.8: Percentages of detection of outliers when p = 5

Next, Figure 4.8 shows the rates of detection of outliers when p = 5. Across Figure 4.8, one can observe severe deterioration in the performance of the robust charts when compared to p = 2. Despite the deterioration, the robust charts still outperform the traditional charts. As the group sizes (m) increase from 50 to 100, there is a sudden increase in the percentage of detection of outliers for the performance of T^2_{tMadn}, followed by a sudden drop back to roughly the initial percentage as m increases to 150. This pattern occurs for both graphs in Figure 4.8a and 4.8b. This pattern indicates that the chart has stronger performance in case of moderate group size than in case of small and large group sizes. The other two robust charts show an almost consistent performance across m when ε = 0.1, but exhibit declining patterns when ε = 0.2. The change in ε from 0.1 to 0.2 also show drops in the performance of the investigated charts in general. Despite the inconsistency in the percentages of T^2_{tMadn} chart, the chart turns out to be the best among all the charts.



a



b

Figure 4.9: Percentages of detection of outliers when p =10

As shown in Figure 4.9, when p increases to 10, the ability of the investigated charts to detect outliers continue to drop. The best chart i.e. T^2_{tMadn} produce percentage of detection just slightly above 20% when ε = 0.1 and further deteriorate to below 14% as ε increases to 0.2. However, the robust charts still surpass the performance of the traditional charts. As m increases, the robust chart exhibit the same pattern for both graphs (Figure 4.9a and Figure 4.9b) such that the performance increases as m increases from 50 to 100 and declines when m equals to 150. No specific pattern could be identified in T_0^2chart, but for T_1^2 there is a positive effect in performance as m increases. However, this chart shows the worst performance among other investigated charts with percentage of less than 8%.

2.3.2 False alarm rates and percentage of detecting outliers at α = 1%

Tables 4.7 – 4.8 display the performance of the investigated charts for Case B at α = 1%, followed by graphical presentations of the charts ability in detecting outliers in Figures 4.10 - 4.12.

Table 4.7: False alarms rates (percent) under dependent case for α = 1%.

			P=2					P=5					P=10				
m	ε	μ	T_o^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_o^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_o^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	1.1	1.1	1.5	2	1.7	0.8	0.9	1.3	1.1	0.8	0.8	1.2	1.4	1.3	1.3
	0.1	5	0.6	0.3	0.8	0.5	0.7	0.6	0.8	0.6	0.5	0.4	0.9	1.2	0.9	0.5	0.6
	0.2	5	0.6	0.4	0.3	0.4	0.3	0.6	0.6	0.4	0.7	0.8	0.8	1.1	0.8	0.6	0.7
100	0	0	1.4	1.6	1.2	1.1	1	0.8	0.4	0.8	0.5	0.8	0.9	1.4	0.9	0.9	1.1
	0.1	5	0.3	0.6	0.9	0.8	0.8	0.6	0.2	0.8	0.3	0.6	0.7	0.7	0.9	0.7	1
	0.2	5	0.3	0.7	0.4	0.5	0.4	0.6	0.5	0.6	0.4	0.3	0.5	0.8	0.5	1.1	0.7
150	0	0	1	1.4	1	1	0.9	0.7	1.4	0.5	0.5	0.8	0.8	1.1	0.9	0.8	0.7
	0.1	5	0.4	0.6	0.5	0.5	0.3	0.5	0.9	0.5	0.7	0.7	0.7	0.9	0.5	0.8	0.5
	0.2	5	0.4	0.7	0.2	0	0	0.4	0.6	0.3	0.6	0.6	0.6	1.2	0.7	0.7	0.6
%			56	67	67	56	44	89	89	78	78	89	100	100	100	100	100

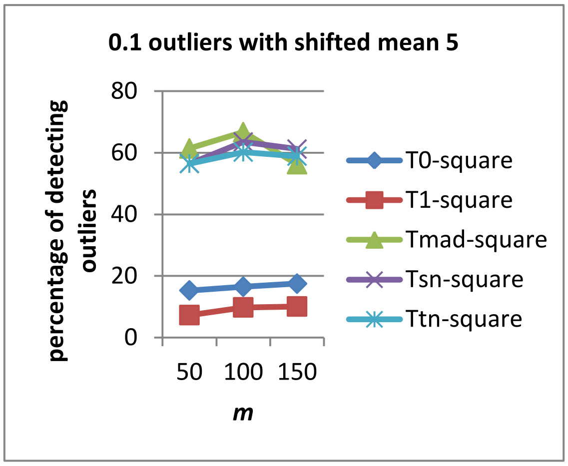
The performance of the control charts in terms of false alarm rates across various values of quality characteristics (p) is shown in Table 4.7. To reiterate, a chart is considered robust if its empirical false alarm rate is within Bradley's interval of (0.5% to 1.5%). For p = 2, when the condition is ideal (ε = 0, $\boldsymbol{\mu}$ = 0), there is inconsistency in the false alarm rates among the three robust charts and the two traditional charts. When the group size is small (m = 50), the rate for the robust charts are slightly liberal. Except for T_{tMadn}^2, the false alarms rates produce by the other two robust charts are beyond the upper Bradley's interval limit. For p = 5 and 10, one can observe that the false alarm rates for all the robust charts and T_0^2 chart are within Bradley's robust criterion. Under the influence of outliers and shifted mean (non-ideal), the robust charts are still in control of their false alarm rates when $\varepsilon = 0.1$ but generally fail to perform when $\varepsilon = 0.2$ under p = 2, while when p = 5 and 10 the performance of the robust charts are in control of false alarm rates for most of the conditions.

As in the ideal condition, no pattern of changes in the false alarm rates can be identified among the groups sizes m. Nonetheless, the change in the proportion of outliers (from 0.1 to 0.2) does have some negative effect on the performance. According to the percentages of the number of robust conditions, T_{tMADn}^2 chart performs the best at p = 2 while for p = 5 and 10, T_{tTn}^2 chart has the best performance.

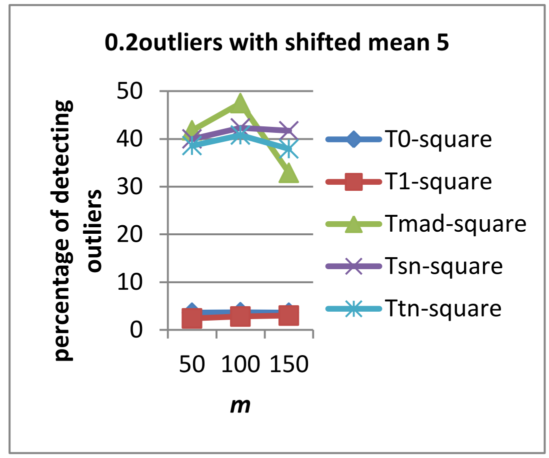
Table 4.8: Percentages of detecting outliers for dependent case at α = 1%.

			p=2					p=5					p=10				
m	ε	μ	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2	T_0^2	T_1^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
50	0	0	1.1	1.1	1.5	2	1.7	0.8	0.9	1.3	1.1	0.8	0.8	1.2	1.4	1.3	1.3
	0.1	5	15.3	7.3	61.4	56.5	55.1	4.2	4.4	18.8	15.7	12.5	1.7	2.1	6.2	3.6	5.2
	0.2	5	3.6	2.4	41.8	40	38.6	1.9	2.1	10.3	7.5	6.5	1	1.7	4	1.6	2.9
100	0	0	1.4	1.6	1.2	1.1	1	0.8	0.4	0.8	0.5	0.8	0.9	1.4	0.9	0.9	1.1
	0.1	5	16.5	9.8	66.7	63.5	60.2	4.6	5.6	18.3	11.6	11.4	2.4	1.7	7.9	5	3.6
	0.2	5	3.7	2.8	47.5	42.3	40.8	1.9	2.4	8.6	5.9	4.9	1.8	1.4	4	2.4	1.9
150	0	0	1	1.4	1	1	0.9	0.7	1.4	0.5	0.5	0.8	0.8	1.1	0.9	0.8	0.7
	0.1	5	17.5	10.1	56.3	61.2	58.9	4.6	4.7	9.2	14.7	12.2	2.4	2	6.6	4.4	3.4
	0.2	5	3.6	3	32.9	41.7	37.9	2.6	2.7	3.6	7	6	1.5	1.6	3.6	2.1	1.6

The performance of the charts in terms of outliers detection for α = 1% is exhibited in Table 4.8 and graphically presented in Figure 4.10 to Figure 4.12.



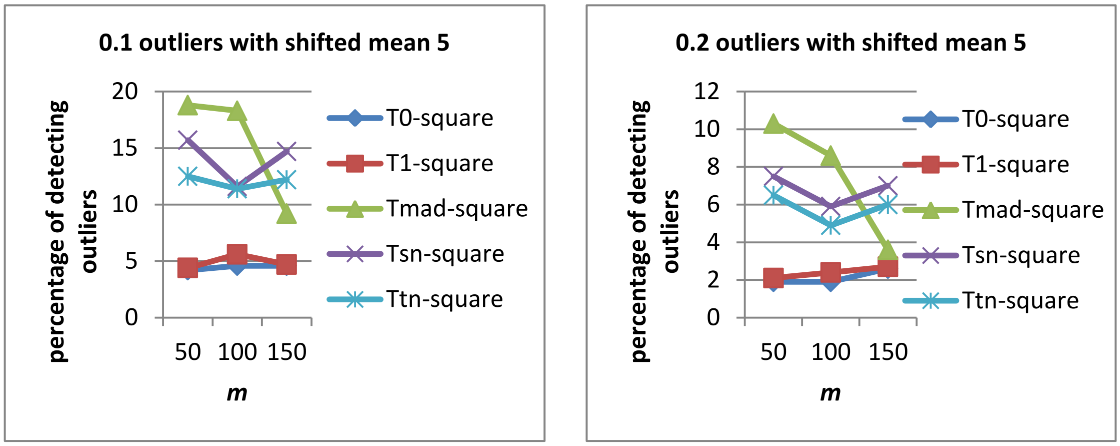
a



b

Figure 4.10: Percentages of detection of outliers when $p = 2$

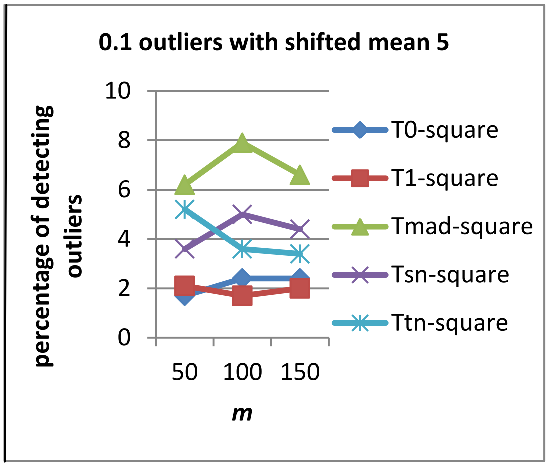
The performance of the charts for p = 2 is illustrated in Figure 4.10. Figure 4.10a which represents the charts when the condition of moderate contamination ε = 0.1 shows moderate performance of the robust charts with values between 55.1% and 66.7%. Despite the moderate performance, these charts still surpass the performance of the traditional, which produce percentage of detection lower than 20%. At ε = 0.1, the performance of the robust charts are on par with each other except for T^2_{tMadn} which plunges when m is large. However, among the traditional charts, there is an obvious gap between T_0^2 and T_1^2. T_1^2 Chart performs badly with values of no more than 10%. When extreme contamination and ε increases, as shown in Figure 4.10b, we observe a negative effect in the performance of all charts especially for the traditional charts with their performance drop to less than 5%. While for the robust charts, even though their performance deteriorated, they still have strong performance if compared to the traditional charts.



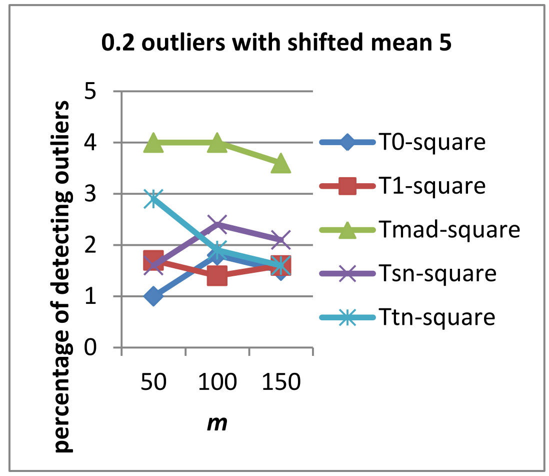
a b

Figure 4.11: Percentages of detection of outliers when p =5

Figure 4.11 presents the performance for p = 5. The figure shows deterioration in the performance for all charts. A crisscross pattern between the lines representing the robust charts could be observed across the group sizes (m), which indicates inconsistencies (effects) in the performance of the charts when there are changes in the group sizes. As for the traditional charts, despite the very low values produce by T_1^2, the chart outperform T_0^2at this stage. As ε increases (Figure 4.11b), the performance of all the charts deteriorates further. The highest value for the robust charts is slightly above 10% which belongs to T_{tMadn}^2while for the traditional charts, the highest value is almost 3%.



a



b

Figure 4.12: Percentages of detection of outliers when p =10

The charts' ability in detecting outliers dwindles further as the number of quality characteristics increases (p = 10) as depicted in Figure 4.12. As we move from Figure 4.12a to Figure 4.12b, where ε increases, a negative effect could be detected on the performance for all charts. Looking across the group sizes, we can see that both figures present inconsistent performances among the charts, thus indicating that the changes in group sizes offer no obvious pattern (effect) on the performance of the charts at this condition.

4.4 Analysis on Real Data

To evaluate on the performance of the investigated control charts on real situation, this study continues with the analysis on real data. Real data were provided to us by Asian Composites Manufacturing Sdn Bhd (ACM), an international joint venture Company between Boeing and Hexcel, USA, manufacturing composites for secondary structures of commercial aircraft. The plant is located at Bukit Kayu Hitam Industrial Estate, Kedah Darul Aman. ACM had provided us the real data on spoilers which consists of several features such as trim edge (X_1), trim edge spar (X_2), and drill hole (X_3). Spoilers are vital devices in an airplane. Their function is to increase lifts when the airplane is flying. The products are used in civilian, defense, and space applications, which cannot compromise any mistakes, albeit a minor one. Thus, careful monitoring is required to ensure that no variation occur in the process. Any slight mistake could risk a human life. A total of 47 products (spoilers) collected throughout year 2009 and 2010 was furnished to us. Out of the total, 21 products were collected in 2009 while the rest were collected in 2010. The collected data in 2009 has been used as historical data in Phase I. The data collected in 2010 has been considered as future observations used in phase II. The details of the historical data set is shown in Table 4.9 below. The calculated values of the upper control limits (*UCL*) are displayed in Table 4.10 below. The calculation of the *UCLs* for the three robust statistics obtained from the simulation method explained in section 3.8.2, while the *UCLs* for the traditional statistics were calculated using formula 3.35 . The values of future observations appeared in the first three columns, while the value of the T^2 statistic corresponding to each control chart is shown in the last four columns of Table 4.11.

The study on real data focused on significance level of 5% (α = 5%) only since the performance of the investigated T^2 statistics based on the simulated data are generally better under this level. By comparing the values of T^2 statistic in Table 4.11 with the corresponding control limit (UCL) values in Table 4.10, two similar observations (20 and 25) across the control charts are highlighted to point out that these observations are considered as out-of-control points. The detection of two similar points regardless of control charts shows that the performance of the investigated charts, be it traditional (T_0^2) or robust charts (T_{tMADn}^2, T_{tSn}^2 and T_{tTn}^2) are on par for this real data problem.

As we could observe, the Hotelling's T^2 values for the three robust charts are the same. The similarities could be due to the large trimming process (40%) and the choices of measure of dispersion. Even though Mad_n, S_n and T_n are different measures of dispersion, which are supposed to produce different scatter matrices, but their almost the same approach might produce the same dispersion value, and this situation is aggravated when the already small observations (<30) was further trimmed by 40%.

The study on simulated data revealed that the traditional chart generally underperformed in detecting outliers, nevertheless, when applied to this real data problem, the chart performed as good as the robust charts. The result concurs with the finding in simulated study such that under small number of quality characteristics (not more than 5), the performance of the traditional chart and the robust charts are almost the same. The robust charts work better under higher number of quality characteristics.

Table 4.9: Historical data set (Phase I data)

Product No.	**Trim edge (x_1)**	**Trim edge spar (x_2)**	**Drill hole (x_3)**
1	-0. 0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	-0.0017	0.0109	0.0177
5	-0.0005	-0.0010	0.0106
6	0.0016	-0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	-0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	-0.0067	0.0147
16	0.0110	-0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 4.10: The values of the upper control limits for the three robust and one traditional charts.

Types of Control Chart	Upper Control Limit $\alpha = 5\%$
T_O^2	11.035
T_{tMadn}^2	20.6053
T_{tSn}^2	20.0658
T_{tTn}^2	6.724

*Table 4.11: The Hotelling's T^2 values for the future (*Phase *II) data*

Product No.	x_1	x_2	x_3	T_0^2	T_{tMadn}^2	T_{tSn}^2	T_{tTn}^2
1	0.0041	0.0087	0.0129	0.5582	0.4796	0.4796	0.4796
2	0.0047	0.0109	0.0124	0.9003	0.71297	0.71297	0.71297
3	0.0031	0.0057	0.0096	0.4992	0.21675	0.21675	0.21675
4	0.0035	-0.0020	0.0101	0.5463	0.0699	0.0699	0.0699
5	0.0040	-0.0028	0.0125	0.4592	0.1086	0.1086	0.1086
6	0.0031	0.0008	0.0061	0.9013	0.1658	0.1658	0.1658
7	-0.0019	0.0101	0.0112	3.0933	1.9958	1.9958	1.9958
8	0.0009	0.0039	0.0082	0.8061	0.3030	0.3030	0.3030
9	-0.0052	0.0090	0.0203	7.3602	4.9852	4.9852	4.9852
10	-0.0008	0.0110	0.0184	3.6198	2.6713	2.6713	2.6713
11	-0.0021	0.0139	0.0170	5.3839	3.7456	3.7456	3.7456
12	-0.0017	0.0092	0.0061	2.7387	1.5149	1.5149	1.5149
13	-0.0010	0.0133	0.0138	3.8058	2.6230	2.6230	2.6230
14	-0.0030	0.0002	0.0053	2.0548	0.8457	0.8457	0.8457
15	0.0016	0.0134	0.0151	2.5073	1.8910	1.8910	1.8910

16	0.0027	0.0086	0.0070	1.1976	0.5881	0.5881	0.5881
17	0.0004	0.0086	0.0087	1.5798	0.9143	0.9143	0.9143
18	-0.0036	0.0136	0.0129	5.7910	3.7964	3.7964	3.7964
19	-0.0028	0.0003	0.0078	1.8304	0.8318	0.8318	0.8318
20	0.0120	0.0123	0.0768	**38.1397**	**26.121**	**26.121**	**26.121**
21	-0.0015	0.0004	0.0115	1.2651	0.6632	0.6632	0.6632
22	0.0009	0.0232	0.0202	8.4181	5.9358	5.9358	5.9358
23	-0.0035	0.0088	0.0107	3.7588	2.3489	2.3489	2.3489
24	0.0016	0.0061	0.0066	1.0602	0.4303	0.4303	0.4303
25	-0.0228	-0.0466	0.0231	**42.8447**	**25.112**	**25.112**	**25.112**
26	0.0037	-0.0038	0.0147	0.4832	0.2047	0.2047	0.2047

4.5 Comparison among the robust Hotelling's T^2 charts

The results of the analysis on the robust Hotelling's T^2 charts in Sections 4.2 and 4.3 proved that the proposed robust Hotelling's T^2 charts outperform the traditional Hotelling's T^2 charts in most conditions of independent and dependent cases.

In this section we will specifically compare the performance among the robust charts. For the purpose of comparison, Table 4.12 summarized the conditions that are considered robust and those, which have at least 80% probability of detecting outliers for each robust chart across the two significance levels in case of independent characteristics, while at least 50% probability of detection for the robust charts in case of dependent characteristics. Under the dependency characteristics, the table is further divided into the two performance measurement i.e. false alarm (FA) and probability of detection (POD). The comparison covers both the levels of significance α.

Table 4.12: Overall performance for independent and dependent cases

Types of Charts	Independent (trimmed)				Dependent (trimmed)			
	α = 5%		α = 1%		α = 5%		α = 1%	
	FA	POD	FA	POD	FA	POD	FA	POD
T^2_{tMadn}	67%	89%	44%	69%	78%	17%	78%	0
T^2_{tSn}	62%	89%	58%	67%	70%	17%	81%	0
T^2_{tTn}	62%	89%	53%	72%	70%	17%	74%	0

With respect to independent case, we can identify that T^2_{tMadn} produce the highest percentage of robust conditions when α = 5%. In contrast, when α = 1%, T^2_{tSn} chart outperforms the other two robust charts leaving T^2_{tTn} with quite a difference in percentage. However, for both significance levels, T^2_{tTn} scores neither the best nor the worst. Thus, it is hard to determine which chart is the best in controlling false alarm rates since all the charts perform reasonably well for both levels of significance. In terms of their capability in detecting outliers, all robust charts show equally strong performance at α = 5% but T^2_{tTn} chart slightly outperform the other charts when at α = 1%.

Under dependent case, the percentage of the robust conditions for each chart increases as compared to the independent case, and further increases when α = 1%. The smaller the significance level, the better is the performance of the charts, and this is in contrast with the independent case. The ranking of the charts is the same as in the independent case where T^2_{tMadn} performs the best for α = 5% and T^2_{tSn} is the best for α = 1%, while T^2_{tTn} is neither the best nor the worst. In detecting outliers, the percentages of the three robust charts drop to 17% and further worsen under smaller significance level. In general, the charts seem to be almost on par with each other in terms of false alarm rates and probability of detection for dependent case.

4.6 Summary

This chapter developed and evaluated three proposed robust multivariate Hotelling's T^2 charts for location based on trimmed mean estimator, and for scale based on the trimmed covariance matrix. The evaluations, which were presented under independent and dependent cases in Section 4.2 and Section 4.3 respectively, revealed that in general the robust Hotelling's T^2 charts perform as good as T_1^2 and better than T_0^2 in terms of controlling false alarm rates. However, in terms of detecting outliers, they perform remarkably well, whereas the traditional charts fail to perform under most conditions.

Under dependent case, when the traditional charts fail to control the false alarm rates, the robust charts seem to be unperturbed with the situation. Even though the capability of the robust charts in detecting outliers dwindles under this case, the charts still outperform both the traditional charts. Without the effect of correlation, the robust charts are superior in detecting outliers and perform reasonably well in controlling false alarm rates.

The increase in p values show improvement in the robust charts as well as the traditional charts with respect to controlling false alarm rates, but show no effect in terms of detecting outliers. When ε changed from 0.1 to 0.2, the robust charts were not obviously affected in terms of false alarm; however, the change negatively affected the probability of detection. As the mean shifted to a larger value, the robust charts seemed to be able to control their false alarm rates and simultaneously increased the probability of detecting outliers. If we compare the performance of the robust charts for $\alpha = 5\%$ and $\alpha = 1\%$, we observed more robust conditions (able to control false alarm rates) at $\alpha = 1\%$, however, for the probability of detection, $\alpha = 5\%$ produced better results.

CHAPTER FIVE

MODIFIEDHOTELLING'ST^2CONTROL CHARTS USING WINSORIZED MOM WITH WINSORIZED COVARIANCE MATRICES

3.1 Introduction

This chapter comprises of three robust Hotelling's T^2 control charts whereby each control chart uses different winsorized modified one-step *M*-estimators(*MOM*) as the location estimators and its respective winsorized covariance matrix as the scatter matrix. The difference between each winsorized *MOM* is due to the robust scale estimators used in the trimming criterion. These scale estimators are $MAD_{n,}$ S_n and $T_n$and are thoroughly explained in Chapter 3. The steps to the construction of the three robust control charts are explained as follows:

Step 1: Obtain the original data (with outliers)

Step 2: Trim the original data using one of the three *MOM* criterions. By default, the *MOM* criterion uses the robust scale estimator MAD_n. MAD_nwill be replaced by S_n and $T_n$to generate different criteria.

Step 3: Winsorized the trimmed data to obtain winsorized sample.

Step 4: Compute winsorized mean vector

Step 5: Compute winsorized covariance matrix

Step 6: Compute Hotelling's T^2 statistics

For a more detail explanation on the process, please refer to Section 3.4.The proposed robust Hotelling's T^2charts are denoted as T^2_{wMadn}, $T^2_{wSn}$and T^2_{wTn}.

Like in chapter 4, the performance of these robust Hotelling's T^2 charts is compared to the two traditional charts, T_0^2 and T_1^2, in terms of their robustness (false alarm) and the percentage of detecting outliers. The organization of this chapter is similar to Chapter 4. Each of the three proposed charts was tested on different scenarios (conditions) which were created by combining the number of quality characteristics (*p*), number of group sizes (*m*), proportion of outliers (*ε*), and the changes (shift) in the mean vectors ($\boldsymbol{\mu}$). The results, which are in the form of false alarm rates and percentage of detection of outliers is presented in tables and figures as in the previous chapter.

Correspondingly, the charts are considered robust if their empirical false alarms rates are within Bradley's interval (1978). The closer the values of the empirical false alarm rate to the nominal level (*α*), the better the performance of the chart in terms of controlling false alarm rates. The analysis of the results is divided into two major sections based on the nature of the quality characteristics i.e. independent and dependent, which are represented by Case A and Case B, respectively.

3.2 Independent Variables Case (A)

The following subsections present the performances for the charts according to the nominal false alarms rates α. Tables 5.1- 5.2 record the performance (in percentage) of the charts according to the two measurements i.e. false alarms rates and percentage of detecting outliers. Figure 5.1-5.6 later helps us to visualize the performance regarding the charts ability in detecting outliers.

3.2.1 False alarm rates and Percentage of detecting outliers at α = 5%

Table 5.1 represents the performance of the robust and traditional Hotelling's T^2 control charts in terms of false alarm rates with respect to the number of quality characteristics, *p*. The false alarm rates that satisfy the Bradley's robust criterion (which is in between 2.5% to 7.5%) are represented by the shaded cells.

Table 5.1: False alarms rates (percent) under independent case for α = 5%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T^2_O	T^2_1	T^2_{wMa}	T^2_{wSn}	T^2_{wTn}	T^2_O	T^2_1	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}	T^2_O	T^2_1	T^2_{wM}	T^2_{wSn}	T^2_{wTn}
25	0	0	5.9	5.7	4.7	5.4	5.8	5.1	5.4	4.7	4.9	4.5	4.9	6.3	4.1	4.9	4
	0.1	3	2.9	5.6	3.8	3.8	3.7	3	3.5	3.9	3.4	4.1	3.1	4.4	3.5	4.1	3.6
		5	2.3	5.6	4	4	3.6	3.1	3.4	4.1	3.8	4	3.2	3.9	3.5	4.2	3.9
	0.2	3	2.8	3.1	3.3	3.2	3	3.2	3.3	3.3	3.4	3.6	3.3	4.8	4.4	4	4.5
		5	2.4	2.5	3.7	4.4	3.8	3.3	3.6	4.2	5	5.7	3.2	4.9	8.3	8.8	9.3
50	0	0	5.6	6	5.8	5.6	6.3	5.3	5.6	5.2	5.8	4.8	5.7	5.7	5.5	5.8	6.1
	0.1	3	2	5.7	2.9	3.4	3.5	2.7	3.6	3.5	3.5	3.6	4.1	3.5	5	5	5.5
		5	1.6	5.7	3.6	3.6	3.7	2.6	3.7	3.6	4.3	3.7	3.8	3.4	6.2	6.5	6.3
	0.2	3	2.1	3.1	2.5	2.8	2.9	2.6	3.1	3.6	3.4	4	4.2	4.8	6.2	5.6	5.9
		5	1.6	2.5	3.6	4.1	4.2	2.5	3.6	5.4	5.7	6.2	4.0	4.8	9.9	10.4	10.9
100	0	0	5.5	6.3	4.5	4.6	4.8	5.4	5	4.4	4.9	4	5.5	5.8	4.9	4.9	4.6
	0.1	3	2.1	5.7	2.8	2.5	2.7	2.9	4.3	3.4	3.2	3.3	3.3	4.3	4.3	4.4	4.7
		5	1.6	5.7	2.7	2.7	2.8	2.8	3.8	3.5	4	3.5	3.4	4.4	5.4	6	5.5
	0.2	3	2.1	3.2	2.5	2.3	2.3	3	4	4.2	3.4	4	3.5	4.9	5.4	4.3	5.5
		5	1.6	2.6	2.8	3.2	3.2	2.9	3.5	5.8	5.7	5.6	3.4	4.6	8.7	9.3	10.2
%			33	100	100	93	93	100	100	100	100	100	100	100	80	80	80

In general, under ideal condition($\varepsilon = 0, \mu = 0$), the false alarm rates for all the control charts regardless of *p* are within the Bradley's interval. For most of the charts, the results improve (approaching nominal level of 5%) when *p* increases, but not in the case of T_1^2 chart where the values are slightly liberal (above 5%). In contrast, under non-ideal condition (with outliers and shifted mean), the performance of the robust charts is in control except when *p* is large and under extreme contamination (*m* = 100, $\boldsymbol{\mu}$ = 5 and ε = 0.2). Even we can observe out of control false alarm rates for T^2_{wSn} and T^2_{wTn} charts when *p* = 2 under such conditions. The false alarm rates for the traditional T_1^2chart are in control regardless of the conditions while for the other traditional chart i.e. T_0^2, there are some fall outs occur when p = 2.

Specifically, when p = 2, the three robust charts and T_1^2chart are in control of their false alarm rates but not in the case of T_0^2 chart. As the group size (m) increases, we can observe a declining effect on the false alarm rates for all the charts, except in the case of T_1^2 chart. When the mean shifts from 3 to 5 the false alarm rates for the robust control charts inflate. Simultaneously, the increasing of proportion of outliers from 0.1 to 0.2 also has some effect on the false alarm rates of the three robust charts where a decreasing trend for all conditions can be observed.

As we move to p = 5, we observe that under non-ideal condition, the performance of the three robust charts and the traditional charts remains robust across all the conditions. However, the ability to control false alarm rates is more prominent in robust charts. Most of the false alarms rates are closer to the nominal value of 5% when compared to the traditional charts, especially when ε = 0.2 and $\boldsymbol{\mu}$ = 5. The nearer the values to the nominal level α, the better the chart's ability to control the false alarm. When the proportion of outliers (ε) increases, there is a positive effect on the robust charts but the other way round for the traditional charts. It can be noticed that all the robust charts perform well where 100% of their cells are shaded.

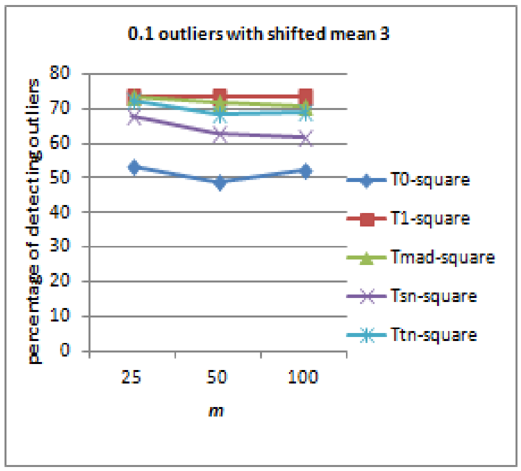
As p increases to 10, even though we can observe some improvement in the performance of the three robust charts under non ideal condition, there are a few non robust values detected when m = 100, $\boldsymbol{\mu}$ = 5 and ε = 0.2. The values inflated to almost 11%, which signify that the charts fail to control the false alarm at this state. In contrast, the performance of T_0^2 and T_1^2 charts are consistent and in control of the false alarm.

Overall, the performance of the robust charts are detected to be stronger than the traditional charts, except at ε = 0.2 and $\boldsymbol{\mu}$ = 5. We can also detect that as group sizes (m) increase; there is a slight increase in false alarm rates for the robust charts, but no effect on the traditional charts. Correspondingly, the effect is negative when there is an increase in the proportion of outliers ε for all the charts. According to the comparison among the robust charts and the percentage of the number of the in control values of false alarms rates across all p, the T_{wTn}^2 chart has the strongest performance.

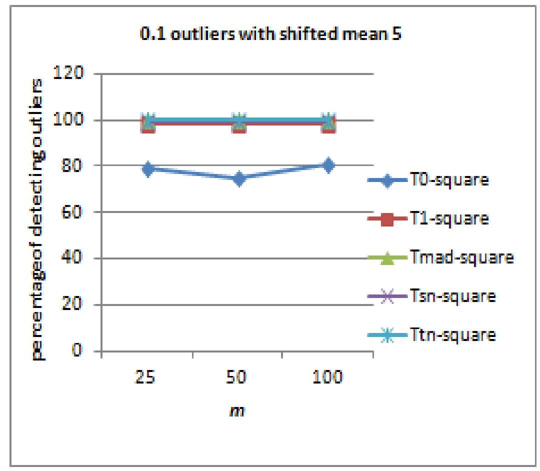
Table 5.2: Percentages of detecting outliers for independent case at α =5%

			P = 2					p = 5					p = 10				
m	ε	μ	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
25	0	0	5.9	5.7	4.7	5.4	5.8	5.1	5.4	4.7	4.9	4.5	4.9	6.3	4.1	4.9	4
	0.1	3	53.3	73.7	73.3	67.9	72.3	40.5	14.9	78.2	65.8	76.6	24.4	16.2	56.3	46.1	53.2
		5	79.4	98.6	99.9	99.6	99.9	46.8	15.7	100	99.9	100	25.9	16.4	100	100	99.9
	0.2	3	14.5	12.2	29.3	18	24.9	9.9	7	24	12.3	19.6	8.2	8.1	23.5	12.3	18.8
		5	13.5	10.5	90.7	63.9	85.7	10.1	6.8	96.9	78	94.6	8.2	7.7	96.9	88.3	96.5
50	0	0	5.6	(6)	5.8	5.6	6.3	5.3	5.6	5.2	5.8	5.6	5.7	5.7	5.5	5.8	6.1
	0.1	3	49.1	73.6	72.1	62.7	68.3	36.4	15.2	75.3	59.8	70.4	24.7	23	60.7	42	53.3
		5	74.7	98.7	100	99.6	99.9	44.2	20.6	100	99.9	100	26.5	25.8	100	99.9	100
	0.2	3	17.6	12.6	27.9	20.1	23.7	11.9	7.1	22	15.2	19.8	10.6	9.9	21.9	13.4	17.5
		5	17.1	10.7	91.9	59.7	86.2	12.3	9.4	95.5	63.5	90.8	10.6	10.1	96.8	75.2	94.4
100	0	0	5.5	6.3	4.5	4.6	4.8	5.4	5	4.4	4.9	4	5.5	5.8	4.9	4.9	4.6
	0.1	3	52.2	73.6	70.8	61.7	69	41.7	20	74.2	55.2	69.3	29.4	27.6	63.6	43.3	56.3
		5	80.6	98.7	100	99.5	99.9	50.7	20.9	100	100	100	31.4	29.1	100	99.7	100
	0.2	3	17.5	12.6	24	18.2	22.7	11.8	11.2	18	11.7	15.8	10.7	10.1	18.1	11.4	16.4
		5	17.5	10.8	91.4	52	85.5	12	9.7	90.3	43.6	82.6	10.8	10.1	92	46.6	85.8

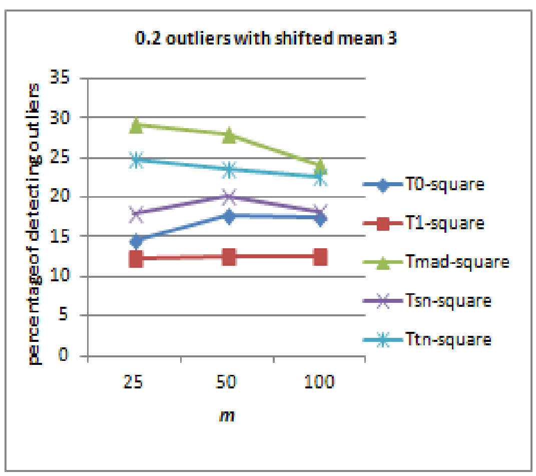
Table 5.2 records the percentage of detecting outliers for all charts. However, for the ease of comparison, we will refer to the visual presentation in Figure 5.1 to Figure 5.3 and alternately refer to the numerical values in Table 5.2.



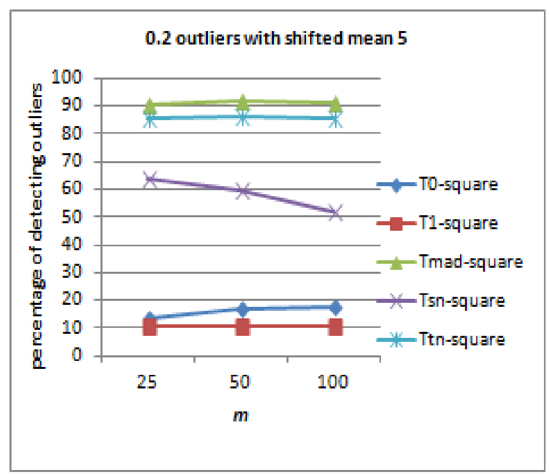
a



b



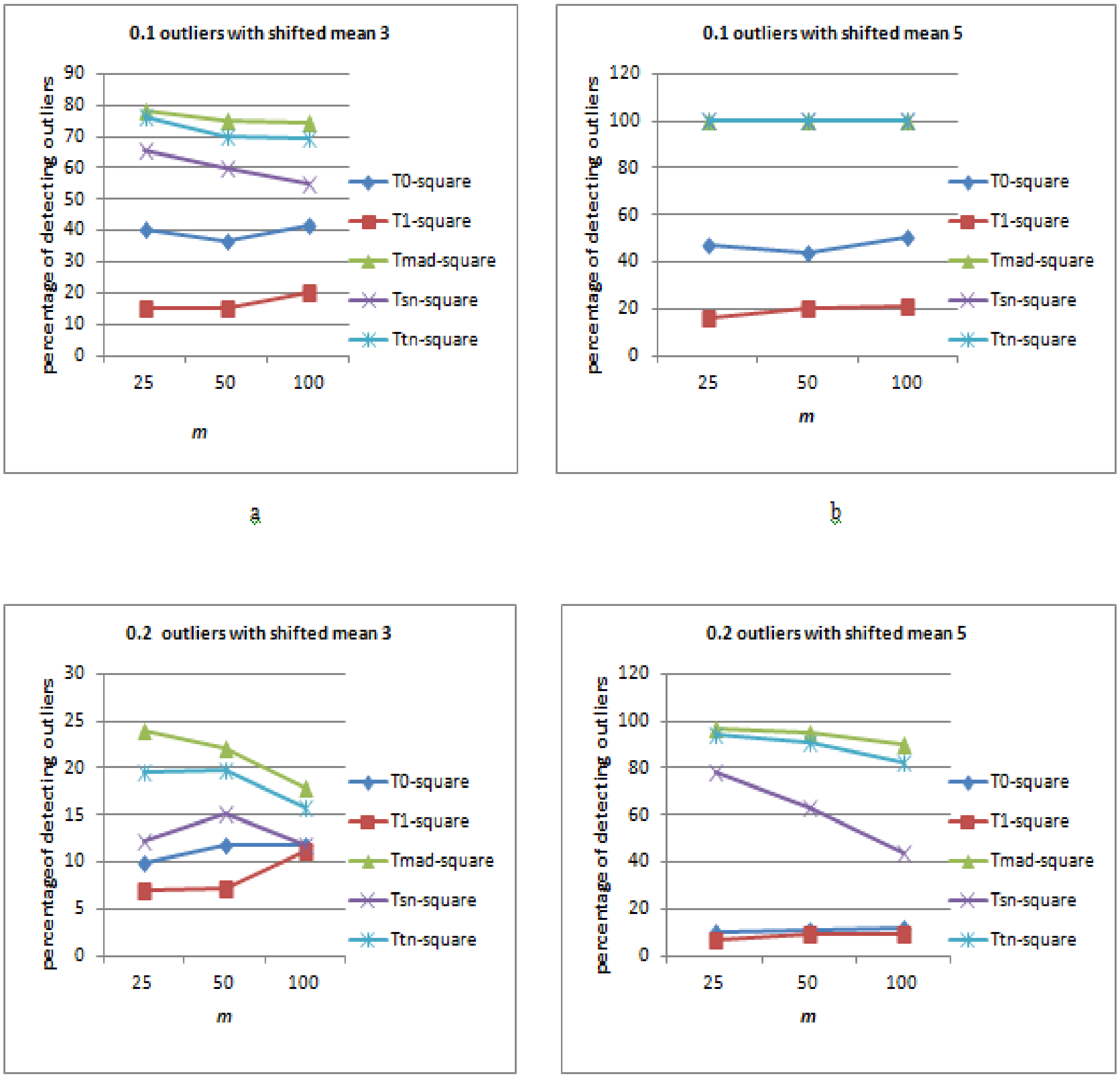
c



d

Figure 5.1: Percentages of detection of outliers when p = 2

Figure 5.1 presents the performance of the investigated charts in terms of their ability in detecting outliers when quality characteristics (p) equals to 2. As shown in Figure 5.1a, when the mild contamination ($\varepsilon = 0.1, \boldsymbol{\mu} = 3$), the T_1^2 chart appears to have the best performance, followed by the robust charts while T_O^2 chart performs theworst. However, in Figure 5.1b, when the mean shifts to 5 while the proportion of outliers stays constant at 0.1, all the charts show improvement in their ability in detecting outliers with values almost reaching the perfect 100% detection except for T_O^2 with values less than 80%. When ε increases to 0.2 (Figure 5.1c), the performance of all the charts suddenly drops. The highest value which is less than 30% belongs to T_{wMadn}^2 followed by the other two robust charts, T_{wTn}^2 andT_{wSn}^2. However, the performance for the robust charts bounces back to above 90%when the extreme contamination ($\varepsilon = 0.1, \boldsymbol{\mu} = 3$) as demonstrated in Figure 5.1d. The ranking for the three robust charts maintains as before but not for the traditional charts where T_1^2 chart records the worst with percentage of detection as low as 11%. The changes in group sizes (m) do not show much effect on the performance of the control charts in detecting outliers.

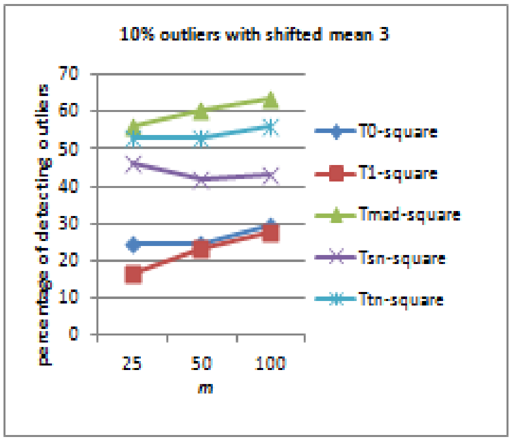


a b

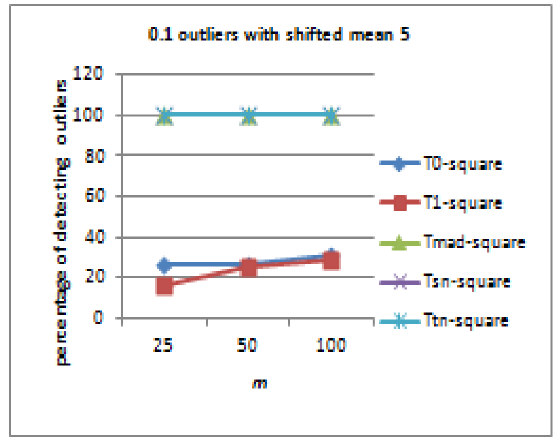
c d

Figure 5.2: Percentages of detecting outliers when p = 5

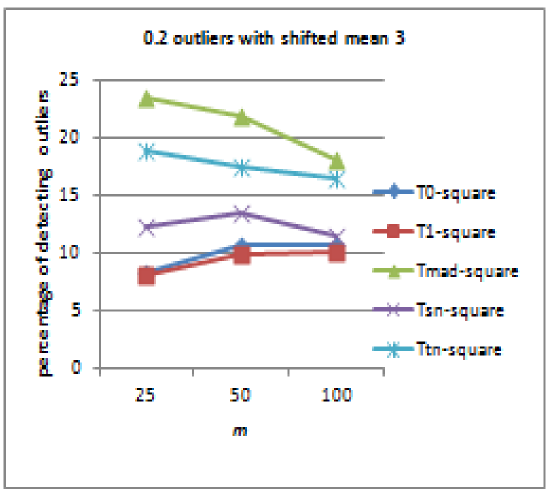
Next, Figure 5.2 represents the performance of the investigated charts for various conditions at p = 5. Throughout the figures, it could be easily observed that T_1^2 has the worst performance with percentage of detection only reaches 20% while the robust chart T_{wMadn}^2 always at the highest with values reaching the 100% detection. At ε = 0.1 and μ = 3 as displayed in Figure 5.2a, the percentage of detection among the robust charts show some differences with T^2_{wMadn} record thehighest detection (74.2% -78.2%) followed by T^2_{wTn} (69.3% -76.6%) and T^2_{wSn} (55.2% - 65.8%). As the mean shifts to 5 (Figure 5.2b), the performance of all the three robust charts suddenly increases to almost the perfect 100% detection. However, the performance of the traditional charts stays the same. Similar to the situation at p = 2, when the value of ε increases to 0.2, there is an abrupt drop in the performances of all the charts as shown in Figure 5.2c, but the percentage bounces back to as high as 96.9% when the mean shifts from 3 to 5 (Figure 5.2d). The changes in group sizes do not show much effect on the charts, however under severe condition (ε = 0.2 and $\boldsymbol{\mu}$ = 5) we can observe a declining effect for T^2_{wSn} as the group sizes increase. Among the robust charts, T^2_{wSn} chart's performance is always at the bottom.



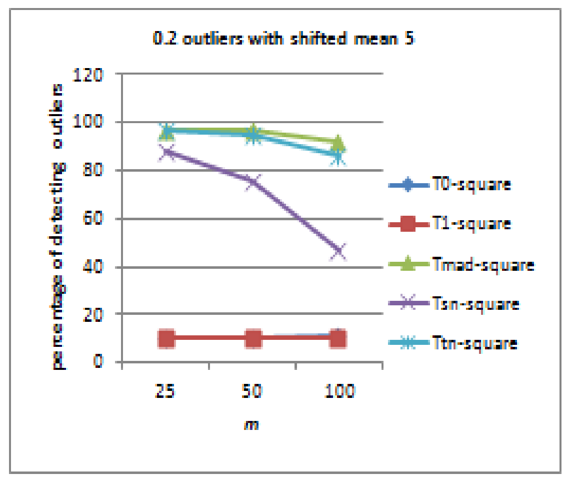
a



b



c



d

Figure: 5.3:Percentages of detection of outliers when p = 10

Figure 5.3 represents the performance of the control charts when p = 10. The patterns are observed to be almost similar to p = 2 and 5 with T^2_{wMadn} produced the best control charts followed by T^2_{wTn} and T^2_{wSn} while the traditional charts especially the T_1^2 charts still records the worst. When the proportion of contamination (ε) and mean shift ($\boldsymbol{\mu}$) is small i.e. 0.1 and 3 respectively (Figure 5.3a), the highest percentage of detection for the best chart $\mathrm{T}^2_{\mathrm{wMadn}}$ is 63.6%. This occurs when *m* = 100. The lowest for the robust charts which belongs to $\mathrm{T}^2_{\mathrm{wSn}}$ is 46.1% at *m* = 25. In contrast, the best percentage of detection for the traditional chart which is produced by T^2_0 is only 29.4% i.e. at *m* = 100. As we move to Figure 5.3b, i.e. when $\boldsymbol{\mu}$ changes to 5 while ε maintains at 0.1, we notice that the performance of the robust charts leaps to the highest of 100%. However, the performance drops tremendously in Figure 5.3c when ε increases in value to 0.2 even though with lower value of $\boldsymbol{\mu}$ (i.e. 3). We could also observe declining in performance for the robust charts as group sizes increase. In Figure 5.3d, when $\boldsymbol{\mu}$ shifts to 5, the performance of the robust charts, T^2_{wMadn} and T^2_{wTn} leaps again to almost 97% but the T^2_{wSn}chart only able to get up to 88.3%. The T^2_1 chart still fares worst amongst other charts.

3.2.2 False alarm rates and Percentage detecting outliers atα = 1%

The following subsections discuss the performance of the investigated control charts according to the false alarms rates and percentage of detecting outliers when α = 1%. The false alarm rates produced by the charts are depicted in Table 5.3. To reiterate, a chart is considered robust or in control of false alarm if the empirical rate falls between 0.5% and 1.5%. Under ideal condition (ε = 0 and $\boldsymbol{\mu}$ = 0), all the robust charts are in control of false alarm rates regardless of the number of quality characteristics (*p*) and number of group sizes (*m*) except for T^2_{wMadn}, where the false alarm rate is slightly above the upper interval limit (i.e. 1.6%) when *p* = 2 and *m* = 100. As for the traditional charts, T_O^2 chart is considered robust under ideal condition, but not in the case of T_1^2 chart as we can detect a few non-robust rates in the table.

Table ***5.3****: False alarms rates (percent) under independent case for α = 1%.*

			p=2					*p*=5					*p*=10				
m	ε	μ	T_O^2	T_1^2	T_{wMad}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMad}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
25	0	0	1.3	1.4	1.4	1.1	1.3	0.9	1.1	0.9	0.7	0.8	1.1	1.8	0.6	0.7	0.5
	0.1	3	0.7	0.5	0.7	0.7	0.6	0.8	0.9	0.7	0.8	0.6	0.8	1.6	0.9	0.9	0.9
		5	0.5	0.5	0.7	0.8	1	0.8	0.9	0.6	0.6	0.6	0.8	1.7	0.8	1.3	0.8
	0.2	3	0.4	0.6	0.5	0.5	0.6	0.4	0.9	1.1	0.9	0.6	0.7	1.4	0.7	0.9	1.3
		5	0.4	0.6	0.7	1	1	0.4	0.9	1.2	1.1	1.1	0.7	**1.7**	**1.7**	**2.6**	**2.5**
50	0	0	1.1	1.1	1.1	1.1	1.1	0.8	0.4	0.6	0.9	1.3	0.8	1.2	0.8	0.8	0.9
	0.1	3	0.5	0.4	0.5	0.6	0.7	0.4	0.7	0.9	0.7	0.7	0.6	0.8	0.9	0.9	1.2
		5	0.5	0.5	0.5	0.6	0.8	0.4	1.1	0.8	0.9	0.8	0.6	0.8	1.1	1.6	1.3
	0.2	3	0.5	0.8	0.6	0.6	0.7	0.4	0.4	0.9	0.6	1	0.6	1	1.3	1	1.4
		5	0.5	0.4	0.6	0.8	1	0.4	0.5	1.5	1.9	1.7	0.6	1	**2.4**	**3.3**	**2.6**
100	0	0	1.4	1.6	1.6	1.4	0.8	0.8	0.4	1	0.8	1.1	0.8	1.4	0.7	0.9	1.1
	0.1	3	0.3	0.6	0.4	0.4	0.6	0.5	1	0.6	0.7	0.6	0.5	0.8	1	0.9	1.2
		5	0.3	0.5	0.5	0.4	0.8	0.6	0.8	0.6	0.7	0.6	0.5	1	1	1.4	1.3
	0.2	3	0.3	0.6	0.4	0.3	0.6	0.4	0.4	0.5	0.5	0.7	0.4	1.1	1.1	0.9	1.3
		5	0.3	0.6	0.5	0.6	0.6	0.4	0.3	1	1.1	1.2	0.4	1	1.6	**2.3**	1.3
%			60	80	80	80	100	47	27	100	93	93	87	73	80	80	87

Under non-ideal condition, when p = 2, we observe that T_{wTn}^2 is the only chart that is able to control false alarm rates regardless of the conditions. The other two robust charts lost control of the false alarm rates at m = 100, especially when $\boldsymbol{\mu}$ = 3. At this group size (m = 100), the traditional chartT_O^2 is also non-robust regardless of ε and $\boldsymbol{\mu}$. We can also detect two non-robust rates for T_1^2 at m = 50.

As we move to p = 5, we can see a series of unshaded cells under T_O^2 column, which indicates that this chart produce a series of non-robust conditions. A few unshaded cells could also be observed under T_1^2columns but the number is less than those produce by T_O^2. However, under T_{wMadn}^2 column, all the cells are shaded, implying that this chart is robust regardless of the conditions. The other two robust charts, T^2_{wSn} and T^2_{wTn} seem to be able to control their false alarm rates except for the condition when m = 50, ε = 0.2 and $\boldsymbol{\mu}$ = 5.

Next, at p = 10, the traditional charts T^2_O and T^2_1show non-robust performance at different group sizes (m). T^2_1chart is non-robust for small m (25) whiles the opposite for T^2_0 chart where the chart lost control of false alarm rate when m is large (100). In the case of robust charts, they become non-robust when the condition turns extreme when ε = 2 and $\boldsymbol{\mu}$ = 5. The false alarm rates inflated to above the interval limit. Under p = 10, the control chart that produce the highest percentage of robust conditions is T^2_{wTn}.

From the analysis, we can summarize that no clear pattern of changes in the false alarm rates could be detected when m, ε or $\boldsymbol{\mu}$ increase. The performance of the robust charts is found to be better than the traditional charts for every value of p such that T^2_{wTn} chart is the best when p = 2 and 10, while T^2_{wMadn} is the best when p = 5.

Table 5.4:Percentages of detecting outliers for independent case at α = 1%

				p=2					p=`5					p=10			
m	ε	μ	T_O^2	T_1^2	T^2_{wMad}	T^2_{wSn}	T^2_{wTn}	T_O^2	T_1^2	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}	T_O^2	T_1^2	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}
25	0	0	1.3	1.4	1.4	1.1	1.3	0.9	1.1	0.9	0.7	0.8	1.1	1.8	0.6	0.7	0.5
	0.1	3	19.8	29	43.1	35	37.1	9.9	7.7	43.3	30.5	39.8	5.3	8.6	17.2	17.2	18.3
		5	29.8	29.8	96	96.1	98.3	11.6	8.2	99	98.7	99.4	5.8	8.9	90.4	96.3	96.4
	0.2	3	2.9	1.6	11.3	4.8	7.1	1.7	2.4	6.7	3.3	7.1	1.5	2.2	4.2	3.1	5.3
		5	2.3	1.8	63.9	45.6	68.1	1.9	2	53.6	63.1	80.4	1.4	2.8	32.4	72.5	85.8
50	0	0	1.1	1.1	1.1	1.1	1.1	0.8	0.4	0.6	0.9	1.3	0.8	1.2	0.8	0.8	0.9
	0.1	3	17.1	7.5	45.1	30.3	37.5	9.7	10.9	53.1	24	33.5	6.6	5.3	36.1	15.3	21.2
		5	24.2	5.4	98.3	95.5	98.1	10.4	11.6	100	99.2	99.7	7.6	4.6	99.9	97.2	99.3
	0.2	3	3.4	2.8	11.1	4.6	6.9	1.8	2.7	9.4	3.4	4.8	2.2	8	8.7	3.4	5.2
		5	2.5	1.7	76.3	30.7	63.8	2	1.9	45	43	74.8	2.3	2.7	55.5	54.5	83.2
100	0	0	1.4	1.6	1.6	1.4	0.8	0.8	0.4	1	0.8	1.1	0.8	1.4	0.7	0.9	1.1
	0.1	3	20.2	10.9	52.5	29.9	38.5	11.6	12.5	60.2	23.4	32.8	7.7	6.8	50.5	15.2	23.3
		5	30.3	7.6	99.3	95.8	98.4	13	12.7	98.4	98.5	99.8	8.2	7.5	100	98.0	99.9
	0.2	3	3.7	2.5	12.3	4.1	5.8	2.1	2.4	6.9	2.7	4.2	2	2.1	6.3	3.3	4.2
		5	2.8	2.3	83.8	20.4	56.7	2	2.6	80.6	21.8	57.7	2.2	2	65.4	25.4	63.8

The performance of the investigated control charts in terms of detecting outliers displayed numerically and visually illustrated in Table 5.4 and Figure 5.4 to Figure 5.6 respectively. First, let us focus on the performance when p = 2 which is exhibited in Figure 5.4a to Figure 5.4d.

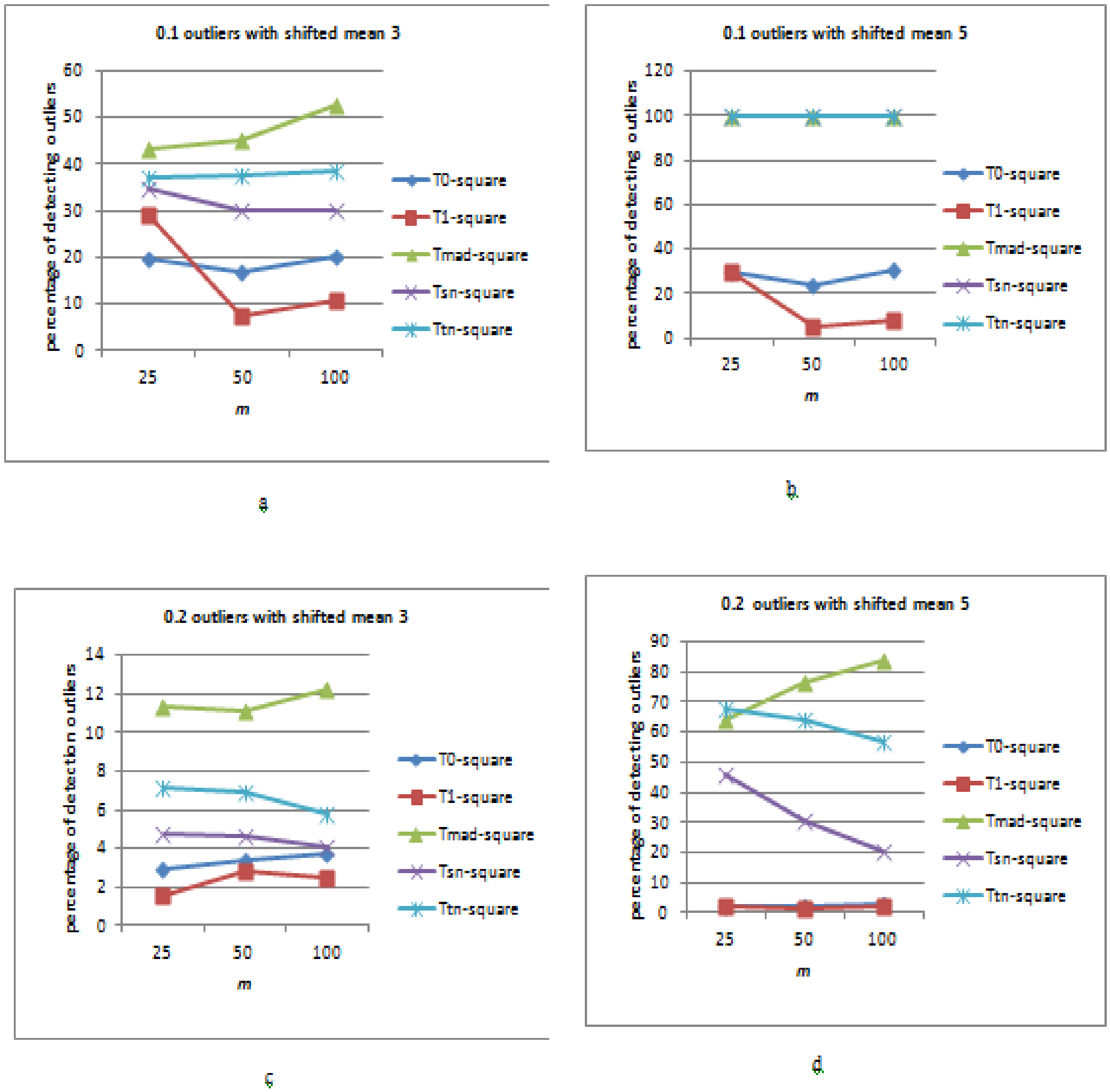


Figure 5.4:Percentages of detection of outliers when p = 2

The lines representing different control charts clearly show that the robust charts always outperform the traditional charts. In Figure 5.4a, with mild condition $\varepsilon = 0.1$ and $\boldsymbol{\mu} = 3$, the highest percentage of detection belongs to T^2_{wMadn} with the value of 52.5% followed by T^2_{wTn} and T^2_{wSn} while the highest percentage for the traditional chart (T^2_1) is only at 29%. A large disparity between the robust and traditional charts can be observed when

$\boldsymbol{\mu}$increases to 5 as shown in Figure 5.4b. At this condition, the performance of the robust charts is consistent with each other across group sizes, achieving almost perfect (100%) detection. When ε increases to 0.2 while $\boldsymbol{\mu}$ = 3 (Figure 5.4c), the performance of all charts drops abruptly. The highest for the robust charts is only 12.3% which still belongs to T^2_{wMadn}. When $\boldsymbol{\mu}$ increases to 5 while ϵ maintains at 0.2 (Figure 5.4d), all the robust charts show great improvement but not for the traditional charts. The changes in group sizes do have some effect on the robust charts under extreme condition (ε = 0.2 and $\boldsymbol{\mu}$ = 5) but not so much under other conditions and vice versa for the traditional chartT_1^2. The changes in ε also show some effect.

Overall, T^2_{wMadn} control chart is the best performer at p = 2. Despite the underperformance of T^2_{wSn} among the robust charts, the chart still can surpass the performance of the traditional chart. The traditional chart, T_1^2 perform the worst among all charts with some values almost as low as the ideal condition. This implies that T_1^2 fails in detecting outliers.

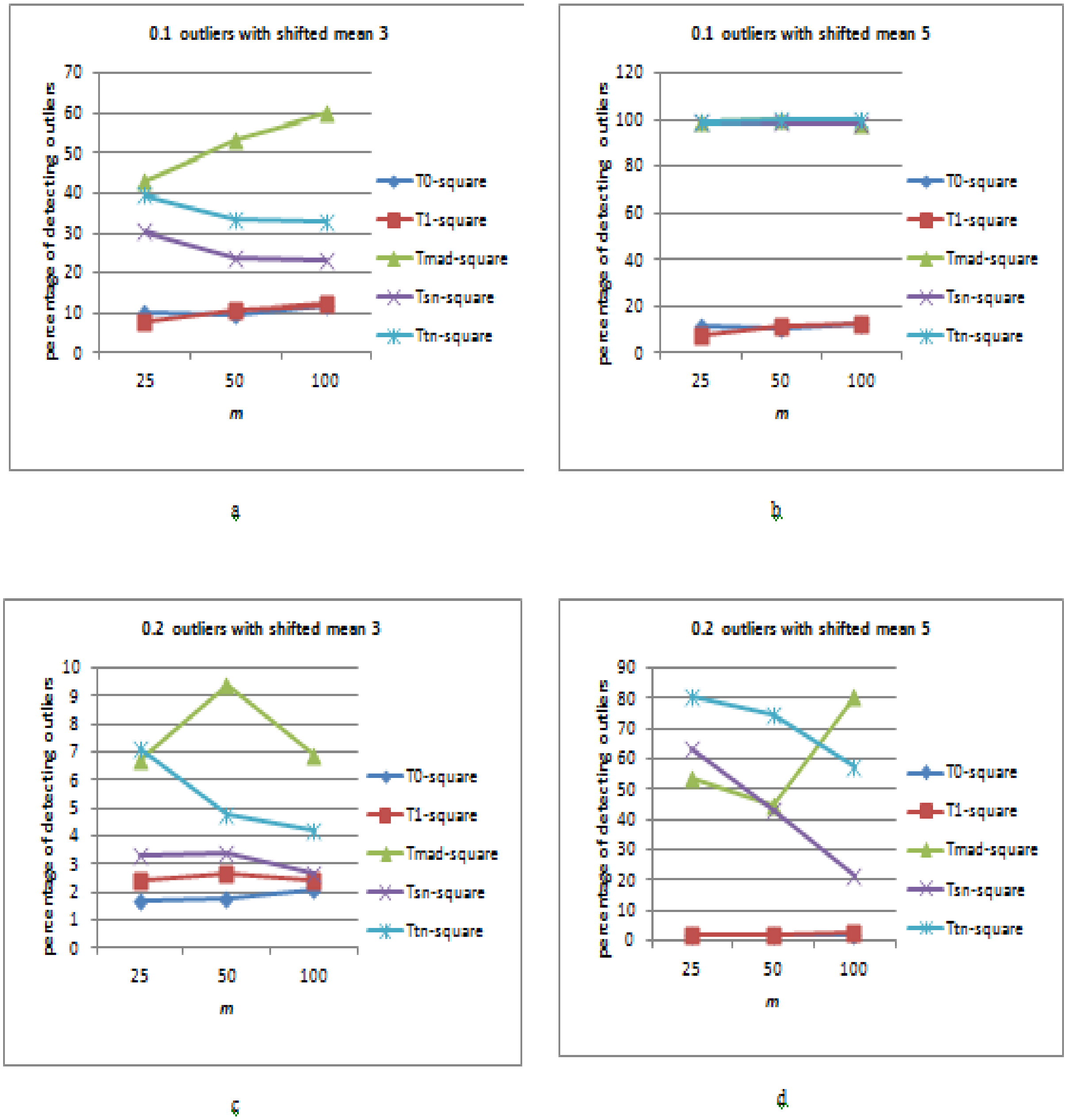


Figure 5.5: Percentages of detection of outliers when p = 5

Figure 5.5 visually shows the performance of the investigated charts when the number of quality characteristics (p) is 5. If we begin with Figure 5.5a which represents the condition when ε = 0.1 and $\boldsymbol{\mu}$ = 3, we notice that T^2_{wMadn} chart outperforms all the other charts. Both the traditional charts are on par with each other, but their performance is still below the three robust charts. The highest percentage recorded by T^2_{wMadn} is 60.2% while the lowest produced by T^2_1 with value as low as 7.7% (refer to Table 5.4). When the value of the mean shift is increased to 5 as shown in Figure 5.5b, there is a sudden jump to almost perfect detection (100%) for the three robust charts, but the traditional charts stay at the same level of performance. As we move to Figure 5.5c which represents the moderate condition when ε = 0.2 with $\boldsymbol{\mu}$ = 3, the performance of all the charts dwindle. The almost perfect scores produced by the three robust charts drop to less than 10%. The performance improves again for the robust charts but not for the traditional charts when the mean shifts to 5 as depicted in Figure 5.5d. At this point, both the traditional charts fail to detect the existence of outliers as explained by the low percentages, which are almost equal to the ideal condition.

Under mild condition (ε = 0.1 with $\boldsymbol{\mu}$ = 3), T^2_{wMadn} is positively affected by the changes in group sizes (*m*), but no obvious effect could be detected for other control charts. When the condition becomes extreme (ε = 0.2 with $\boldsymbol{\mu}$ = 5) we observe a contrast pattern between the robust charts as m increases to 100. There is a positive effect on T^2_{wMadn}chartwhile a negative effect onT^2_{wSn} and T^2_{wTn} charrts.The reason of the difference behavior among the three robust charts is due to the shape of the data, where MAD_n estimator works well when the distribution of the data is symmetric. While the other two robust estimators S_n and T_n work well when the distribution of the data is asymmetric.

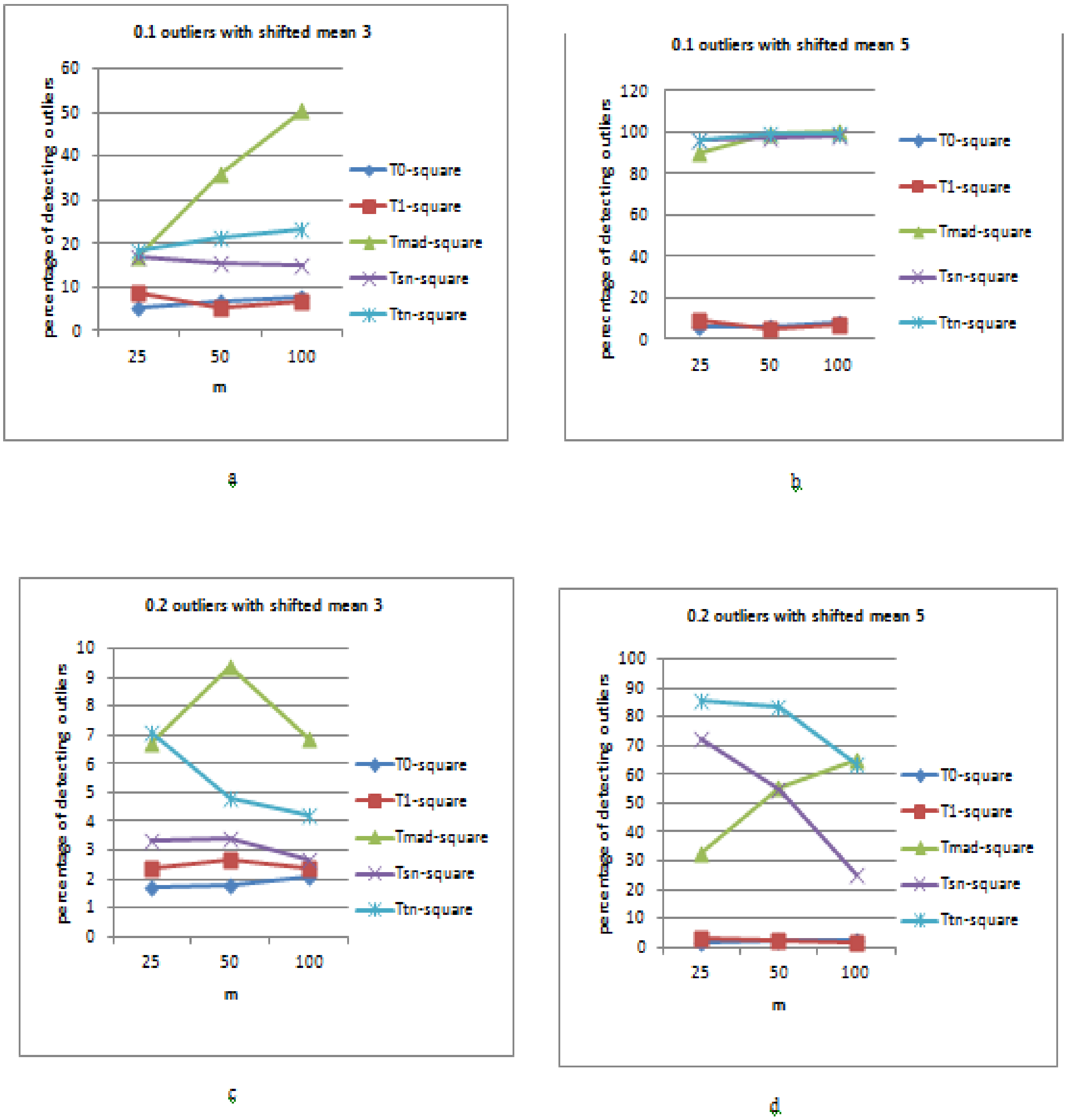


Figure 5.6:Percentages of detection of outliers for p = 10

Figure 5.6 exhibits the performance of the investigated control charts when p = 10. As in p = 5, the robust charts outperform the traditional charts for all conditions investigated. The first figure which represents the performance of the charts under mild condition ε = 0.1 and $\boldsymbol{\mu}$ = 3 shows a sudden leap in the performance of T^2_{wMadn} from

17.2% to 50.5% as m increases from 25 to 100, while other charts do not exhibit obvious effect under this condition.

When $\boldsymbol{\mu}$ shifts to 5, there is a huge increase in the performance for all the robust charts to above 90% and continues to increase to almost 100% as m increases. However, when ε changes to 0.2 while $\boldsymbol{\mu} = 3$ (Figure 5.6c), the performance of all charts drops with the best performance does not exceed the 10% level. The performance improves again when $\boldsymbol{\mu} = 5$ (5.6d), but the improvement is not as good as when $\varepsilon = 0.1$ with the same mean shift.

We can observe some obvious effect on the performance of the robust charts when m changes under mild ($\varepsilon = 0.1$ and$\boldsymbol{\mu} = 3$) and extreme ($\varepsilon = 0.2$ and $\boldsymbol{\mu} = 5$) conditions. However, for the traditional charts, no obvious effect could be detected regardless of conditions.

For $p = 10$, we cannot identify a specific robust chart that can represent the best chart for the whole conditions since different conditions produce different best charts.

3.3 Dependent variables

The following subsections analyze on the performance of the investigated charts in the case of dependent variables. The performance is measured in terms of false alarm rates and percentage of detecting outliers as in the previous case. The text is arranged according to the number of characteristics variables, p.

3.3.1 False alarm rates and Percentage of detecting outliers at α = 5%

Table 5.5 records the false alarm rates for the investigated charts. Under ideal condition, all the charts are in control of their false alarm rates. Nevertheless, in general, the rates for the robust charts are better (closer to but not exceeding the nominal value of 5%) than the traditional charts. When the condition is non-ideal, we observe a few cells with values below the lower limit of Bradley's interval. Only two charts i.e. T_1^2 and T_{wTn}^2 produce robust values regardless of the conditions. At p = 2, the worst chart is T_O^2 which produce the most number (i.e. 4) of non-robust rates followed by T_{wSn}^2 (i.e. 2) and T_{wMadn}^2 (1). The number of robust cells (shaded) increases for all the charts when p changes to 5 except for T_{wMadn}^2. When the number p further increases, we notice that all the charts are in control of their false alarm rates despite of the decreasing in values for most of the charts. Investigation on the change in the proportion of outliers (ε) also shows a dwindling in values of the false alarm rates when ε is increased. Overall, for this case, T_1^2 and T_{wTn}^2 can be considered as the best control charts since they are robust across all conditions investigated.

Table 5.5: False alarms rates (percent) under dependent case for α = 5%.

			P=2					P=5					P=10				
m	ε	μ	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
25	0	0	5.9	5.7	5.3	4.8	4.6	5.1	5.4	4.1	4.1	3.8	4.9	6.3	4	4.3	4
	0.1	5	2.9	5.6	4.7	3.8	4.2	3.9	4.3	3	3.3	3.8	4.4	5.7	4.1	3.6	4.2
	0.2	5	2.8	3	2.4	2.8	2.7	4	3.8	2.6	3.3	3.2	4.2	4.3	4.3	3.8	4.1
50	0	0	5.6	6	5.2	5.3	5.5	5.3	5.6	3.7	4.9	4.3	5.7	5.7	3.4	4.8	3.8
	0.1	5	2.3	5.8	3.7	3.3	3.7	3.8	4.7	3.2	3.3	3.4	4.8	4.6	3.7	4.2	3.7
	0.2	5	2.3	3.8	2.8	2.5	2.6	3.6	3.8	2.3	3	2.7	5.8	3.8	2.7	3.7	3.8
100	0	0	5.5	6.3	4.3	4.5	4.4	5.4	5	2.8	3.5	3.4	5.5	5.8	3.3	4.4	3.9
	0.1	5	2.4	5.7	2.9	2.2	2.8	3.8	4.8	2.2	2.8	2.7	4.6	4.9	2.9	4	3.5
	0.2	5	2.2	3.1	2.3	1.6	2.6	3.7	4	2.2	2.8	2.6	4.4	4.7	2.7	3.5	3.5
%			**56**	**100**	**89**	**78**	**100**	**100**	**100**	**78**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

The following table and figures present the performance of the investigated charts in terms of their ability in detecting outliers.

Table 5.6: Percentages of detecting outliers for dependent case at α =5%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_0^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
25	0	0	5.9	5.7	5.3	4.8	4.6	5.1	5.4	4.1	4.1	3.8	4.9	6.3	4	4.3	4
	0.1	5	45.9	62.2	73.3	66.1	70.8	17.6	10.4	24.9	22.1	22.6	7.7	8	9.4	8.1	8.6
	0.2	5	14.7	12.5	42.4	27.5	36.2	8.6	6.3	13.3	9.3	12.2	6.3	6.3	7.4	6.2	7.3
50	0	0	5.6	6	5.2	5.3	5.5	5.3	5.6	3.7	4.9	4.3	5.7	5.7	3.4	4.8	3.8
	0.1	5	42.9	62.3	74.3	67.7	71.8	15.3	12.6	24.5	20.3	22.9	10.6	7.5	10.1	10.2	9.5
	0.2	5	17.3	12.7	47.3	28.9	38.7	9.1	6.4	12	10.1	11.9	8	5.6	6.5	6.8	6.5
100	0	0	5.5	6.3	4.3	4.5	4.4	5.4	5	2.8	3.5	3.4	5.5	5.8	3.3	4.4	3.9
	0.1	5	45.7	62.5	74	63.1	70.6	18.2	12.7	23.1	19.6	19.9	9.3	8.4	9.1	9.6	9.7
	0.2	5	17.4	12.9	39.3	23	34	9.4	6.6	10.8	8.9	10.2	7.8	7.1	6.7	6.8	6.3

Based on the percentages in Table 5.6, at p = 2, the performance of all charts is moderate when$\varepsilon = 0.1$ and weakens when the proportion of outliers is increased to $\varepsilon = 0.2$. However, the performance of the three robust charts is still stronger than the performance of the two traditional charts as visually shown in Figure 5.7.

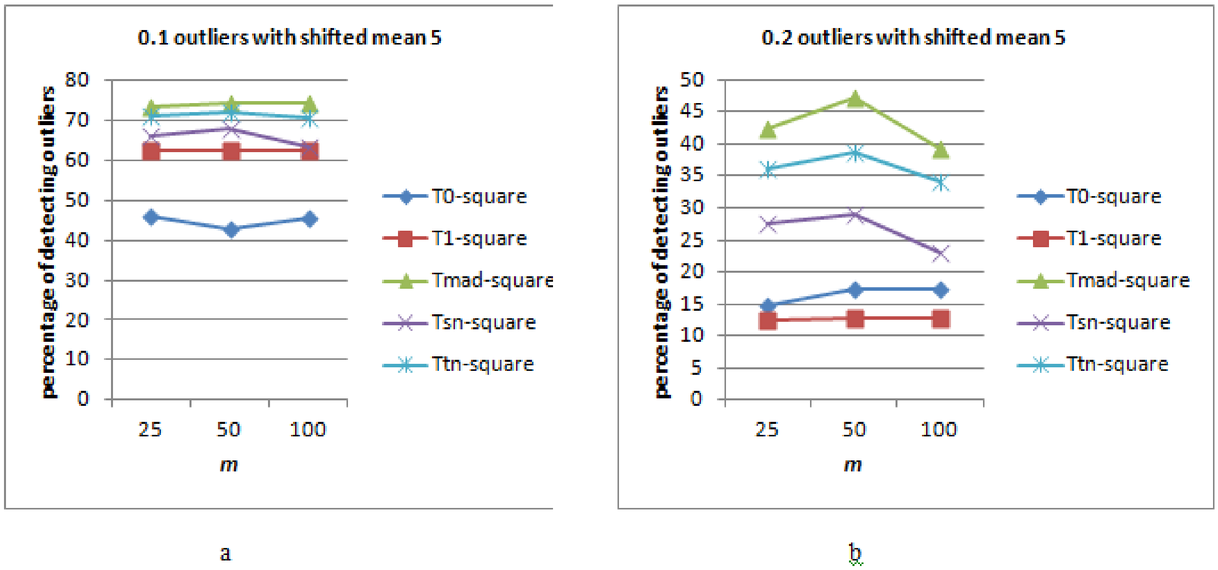


Figure 5.7:*Percentages of detection of outliers when p = 2*

Figure 5.7a shows the moderate detection for all charts with highest value less than 75% when ε = 0.1. The performance of the robust charts is quite close to each other with percentages ranging from 63.1% to 74%. Even the performance of the traditional chart,T_1^2, is close to the robust charts leaving T_O^2 chart at the lowest with values below 46%. The performance worsens as ε increases as shown in Figure 5.7b. The highest percentage scores by the best chart, i.e. T_{wMadn}^2 falls below 50% (from 74%). Unlike when ε is smaller, there is inconsistency in the performance of the robust charts across the group sizes (m) but the traditional charts stay consistent especiallyT_1^2. However, under this condition, T_1^2 performs the worst with value as low as 12.5%.

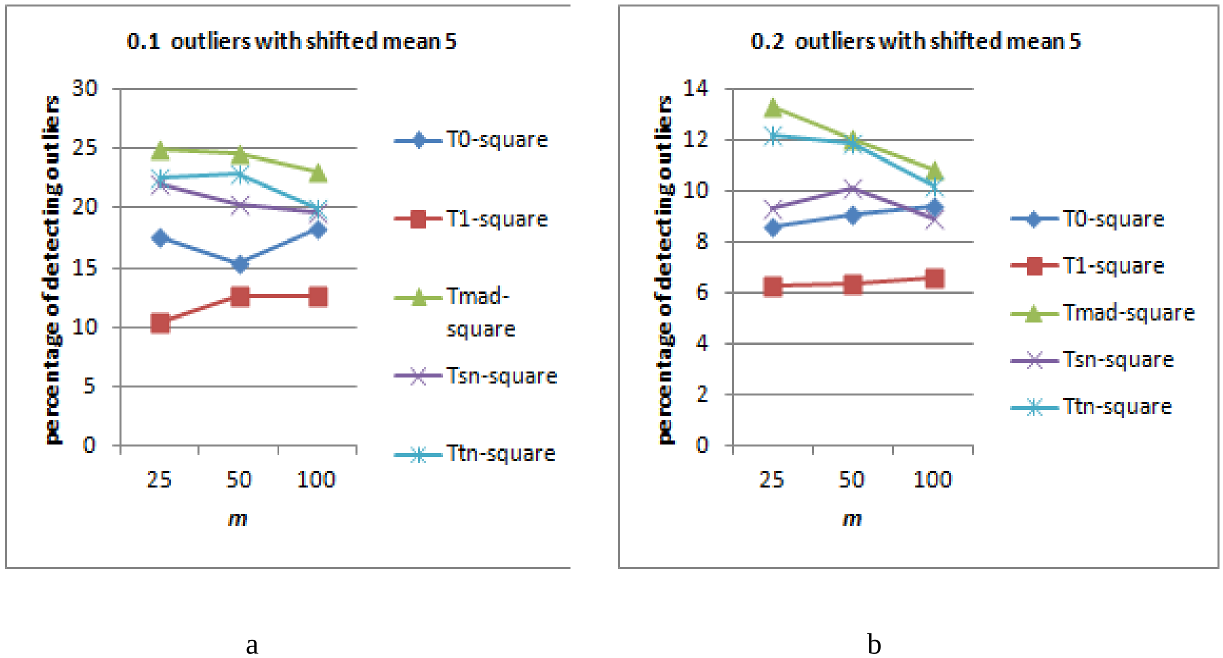
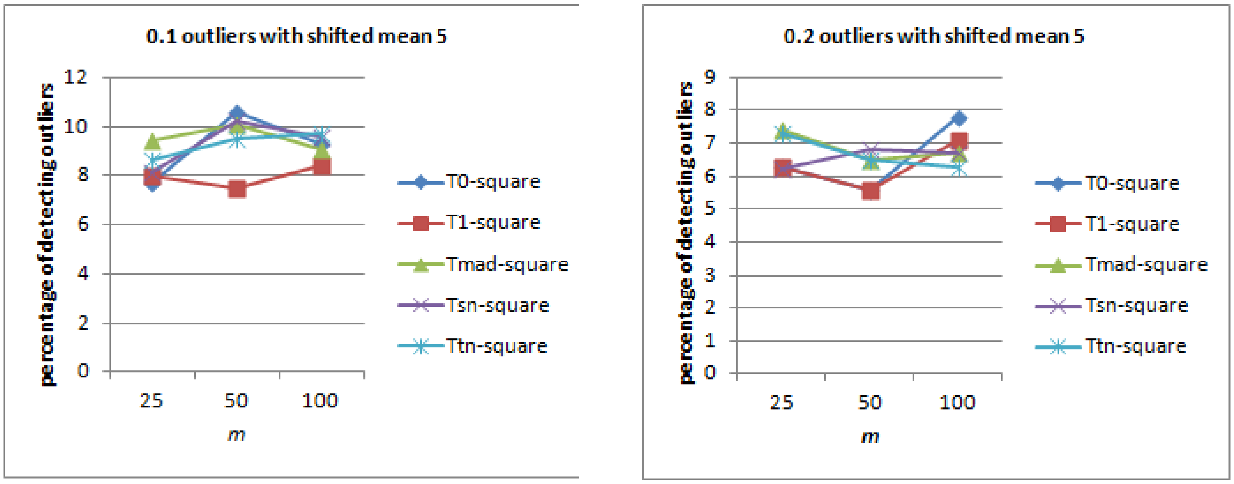


Figure 5.8:Percentages of detection of outliers when p = 5

The performance of the charts further deteriorates as the number of p increases as depicted in Figure 5.8. At ε = 0.1, the highest percentage is only at 24.9% produced by T^2_{wMadn} (refer to Table 5.6). We can also observe negative effect on the robust charts as m increases. By increasing the proportion of outliers (ε = 0.2), the performance of the charts from bad turns to worse as shown in Figure 5.8b. All charts show very weak performance with highest percentage of less than 14%. T^2_{wMadn} chart maintains to be the best among all, despite sharp declines in values as m increases, while T^2_1 stays at the lowest.



a b

Figure 5.9:Percentages of detection of outliers when p = 10

As could be observed in Figure 5.9a and Figure 5.9b which illustrate the performance when p = 10, the lines representing the charts are stacking on each other which signify that their performance are on par with one another. Even their performance for both conditions is quite consistent across m. However, the charts ability in detecting outliers drops considerably when p increases and further deteriorates when the proportion of outliers gets larger. The effect is more prominent in the robust charts. The highest percentage does not exceed 11% while the lowest percentage is less than 6%.

3.3.2 False alarm rates and Percentage detecting outliers at α = 1%

This section narrates the performance of the investigated charts for dependence case under α = 1%**.** The performance measured in terms of false alarm rates and percentage of detecting outliers is presented in Table 5.7 and Table 5.8 respectively. In addition, Figures 5.10 - 5.12 graphically report the results in 5.8 with respect to the number of quality characteristics, *p*.

Table 5.7: False alarms rates (percent) under dependent case for α = 1%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2	T_O^2	T_1^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
25	0	0	1.3	1.4	0.9	1.1	1.3	0.9	1.1	1.2	0.8	0.7	1.1	1.8	0.9	0.5	0.8
	0.1	5	0.7	0.5	0.6	0.8	0.7	1	0.9	1.1	1	1.1	1.1	1.7	1	0.9	0.9
	0.2	5	0.5	0.5	0.6	0.6	0.6	0.8	0.9	0.9	0.7	0.8	0.9	1.6	1.3	1	1.1
50	0	0	1.1	1.1	0.7	0.9	1.2	0.8	0.9	0.5	0.8	.5	0.8	1.2	0.3	0.7	0.5
	0.1	5	0.6	0.3	0.7	0.6	0.8	0.6	0.8	0.3	0.5	0.5	0.9	1.2	0.3	0.7	0.5
	0.2	5	0.6	0.4	1	0.7	0.7	0.6	0.6	0.3	0.5	0.8	0.8	1.1	0.3	0.5	0.5
100	0	0	1.4	1.6	0.8	1.3	1.1	0.8	0.4	0.5	0.6	0.7	0.9	1.4	0.2	0.7	0.5
	0.1	5	0.3	0.6	0.7	0.3	0.5	0.6	0.2	0.6	0.5	0.5	0.7	0.7	0.5	0.7	0.6
	0.2	5	0.3	0.7	0.7	0.3	0.5	0.6	0.5	0.5	0.5	0.5	0.5	0.8	0.3	0.7	0.7
%			78	67	100	78	100	100	78	78	100	100	100	78	44	100	100

To reiterate, based on Bradley's robust criterion, a chart is considered robust at α = 1% if the empirical false alarm rates fall between 0.5% and 1.5%. We start with the ideal condition, where we observe that all the robust charts and $T_0^2$are in control of their false alarm rates, while T_1^2 chart fails to control the rate when *m* is large (100). When *p* increases to 5, the false alarm rates for the robust charts and $T_0^2$are still under control, but the rate for theT_1^2 chart at *m* = 100 drops from above (1.6%) the upper limit of the robust interval to below (0.4%) the lower limit. As we move to *p* = 10, the pattern changes. The T_1^2chart is found to be robust when *m* is large (100) and becomes non-robust when *m* is small. T_{wMadn}^2chart is also found to be non-robust except when *m* is small.

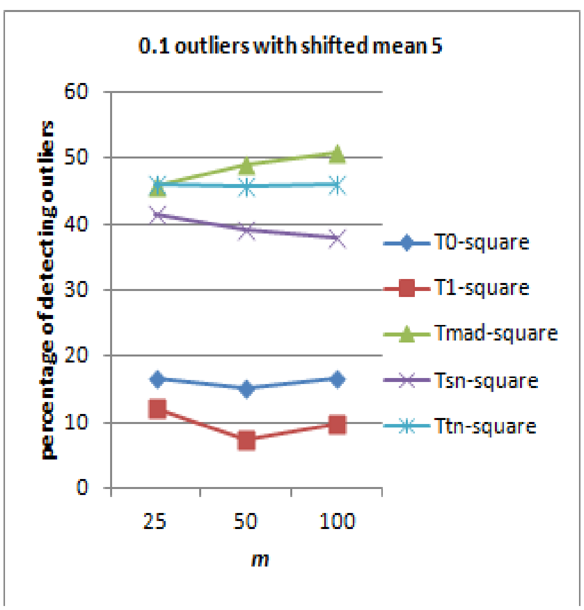
Next, under non-ideal condition, at p = 2, the performance is good for all the charts when m is small and moderate. However, when m is large, the performance of T_0^2 and T_{wMadn}^2 charts drops below the robust interval limits. The change in ε seems to have no effect on the performance of the charts. As the number of quality characteristics (p) increases to 5, all the cells of the three robust charts are shaded except for T_{wMadn}^2 at m = 50. Regarding the traditional charts T_1^2appears to be non-robust when m is large with ε = 0.1 while $T_0^2$is robust regardless of m and ε. We could also observe that the performance of the charts in general dwindles as m increases even though majority is in control of the false alarm rates. The three charts that are robust regardless of m and ε are T_0^2, T_{wSn}^2 and T_{wTn}^2. These three charts maintain to be robust regardless of m and ε when p changes to 10. The performance of the other two charts, $T_1^2$and T_{wMadn}^2 worsens with more number of non-robust cells under their columns. Even though there is no obvious pattern on the performance due to the changes in m, we could observe that the false alarm rates are larger when m is small as compared to moderate and large. About ε, the change does not have any effect on the performance. In Table 5.7, the only column that all its cells are shaded is T_{wTn}^2 , thus we can say that this chart is the best in controlling false alarm rates under dependent case given α = 1%.

Analysis on the performance of the investigated charts continues with regards to the charts' ability in detecting outliers. The results are presented numerically in Table 5.8 and graphically in Figures, 5.10 to 5.12. A glance across the table shows that the percentages produced by the charts are quite low. However, there are a few cells with values around 50%, which belong to T^2_{wMadn}. None of the traditional charts scores above 17%.

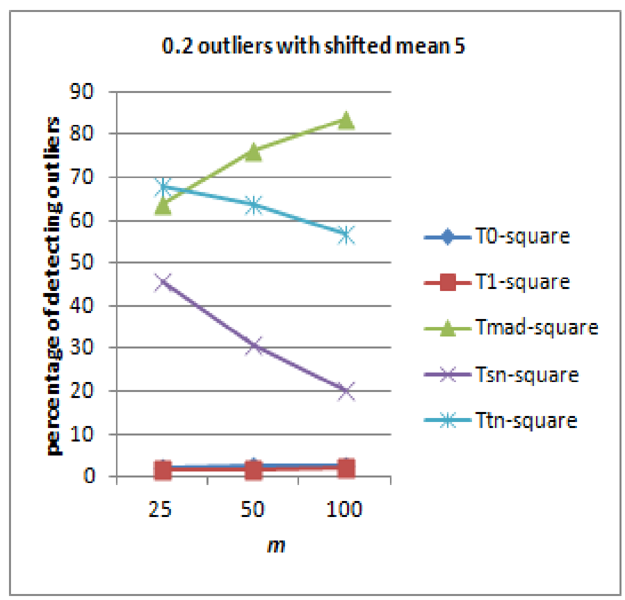
Table 5.8: Percentages of detecting outliers for dependent case at α = 1%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T^2_0	T^2_1	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}	T^2_0	T^2_1	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}	T^2_0	T^2_1	T^2_{wMadn}	T^2_{wSn}	T^2_{wTn}
25	0	0	1.3	1.4	0.9	1.1	1.3	0.9	1.1	1.2	0.8	0.7	1.1	1.8	0.9	0.5	0.8
	0.1	5	16.7	12.3	45.8	41.6	46.1	4.3	4.2	10.7	6.6	7.9	1.8	2	51.5	2.3	3.1
	0.2	5	3	1.6	23.2	11.4	17.5	1.8	1.8	4.6	2.1	3.9	1.2	1.7	49.6	1.5	1.8
50	0	0	1.1	1.1	0.7	0.9	1.2	0.8	0.9	0.5	0.8	0.5	0.8	1.2	0.3	0.7	0.5
	0.1	5	15.3	7.3	49	39.1	45.8	4.2	4.4	9.1	6.4	8.2	1.7	2.1	24.5	2.2	2.1
	0.2	5	3.6	2.4	23.6	9.7	17.9	1.9	2.1	4	2.9	3.3	1	1.7	19.1	1	1.1
100	0	0	1.4	1.6	0.8	1.3	1.1	0.8	0.4	0.5	0.6	0.7	0.9	1.4	0.2	0.7	0.5
	0.1	5	16.5	9.8	50.7	37.9	46.1	4.6	5.6	8.5	5	7.4	2.4	1.7	20.1	2.7	2.8
	0.2	5	3.7	2.8	20.2	6.5	12.4	1.9	2.4	3.5	3.1	2.9	1.8	1.4	13.2	1.9	1.8

Figures 5.10 to 5.12 graphically present the results in Table 5.8 according to the values of *p*.



a

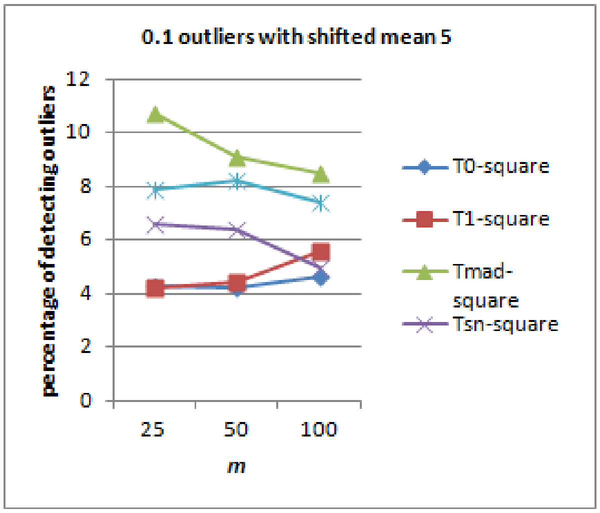


b

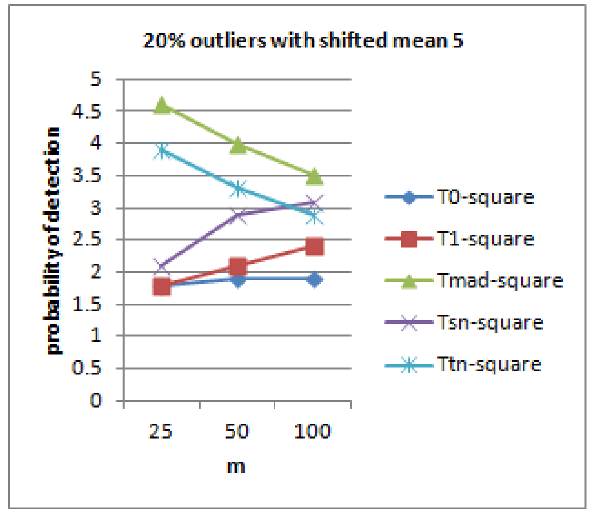
Figure 5.10:Percentages of detection of outliers when p = 2

From Figure 5.10a, which represent the performance of the charts when p = 2, we observe that there is quite a large difference in performance between the robust and traditional charts. The robust charts outperform the traditional with T^2_{wMadn} chart shows the strongest detection (≈ 50%) followed by T^2_{wTn}(≈ 46%) and T^2_{wSn} (≈ 40%). For the traditional charts, the percentage of detection is less than 17% for T^2_0 chart and approximately 11% for T^2_1 chart.T^2_{wMadn}chart maintains to have the strongest performance across the group sizes, and it is clearly shown in the graph that it is affected positively when there is an increase in the group sizes, and this is in contrast with T^2_{wTn} and T^2_{wSn} charts.

When the proportion of outliers increases, the probability of detection of outliers for T^2_{wMadn} and T^2_{wTn} charts inflates to above 60% with T^2_{wTn} chart almost reaching the 70% level. However, as m increases, there is a declining effect in the performance of T^2_{wTn} but the T^2_{wMadn} chart continues to increase. Even though T^2_{wSn} perform quite badly under this condition; the chart still outperforms both the traditional charts, which can only perform up to less than 3% level.



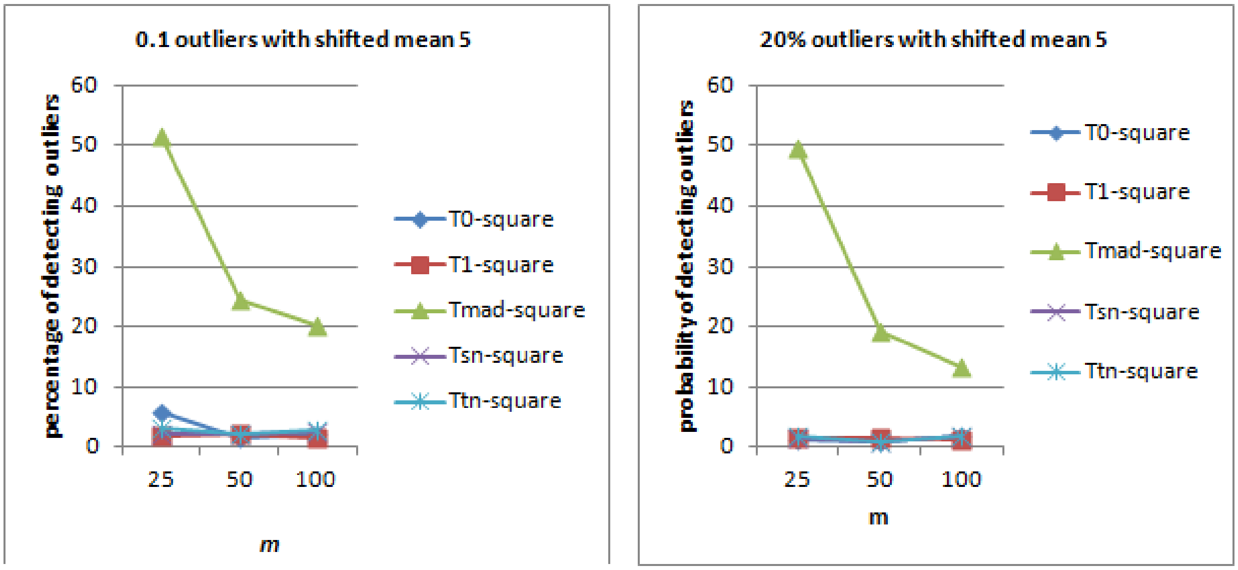
a



b

Figure 5.11:Percentages of detection of outliers when p = 5

Figure 5.11a shows more deterioration in the performance of all charts. The rates for the robust charts suddenly drop to lower than 12%.The T^2_{wMadn} chart still has the best performance but it is negatively affected when m increases. As usual, the traditional charts show the worst performance among all charts. All the charts deteriorate further when ε increases to 0.2.



a b

Figure 5.12:Percentages of detection of outliers when p =10

As p increases to 10, the probability of detection for all the robust charts drops terribly, except for T^2_{wMadn} chart as shown in Figure 5.12. The performance continues to worsen when ε increase. Nevertheless, the performance of T^2_{wMadn} dwindles as m gets larger while other charts do not show obvious effect in the changes of m.

5.4 Analysis on Real Data

As mentioned in chapter four, to further evaluate on the performance of the three robust charts T^2_{wMADn},T^2_{wSn} and T^2_{wTn} and their comparison with the traditional chart, T^2_0, the investigation continued on real data. The same set of real data from 47 airplane spoilers (m = 47) were provided by Asian Composites Manufacturing (ACM) Sdn. Bhd. This real data has several features, namely; trim edge (X_1), trim edge spar (X_2) and drill hole (X_3) quality characteristics. The measurements of 21 products were collected in 2009 while the rest were collected in 2010. Hence, we decided to use the 2009 data as historical data, and considering those from 2010 as future observations. The details of the historical data sets are displayed in Table 5.9 below. The location vector ($\bar{\boldsymbol{x}}$) and scatter matrix ($\boldsymbol{S}$) presented in Table 5.10 below are both estimators for the historical observation in phase I. The upper control limits (*UCL*) shown in the last column in table 5.10 were obtained via simulation method which has been explained in section 3.8.2. The future observations appear in the first three columns of Table 5.11, while the values of the T^2 statistics calculated based on the estimators in phase I are mentioned in the last four columns of the table.

As mentioned in Chapter 4, the analysis on real data only focused on 5% significance level (α = 5%). The comparison of the T^2 statistics in Table 5.11 with the corresponding upper control limits in Table 5.10 observed that the three robust statistics T^2_{wMadn}, $T^2_{wSn}$and T^2_{wTn} charts signaled observations 20, 22 and 25 as out-of-control points, while the traditional T^2_0 chart signaled observations 20 and 25 as out-of-control points. The results for T^2_O was expected as the analysis on the probability of the detection of outliers using simulated data showed that T^2_Ochart was not as effective as the other robust charts in detecting the outliers. The analysis on simulated data also showed that the ability of the proposed robust charts in detecting outliers as well as controlling false alarm rates were good and perform almost on par with each other (Refer to section 5.2 and 5.3)

*Table 5.9: Historical data set (*Phase *I data)*

Product No.	**Trim edge (x_1)**	**Trim edge spar (x_2)**	**Drill hole (x_3)**
1	-0. 0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	-0. 0017	0.0109	0.0177
5	-0. 0005	-0. 0010	0.0106
6	0.0016	-0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	-0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	-0.0067	0.0147
16	0.0110	-0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 5.10: The values of the upper control limits for the three robust and one traditional charts.

Types of Control Chart	Upper Control Limit (UCL) α = 5%
T_0^2	11.035
T_{wMADn}^2	14.22
T_{wSn}^2	11.83
T_{wTn}^2	12.77

Table 5.11: The values of future observations and hotelling T^2 statistics.

Product No.	x_1	x_2	x_3	T_0^2	T_{wMadn}^2	T_{wSn}^2	T_{wTn}^2
1	0.0041	0.0087	0.0129	0.5582	1.0559	1.0815	1.0559
2	0.0047	0.0109	0.0124	0.9003	1.9673	2.0187	1.9673
3	0.0031	0.0057	0.0096	0.4992	0.6538	0.6161	0.6538
4	0.0035	-0.0020	0.0101	0.5463	1.0679	0.9930	1.0679
5	0.0040	-0.0028	0.0125	0.4592	0.9427	0.9726	0.9427
6	0.0031	0.0008	0.0061	0.9013	1.6086	1.1409	1.6086
7	-0.0019	0.0101	0.0112	3.0933	4.0206	4.0406	4.0206
8	0.0009	0.0039	0.0082	0.8061	1.2320	1.0386	1.2320
9	-0.0052	0.0090	0.0203	7.3602	9.7286	9.0709	9.7286
10	-0.0008	0.0110	0.0184	3.6198	5.0513	4.5005	5.0513
11	-0.0021	0.0139	0.0170	5.3839	7.4084	7.0272	7.4084
12	-0.0017	0.0092	0.0061	2.7387	4.0369	3.7012	4.0369
13	-0.0010	0.0133	0.0138	3.8058	5.2840	5.2049	5.2840
14	-0.0030	0.0002	0.0053	2.0548	4.4919	3.8560	4.4919
15	0.0016	0.0134	0.0151	2.5073	4.1043	3.9081	4.1043

16	0.0027	0.0086	0.0070	1.1976	1.9270	1.7154	1.9270
17	0.0004	0.0086	0.0087	1.5798	2.1935	2.0945	2.1935
18	-0.0036	0.0136	0.0129	5.7910	7.6287	7.6541	7.6287
19	-0.0028	0.0003	0.0078	1.8304	3.6994	3.4332	3.6994
20	0.0120	0.0123	0.0768	**38.1397**	**99.765**	**50.508**	**99.765**
21	-0.0015	0.0004	0.0115	1.2651	2.2934	2.2895	2.2934
22	0.0009	0.0232	0.0202	8.4181	**15.424**	**14.117**	**15.424**
23	-0.0035	0.0088	0.0107	3.7588	4.9187	4.9660	4.9187
24	0.0016	0.0061	0.0066	1.0602	1.6460	1.3250	1.6460
25	-0.0228	-0.0466	0.0231	**42.8447**	**119.80**	**122.38**	**119.80**
26	0.0037	-0.0038	0.0147	0.4832	1.2069	1.2130	1.2069

5.5 Comparison among the robust Hotelling's T^2 charts

As in chapter 4, the percentages of robust conditions of each investigated control charts were calculated. Also, the percentages of conditions in which the robust charts have at least 80% ability to detect outliers, and at least 50% probability of detection for the robust charts in case of dependent characteristics were computed. Under the dependency characteristics, the table is further divided into the two performance measurement i.e. false alarm (FA) and probability of detection (POD). The comparison covers both the levels of significance α. Both percentages are presented in Table 5.12.

Table 5.12: Overall performance for independent and dependent cases.

Types of Charts	Independent				Dependent (winsorized)			
	α = 5%		α = 1%		α = 5%		α = 1%	
	FA	POD	FA	POD	FA	POD	FA	POD
T^2_{wMadn}	93%	50%	87%	61%	85%	0	74%	0
T^2_{wSn}	91%	31%	84%	50%	89%	0	85%	0
T^2_{wTn}	91%	50%	93%	67%	100%	0	100%	0

In terms of independent case, the table shows that the robust charts are robust in most conditions of independent or dependent cases. With respect to independent case, it can be identified that T^2_{wMadn} chart produced more robust conditions than the other two robust charts at α = 5%. In contrast, at α = 1%, T^2_{wTn} chart produced more robust conditions. In detecting outliers, all the charts show high percentages of robust conditions, and the percentages become even higher when α = 1%.

As we moved to the dependent case, we observe highest percentages in the number of robust conditions for T^2_{wTn} producing 100% robust conditions for both significance levels. The other charts also able produce high percentages of robust conditions but the percentages slightly drop at α = 1%. However, all the charts, including the best performer, T^2_{wTn}, unable to achieve the 80% level of detection under dependent case.

5.6 Summary

Basically, this chapter delineated the performance of another three robust multivariate Hotelling's T^2charts which used winsorized modified one-step *M*-estimator (*MOM*) as the location measure and winsorized covariance matrix as the scale measure. The evaluations for the independent and dependent cases indicated that the robust charts performed better than the traditional charts especially in detecting outliers. Among the robust charts, it has been found that the T^2_{wMadn} chart has shown better performance than the T^2_{wSn} chart and the T^2_{wTn} chart under independent case at α = 5%. However, T^2_{wTn} chart is the best in controlling false alarm when α = 1%. In the case of dependent, T^2_{wTn} chart maintains as the best performer regardless of the levels of significance. In terms of detection of outliers, the results have shown strong performance for all robust charts in the case of independent. For dependent case, even though the robust charts' performance in detecting outliers cannot achieve the 80% level, their ability to detect outliers is always higher than the traditional charts.

The results also revealed that the robust charts in general improve in controlling false alarm when the value of p increases under both dependent and independent cases, but in terms of detecting outliers, the changes in p especially when p reaches to 10 be positively affected while it is negatively affected under dependent case. While for traditional charts the increase of p affected positively in term of false alarm while its effect negatively in detecting outliers, regardless of the values of level of significance and for independent and dependent cases.

The results show that the robust charts maintain on their behaviors as the proportion of outliers increases while the affect is negative on the performance of robust charts in term of detection of outliers in independent and dependent cases. While for the traditional charts, especially T_0^2 charts showed negatively affected in terms of false alarms and detection of outliers, regardless of the values of α and the dependency of the characteristics. However, the robust charts, positively affect can be seen towards the status of the process whether it is in control or out of control. In contrast, it affects positively in term of detection of outliers. Respecting to the traditional charts the increase of the shifted mean affected negatively in most cases in terms of false alarms and detection of outliers regardless of the dependency of the characteristics.

With respect to the robustness, the results as given and discussed in this chapter have proven that the robust charts have more ability to control on false alarms when the evaluation was performed at α = 5% than α = 1%. Likewise in terms of detection of outliers, the performance of the robust charts when α = 5% are stronger than the performance when α = 1% in independent variable case.

Investigation on real data also demonstrate that the three robust charts are on par with each other, and the charts were able to detect out-of-control observations better than the traditional chart.

CHAPTER SIX

MODIFIED HOTELLING T^2 CHARTS USING HODGES – LEHMANN ESTIMATOR WITH ROBUST SCALE ESTIMATORS

4.1 Introduction

This chapter discusses the results of the modified Hotelling's T^2 charts based on the Hodges-Lehmann as a robust location estimator with three robust scale estimators namely $Mad_{n,}$ S_n and T_n. The modified control charts are constructed by replacing the Hodges–Lehmann estimator to the maximum likelihood mean vector of the traditional Hotelling's T^2 chart. Whilst, we replace the maximum likelihood covariance matrix with each of the robust scale estimators, $Mad_{n,}$ S_n and T_n. In general, the construction involves the following steps:

Step 1: Get original data (with outliers)

Step 2: Compute the Hodges - Lehmann estimator

Step 3: Compute robust scale estimator, any of $Mad_{n,}$ S_n and T_n.

Step 4: Compute Hotelling's T^2 statistics

The robustness of these modified control charts measured and compared to the traditional Hotelling's T^2 chart based on two measurements, which are the false alarm rates and the percentage of detecting of outliers. Similar to the investigations conducted in Chapter 4 and Chapter 5, each of these new proposed control charts is tested on two different settings of dependency among the variables: (i) Case A refers to independent of variables and (ii) Case B refers to dependent variables. In addition, the investigations maintained on evaluating the charts based on the individual observation, group sizes (m), proportion of outliers (ε), the shifted means (μ) and the number of variables (p). Subsequently, these settings will allow us to highlight the strengths and weaknesses of the proposed control charts.

The discussion of this chapter is arranged into two main subsections according to the settings of variables (Case A and Case B) where each subsection will further investigates the control charts based on the false alarm rates (α = 5% and α = 1%) and percentage of detecting outliers. The discussion cover the performance of the proposed Hotelling's T^2 charts based on the Hodges-Lehmann robust location estimator with three robust scale estimators, namely $MAD_{n,}$ S_n and T_n. Comparison to the traditional control charts are made for all settings to identify whether the proposed charts are better than or at least at par to the traditional ones.

4.2 Independent Variables (Case A)

The investigation of the control charts for the independent variables is based on two measurements namely false alarm rates and percentage of detecting outliers.

4.2.1 False alarm rates and percentage detecting of outliers at α = 5%

The results of the analysis of the control charts based on the false alarm rates and percentage detecting of outliers at α = 5% are summarized in Tables 6.1 and 6.2.

Table 6.1: False alarms rates (percent) under independent case for α = 5%.

			P=2					P=5					P=10				
m	ε	μ	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	5.6	6	6	5.2	5.4	5.3	5.6	5.6	5.5	5.6	5.7	5.7	6.5	6.6	6.6
	0.1	3	2	5.7	4.1	3.3	3.4	2.7	3.6	3.3	2.3	2.8	4.1	3.5	5.3	4	5
		5	1.6	5.7	4	3.1	3.4	2.6	3.7	2.0	2.3	2.8	3.8	3.4	5.6	3.7	5.2
	0.2	3	2.1	3.1	2.1	2.1	2.3	2.6	3	2.8	2.3	2	4.2	4.8	5.2	3.7	5.4
		5	1.6	2.5	2.1	1.9	2.4	2.5	3.6	1.3	2.3	3.1	4.0	4.8	5.8	3.5	6
100	0	0	5.5	6.3	5.2	5	5.3	5.4	5	4.1	4.2	3.9	5.5	5.8	5.2	4.7	4.9
	0.1	3	2.1	5.7	3.0	2.1	3.0	2.9	3	3	2	2.8	3.3	4.3	4.6	2.8	3.8
		5	1.6	5.7	3.1	1.9	2.8	2.8	3.	3.1	2	2.8	3.4	4.4	5.1	3	4
	0.2	3	2.1	3.2	2.0	1.5	1.9	3.0	4	2	1.4	2.7	3.5	4.9	5.9	2.2	3.6
		5	1.6	2.6	2.2	1.5	2.0	2.9	3.5	2.1	1.1	3.1	3.4	4.6	4.4	2.1	3.7
150	0	0	5.3	6.4	4.2	4.9	4.7	4.3	4.4	4.2	4	4.6	4.6	5.4	4.6	4.6	4.7
	0.1	3	2.3	5.7	3	2.7	2.8	2.1	4.4	3	2.3	2.8	3.7	3.8	4	3.1	3.8
		5	1.9	5.7	3.1	2.3	2,8	2.1	3.7	3.2	2.3	3	3.6	3.5	4.4	3	3.9
	0.2	3	2.2	3.1	2	1.5	2.1	2	3.6	2.5	1.4	2.3	3.6	3.6	3.6	2.4	3.8
		5	1.9	2.7	2.1	1.4	2.1	2.1	3.9	2.8	1.2	2.7	3.5	4.8	4.2	1.8	4
%			20	100	60	40	60	73	100	73	20	80	100	100	100	73	100

Table 6.1 displays the results for each condition of group sizes (*m = 50, 100* and *150*), proportion of outliers (ε = *0, 0.1* and *0.2*), the shifted means (μ = 0, 3and 5) and the number of variables (*p = 2, 5* and *10*). We shaded the in-control false alarm on Table 6.1 to make an ease view. In general, the Hotelling's T^2 without outliers control chart (T_1^2) scores better false alarm rates, which are much closer to α = 5%, compare to the other control charts in most conditions. Such result is expected as the chart eliminates outliers prior to the construction of control chart, while the robust charts reduce the effect of outliers via the estimation process. Besides, all control charts are having similar performance in controlling false alarm under ideal condition ($\boldsymbol{\mu}$ = 0 and ε = 0) but the performance is varies under non-ideal conditions.

Some patterns can be detected under the non-ideal conditions. In the case of p = 2, all robust charts are recorded as in control regardless of group sizes (m) and mean shifts ($\boldsymbol{\mu}$). Nevertheless, the charts are out of control when the proportion of outliers (ε) increases to 0.2. Such results change when the number of variables increases to 5 and 10 where the robust control charts are recorded as in control in term of false alarm rates regardless of m, ε and μ except for the T^2_{HSn}. Nevertheless, the performance of the robust control charts is much better than the traditional chart under the non-ideal conditions. Both T^2_{HMADn} and T^2_{HTn} are competitive in most conditions.

Next, we will look into the other measurement of performance based on the percentage detecting of outliers. The percentage for each condition and control chart for α = 5% is tabulated in Table 6.2. These values are plotted in line charts as in Figure 6.1 for easy view.

Table 6.2: Percentages detecting outliers for independent case at α = 5%.

			p=2					p=5					p=10				
m	ε	μ	T^2_0	T^2_1	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}	T^2_0	T^2_1	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}	T^2_0	T^2_1	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}
50	0	0	5.6	6	6	5.2	5.4	5.3	5.6	5.6	5.5	5.6	5.7	5.7	6.5	6.6	6.6
	0.1	3	49.1	73.6	77.9	77.4	78.3	36.4	15.2	84.6	81.6	83.7	24.7	23	75.7	70.1	75.5
		5	74.7	98.7	100	100	100	44.2	20.6	100	100	100	26.5	25.8	100	100	100
	0.2	3	17.6	12.6	38.9	32.6	42	11.9	7.1	30.7	23.7	29.6	10.6	9.9	23.9	18.2	23.5
		5	17.1	10.7	96.8	94.1	97.1	10.9	9.4	97.4	92.6	96.9	10.6	10.1	86.3	74.4	87.3
100	0	0	5.5	6.3	5.2	5	5.3	5.4	5	4.1	4.2	3.9	5.5	5.8	5.2	4.7	4.9
	0.1	3	52.2	73.6	81.8	78.6	81.4	41.7	20	89.2	86.9	88.7	29.4	27.6	88.9	83.6	87.7
		5	80.6	98.7	100	100	99.2	50.7	20.9	100	100	100	31.4	29.1	100	100	100
	0.2	3	17.5	12.6	43.2	35.8	44.1	11.8	11.2	27.8	23.4	29.6	10.7	10.1	26	16.9	24.1
		5	17.5	10.8	97.5	95.4	97.8	12.0	9.7	98.8	95.7	98.8	10.8	10.1	95.1	85	96.1
150	0	0	5.3	6.4	4.2	4.9	4.7	4.3	4.4	4.2	4	4.6	4.6	5.4	4.6	4.6	4.7
	0.1	3	53.7	73.6	82.2	80	82	45.5	20.5	90.6	88.6	90.6	32	15.2	91.9	87.9	91.2
		5	80.4	98.7	100	100	100	54.2	16.4	100	100	100	33.7	15	100	100	100
	0.2	3	18.3	12.7	43.8	36.4	44.4	12.2	10.9	31.7	24.9	34.2	11.8	8.1	25.8	17.8	28.9
		5	18.3	11	98.2	96.2	98.2	12.5	13.5	92.4	96.8	99.4	12.1	9.8	96.7	89.7	97.4

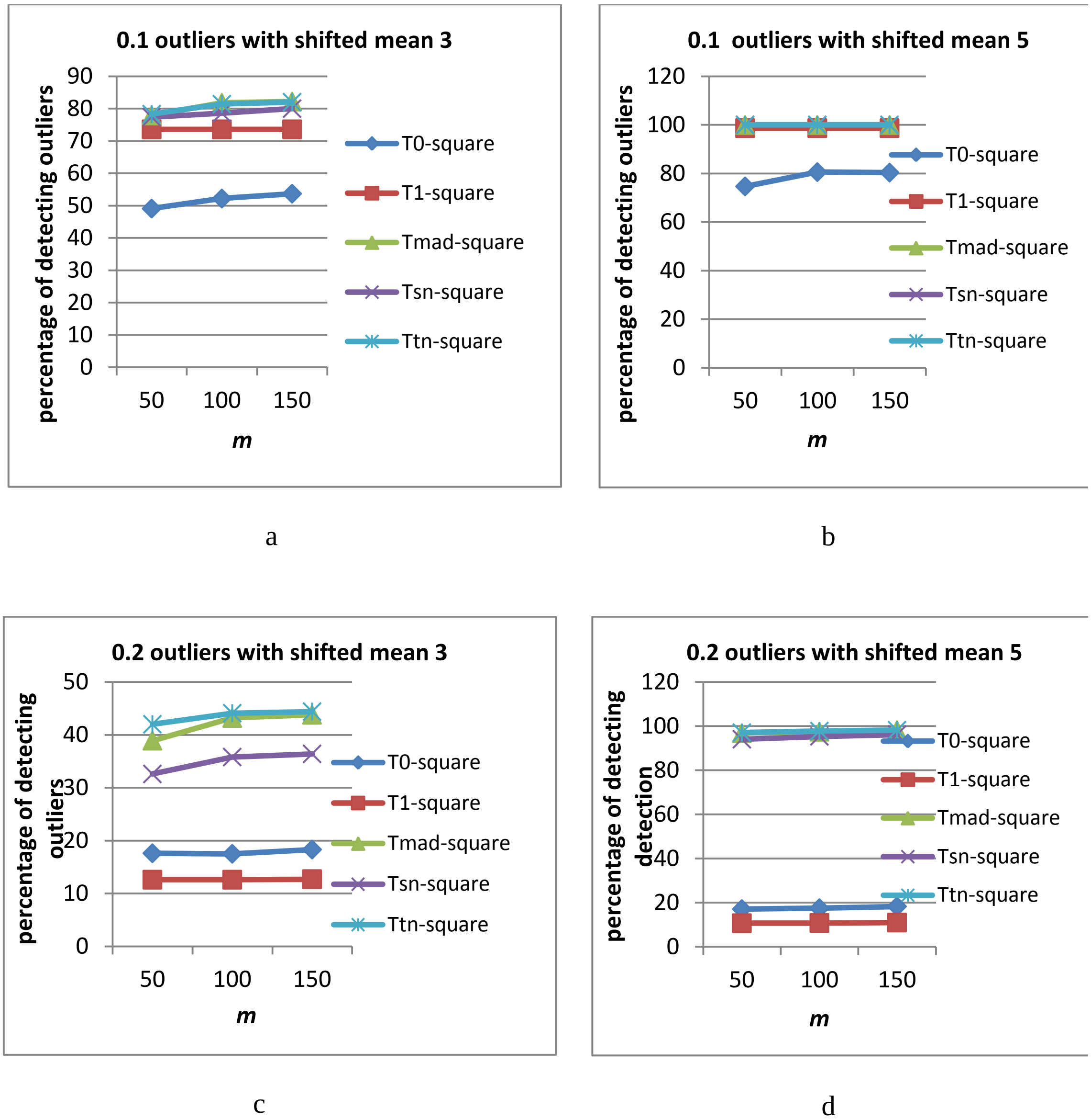
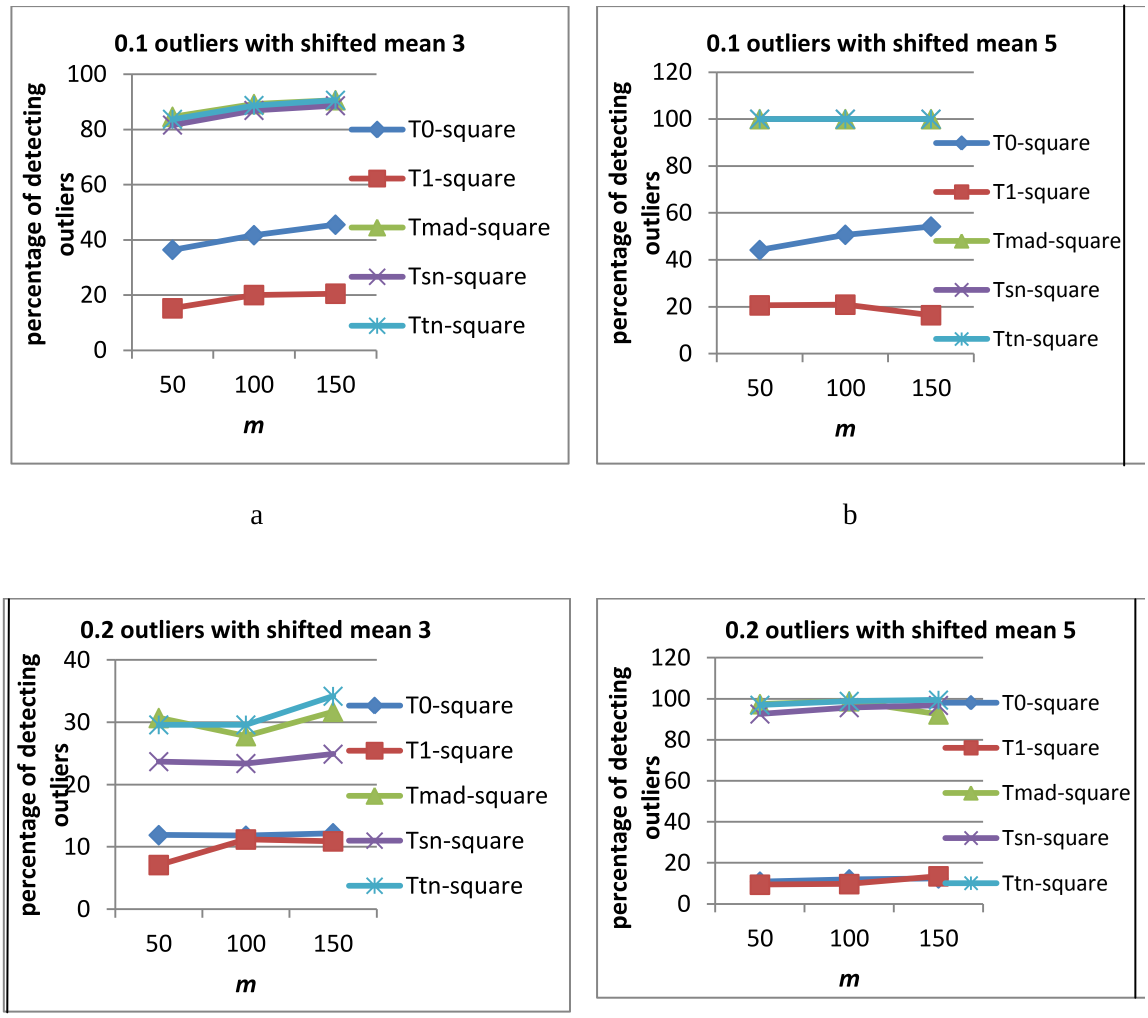


Figure 6.1: Percentages of detection of outliers when p = 2

Each chart in Figure 6.1 shows the performance of the control charts based on the percentage detecting of outliers as the group sizes increases when p = 2. We put altogether the charts for various conditions on mean shifts ($\boldsymbol{\mu}$) and proportions of outliers (ε) to identify some explainable behavior of the charts. Overall, the percentage detecting of outliers for the traditional chart is among the worst as it scores the furthest from 100 percent detection. This includes the T_1^2 chart where it scores the lowest percentage detecting of outliers when the proportion of outliers is 20%. Meanwhile, robust charts always score close to 100 percent detecting of outliers except for the moderate condition when $\boldsymbol{\mu}$ = 3 and ε = 0.20.

In the case when there are about 10 percent outliers in the data (ε = 0.10), the robust charts are able to identify more than 75 percent of them and it increases to more than 80 percent when the group sizes increases regardless of the mean shifts. However, the performance of the robust charts is bad for the case of moderate contamination ε = 0.20 and mean shift to 3 before they improve to good performance when the mean shift to 5.

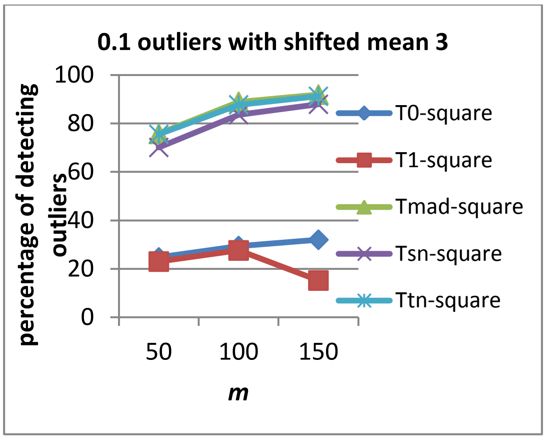
Similar performance of the robust charts is detected when the number of variables increases to 5 and 10 (see Figure 6.2 and Figure 6.3). In general, the traditional control charts (T_0^2 and T_1^2) are the worst in all conditions as the recorded percentages detecting of outliers are always less than 50 percent. Such results indicate that the traditional charts are unable to determine the existing of outliers in the data. Meanwhile, the robust charts are capable to identify more than 80 percent of the outliers in many conditions except when there are 20 percent outliers in the data and the mean of the process shift to 3. In this case, the performance is improved when the mean of the process is shifted to 5.



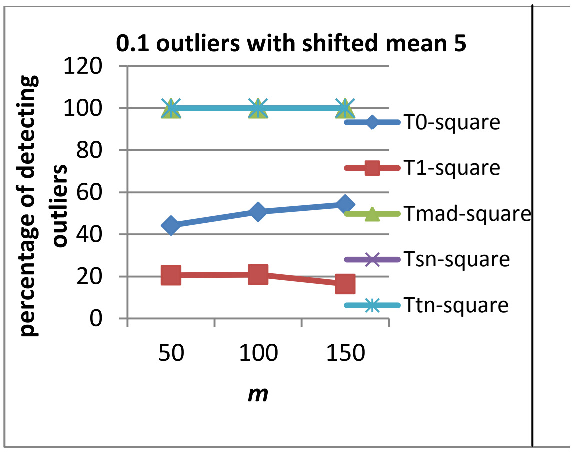
a b

c d

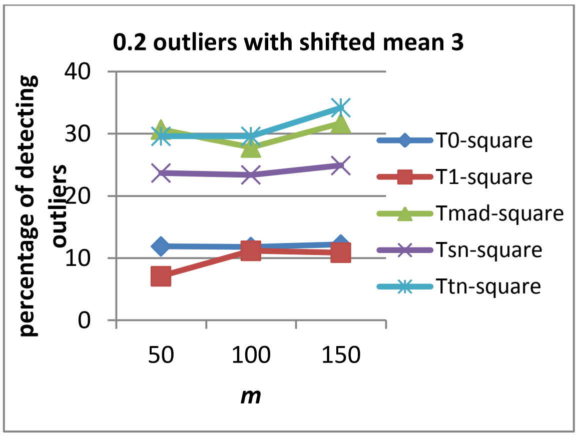
*Figure 6.2:**P**ercentageof detection of outliers when p =5*



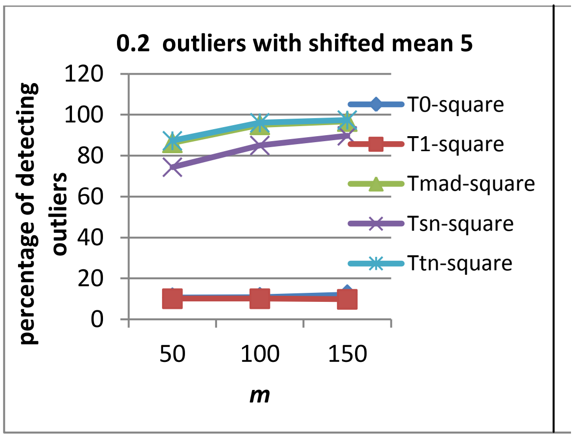
a



b



c



d

Figure 6.3:Percentage of detection of outliers when p = 10.

By carefully investigating the results in Table 6.2 and Figures 6.1 to 6.3, we can identify slight increment on the percentage of detecting the outliers in the data by the robust charts as the number of variables increases and the mean shift further from 0. Such behavior has contributed to bigger differences of performance between the robust charts and the traditional charts. Next, we will investigate the performance of the control charts when α = 1%.

4.2.2 False alarm rates and percentage detecting of outliers at α = 1%

We carry out similar investigation as discussed in subsection 6.1.1 on the control charts but when α = 1%. Table 6.3 and Table 6.4 display the observed false alarm rates and percentage detecting of outliers for all control charts for each combination of conditions on group sizes (m = 50, 100and 150), proportion of outliers (ε = 0, 0.1and 0.2), the shifted means (**μ=** 0, 3and 5) and the number of variables (p = 2, 5and10). In addition, we shaded the in-control false alarm on Table 6.3 for easy view.

Table 6.3: False alarms rates (percent) under independent case for α = 5%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	1.1	1.1	1.4	1.5	1.3	0.8	0.4	1	1.3	1.2	0.8	1.2	0.9	1	0.9
	0.1	3	0.5	0.4	0.7	0.5	0.7	0.4	0.7	0.9	0.7	0.7	0.6	0.8	0.8	0.6	0.6
		5	0.5	0.5	0.7	0.4	0.7	0.4	1.1	0.9	0.7	0.7	0.6	0.8	0.9	0.6	0.6
	0.2	3	0.5	0.8	0.6	0.3	0.5	0.4	0.4	1.1	0.8	1	0.6	1	0.7	0.6	0.8
		5	0.5	0.4	0.5	0.3	0.5	0.4	0.5	1	0.6	1	0.6	1	0.9	0.4	0.8
100	0	0	1.4	1.6	0.9	1	1.2	0.8	0.4	0.7	0.6	0.8	0.8	1.4	0.5	0.7	0.8
	0.1	3	0.3	0.6	0.4	0.3	0.6	0.5	1	0.5	0.2	0.5	0.5	0.8	0.6	0.3	0.6
		5	0.3	0.5	0.5	0.3	0.4	0.6	0.8	0.5	0.2	0.5	0.5	1	0.8	0.3	0.6
	0.2	3	0.3	0.6	0.4	0.3	0.4	0.4	0.4	0.4	0.3	0.4	0.4	1.1	0.5	0.4	0.5
		5	0.3	0.6	0.4	0.3	0.5	0.4	0.3	0.5	0.3	0.4	0.4	1	0.7	0.3	0.5
150	0	0	1	1.4	0.6	0.7	0.8	0.7	1.4	0.8	0.7	0.5	0.8	1.1	1	0.9	1
	0.1	3	0.4	0.5	0.3	0.3	0.3	0.3	0.7	0.6	0.1	0.5	0.6	1.3	0.9	0.5	0.8
		5	0.4	0.6	0.3	0.3	0.3	0.3	1	0.6	0.2	0.5	0.6	0.9	1	0.6	0.8
	0.2	3	0.4	0.5	0.3	0.3	0.3	0.4	0.3	0.3	0.1	0.3	0.6	1.1	0.6	0.3	0.8
		5	0.4	0.8	0.3	0.3	0.3	0.4	0.3	0.3	0.1	0.4	0.6	1	0.7	0.4	0.8
%			47	87	53	27	60	33	53	87	47	73	87	100	100	60	100

The recorded false alarm rates for each control chart under the ideal condition (**μ** = 0 and ε = 0) is closer to 1% except for two results for T_1^2 where the false alarm rate is

0.4 when the m = 50 and m =100. This result gives a signal that the traditional Hotelling's T^2 may be deteriorated in small group sizes and stringent false alarm rate. In contrast, the robust charts still in control although α has been tightened.

Several interesting findings are discovered for the investigated control charts under the non-ideal conditions ($\boldsymbol{\mu} \neq 0$ and $\varepsilon \neq 0$). In the case when p = 2, the recorded false alarm rate for the robust charts is deteriorating further from α = 1% as the group sizes increases from 50 to 150. Table 6.3 shows that 9 out of 12 results of the robust charts for p = 2 and m = 50 are in control, but this result decreases to 3 out of 12 robust charts in control when m = 100 and none of the robust charts are in control when m = 150. In this same condition of p = 2, we also spotted that the number of in control robust charts is decreasing when the proportion of outliers, ε, is increasing from 0.10 to 0.20. However, there is no clear pattern can be identified when the mean process shifts from 3 to 5.

Similar findings are discovered when the number of variables increases to 5 and 10. More charts that are robust are in control when the group sizes m small and few outliers in the data. Meanwhile, the mean shift does not explain much about the performance of the robust charts. Careful investigation on Table 6.3 across the columns shows that more charts that are robust are in control when the number of p increases from 2 to 10. In contrast, the traditional chart (T_0^2) is always out of control in non-ideal conditions although there are adequate number of variables and group sizes.

All these evidences signify that the performance of the robust charts is promising when either the data are contaminated with outliers or the traditional chart (T_0^2) fails to perform. The next discussion will analyze the control charts on their ability to identify outliers in the data.

Table 6.4: Percentages of detecting outliers for independent case at α = 1%.

			P=2					P=5					P=10				
m	ε	μ	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	1.1	1.1	1.4	1.5	1.3	0.8	0.4	1	1.3	1.2	0.8	1.2	0.9	1	0.9
	0.1	3	17.1	7.5	44	43.9	47.8	9.7	10.9	55	54.3	56.6	6.6	5.3	35.9	30.3	33.5
		5	24.2	5.4	98.4	97.9	98.9	10.4	11.6	100	100	100	6.6	4.6	99.9	99.9	100
	0.2	3	3.4	2.8	11.2	9.9	12.1	1.8	2.7	10.2	7.7	9.8	2.2	2.8	6.5	4.6	6.5
		5	2.5	1.7	74	69.7	78.9	2	1.9	74	65.1	75.7	2.3	2.7	50.6	36.3	49.9
100	0	0	1.4	1.6	0.9	1	1.2	0.8	0.4	0.7	0.6	0.8	0.8	1.4	0.5	0.7	0.8
	0.1	3	20.2	10.9	53.4	51.5	54.8	11.6	12.5	61.9	57.2	63.1	7.7	6.8	56.6	46.8	54.7
		5	30.3	7.6	99.3	99.3	99.4	13	12.7	99	100	100	68.2	7.5	100	100	100
	0.2	3	3.7	2.5	13.5	10.9	15.1	2.1	2.4	7.8	4.8	9	2	2.1	7.5	4.4	7.7
		5	2.8	2.3	84.2	76.8	86.6	2	2.6	81.1	69.4	86	2.2	2	68.2	44.0	68.4
150	0	0	1	1.4	0.6	0.7	0.8	0.7	1.4	0.8	0.7	0.5	0.8	1.1	1	0.9	1
	0.1	3	19.7	11.1	50.7	48.5	53.6	11.8	15.4	69.4	62.2	67.8	9	8.2	66.1	57.7	62.6
		5	32.1	9.6	99.5	99.5	99.6	14	11.4	100	100	100	9	9.6	100	100	100
	0.2	3	3.4	2.5	11.9	9.5	13.6	2.5	3	10.4	5.9	9.3	2.2	1.9	7.7	4.4	7.6
		5	2.9	2.1	83.2	74.5	87.2	2.1	2.9	88.3	73.9	88.3	2.2	1.7	75.7	74	76.3

Table 6.4 displays the recorded percentage detecting of outliers for control charts and Figures 6.4 to 6.6 summarize these percentages in much better view. The behavior of the charts to identify outliers in the data when p = 2 is displayed in Figure 6.4. Each line chart in Figure 6.4 displays the relationship between the percentages detecting of outliers for each control chart and the group sizes (m). Overall, the percentage detecting of outliers increases as the group size increases. In addition, it is clearly plotted that both T_0^2 and T_1^2 are always the worst with the lowest percentage detecting of outliers in the figure. These two findings indicate that all control charts may identify more outliers as there are more objects in the data but

the traditional charts have lower capability to identify outliers compared to the robust charts.

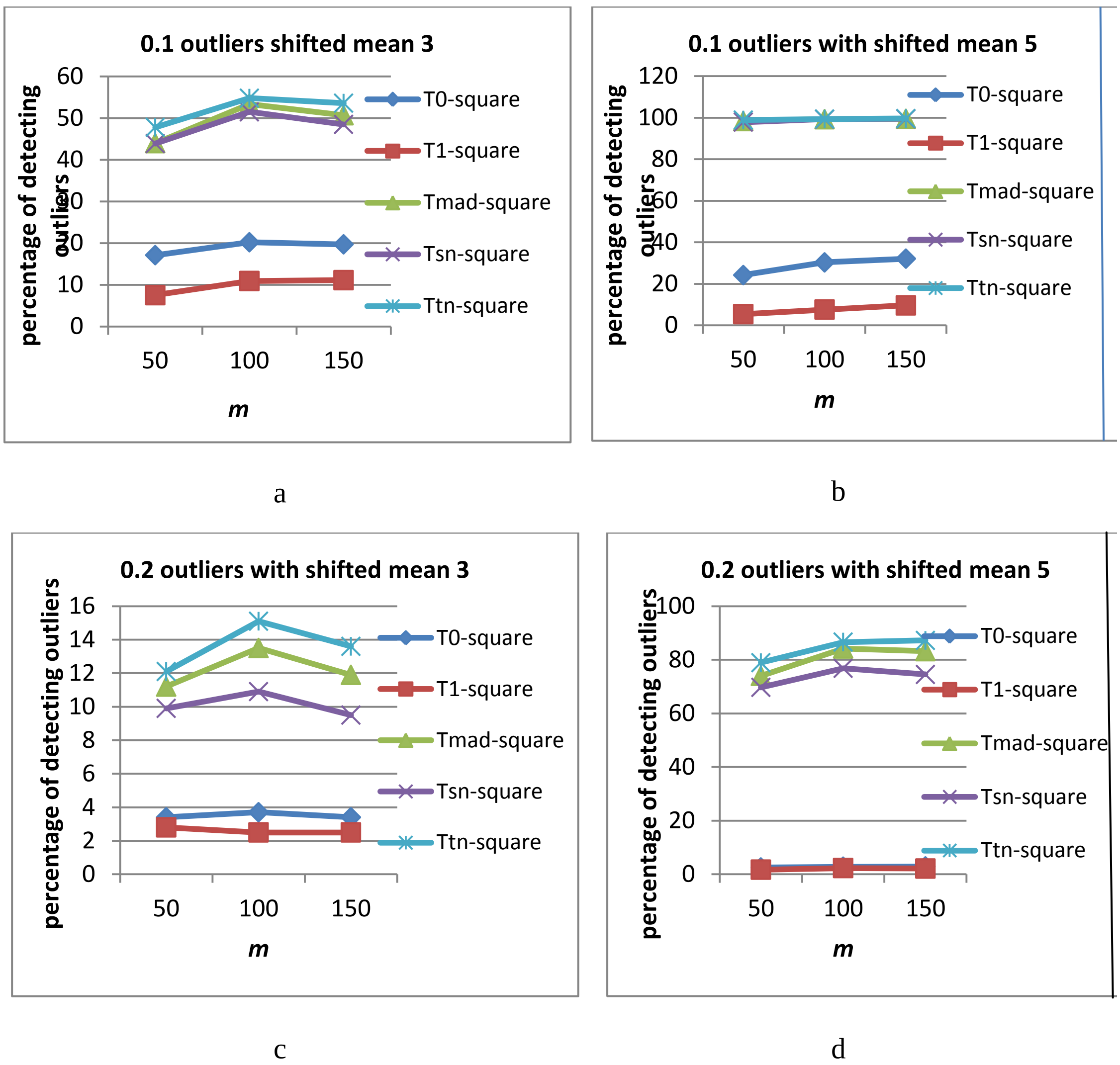


Figure 6.4:Percentages of detection of outliers when p = 2

The robust charts are able to identify about 70 percent of the outliers when mean shift is 5 (see Figures 6.4(b) and 6.4(d)) but they are poor when the mean shift is 3 (see Figures 6.4(a) and 6.4(c)). Such results show that small shifted from the mean of zero may give some hurdles for identifying the outliers. Among the robust charts, T^2_{HTn} has identified more outliers than the other two robust charts. Meanwhile, the increment of proportion of outliers (ε) in the data does not explain much about the behavior of the robust charts.

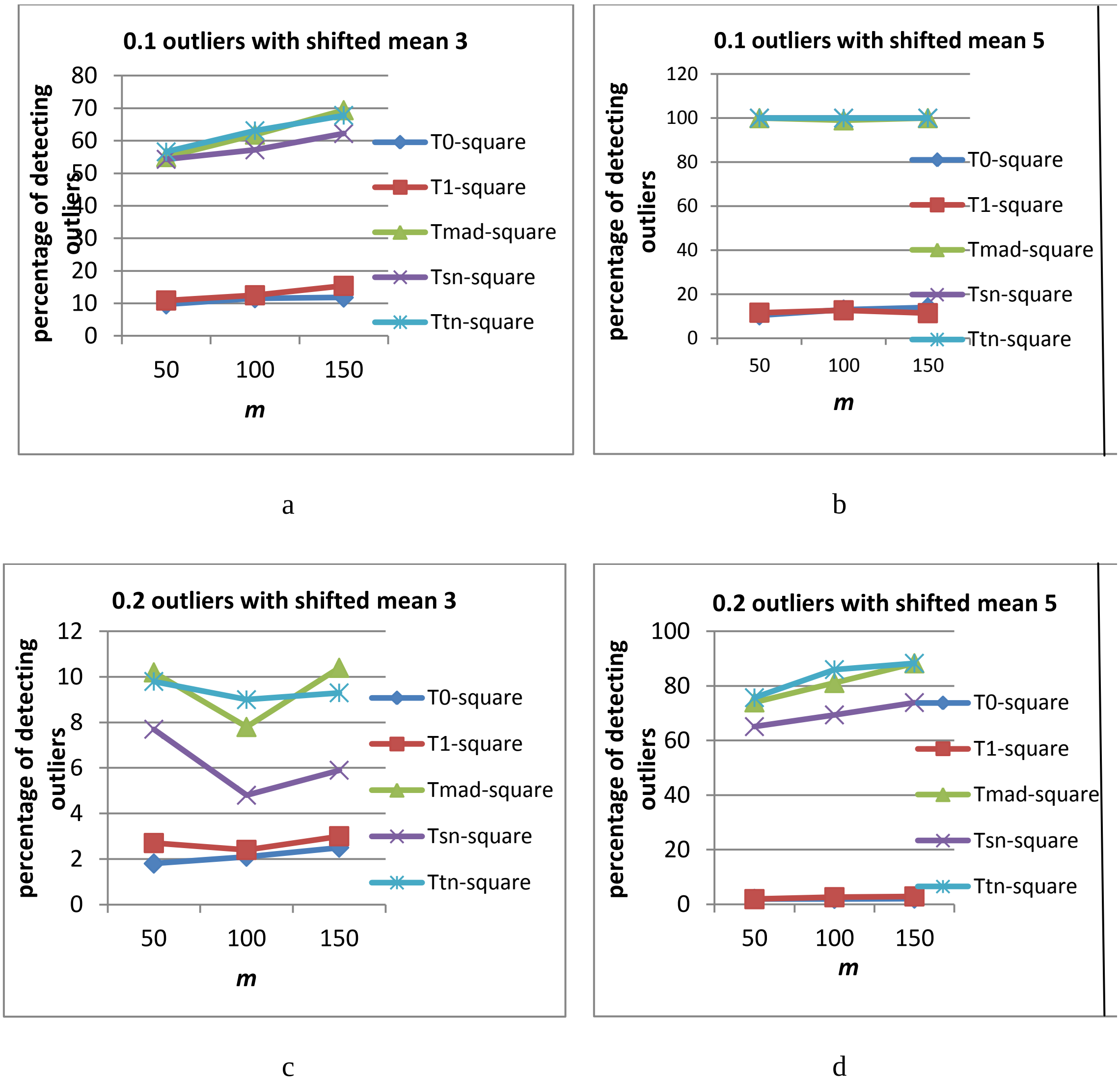
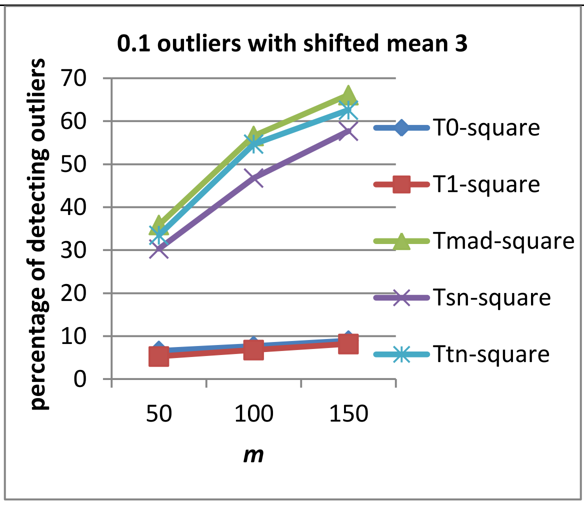
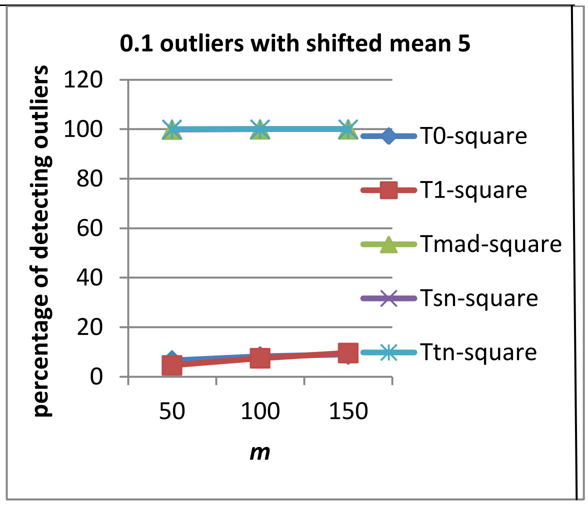


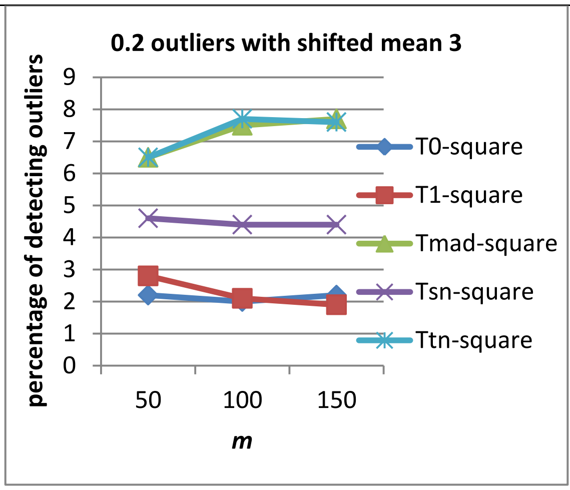
Figure 6.5: Percentages of detection of outliers when p=5



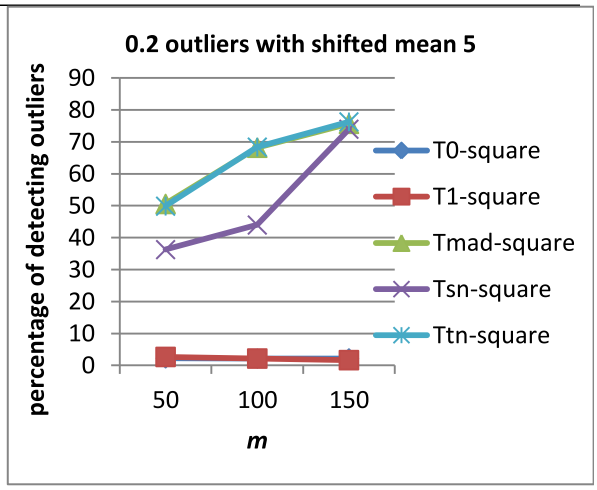
a



b



c



d

Figure 6.6:Percentages of detection of outliers when p=10

Behavior of control charts in detecting outliers when p = 5 and p = 10 is given in Figure 6.5 and Figure 6.6 respectively. Overall, similar behaviors as when p = 2 are identified where the ability to identify outliers in the data increases when the group size increases, robust charts always outperform the traditional charts and robust charts are good when mean shift is greater than 3. The investigation on percentage detecting of outliers towards the number of variables reveals that the robust charts show slightly increment in the percentage of outliers detected as the p increases from 2 to 5 but accelerates in the case $p = 10$.

Previous discussion has limited the investigation on control charts when the observed variables are independent. The next section will discuss the results when the variables are dependent.

4.3 The Dependent Variables (Case B)

The investigation on the control charts for dependent variables conditions are conducted based on two measurements namely the false alarm rates and percentage detecting of outliers which are tested at $\alpha = 5\%$ and $\alpha = 1\%$. The conditions of the data are similar to the ones given in Section 6.2.

4.3.1 False alarm rates and percentage detecting of outliers at $\alpha = 5\%$

The results of the analysis of the control charts based on the false alarm rates and percentage detecting of outliers at $\alpha = 5\%$ are summarized in Table 6.5 and Table 6.6 respectively.

Table 6.5: False alarms rates (percent) under dependent case for at=5%.

			P=2					P=5					P=10				
m	ε	μ	T_0^2	T_1^2	T_{HMad}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMad}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	5.6	6	6.1	5.1	5	5.3	5.6	6.2	3.5	2.2	4.7	5.7	6.7	5.3	2.9
	0.1	5	2.3	5.8	4.7	2.8	2.7	3.8	4.7	5.5	1.9	1.4	3.8	4.6	3.7	4.1	2
	0.2	5	2.3	3.1	2.9	1.4	1.7	3.6	3.8	4.1	1.6	2.2	4	3.8	2.7	4.6	2.3
100	0	0	5.5	6.3	4.2	3.8	3.7	5.4	5	5.4	3	1.9	4.5	5.8	5.2	3.1	1
	0.1	5	2.4	5.7	3.5	2.2	2.1	3.8	4.8	5.4	1.7	1.3	3.6	4.9	2.9	2.3	0.6
	0.2	5	2.2	3.1	2.6	1.2	1.1	3.7	4	4	1	0.6	3.4	4.7	2.7	2.1	0.4
150	0	0	5.3	6.4	4.5	3.8	3.3	4.3	4.4	5.3	2.3	1.8	4.6	5.4	2.7	5.1	1
	0.1	5	2.5	5.7	3.5	2.3	1.9	2.8	4.8	4.1	2	1.1	4.3	5	2.6	2.7	0.7
	0.2	5	2.4	3.2	2.7	1.1	1.7	2.7	3.6	3	1.3	0.6	4.3	4.8	2.6	2.3	0.3
%			44	100	100	44	44	100	100	100	22	0	100	100	100	67	11

Table 6.5 displays the results for each condition of group sizes (m = *50*, *100* and *150*), proportion of outliers (ε = *0, 0.1* and *0.2*) and the number of variables (p = *2, 5* and *10*) with the shifted means ($\boldsymbol{\mu}$ = 0and **5**) for the investigation on dependent variables. Table 6.5 shows that all control charts are in control for the ideal conditions ($\boldsymbol{\mu}$ = 0 and ε = 0) except for the T_{HTn}^2 when p = 5 and in two conditions when p = 10.

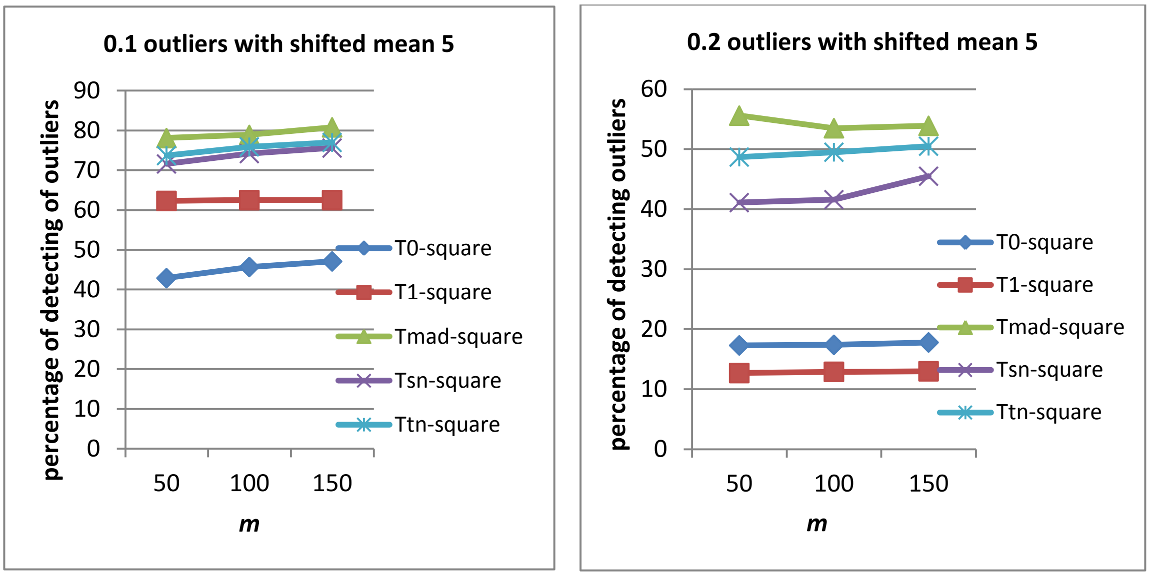
Results on the non-ideal conditions ($\boldsymbol{\mu} \neq 0$ and $\varepsilon \neq 0$) show that T_{HMADn}^2 has good performance in all conditions compare to the traditional chart T_0^2 and the other robust charts, T_{HSn}^2 and T_{HTn}^2. It followed by the traditional chart T_0^2 where it performs better as the number of variables increases from 2 to 10. Both T_{HSn}^2 and T_{HTn}^2 are not performing well when the variables are dependent. In the case when p = 2, there are slightly decrement on the false alarm rates as the group size increases. In addition, the false alarm rates decrease when there are more outliers in the data.

Similar fashions are detected when p = 5 and p = 10. However, the results on Table 6.5 are hardly to explain the behavior of robust charts when the number of variables increases.

Next, we are going to evaluate the capability of the charts to determine the outliers in the data with dependent variables at α = 5%. The observed percentage detecting of outliers are tabulated in Table 6.6 and plotted in line charts given in Figures 6.7 to 6.9.

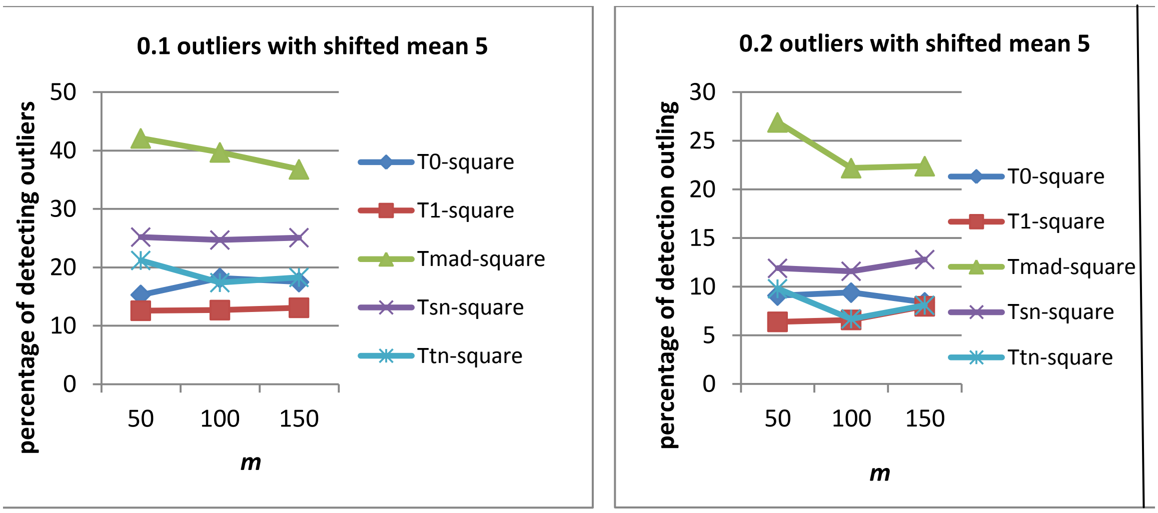
Table 6.6Percentages of detecting outliers for dependent case at α = 5%.

			p=2					p=5					p=10				
m	ε	μ	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	5.6	6	6.1	5.1	5	5.3	5.6	6.2	3.5	2.2	5.7	5.7	6.7	5.3	2.9
	0.1	5	42.9	62.3	78.1	71.6	73.7	15.3	12.6	42.1	25.2	21.2	10.6	7.5	10.1	17.2	10.1
	0.2	5	17.3	12.7	55.6	41.1	48.7	9.1	6.4	26.9	11.9	9.8	8	5.6	6.5	11.6	5.4
100	0	0	5.5	6.3	4.2	3.8	3.7	5.4	5	5.4	3	1.9	5.5	5.8	5.2	3.1	1
	0.1	5	45.7	62.5	78.9	74.2	75.9	18.2	12.7	39.7	24.7	17.4	9.9	8.4	9.1	14.4	5
	0.2	5	17.4	12.9	53.5	41.6	49.5	9.4	6.6	22.2	11.6	6.7	7.8	7.1	6.7	9.3	2.9
150	0	0	5.3	6.4	4.5	3.8	3.3	4.3	4.4	5.3	2.3	1.8	4.6	5.4	2.7	5.1	1
	0.1	5	47.1	62.5	80.8	75.6	77	17.5	13.1	36.8	25.1	18.3	10.6	8.6	9.4	13.5	6.1
	0.2	5	17.8	13	53.9	45.5	50.5	8.4	8	22.4	12.8	8.1	7.6	7.7	6.3	7.6	4.4



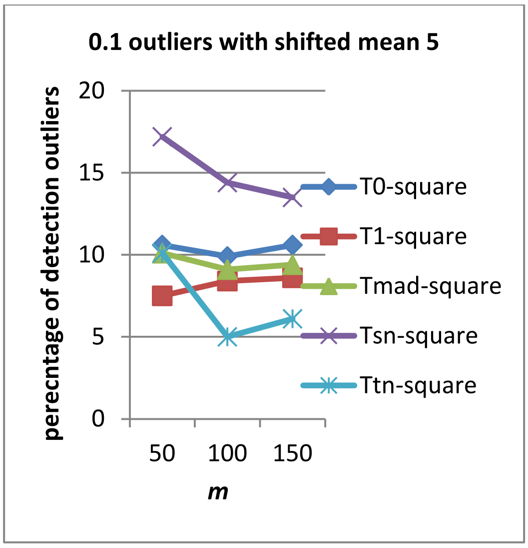
a **b**

Figure 6.7:Percentages of detection of outliers when p=2

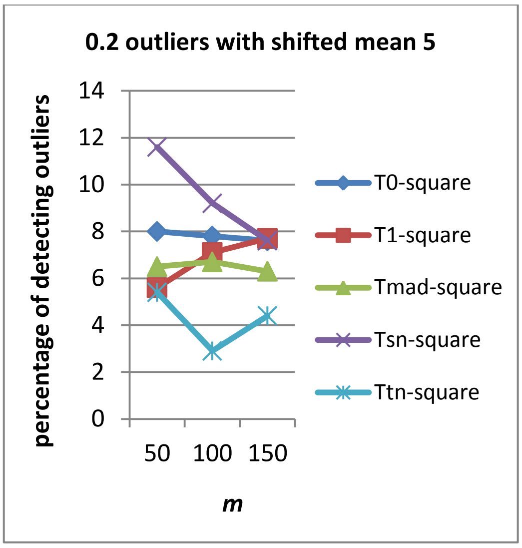


a b

Figure 6.8: Percentages of detection of outliers when p=5



a



b

Figure 6.9:Percentages of detection of outliers when p=10

Figure 6.7 illustrates the relationship between the percentage detecting of outliers and the group size for different conditions on proportion of outliers in the data for p = 2. Both Figure 6.7(a) and Figure 6.7(b) show the increment of percentage detecting of outliers among the control charts when the group size increases. In addition, the figures give evidence that the robust charts are highly capable to detect the outliers in the data compared to the traditional charts. However, the performance of the robust charts is bad when ε = 0.20 compared to their performance when ε = 0.10. This could be influenced by the structure of dependency of the variables, which may hide the outliers in the data.

Different scenarios illustrated by Figure 6.8 and Figure 6.9 when the investigation was carried out on p = 5 and p = 10 respectively. Both figures show that none of the charts is the best with percentage of detection at least 50 percent. As the matter of fact, T^2_{HTn} deteriorates badly when p is large. Results in Table 6.6 show that T^2_{HTn} gets as lowest as 1 percent in detecting the outliers.

4.3.2 False alarm rates and percentage detecting of outliers at α = 1%

The performance of the control charts at more stringent α = 1% is discussed based on false alarm rates (Table 6.7) and percentage detecting of outliers (Table 6.8). The conditions of the data are as given in previous subsection 6.2.1 when the observed variables are dependent on each others.

Table 6.7: False alarms rates (percent) under dependent case for α = 1%.

			p=2					*p*=5					*p*=10				
m	ε	μ	T_0^2	T_1^2	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}	T_0^2	T_1^2	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}	T_0^2	T_1^2	T^2_{HMadn}	T^2_{HSn}	T^2_{HTn}
50	0	0	1.1	1.1	1.2	1.3	1.4	0.8	0.9	0.9	1	0.7	0.8	1.2	0.8	1.3	0.7
	0.1	5	0.6	0.3	0.6	0.8	0.8	0.6	0.8	0.5	0.6	0.3	0.9	1.2	0.8	0.8	0.5
	0.2	5	0.6	0.4	0.5	0.3	0.3	0.6	0.6	0.6	0.3	0.3	0.8	1.1	0.8	0.8	0.5
100	0	0	1.4	1.6	1.3	1.1	0.6	0.8	0.4	1	0.7	0.5	0.9	1.4	1.1	1	0.3
	0.1	5	0.3	0.6	0.4	0.4	0.5	0.6	0.2	0.7	0.6	0.3	0.7	0.7	1	0.5	0.3
	0.2	5	0.3	0.7	0.3	0.3	0.3	0.6	0.5	0.7	0.4	0.4	0.5	0.8	0.8	0.5	0.3
150	0	0	1	0.9	0.7	0.8	0.6	0.7	1.4	0.6	0.9	0.2	0.8	1.1	0.8	0.6	0
	0.1	5	0.4	1.4	0.6	0.4	0.4	0.5	0.9	0.4	0.3	0.3	0.7	0.9	0.3	0.6	**0**
	0.2	5	0.4	0.6	0.4	0.1	0.3	0.4	0.6	0.2	0.2	0	0.6	1.2	1.4	0.7	**0**
%			56	78	67	44	56	89	78	78	56	22	89	100	89	100	**33**

The results of the false alarm rates in Table 6.7 indicate that most control charts are in control at α = 1% when the investigation was performed under ideal conditions

($\boldsymbol{\mu}$= 0 and ε = 0). However, the T^2_{HTn} chart does not perform well when p = 10 and the group size greater than 50.

Some patterns are identified when the investigation is carried out under the non-ideal conditions ($\boldsymbol{\mu} \neq 0$ and $\varepsilon \neq 0$). Table 6.7 shows that more robust charts are in control when the group size is 50 than 150 at p = 2. It means that the robust charts are in control when dealing with small group size. There are also some evidences that false alarm rates for the robust charts decrease when there are more outliers (ε) in the data. Similar behavior of false alarm rates are detected when p = 5 and p = 10 where the robust charts are in control for small group size and less outliers in the data.

The false alarm rates can be seen better when the number of variables increases from p = 2 to p = 10. However, only the T^2_{HMADn} chart has shown some promising results in most conditions compared to the other two robust charts. In contrast, the T^2_{HTn} chart has shown poor performance where it is out of control in 79 percent of the tested conditions, poorer than the performance of the traditional chart, T^2_0. Such results imply that the T^2_{HTn} chart may not suitable for dependent variables with tight false alarm rate.

We will look for more information about the performance of the control charts when the variables are dependent based on the percentage detecting of outliers at α = 1%. These percentages are tabulated in Table 6.8 and plotted in Figures 6.10 to 6.12.

Table 6.8: Percentages of detecting outliers for dependent case a tα = 1%.

			$p=2$					$p=5$					$p=10$				
m	ε	μ	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2	T_0^2	T_1^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
50	0	0	1.1	1.1	1.2	1.3	1.4	0.8	0.9	0.9	1	0.7	0.8	1.2	0.8	1.3	0.7
	0.1	5	15.3	7.3	52	45.2	47.3	4.2	4.4	28.5	13.6	10.2	1.7	2.1	5.9	9.8	2
	0.2	5	3.6	2.4	26.1	18.2	21.1	1.9	2.1	14.2	5.3	3.6	1	1.7	7.2	5.7	4.2
100	0	0	1.4	1.6	1.3	1.1	0.6	0.8	0.4	1	0.7	0.5	0.9	1.4	1.1	1	0.3
	0.1	5	16.5	9.8	56.1	49.7	51	4.6	5.6	24	10.9	6.5	2.4	1.7	3.8	7.9	2.1
	0.2	5	3.7	2.8	26.5	18.8	22.6	1.9	2.4	11.8	3.9	2.4	1.8	1.4	2.9	4	0.9
150	0	0	1	1.4	0.7	0.8	0.6	0.7	1.4	0.6	0.9	0.2	0.8	1.1	0.8	0.6	0
	0.1	5	17.5	10.1	46.2	49.1	50.8	4.6	4.7	9.7	11.6	11.3	2.4	2	15.8	7	1.6
	0.2	5	3.6	3	15.3	16.4	21	2.6	2.7	4	4.4	1.9	1.5	1.6	10.5	2.9	0.5

Visually, the percentages of detection outliers explained through the figures 6.10 to 6.12, according to increase the number of characteristics, *p*.

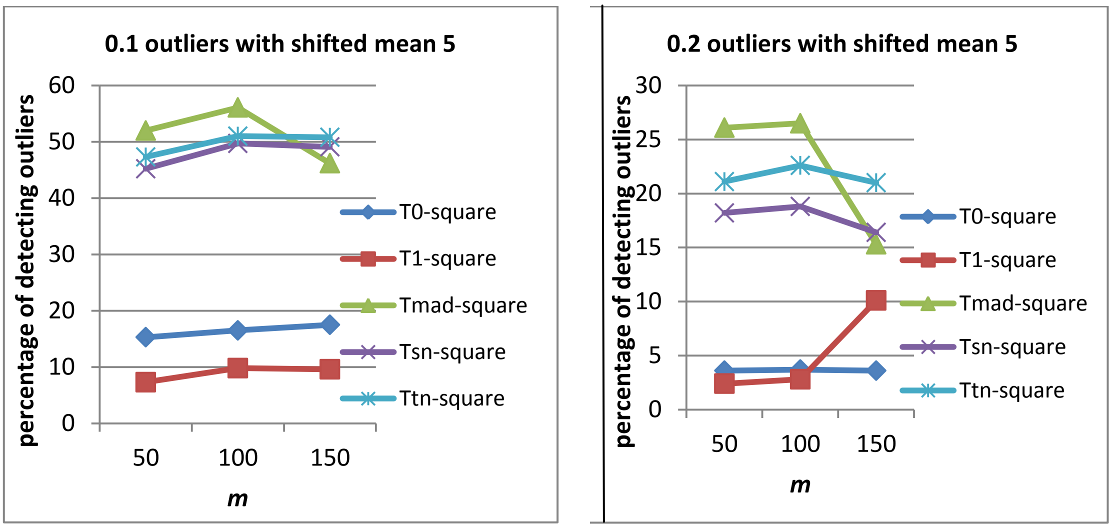
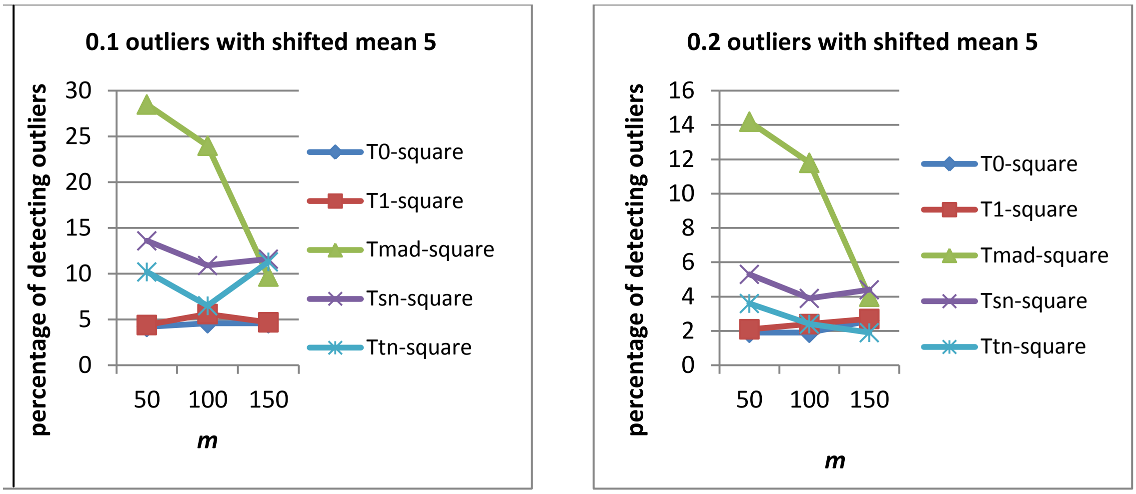
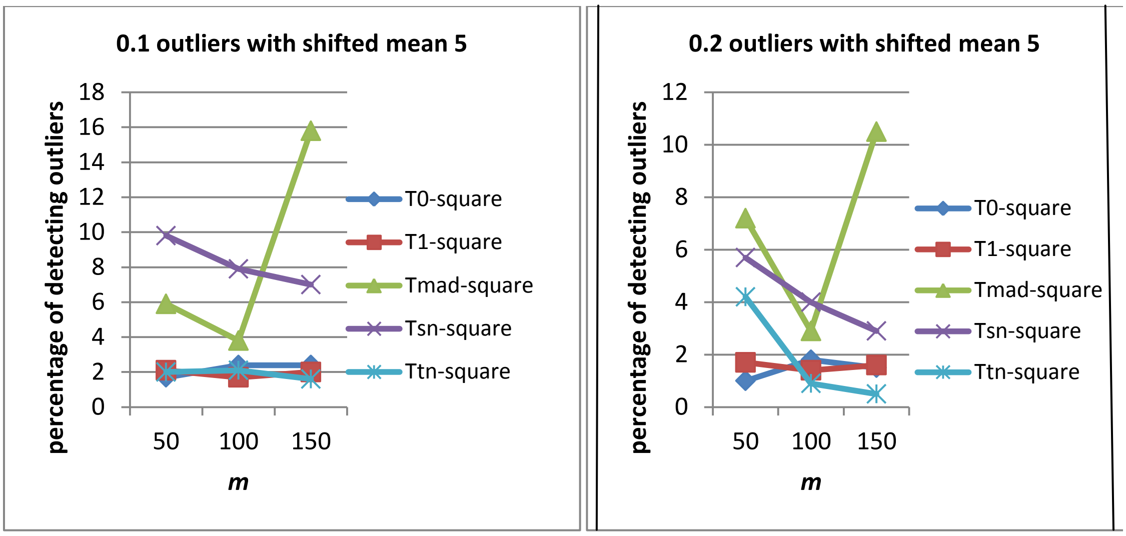


Figure 6.10:Percentages of detection of outliers when p=2



a b

Figure 6.11:Percentages of detection of outliers when p=5



a b

Figure 6.12:Percentages of detection of outliers when p=10

Figure 6.10 gives the relationship between the percentage detecting of outliers and group size for all investigated control charts when p = 2. The Figure 6.10(a) which represent the mild condition and moderate condition as it is shown in Figure 6.10(b) clearly show that the robust charts are outperforming the traditional charts in detecting the outliers in the data. However, the capability of detecting outliers among the robust charts decreases when the proportion of outliers (ε) increases from 0.10 to 0.20. In fact, the percentages of detecting outliers for the T^2_{HMADn} chart decrease when the group size reaches to 150.

Figure 6.11 (p = 5) and Figure 6.12 (p = 10) display similar behaviors as depicted by the Figure 6.10. Both figures give clear plot that the robust charts outperform the traditional ones in detecting the outliers for all conditions. All control charts are more capable to detect outliers when the group size is much bigger. However, despite of such good results of the robust charts, they do not well perform in detecting the outliers as their capability to identify them is less than 50 percent.

Our discussion has focused to evaluate the performance of the robust charts compare to the traditional charts in various interested conditions as explained in early section of this chapter. It is important to determine performance of each robust chart and to highlight similarities, strengths and weaknesses of them to allow better use for the charts. The next section will discuss in details about each of the robust charts pertaining to the investigations that were performed in this study.

6.4 Analysis on Real Data

In this chapter, the performance of the three robust charts T^2_{HMADn}, T^2_{HSn} and T^2_{HTn} on real data and their comparison with the traditional chart T^2_0 will be presented. The same data from the previous chapters (4 and 5) were used again in these charts. The calculations of the upper control limits (*UCL*) based on the estimated values are presented in Table 6.10. The values of future observations are shown in the first three columns, while the values of the T^2 statistics based on the estimated values in Table 6.10 are presented in the last four columns of Table 6.11.

Based on the comparison of the T^2 statistics in Table 6.11 with the corresponding control limits in Table 6.10, we observed that for α = 5%, two robust charts, $T^2_{HSn}$and T^2_{HTn} signaled three similar observations (20, 22 and 25) as out-of-control points. While the traditional T^2_0 chart manage to signal only two observations(22 and 25) as out of control points. In addition;T^2_{HMADn} chart has worse ability for detecting the outliers as well as the traditional T^2_O chart. While for α = 1%, it has been observed that the robust chartsT^2_{HMADn}, $T^2_{HSn}$and$T^2_{HTn}$and the traditional T^2_0 charts have the same signal observation values (20 and 25) which is considered out-of-control points. As it is expected, the result for T^2_O chart is similar to the results of the robust charts, since the analysis on the probability of detection of outliers calculated based on the simulated data. The simulated data showed that T^2_O was not as effective as the other robust charts in term of detecting outliers. According to the previous obtained results in the simulation (sections 6.2 and 6.3),it is clear that the three robust charts approximately have the same performance in terms of false alarms and probability of detection of the outliers, in case that small quality characteristics p is equal to 2. Also, as revealed in the simulation study, the T_0^2 chart has performed well in detecting outliers under low dimension (not more than 5) only.

*Table 6.9: Historical data set (*Phase *I data)*

Product No.	**Trim edge (x_1)**	**Trim edge spar (x_2)**	**Drill hole (x_3)**
1	-0. 0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	-0. 0017	0.0109	0.0177
5	-0. 0005	-0. 0010	0.0106
6	0.0016	-0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	-0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	-0.0067	0.0147
16	0.0110	-0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 6.10: The values of the upper control limits for the three robust and one traditional charts.

Types of Control Chart	Upper Control Limit (UCL) $\alpha = 5\%$
T_O^2	11.035
T_{HMADn}^2	14.22
T_{HSn}^2	11.83
T_{HTn}^2	12.77

Table 6.11: The values of future observations and Hotelling's T^2 statistics.

Product No.	x_1	x_2	x_3	T_O^2	T_{HMadn}^2	T_{HSn}^2	T_{HTn}^2
1	0.0041	0.0087	0.0129	0.5582	0.9915	0.9457	1.2066
2	0.0047	0.0109	0.0124	0.9003	1.7626	1.6819	2.0880
3	0.0031	0.0057	0.0096	0.4992	0.5299	0.4830	0.5205
4	0.0035	-0.0020	0.0101	0.5463	0.7516	0.6989	0.6068
5	0.0040	-0.0028	0.0125	0.4592	0.663	0.6301	0.5710
6	0.0031	0.0008	0.0061	0.9013	1.178	1.0532	1.0225
7	-0.0019	0.0101	0.0112	3.0933	3.271	2.8450	3.1874
8	0.0009	0.0039	0.0082	0.8061	0.9465	0.8023	0.7581
9	-0.0052	0.0090	0.0203	7.3602	7.3284	6.1959	7.2843
10	-0.0008	0.0110	0.0184	3.6198	4.2262	3.7671	4.5167
11	-0.0021	0.0139	0.0170	5.3839	6.1630	5.5186	6.4692
12	-0.0017	0.0092	0.0061	2.7387	3.2730	2.8435	3.0890
13	-0.0010	0.0133	0.0138	3.8058	4.4870	4.0606	4.7056
14	-0.0030	0.0002	0.0053	2.0548	3.3883	2.8313	2.8620
15	0.0016	0.0134	0.0151	2.5073	3.6659	3.4240	4.0858
16	0.0027	0.0086	0.0070	1.1976	1.5746	1.4488	1.6229

17	0.0004	0.0086	0.0087	1.5798	1.8288	1.6261	1.7780
18	-0.0036	0.0136	0.0129	5.7910	6.2007	5.4517	6.2216
19	-0.0028	0.0003	0.0078	1.8304	2.7078	2.2260	2.2046
20	0.0120	0.0123	0.0768	**38.1397**	**84.493**	**75.621**	**93.728**
21	-0.0015	0.0004	0.0115	1.2651	1.6002	1.2943	1.2395
22	0.0009	0.0232	0.0202	8.4181	13.462	**12.630**	**14.838**
23	-0.0035	0.0088	0.0107	3.7588	3.8324	3.2249	3.5605
24	0.0016	0.0061	0.0066	1.0602	1.3103	1.1611	1.2127
25	-0.0228	-0.0466	0.0231	**42.8447**	**87.870**	**78.700**	**87.664**
26	0.0037	-0.0038	0.0147	0.4832	0.8755	0.8323	0.8184

6.5 Comparison among the robust Hotelling's T^2 charts

Evidences and discussions carried out in Sections 6.2 and 6.3shows that the proposed robust charts outperform the traditional Hotelling's T^2 chart in most conditions of independent or dependent variables. Therefore, the issue now is to understand how they are leading each other. To allow for further discussion, we compute the percentages of in control results for all investigated conditions of each robust control chart. The in control results refers to the observed false alarm rate that falls in between: (i) 0.025 (2.5%) and 0.075 (7.5%) when α = 0.05 (5%) and (ii) 0.005 (0.5%) and 0.015 (1.5%) when α = 0.01(1%). We also compute the percentage of conditions in which the robust charts are able to identify at least 80 percent outliers in the data for independent case. While at least 50% probability of detection for the robust charts in case of dependent characteristics. Under the dependency characteristics, the table is further divided into the two performance measurement i.e. false alarm (FA) and probability of detection (POD). The comparison covers both the levels of significance α. These values indicate how good a robust chart to handle outliers is. Both these percentages are presented in Table 6.12.

Table 6.12: Overall performance for independent and dependent cases.

Types of Charts	Independent				Dependent (winsorized)			
	α = 5%		α = 1%		α = 5%		α = 1%	
	FA	POD	FA	POD	FA	POD	FA	POD
T^2_{wMadn}	**78%**	**69%**	**78%**	**36%**	**100%**	**33%**	**78%**	**11%**
T^2_{wSn}	**44%**	**64%**	**42%**	**25%**	**44%**	**17%**	**67%**	**0%**
T^2_{wTn}	**82%**	**69%**	**78%**	**36%**	**22%**	**17%**	**41%**	**11%**

When the variables are independent, it can be identified from Table 6.9 that the T^2_{HTn} chart show greater in control based on false alarm rates at α = 5%. It indicates that this chart has greater power to be in control in multi observed variables. In addition, Table 6.9 shows that both T^2_{HMadn} and T^2_{HTn} charts are performing well although α has been tightened to 1%. Such behaviors do not displayed by the T^2_{HSn} chart, but it shows good performance at the largest investigated number of variables, p = 10. In term of the capability to identify outliers in the data, all robust charts have shown their strength to identify the outliers especially at α = 5%, but they face some difficulties to identify more outliers when α = 1%.

The great dependency among variables gives some limitations for the robust charts to perform as good as when the variables are independent. Table 6.9 gives clear evidence that the T^2_{HMADn} chart is the greatest in control when α = 5% and it does not badly affected although α has been tightened to 1%. The other robust charts are unable to show promising results when α = 5% but show improvement as α has been tightened to 1%which means that these charts control on false alarms more when the study be more conservative. In addition, the T^2_{HMadn}chart identifies more outliers in the data with the dependent variables than$T^2_{HSn}$and T^2_{HTn}charts.

The evaluation based on false alarm rates has identified the T^2_{HMadn}chart as the best multivariate control chart when the variables are independent compared to the T^2_{HSn} chart and the T^2_{HTn} chart. The performance of the T^2_{HMadn} chart does not affect much when the variables are dependent. Meanwhile, the T^2_{HSn} chart shows moderate performance in both independent and dependent variables. Lastly, the T^2_{HTn} chart shows competitive performance to the T^2_{HMadn} chart when the variables are independent but it performs badly when there are some dependencies among the variables.

Similarly, the evaluations based on the percentage detecting of outliers reveal that all the robust charts are able to identify outliers in the data when independent variables but the T^2_{HMadn} chart is good enough to identify outliers when the variables are dependent. Following these results, one may opt to use either the T^2_{HMadn} chart or the T^2_{HTn} chart when the variables are known to be independent and to opt for the T^2_{HMadn} chart when the variables are known to be dependent. However, the T^2_{HMadn} chart is preferable when one does not certain about the dependency of the variables.

6.6 Summary

This chapter has introduced and evaluated three proposed robust multivariate Hotelling's T^2 charts based on the robust location estimator of Hodges-Lehmann with the covariance matrices of robust scale estimators, the choice ofMad_n, S_n and T_n. The evaluations which were presented in the case of independent and dependent variables in Section 6.2 and Section 6.3 indicated to that the robust charts perform better than the traditional Hotelling's T^2 in term of false alarm rate and identifying outliers in the data. Among the robust charts, it has been found that the T^2_{HMadn} chart has shown greater performance than the T^2_{HSn} chart and the T^2_{HTn} chart in both cases, either independent or dependent variables.

The results presented in Table 6.1 to Table 6.8 show some explainable relationships between the interested factors namely the number of variables (p), group sizes (m), proportion of outliers (ε), mean shift ($\boldsymbol{\mu}$) and the nominal false alarm rate (α), and the status of the process of the robust charts, either in control or out of control. The dependency among variables play important scenario to influence the performance of the robust charts and this has been proven in Section 6.3. In general, charts that are more robust are in control false alarms as p increases when the variables are independent. Nevertheless, such scenario is not occurring when the variables are dependent where the performance of robust charts affected negatively at α = 5% while affected positively at α = 1%. Besides, the size of contaminated data may affect the status of the process. The tables presented in Section 6.2 and Section 6.3

show that more robust charts are affected negatively if there are more proportion of outliers as given by ε in the data. However, the positively affect can be seen on the effect of mean shift towards the status of the process.

Another factor, which is important to highlight, is the set of false alarm, α, for the evaluation. Results as given and discussed in this chapter have proven that more robust charts are in control when the evaluation was performed at nominal false alarm rates, α = 5%. While, there are many out of control results occur when the evaluation was performed at rigid nominal false alarm rates, α = 1%. However in practice, the choice of α may varies depends on the application and problems in hand.

CHAPTER SEVEN
CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCHES

5.1 Introduction

This thesis proposed three new approaches of robust Hotelling's T^2 charts. The new robust charts are Hotelling's T^2 with trimmed means (T_t^2), Hotelling's T^2 with winsorized means (T_w^2) and Hotelling's T^2 with Hodges Lehman (T_H^2). For each approach, we introduced three different scale estimators comprising ofMad_n, S_n andT_n. These scale estimators functioned differently for each type of chart. When applied inT_t^2, the scale estimators functioned as the scatter matrix in Mahalonobis distance used to determine the data to be trimmed. InT_w^2, the estimators acted as the trimming criterion before the winsorization of the data could be done, while inT_H^2, they took the role as the scatter matrix in the robust Hotelling T^2 charts. Altogether, a total of nine new robust charts were being investigated. These new charts were shown to have better performance as compared to the traditional Hotelling T^2 charts in the presence of outliers. In this chapter, we summarize the comparison of the performance among the proposed robust charts and the traditional charts based on independent and dependent cases in sections 7.1 and 7.2, respectively. The effect of the charts with regards to variables manipulated is summarized in section 7.2. The discussion on the contributions of this study continues in section 7.3 and finally the future studies related to this thesis are summarized and discussed in section 7.4.

5.2 COMPARISON AMONG THE CONTROL CHARTS

The performance of the robust charts is compared in terms of their false alarms rates and the percentage of detecting outliers. The comparisons between the approaches are discussed separately based on the nature of the quality characteristics i.e. independent and dependent cases.

5.2.1 Independent case.

The results for the comparison under independent case are summarized in Table 7.1. The table is sectioned according to the three types of robust approaches (T_t^2, T_w^2 and T_H^2). Each approach is then divided into two levels of significance (α = 5% and 1%). Under each level of significance, the table is further divided into the two performance measurement i.e. false alarm (FA) and probability of detection (POD). The corresponding robust charts using different scale estimators (Mad_n, S_n and T_n) for each approach are arranged in rows followed by the traditional charts in the last two rows of the table. The percentages under FA column represent the frequency of the robust conditions over the number of conditions studied for a particular chart (denoted as FA hereafter) while the percentages under POD column represent the frequency the chart achieved above 80% level of detection (denoted as POD hereafter) across the conditions.

Table 7.1 Overall performance for independent case

Scale Estimators	T_t^2				T_w^2				T_H^2			
	$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$	
	FA	**POD**	**FA**	**POD**	**FA**	**POD**	**FA**	**POD**	**FA**	**POD**	**FA**	**POD**
Mad_n	**67%**	**89%**	**44%**	**69%**	**93%**	**50%**	**87%**	**61%**	**78%**	**69%**	**78%**	**36%**
Sn	**62%**	**89%**	**58%**	**67%**	**91%**	**31%**	**84%**	**50%**	**44%**	**64%**	**42%**	**25%**
Tn	**62%**	**89%**	**53%**	**72%**	**91%**	**50%**	**93%**	**67%**	**82%**	**69%**	**78%**	**36%**
	Traditional											
T_1^2	**100%**	**8%**	**80%**	**0**	**100%**	**8%**	**76%**	**0**	**100%**	**8%**	**80%**	**0**
T_0^2	**64%**	**6%**	**56%**	**0**	**78%**	**33%**	**64%**	**0**	**64%**	**6%**	**56%**	**0**

As we can observe from Table 7.1,at α = 5%, the robust charts that perform the best in controlling false alarms come from the winsorized approach (T_w^2) where most of the percentages are above 90%. The next best approach is T_H^2 where the approach used no trimming or winsorizing on the data. However, the FA percentages among the robust charts under this approach are not consistent as in the earlier approach. The best among this approach is T_{HTn}^2. The approach that produces the least robust conditions is the one using trimmed mean i.e. T_t^2. The FA percentages among the charts under T_t^2 approach are quite consistent with values ranging from 62% to 67%. Even though this approach produces the least robust conditions, but it has the highest percentage (89%) of POD. The second best goes to T_H^2 with POD percentages around 66% while the least belongs to T_w^2. If we refer to robustness per se, we would recommend the robust charts using winsorized mean. However, if the performance takes into consideration both the robustness and the ability in detecting outliers, the best suggestion would be the T_H^2 chart, specifically T_{HMadn}^2 and T_{HTn}^2.

At α = 1%, T_w^2 still stays on top of the list with regards to FA, even though the percentages slightly drop from α = 5%. Nonetheless, in terms of POD, the percentages for the charts increase, thus making these charts as the better choices than the others.

Next, the comparison continues with the traditional charts. From the table, obviously we can see that T_1^2scores perfect 100% which indicates that all the conditions tested are robust. However, the chart fails to score any points in term of POD. For the case of T_0^2, the chart in general cannot outdo the robust charts in terms of FA, and as expected, perform as bad as T_1^2 with respect to POD.

5.2.2 Dependent Case

The summary for dependent case is presented in Table 7.2. The arrangement of the table is similar to the independent case, but the percentages under POD for this case represent the frequency the chart achieved above 50% level of detection across the conditions. In terms of FA at 5% significance level, the charts belong to T_w^2 approach produce high percentage (at least 85%). One of the percentages, which belong to T_{wTn}^2 reaches the 100% score. Trailing behind is T_t^2 followed by T_H^2. Even though the overall FA percentage for T_H^2 not so promising, but one of the charts, i.e. T_{HMadn}^2 scores perfect 100% and produce quite good percentage (33%) for POD. Among the approaches, T_t^2 charts score the highest (33% - 39%) while those from T_w^2 score the lowest (17%).

As we moved to α = 1%, we observe some improvements in the percentages of FA for a few charts belong to T_t^2 and T_H^2. On contrary, we also observe a drop in percentages for some charts in T_w^2. Despite the drop in percentages of FA, T_w^2 charts are still above the charts from the other approaches, with T_{wTn}^2 scores perfect 100% when α = 5%. However, in terms of POD, the chart fails to score any point. Other charts which are show a balance between FA and POD are from T_t^2.

Table 7.2 Overall performance for dependent case

Scale Estimators	T_t^2				T_w^2				T_H^2			
	α = 5%		α = 1%		α = 5%		α = 1%		α = 5%		α = 1%	
	FA	POD	FA	POD	FA	POD	FA	POD	FA	POD	FA	POD
Mad_n	78%	39%	78%	17%	85%	17%	74%	11%	100%	33%	78%	11%
S_n	70%	33%	81%	17%	89%	17%	85%	0%	44%	17%	67%	0%
T_n	70%	33%	74%	17%	100%	17%	100%	0%	22%	17%	41%	11%
	Traditional											
T_1^2	100%	17%	89%	0%	100%	17%	70%	0%	100%	17%	89%	0%
T_0^2	81%	0%	81%	0%	85%	0%	93%	0%	81%	0%	81%	0%

5.3 The effects of the manipulated variables on the charts

The following subsections will summarize on the general performance of the robust and traditional charts with respect to the manipulated variables.

5.3.1 Independent variables case (A)

The nine robust Hotelling's T^2 charts proposed in this study performed well in terms of controlling false alarm and had high probability of detecting outliers for certain studied conditions. When we compared with the traditional charts, we observed that, even though T_1^2 chart was known to have very good control of false alarm, but the chart failed in detecting outliers unlike the proposed robust charts. These robust charts also outperformed the T_0^2 under most conditions in terms of false alarm and all the time above the T_0^2chart with regards to probability of detecting outliers. We have summarized the performance of all the charts in the previous sections. This section will summarize on the effect of the robust charts in general based on the manipulated variables.

i) Number of quality characteristics (*p*)

Generally, when the number of quality characteristics (*p*) became larger, we observed some positive effect on the false alarm but the changes in *p* showed insignificant effect on the probability of detection. These situations applied to all charts.

ii) Group sizes (*m*)

The increased of group sizes did have some influence on the robust charts. However, different charts portrayed different behaviors, thus making it impossible to determine the common behavior of the false alarm and probability of detection for the robust and traditional charts.

iii) Proportion of outliers (ε)

In terms of proportion of outliers, the increased of this variable negatively affected the false alarm and probability of detection of the robust charts. While, in the case of the traditional charts, the similar changes negatively affected the performance in terms of false alarms but no obvious effect could be identified in terms of detecting outliers.

iv) Shifted mean ($\boldsymbol{\mu}$)

When the mean shifted from 3 to 5, no significant effect could be observed in terms of false alarm for the robust charts, but negative effect could be identified for the traditional charts. In terms of probability detection, the shifts in mean showed positive effect on the performance of the robust charts but no obvious effect on the traditional charts.

5.3.2 Dependent variables case (B)

The effects on false alarms rates, due to the changes in the manipulated variables for dependent case were quite similar to the independent case. Nevertheless, with regards to probability of detection, there was a decline in the values as compared to the independent case.

i) Number of quality characteristics (p)

Generally, when the number of quality characteristics (p) increased, we observed some positive effect on the false alarm and negative effect on the probability of detection for all charts.

ii) Group sizes (m)

Like in the independent case, different patterns were observed for different charts when the group sizes increased. Thus, no common performance pattern could be identified for the charts in terms of false alarm and probability of detection.

iii) Proportion of outliers (ε)

The performance of the robust charts dwindles when the proportion of outliers changes from 0.1 to 0.2 in terms of false alarm and probability of detection and the probability of detection further decreases under dependent case compared to independent. While the traditional charts affected negatively in term of false alarms, however there are no obvious effect in detecting outliers.

7.4 APPLICATION ON REAL DATA

When the investigated charts were applied on real data (the production of airplane spoilers), the robust charts especially those using winsorized *MOM* performed the effectively in detecting out-of-control observations. All the three charts (T^2_{wMadn}, T^2_{wSn} and T^2_{wTn}) were able to detect three out-of-control observations as compared to only two detected by the traditional chart. Meanwhile, two of the robust charts using Hodges Lehamann i.e. T^2_{HSn} and T^2_{HTn}were also able to detect the same number of observations. However, their counterparts, T^2_{HMadn} as well as those charts using trimmed mean, T^2_{tMadn}, T^2_{tSn} and T^2_{tTn} can perform as good as the traditional chart, and not less. The simulated study showed that the traditional chart performed well under small number of quality characteristic (p < 5), which is why the chart is on par with some other charts for this particular problem (p = 3)

Generally, the findings on simulated and real data showed that robust Hotelling's T^2 charts using winsorized *MOM* could be a favorable alternative for multivariate control charts.

3.5 THE CONTRIBUTIONS OF THE THESIS

Our goal is to search for alternative control charts for the traditional Hotelling's T^2 charts when the performance of the latter is arguable due to violations of assumptions. In this final chapter, we would like to share some of the advances that emerged from this study.

Modifications made on the Hotelling's T^2 chart successfully improved the performance of the charts especially in terms of probability of detection. Even though the robust charts generally perform moderately in terms of robustness (false alarm), they perform really well when p is large, which is known to be one of the weaknesses of Hotelling's T^2 statistic. Furthermore, these robust charts are unperturbed even when the means shifted, which indicate that they are robust towards outliers.

To reiterate, this study covered various aspect in dealing with outliers. It introduced an approach, which needs no trimming of data i.e. T_H^2. Even in the presence of outliers, this approach was proven to be able to control false alarm and simultaneously able to detect outliers at desired level. This study also offers the symmetric trimming approach denoted as T_t^2 and the asymmetric winsorizing approachT_w^2. Trimming or no trimming, the approaches produced comparable control charts to traditional Hotelling T^2 under ideal conditions and perform better than the traditional charts under violation of assumptions. With these alternative charts, we hope that the problems which were previously unable to be solved under certain conditions can now be overcome using either one of the alternatives.

5.4 LIMITATION and FUTURE STUDIES

We believe that this study is not a comprehensive study that covers the real situations occur in industries. The study limits the finding based on the simulated data whereby all the conditions are controlled. However, we believe that the outcomes from this study will not be much affected from the real application.

While conducting this study, we encountered a few problems, which lead us to give some ideas on future research. As we all aware, the distributions for the robust statistics are unknown, thus in this study we used simulation method to calculate the control limits. The process takes its toll on the computation time. We need more extensive theoretical research to be done in order to identify the suitable distributions for the statistics. Prior to this study, we are not aware that the proposed method would encounter any problem when the condition of m/p > 5 is not fulfilled. However, along the way, we realized that only one of the approaches i.e. T_w^2 able to perform even when the condition is violated. In future, we would like to further understand why the other two approaches failed to perform. Last but not least, another suggestion that we would like to bring forward for future work is to extend the investigation on distributional shapes.

References

Abu-Shawiesh, M. O. & Abdullah, M. (2001). A new Robust Bivariate Control Chart For Location. *Communication in statistics, 30*, 513 – 529.

Alfaro, J. L. & Ortega, J. F. (2008). A Robust Alternative to Hotelling's T^2 Control Chart Using Trimmed Estimators. *Qual. Reliab. Eng. Int., 24,* 601-611.

Alfaro, J. L.& Ortega, J. F. (2009). A comparison of robust alternative to Hotelling T^2 control chart. *Applied Statistics, 36*(12), 1385-1396.

Alloway, J. A. J. & Raghavachari, M. (1990). Multivariate Control Charts Based on Trimmed Mean. *ASQC Quality Congress Transactions – San Francisco,* 449-453.

Alloway, J. A. & Raghavachari, M. (1991). An introduction to Multivariate Control Charts. *ASQC Quality Congress Transactions – Milwaukee,* 773-781.

Alt, F. B. (1985). Multivariate quality control. *New York, John Wiley & Sons*.

Ansell, J. G., & Margaret. (2009). Robustness of Location Estimators to Asymmetry. *Journal of the Royal Statistical Society. Series C (Applied statistics),* 249-254.

Ben-Gal, I. (2005). Outliers detection. In: Maimon O. and Rockach L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, *Kluwer Academic Publishers, ISBN* 0-387-24435-2.

Bersimis, S., Psarakis, S. & Panaretos, J. (2006). Multivariate Statistical Process Control Charts: An Overview. *Qual. Reliab. Engng. Int., 23, 517–543.*

Besterfield, D. H. (Ed.) (2004). *Quality Control* (7 Ed.): Pearson Education, Inc., Upper Saddle River, New Jersey 07458.

Bickel, P. J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics, 36,* 847-858.

Bluman, & Allan, G. (2009). *Elementary Statistics: A step by step approach* (7 ed.).

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology. 31*, 144-152.

Brooks, A. (1985). Heat treating shows why SPC is no curve-all for manufacturers production Engineering. *32*(9), 66-77.

Brown, B. M. & Kildea, D. G. (1978). Reduced u-statistics and the Hodges-Lehmann estimator. *The Annals of Statistics, 6, (*4), 828-835.

Cheng, L. M., Away, Y. & Hasan, M. K. (2006). The Algorithm and Design of Real-time Multivariate Statistical Process Control System. *Multimedia Cyberspace*, *4*(2), 18-23.

Chenouri, S., Variyath, A., M. & Steiner, S. H. (2009). A Multivariate Robust Control Chart for Individual Observations. *Quality Technology*. *41*(3), 259-271.

Chernobai, A. & Rachev, S. T. (2006). Applying robust methods to operational risk modeling *Operational Risk, 1*(1), 27-41.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality schemes. *Technometrics*, *30*(3), 291-303.

Everitt, B. S. (1979). A Monte Carlo investigation of robustness of Hotelling' one and two sample T- square test. *American statistical Association*, *74*(365), 48-51.

Geyer, C., J. (2006). Breakdown Point Theory Notes. Retrieved from http://www.stat.umn.edu/geyer/s06/5601/theo/break.pdf

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc., 69*, 383-393.

Hampel, F. R. (2000). Robust inference. *Seminar for Statistic. Research Report No. 93.* Switzerland.

Hellerstein, J. M. (2008). Quantitative Data of Cleaning for Large Data basis. *EECS Computer Science Division. UC Berkeley*

Hodges, J. L.& Lehmann, E. L. (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics. 34*, 598-611.

Hodges, R. V. (1967). Some Observations on Robust Estimation. *American Statistical Association, 62(320)*, 1179-1186.

Hogg, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *American Statistical Association. 69(348),* 909-923.

Holmes, D. S. & Mergen, A. E. (1996). Identifying the sources for out of control signals when the T^2 control chart is used. *Quality Engineering, 8*(1), 137-143.

Hotelling, H. & McGraw-Hill (1947). Multivariate Quality Control, Illustrated by the Air Testing of Sampling Bombsights. *Techniques of Statistical Analysis, Eisenhart, C., Hastay, M., Wallis, W. A. (eds.),* 111-184.

Jackson, J. E. (1985). Multivariate Quality Control. *Communication in statistics: Theory and methods*, *14*(110), 2657-2688.

Janacek, G. J., Meikle, S. E. (1997). Control Charts Based on Medians. *The Statistician, 46*(1), 19-31.

Jensen, W. A., Birch, J. B. & Woodall, W. H. (2007). High Breakdown Estimation Methods for Phase I Multivariate Control Charts. *Quality Reliability Engineering International*, 23, 615-629.

Johnson, M. E. (Ed.) (1987). *Multivariate statistical simulation*: New York, John Wiley and sons.

Johnson, N. (2007). *A comparative simulation study of robust estimators of standard error.*, Brigham Young University

Linden, A., Adams, J. L. & Robert, N. (2004). Using an Empirical Method for Establishing Clinical Outcome Targents in Disease Management Programs. *Disease Management. 7(2),93-101.*

Majid, A., Haron, K. & Midi, H. (2010). Hodges Lehmann Estimator for Robust Design Using Dual Response Surface Approach. *Australian Journal of Basic and Applied Science. 4(8), 3418 – 3423, ISSN 1991 – 8178. INSI net publication*

Mahir A., R., & Al-Khazaleh, A. M. (2009). New Method to Estimate Missing Data by Using the Asymmetrical Winsorized Mean in a Time Series. *Applied Mathematical Sciences, 3*(35), 1715 - 1726.

Mason, R. & Young, J. (1999). Improving the sensitivity of the T^2 statistic in multivariate process control. *Quality Technology*, *31*, 155-165.

Mason, R. L., Chou, W., Y. & Young, J. C. (2001). Applying Hotelling's T^2 statistic to Batch Processes. *Quality Technology, 33(4), 466.*

Mason, R. L., Chou, Y., Sullivan, J. H., Stoumbos, Z. G. & Young, J. C. (2003). Systematic patters in T^2 charts. *Quality Technology, 35*, 47-58.

Mason, R. L. & Young, J. C. (2002). Multivariate Statistical Process Control with Industrial Applications. *Vol. 9* of *American Statistical Association-Society for Industrial and Applied Mathematics (ASA-SIAM).*

Math works Matlab Version 7.8 (R2009a).

Mehrotra, K., Jackson, P. & Schick, A. (1991). On choosing an optimally trimmed mean. *Communication in Statistics-Simulation and Computation, 20*(1), 73-80.

Midi H., Shabbak, A., Talib, B. A. & Hassan, M. N. (2009). *Multivariate control charts based on robust Mahalonobis distance.* Proceedings of the 5th Asian Mathematical Conference, Malaysia

Mingxin W. (2006). *Abstract trimmed and winsorized estimator.* PhD. Michigan State University.

Mohammadi M., Midi H., Arasan J. & Al-Talib B. (2011). High Breakdown Estimators to Robustify Phase II Control Charts. *Applied Sciences, 11 (3),* 503-511

Montgomery, D. C. (Ed.) (2005). *Introduction to Statistical Quality Control* (5 ed.): John Wiley and Sons, New York, NY, USA.

Nedumaran, G.& Pignatiello, J. J. (2000). On constructing T^2 control charts for retrospective examination. *Communication in Statistics-Simulation and Computation*, *29*, 621-632.

Oktay A. & Aricigil A. (2001). Multivariate quality control: A historical perspective. *Statistics statistical techniques*, 54-65.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R. & Fradette, K. (2004). Comparing measures of the 'typical' score across treatment groups. *British Journal of Mathematical and Statistical Psychology*. *57,* 215 - 234.

Pei-Chen, W. (2002). *The central limit theorem and comparing means, trimmed means, one step M-estimators and modified one step M-estimators under non normality.* Southern of California, Los Angeles, California.

Peña& Prieto, F. J. (2001). Multivariate Outlier Detection and Robust Covariance Matrix Estimation. *Technometrics August, 43*(3), 286-301.

Pei-Chen. W. (2007). Modern One-Way ANOVA F Methods: Trimmed Means, One Step M-estimators. *Quantitative Research, 12(1),* 151-169.

Randles, R. H., Wolf, D. A. (Ed.) (1979). *Introduction to the theory of nonparametric statistics*. John Wiley & Sons. ISBN: 0471042455.

Rock, D. M., Downs, G. W. & Rock, A. J. (1982). Are robust estimators necessary? *Technimetrics, 24*, 95 - 101.

Rocke, D. M. & Woodruff, D. L. (1998). Some Statistical Tools for Data Mining Applications. *Center for Image Processing and Integrated Computing.*

Rosenberger, J. L.& Gasko & (1983). Comparing location estimators: Trimmed means, medians and trimmed mean. In Hoaglin, D.C., Mosteller, F. & Tukey, J.W. (Eds). *Understanding robust and exploratory data analysis, New York: John Wiley & Sons*, 297-336.

Rousseeuw, P. J.& Croux, C. (1992). Time-efficient algorithms for two highly robust estimators of scale. *Computational Statistics, 1*, 411-428.

Rousseeuw, P. J.& Croux, C. (1993). Alternatives to the Median Absolute Deviation. [Theory and Methods]. *American Statistical Association. 88(424).*

Rousseeuw, P. J. & Van Zomeren, R. C. (1990). Unmasking multivariate outliers and leverage points. *American statistical Association, 85(633-651).*

Sepulveda, A.& Nicholas, J. A. (1997). A Simulation Approach to Multivariate Quality Control. *Computers in. Engng, 33 (*No's I-2), 113-I 116.

Siegel, A. F. (Ed.). (1988). *Statistics and data analysis: an introduction*: New York: John Wiley & Sons.

Steiner, S. H. (1994). *Quality Control and Improvement Based on Grouped Data*. (Doctoral dissertation). Retrieved from http//digitalcommons.mcmaster.ca/open dissertations. Paper 3883.

Sullivan, J. H.& Woodal, W. H. (1996). A comparison of Multivariate Control Charts for Individual Observations. *Quality Technology, 28*, 398 - 408.

Tiku, M. L.& Singh, M. (1982). Robust statistics for testing mean vectors of multivariate distribution. *Communication in statistics - theory and methods, 11*(9), 985-1001.

Thompson, J.R. and Koronacki, J. (1993). *Statistical Process Control for Quality Improvement*. Chapman & Hall, 391 pgs.

Tracy, N. D., Young, J. C.& Mason, R. L. (1992). Multivariate control charts for individual observations. *Quality Technology, 24*, 88-95.

Vargas, J. A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observation. *Quality Technology, 35*, 367 - 376.

Wei, G. C. (2007). *Comparison Between a Robust Control Chart Based on Hodges Lehmann Estimator and Classical Hotelling T square Chart*. Thesis of Master of Applied Statistics. Department of Mathematics, Institute of Mathematical Research, University Putra Malaysia.

Wilcox, R. (2003). Multiple Comparisons Among Dependent Groups Based on a Modified One -step M-estimator. *Biometrical, 44*, 466-477.

Wilcox, R., & Keselman, H. J. (2003). Repeated Measures ANOVA Based on a Modified One-Step M-Estimator. *British Mathematical and Statistical Psychology, 56(1),* 15 - 25.

Wilcox, R.R. & Keselman, H.J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods 8 (3)*: 254–274.

Wilcox, R. R. (1995). ANOVA: A paradigm for low power and misleading measures of effect size? *Review of Educational Research, 65(1),* 51-77.

Wilcox, R. R. (1997). Introduction to Robust Estimation and Hypothesis Testing. *Academic Press*.

Williams, J. D., Sullivan, J. H. & Birch, J. B. (2006). Maximum Value of Hotelling's T^2 Statistics Based on the Successive Differences Covariance Matrix Estimator. *Communication in statistics-theory and Method. 38(4),* 471 - 483.

Williams, J. D. Woodall, W. H., Birch, J. B. & Sullivan, J. H. (2006). Distribution of Hotelling's T^2 Statistics Based on the Successive Differences Estimator. *American Statistical Association, 38(3), 217 - 229*

Woodall, W. H., & Montgomery, D. C. (1999). Research Issues and Ideas in Statistical Process Control. *Quality Technology. 31*, 376-386.

Syed Yahaya, S. S., Ali, H. & Omar, Z. (2011). Alternative Hotelling T^2 Control Charts Based on Minimum Vector Variance (MVV). *Modern Applied Science*. 5(4), ISSN 1913-1844 (print) ISSN 1913-1852 (online).

Syed Yahaya, S. S., Othman, A. R., & Keselman, H. J. (2006). Comparing the Typical Score Across Independent Groups Based on Different Criteria for Trimming. *Metodoloski zvezki, 3*(1), 49-62.

Syed Yahaya, S. S., Othman, A. R., &Keseman, H. J. (2004). Testing the equality of location parameters for skewed distributions using S_1 with high breakdown robust scale estimators. In M. Hubert, G. Pison, A. Struyf and S. Van Aelst (Eds.), *Theory and Applications of Recent Robust Methods, Series: Statistics for Industry and Technology* (319-328). Basel, Switzerland: Birkhauser.

Appendix A

Programs Calculate The Modified Hotelling’s T-Square Charts

1) Program calculates the false alarms and the probability of detection for the traditional Hotelling's T-square control charts without clean data

```
clear all;    % to clear the previous calculations
R=5000;       % number of iterations
R1=1000;
N=150;        %Total number of rows
P=2;          % Number of characteristics variables
pi=0.2;
MeanData1=zeros(R1,P);
meanminusmean=zeros(R,P);
%The values of shifted means
m=[5 5];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.0;
end
end
end
COVARIC=zeros(P,P,R1);
COVARIC1=zeros(P,P,R);
ROUND=floor(N*pi);
T4=zeros(R,1); %location for T-square
```

```
% PhaseI
for r1=1:R1        % loop for the iterations R
    seed = 95395+r1; % Seed number to fix the data
    rand('seed',seed)
    randn('seed',seed)
    Z=randn(N,P);   % Generate the data from standard normal distribution
% Put the outlier in the data

Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;
    MeanData1(r1,:)=mean(Data11);  % Calculate the mean
    COVARIC(:,:,r1)=cov(Data11);   % Calculate the covariance matrix
end
%Phase II
T91=zeros(R1,1);
T92=zeros(R1,1);
for r1=1:R1
    seed = 15391+r1;
    rand('seed',seed);
    randn('seed',seed);
    Z1=randn(1,P)*sigma0;   % Generate the observation from standard normal
distribution
    Z2=Z1+m;                % Generate the observation from out of control distribution
% Calculate the T-square in ordere to calculate the false alarms and
% probabilit of detection outliers, respectively
    MEANminusmean1=Z1-MeanData1(r1,:);
    MEANminusmean2=Z2-MeanData1(r1,:);
    T91(r1)= MEANminusmean1/COVARIC(:,:,r1)*(MEANminusmean1)';
    T92(r1)= MEANminusmean2/COVARIC(:,:,r1)*(MEANminusmean2)';
end
```

```
UCL2=(P*(N+1)*(N-1))/(N^2-N*P)*finv(0.99,P,N-P);

Count1=0;
Count2=0;
for i=1:R1
if( T91(i)>UCL2)
     Count1=Count1+1;
end
if( T92(i)>UCL2)
     Count2=Count2+1;
end
end
 typeerror1=Count1/R1;   % Calculate the false alarm
 typeerror2=Count2/R1;   % Calculate the probability of detection outliers
```

2) Program calculates the false alarms and the probability of detection outliers using the historical data sets

```
function [typeerrocount3,typeerror2]=test
[Data11,MeanD,Cov2,UCL2] = mytest;
%ZR = [5 5] ;
ZR = [5 5 5 5 5] ; % shift mean to create outliers
%ZR = [0 0 0 0 0];
Co = 5 ;Ob = 1000;Ch = Co;
T91=zeros(Ob,1);
T92=zeros(Ob,1);
for count3=1:Ob
    seed = 15391+count3;
    rand('seed',seed);
    randn('seed',seed);
    Z1=randn(1,Ch);
```

```
    Z2=Z1+ZR;
    MEANminusmean1=Z1-MeanD(count3,:);
    MEANminusmean2=Z2-MeanD(count3,:);
    T91(count3)= MEANminusmean1/Cov2(:,:,count3)*(MEANminusmean1)';
    T92(count3)= MEANminusmean2/Cov2(:,:,count3)*(MEANminusmean2)';

end
Count1=0;
Count2=0;
for i=1:Ob
if( T91(i)>UCL2(i))
     Count1=Count1+1;
end
if( T92(i)>UCL2(i))
     Count2=Count2+1;
end
end
 typeerrocount3=Count1/Ob;
 typeerror2=Count2/Ob;
% This subroutine calculates the historical data sets

function  [Data11,Mean,Cov,UCL2] = mytest
Ob = 1000;
Ro = 50; % Ro # of rows
Co = 5 ; % Co #  of coloumn
Pr = 0.1 ; % persntage of outliers data to be add
%ZR = [ 5 5 ] ; % shft mean to creat outliers
ZR=[5 5 5 5 5];
%ZR=[0 0 0 0 0];
Round = floor(Ro * Pr); % # of outliers
Mean=zeros(Ob,Co);Cov=zeros(Co,Co,Ob);UCL2=zeros(Ob,1);
```

```
for Count = 1 : Ob % end of for loop will be changed && Start * loop
seed = 95395 + Count; % fix of our data
rand('seed',seed);randn('seed',seed); % randomization to fix the data
Z = randn(Ro,Co); % Z is a standard normal distrubution
Data11 = [ Z(1:Round,:) + repmat(ZR, Round ,1); Z(Round+1 : Ro,:)];% Main
% matrix affter data put outliers
MeanD1 = mean(Data11); % to callculate the mean
Coveriance = cov(Data11); % to callculate the coveriance
UCL = (((Ro-1)^2)/Ro)* betainv(0.95,Co/2,((Ro - Co - 1 ) /2) );
out = T4c(MeanD1,Coveriance,Data11,UCL);
[k ,w]= size(Data11);Ch = Co;
UCL2(Count) = ((Ch*(k+1)*(k-1))/(k*(k-Ch)))*finv(0.95,Ch,k-Ch);
Data11 = out;
Mean(Count,:) = MeanD1;
Cov(:,:,Count) = Coveriance;
end
end
function out = T4c(Mean,Cov,Data,UCL)
[ro,co] = size(Data);
k = 0 ;
for i = 1 : ro
   x  = Data(i,:) - Mean;
   y{i} = ( x/Cov(:,:)) * x';
if y{i} < UCL
       k = k + 1;
       DData(k,:) = Data(i,:);
end
end
out = DData;
end
```

3) Program calculates the false alarms and the probability of detection for the modified Hotelling T-square for winsorized *MOM* with *MADn*

```
clear all; % to clear the previous calculations
R=5000;      % number of iterations
R1=1000;
N=50;      % number of samples of size 1
P=10;       % Number of characteristics variables
pi=0.1;
m=[3 3 3 3 3 3 3 3 3 3];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.0;
end
end
end
MeanMADMatCov1=zeros(P,P,R1);
MeanMADMatCov=zeros(P,P,R);
MomwinsDatamad=zeros(R,P);
meanminusmean=zeros(R1,P);
MomwinsDatamad1=zeros(R1,P);
MOMPARA=0;
ROUND=floor(pi*N);% number of the outliers
T4=zeros(R,1);
for r=1:R
   seed = 3985+r;  % to fix all random data sets
```

```
    rand('seed',seed);% generate from uniform distribution
    randn('seed',seed);%generate from standdard normal distribution
    Z=randn(N+1,P);% generate data set from the standard normal distribution
% calculate the winsorized mean
for j=1:P
   MomwinsDatamad(r,j)=mean(WMADn_sample(Z(1:N,j)));
end
% calculate the MAD  covariance matrix
   MeanMADMatCov(:,:,r)=MadCov1(Z(1:N,:));
% calculate the hotelling t-square
   meanminusmean(r,:)= Z(N+1,:)-MomwinsDatamad(r,:);
   T4(r,1)=meanminusmean(r,:)/MeanMADMatCov(:,:,r) * (meanminusmean(r,:))';
end
UCL=prctile(T4,95); % upper control limit
%Phase I
for r1=1:R1
    seed = 95395+r1;
    rand('seed',seed);
    randn('seed',seed);
    Z=randn(N,P);  % generate matrix of row N and column P from standard normal
distribution
%  put number ROUND of the outliers in data

Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;
% calculate the mean for the winsorized sample
for j=1:P
   MomwinsDatamad1(r1,j)=mean(WMADn_sample(Data11(1:N,j)));  % calculate
the winsorized mean
end
% calculate the covariance for the winsorized sample
```

```
  MeanMADMatCov1(:,:,r1)=MadCov1(Data11(1:N,:));  % calculate covariance
matrix using the subroutine function MadCov1
end
%Phase II
T41=zeros(R1,1); % location for the values of the Hotelling t square
T412=zeros(R1,1);% location for the values of the Hotelling t square
for r1=1:R1
   seed = 15391+r1; % seed number to fix the data
   rand('seed',seed); %#ok<RAND>
   randn('seed',seed); %#ok<RAND>
   Z=randn(1,P)*sigma0;    % generate a new observation vector from standard
normal distribution
    Z1=Z+m;        %generate a new observation vector from  normal distribution
with shifted mean m
    momDataMinusMean=Z-MomwinsDatamad1(r1,:);
% calculate the Hotelling T square to calculate the type I error
   T41(r1)=
momDataMinusMean/MeanMADMatCov1(:,:,r1)*(momDataMinusMean)';
   momDataMinusMean1=Z1-MomwinsDatamad1(r1,:);
% calculate the Hotelling T square to calculate probability of
% detection outliers
   T412(r1)=
momDataMinusMean1/MeanMADMatCov1(:,:,r1)*(momDataMinusMean1)';
end
% check whether the values of T square in T41, T412 are greater than UCL
Count1=0;
Count2=0;
for i=1:R1
if( T41(i)>UCL)
      Count1=Count1+1;
end
```

```
if( T412(i)>UCL)
      Count2=Count2+1;
end
end
 typeerror1=Count1/R1;    % false alarms rates
 probdetection=Count2/R1; % percentage of probability of detection
```

4) Program calculates the false alarms and the probability of detection for the modified Hotelling T- square for winsorized MOM with Sn.

```
clear all; % to clear the previous calculations before the new run
R=5000;  % number of iterations
R1=1000;
N=100;      % number of samples of size 1
P=2;        % Number of characteristics variables
pi=0.2;
ROUND=floor(pi*N);% number of the outliers
m=[3 3];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
         sigma0(i,j)=1;
else
         sigma0(i,j)=0.0;
end
end
end
MeanSNMatCov1=zeros(P,P,R1);
MeanSNMatCov=zeros(P,P,R);
```

```
MomwinsDatasn=zeros(R,P);
meanminusmean=zeros(R1,P);
MomwinsDatasn1=zeros(R1,P);
T7=zeros(R,1);
for r=1:R       % loop for the iterations R
  seed = 3985+r;% to fix all random data sets
  rand('seed',seed)% generate from uniform distribution
  randn('seed',seed)%generate from standdard normal distribution
  Z=randn(N+1,P);% generate data set from the standard normal distribution
% calculate the winsorized mean
for j=1:P    % p denotes to the number of the random variables
     MomwinsDatasn(r,j)=mean(WSn_sample(Z(1:N,j)));
end
% calculate the Sn variance covariance matrix
  MeanSNMatCov(:,:,r)=SnCov1(Z(1:N,:));
% calculate the hotelling t square
  meanminusmean(r,:)= Z(N+1,:)-MomwinsDatasn(r,:);
  T7(r,1)=meanminusmean(r,:)/MeanSNMatCov(:,:,r) * (meanminusmean(r,:))';
end
 UCL=prctile(T7,95);
%Phase I
for r1=1:R1       % loop for the number of iterations
  seed = 95395+r1;
  rand('seed',seed)
  randn('seed',seed)
  Z=randn(N,P);% generate matrix of row N and column P from standard normal %
distribution
Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;
%  put number ROUND of the outliers in data
for j=1:P    % p denotes to the number of the random variables
```

```
    MomwinsDatasn1(r1,j)=mean(WSn_sample(Data11(1:N,j
end
% calculate covariance matrix using the subroutine function SnCov1
    MeanSNMatCov1(:,:,r1)=SnCov1(Data11(1:N,:));
end
T712=zeros(R1,1); % location for the values of the Hotelling t square T712
%Phase II
 T71=zeros(R1,1);% location for the values of the Hotelling t square T71
for r1=1:R1
   seed = 15391+r1;
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(1,P)*sigma0; % generate a new observation vector from
% standard normal distribution
   Z1=Z+m;      %generate a new observation vector from
% normal distribution with mean m
% calculate the Hotelling T square to calculate the type I error
   momDataMinusMean=Z-MomwinsDatasn1(r1,:);
   T71(r1)=
momDataMinusMean/MeanSNMatCov1(:,:,r1)*(momDataMinusMean)';
% calculate the Hotelling T square to calculate probability of
% detection outliers
   momDataMinusMean1=Z1-MomwinsDatasn1(r1,:);
   T712(r1)=
momDataMinusMean1/MeanSNMatCov1(:,:,r1)*(momDataMinusMean1)';
end
Count1=0;
Count2=0;
%check whether the values of T square in T71, T712 are greater than UCL
for i=1:R1
if( T71(i)>UCL)
```

```
        Count1=Count1+1;
end
if( T712(i)>UCL)
        Count2=Count2+1;
end
end
  typeerror1=Count1/R1;       % false alarms rates
 probdetetion=Count2/R1;      % percentage of probability of detection
```

5) Program calculates the false alarms and the probability of detection outliers for the modified Hotelling T square for winsorized MOM with the scale estimator Tn

```
clear all; % to clear the previous calculations
R=5000;      % number of iterations
R1=1000;     % number of iteration to calculate the Hotelling t-square
N=25;      % number of group samples
P=5;       % Number of characteristics variables
% The following vectors represent the values of mean vectors
m=[3 3 3 3 3];
pi=0.1;
ROUND=floor(pi*N);% number of the outliers
MeanTnMatCov1=zeros(P,P,R1);
MeanTnMatCov=zeros(P,P,R);
MomwinsDatatn=zeros(R,P);
MomwinsDatatn1=zeros(R1,P);
meanminusmean=zeros(R,P);
meanDatamean1=zeros(R1,P);
sigma0=zeros(P,P);
for i=1:P
```

```
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.0;
end
end
end
T4=zeros(R,1);
for r=1:R
  seed = 3985+r;% to fix all random data sets
  rand('seed',seed)% generate from uniform distribution
  randn('seed',seed)%generate from standdard normal distribution
  Z=randn(N+1,P);% generate data set from the standard normal distribution
% calculate the winsorized MOM
for j=1:P
  MomwinsDatatn(r,j)=mean(WTn_sample(Z(1:N,j))); % calculate the winsorized %
mean
end
  MeanTnMatCov(:,:,r)=TnCov1(Z(1:N,:),N);% calculate the Tn  covariance matrix
% Calculate the Hotelling T-square
  meanminusmean(r,:)= Z(N+1,:)-MomwinsDatatn(r,:);
  T4(r)=meanminusmean(r,:)/MeanTnMatCov(:,:,r) * (meanminusmean(r,:))';
end
UCL=prctile(T4,95);
%Phase I
for r1=1:R1
  seed = 95395+r1;
  rand('seed',seed)
  randn('seed',seed)
```

```
  Z=randn(N,P);% generate matrix of row N and column P from standard normal %
distribution
% put number ROUND of the outliers in data

Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;
% Calculate the winsorized MOM
for j=1:P
  MomwinsDatatn1(r1,j)=mean(WTn_sample(Data11(1:N,j)));% calculate the
%winsorized mean
end
  MeanTnMatCov1(:,:,r1)=TnCov1(Data11(1:N,:),N);% calculate covariance
%matrix using the subroutine function TnCov1
end
%Phase II
T412=zeros(R1,1);% location for the values of the Hotelling t square T412
T41=zeros(R1,1);% location for the values of the Hotelling t square T41
for r1=1:R1
   seed = 15391+r1;
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(1,P)*sigma0; % generate a new observation vector from standard
%normal distribution
    Z1=Z+m;     %generate a new observation vector from  normal %distribution
with mean m
% calculate the Hotelling T square to calculate the type I error
    momDataMinusMean=Z-MomwinsDatatn1(r1,:);
   T41(r1)= momDataMinusMean/MeanTnMatCov1(:,:,r1)*(momDataMinusMean)';
% calculate the Hotelling T square to calculate probability of
% detection outliers
   momDataMinusMean1=Z1-MomwinsDatatn1(r1,:);
```

```
   T412(r1)=
momDataMinusMean1/MeanTnMatCov1(:,:,r1)*(momDataMinusMean1)';
end
Count1=0;
Count2=0;
%check whether the values of T square in T41, T412 are greater than UCL
for i=1:R1
if( T41(i)>UCL)
     Count1=Count1+1;
end
if( T412(i)>UCL)
     Count2=Count2+1;
end
end
  typeerror1=Count1/R1;       % false alarms rates
 probdetection=Count2/R1;      % percentage of probability of detection
```

6) Program calculates the false alarms and the probability of detection outliers for the modified Hotelling T square for Hodges-Lehmann with MADn

```
clear all;% to clear the previous calculations before the run
R=5000;       % number of iterations
R1=1000;
N=150;       % number of samples of size 1
P=10;        % Number of characteristics variables
m=[0 0 0 0 0 0 0 0 0 0 ];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
```

```
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.0;
end
end
end
pi=0.0;

Data1=zeros(N,P);
Datatn=zeros(N,P);
Mattn=zeros(N,P);
HOGSLMANmadMatCov=zeros(P,P,R);
HOGSLMANmadMatCov1=zeros(P,P,R1);
HOGDESLEMANamad=zeros(R,P);
HOGLEMANminusmean=zeros(R,P);
HOGDESLEMANamad1=zeros(R1,P);
ROUND=floor(pi*N);% number of the outliers
T4=zeros(R,1);
for r=1:R
   seed = 3985+r; % to fix all random data sets
   rand('seed',seed);% generate from uniform distribution
   randn('seed',seed);%generate from standdard normal distribution
   Z=randn(N+1,P);   %generate data set from the standard
% normal distribution
for j=1:P
  HOGDESLEMANamad(r,j)=HL(Z(1:N,j));% calculate the Hodges-Lehmann
end
  HOGSLMANmadMatCov(:,:,r)=MadCov11(Z(1:N,:)); % calculate the MAD
% variance covariance matrix
% calculate the Hotelling t square
  HOGLEMANminusmean(r,:)= Z(N+1,:)-HOGDESLEMANamad(r,:);
```

```
T4(r)=HOGLEMANminusmean(r,:)/HOGSLMANmadMatCov(:,:,r)*(HOGLEMAN
minusmean(r,:))';
end
UCL=prctile(T4,95);
%Phase I
for r1=1:R
   seed = 95395+r1;
   rand('seed',seed)
   randn('seed',seed)
   Z=randn(N,P); % generate matrix of row N and column P from standard
% normal distribution
%  put number ROUND of the outliers in data

Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;   for j=1:P
   HOGDESLEMANamad(r1,j)=HL(Data11(1:N,j));% calculate the Hodges-
Lehmann
end
   HOGSLMANmadMatCov(:,:,r1)=MadCov11(Data11(1:N,:));% calculate the
MAD variance covariance matrix
end
%Phase II
T412=zeros(R1,1); % location for the values of the Hotelling t square T412
T41=zeros(R1,1);  % location for the values of the Hotelling t square T41
for r1=1:R1
    seed = 15391+r1;
    rand('seed',seed);
    randn('seed',seed);
    Z=randn(1,P)*sigma0;% generate a new observation vector from
% standard normal distribution
    Z1=Z+m;   %generate a new observation vector from  normal
```

```
% distribution with mean m
   HOGLEMANminusmean1=Z-HOGDESLEMANamad(r1,:);
% calculate the Hotelling T square to calculate the type I error
   T41(r1)= HOGLEMANminusmean1
/HOGSLMANmadMatCov(:,:,r1)*(HOGLEMANminusmean1)';
   HOGLEMANminusmean2=Z1-HOGDESLEMANamad(r1,:);
% calculate the Hotelling T square to calculate probability of detection
   T412(r1)= HOGLEMANminusmean2
/HOGSLMANmadMatCov(:,:,r1)*(HOGLEMANminusmean2)';
end
Count1=0;
Count2=0;
% check whether the values of T square in T41, T412 are greater than UCL
for i=1:R1
if( T41(i)>UCL)
        Count1=Count1+1;
end
if( T412(i)>UCL)
        Count2=Count2+1;
end
end
  typeerror1=Count1/R1;       % false alarms rates
 probdetection=Count2/R1;     % percentage of probability of detection
```

7) Program calculates the false alarms and probability of detection outliers for the Hotelling T- square of Hodges-Lehman with scale estimator *Sn*.

```
clear all; % to clear the previous calculations
R=5000;  % number of iterations
R1=1000; % number of iterations
```

```
N=150;  % number of samples of size 1
P=10;        % Number of characteristics variables
pi=0.0;
m=[0 0 0 0 0 0 0 0 0 0];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
          sigma0(i,j)=1;
else
          sigma0(i,j)=0.0;
end
end
end
HOGLEMANSNMatCov1=zeros(P,P,R1);
HOGLEMANSNMatCov=zeros(P,P,R);
HOGDESLEMANsn=zeros(R,P);
HOGDESLEMANsn1=zeros(R1,P);
HOGLEMANminusmean=zeros(R,P);
ROUND=floor(pi*N); % number of the outliers
T8=zeros(R,1);
for r=1:R       % loop for the iterations R
     seed = 3985+r; % to fix all random data sets
     rand('seed',seed)% generate from uniform distribution
     randn('seed',seed)%generate from standdard normal distribution
     Z=randn(N+1,P); % generate data set from the
 % standard normal distribution
for j=1:P
       HOGDESLEMANsn(r,j)=HL(Z(1:N,j));% calculate the Hodges-Lehmann
end
```

```
   HOGLEMANSNMatCov(:,:,r)=SnCov11(Z(1:N,:));% calculate the Sn
% covariance matrix
% calculate the Hotelling t square
   HOGLEMANminusmean(r,:)= Z(N+1,:)-HOGDESLEMANsn(r,:);

T8(r,1)=HOGLEMANminusmean(r,:)/HOGLEMANSNMatCov(:,:,r)*(HOGLEMA
Nminusmean(r,:))';
end
UCL=prctile(T8,99);
%Phase I
for r1=1:R1       % loop for the iterations R
     seed = 95395+r1;
     rand('seed',seed)
     randn('seed',seed)
     Z=randn(N,P); % generate matrix of row N and column P from
% standard normal distribution

%  put number ROUND of the outliers in data

Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;
% calculate the Hodges and lehmann
for j=1:P
      HOGDESLEMANsn1(r1,j)=HL(Data11(1:N,j));% calculate the
% Hodges-Lehmann for Data11
end
% calculate the covariance matrix for Sn
   HOGLEMANSNMatCov1(:,:,r1)=SnCov11(Data11(1:N,:));% calculate
% the Sn variance covariance matrix for Data11
end
T812=zeros(R1,1);% location for the values of the Hotelling t square T812
```

```
%Phase II
T81=zeros(R1,1);% location for the values of the Hotelling t square T81
for r1=1:R1
    seed = 15391+r1;  % fix the data
    rand('seed',seed);
    randn('seed',seed);
    Z=randn(1,P)*sigma0; % generate a new observation vector from
% standard normal distribution
    Z1=Z+m;   %generate a new observation vector from
% normal distribution with mean m
% calculate the Hotelling T square to calculate the type I error
    HOGLEMANminusmean1=Z-HOGDESLEMANsn1(r1,:);
    T81(r1)=
HOGLEMANminusmean1/HOGLEMANSNMatCov1(:,:,r1)*(HOGLEMANminus
mean1)';
% calculate the Hotelling T square to calculate probability of detection
   HOGLEMANminusmean2=Z1-HOGDESLEMANsn1(r1,:);
   T812(r1)=
HOGLEMANminusmean2/HOGLEMANSNMatCov1(:,:,r1)*(HOGLEMANminus
mean2)';
end
Count1=0;
Count2=0;
% check whether the values of T square in T81, T812 are greater than UCL
for i=1:R1
if( T81(i)>UCL)
        Count1=Count1+1;
end
if( T812(i)>UCL)
        Count2=Count2+1;
end
```

```
end
 typeerror1=Count1/R1; % percentage of false alarms
 probdetection=Count2/R1;% percentage of probability of detection
```

8) Program calculates the false alarms and probability of detection outliers for the Hotelling T square of Hodges-Lehman and scale estimator *Tn*.

```
clear all; % to clear the previous calculations before the run
R=5000;       % number of iterations to calculate UCL
R1=1000;      % number of iteration of Hotelling T-square to calculate
%Type I error and probability of detection
N=150;    % number of samples of size 1
P=2;        % Number of characteristics variables
%m=[3 3];
%m=[5 5];
m=[0 0];
pi=0.0;
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.9;
end
end
end
Mattn=zeros(N,P);
HOGLEMANTnMatCov1=zeros(P,P,R1);
HOGLEMANTnMatCov=zeros(P,P,R);
```

```
HOGDESLEMANTN=zeros(R,P);
HOGDESLEMANTN1=zeros(R,P);
HOGLEMANminusmean=zeros(R,P);
MEDIANDatatn1=zeros(R1,P);
ROUND=floor(pi*N);% number of the outliers data
T5=zeros(R,1);
for r=1:R
  seed = 3985+r; % to fix all random data sets
  rand('seed',seed)% generate from uniform distribution
  randn('seed',seed)%generate from standdard normal distribution
  Z=randn(N+1,P); % generate data set from the standard normal distribution
for j=1:P
   HOGDESLEMANTN(r,j)=HL(Z(1:N,j));% calculate the Hodges-Lehmann
end
   HOGLEMANTnMatCov(:,:,r)=TnCov11(Z(1:N,:),N);% calculate the Tn
% covariance matrix
% calculate the Hotelling t square
   HOGLEMANminusmean(r,:)= Z(N+1,:)-HOGDESLEMANTN(r,:);

   T5(r,1)=HOGLEMANminusmean(r,:)/HOGLEMANTnMatCov(:,:,r) *
(HOGLEMANminusmean(r,:))';
end
UCL=prctile(T5,99);
%Phase I
for r1=1:R
  seed = 95395+r1;
  rand('seed',seed)
  randn('seed',seed)
  Z=randn(N,P); % generate matrix of row N and column P from
% standard normal distribution
% put number ROUND of the outliers in data
```

```
Data11=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0]
;        for j=1:P
       HOGDESLEMANTN1(r1,j)=HL(Data11(1:N,j));% calculate the
% Hodges-Lehmann for Data11
end
  HOGLEMANTnMatCov1(:,:,r1)=TnCov11(Data11(1:N,:),N);% calculate
% the Tn covariance matrix for Data11
end
T51=zeros(R1,1);% location for the values of the Hotelling t square T51
T512=zeros(R1,1);% location for the values of the Hotelling t square T512
for r1=1:R1
   seed = 15391+r1;
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(1,P)*sigma0; % generate a new observation vector from
% standard normal distribution
   Z1=Z+m;%generate a new observation vector from  normal distribution
% with mean m
% calculate the Hotelling T square to calculate the false alarms
% calculate the Hotelling T square to calculate probability of detection
   HOGLEMANminusmean1=Z-HOGDESLEMANTN1(r1,:);
   T51(r1)=
HOGLEMANminusmean1/HOGLEMANTnMatCov1(:,:,r1)*(HOGLEMANminusm
ean1)';
   HOGLEMANminusmean2=Z1-HOGDESLEMANTN1(r1,:);
   T512(r1)=
HOGLEMANminusmean2/HOGLEMANTnMatCov1(:,:,r1)*(HOGLEMANminusm
ean2)';
end
% end
```

```
Count1=0;
Count2=0;
% check whether the values of T square in T51, T512 are greater than UCL
for i=1:R1
if( T51(i)>UCL)
        Count1=Count1+1;
end
if( T512(i)>UCL)
        Count2=Count2+1;
end
end
 typeerror1=Count1/R1;      % false alarms rates
 probdetection=Count2/R1;   % percentage of probability of detection
```

9) This program calculates the false alarms and the probability of detectionoutliers for the Hotelling T square of trimmed mean with the trimmed covariance matrix.

```
clear all;      % to clear the previous calculations before the run
R=5000;         % number of iterations
R1=1000;
N=150;          %Total number of rows
P=2;            % Number of characteristics variables
pi=0.0;         % percentage of outliers
per=0.40;       % percentage of the trimming the outliers
K=floor(N*per); % the number of outliers points

TRIMmeanminusmean=zeros(1,P);
MEDIANDatamad=zeros(1,P);
TRIMVARIANC=zeros(P,P,R);
sigma0=zeros(P,P);
```

```
MEDIANmadMatCov1=zeros(P,P);
MEDIANmadMatCov=zeros(P,P);
MEDIANminusmean1=zeros(1,P);
MEDIANDatamad1=zeros(R1,P);
Delta=zeros(N,1);
TRIMVARIANC1=zeros(P,P,R1);
TRIMSTANDARD1=zeros(P,P,R1);
TRIMmeanminusmean1=zeros(1,P);
TRIMmeanminusmean2=zeros(1,P);
MEANTRIM=zeros(R,P);
MEANTRIM1=zeros(R1,P);
ROUND=floor(pi*N);
m=[0 0];
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.9;
end
end
end
T4=zeros(R,1); % location for the values of Hotelling T-square
for r=1:R
  seed = 3985+r;    % seed number to fix the data
  rand('seed',seed)
  randn('seed',seed)
  Z2=randn(N+1,P);   % generate the data from standard normal
  MEDIANDatamad=median(Z2(1:N,:));   % Calculate the median for the data set
  Delta=zeros(N,1);
  MEDIANmadMatCov=MadCov1(Z2(1:N,:));  % calculate the covariance matrix
```

```
% for the MADn
   MEDIANminusmean=zeros(N,P);
% Calculate the Mahalanobis distance for each row of the data matrix
for j=1:N
    MEDIANminusmean= Z2(j,:)-MEDIANDatamad;
    Delta(j)=MEDIANminusmean/MEDIANmadMatCov*(MEDIANminusmean)';
end
RANKDELTA=sort(Delta);    % sort the values of Mahalonobis distance
MAX=zeros(K,1);
MIN=zeros(K,1);
for i=1:K
for j=1:N
if (Delta(j)==RANKDELTA(N-i+1))
        MAX(i)=j;
end
if (Delta(j)==RANKDELTA(N-K-i+1))
        MIN(i)=j;
end
end
   Z2(MAX(i),:)=Z2(MIN(i),:);
end
% Calculate the Hotelling T-square
MEANTRIM(r,:)=(sum(Z2(1:N,:))-sum(Z2(MAX(1:K),:)))/(N-K);
TRIMVARIANC(:,:,r)=(N-1)/(N-K-1)*cov((Z2(1:N,:)));
TRIMmeanminusmean(r,:)= Z2(N+1,:)-MEANTRIM(r,:);
T4(r,1)=TRIMmeanminusmean(r,:)/TRIMVARIANC(:,:,r)*(TRIMmeanminusmean(
r,:))';
end
%  Calculate the upper control limit
UCL=prctile(T4,99);
% Phase I
```

```
T41=zeros(R1,1);
T42=zeros(R1,1);
for r1=1:R1
    seed = 95395+r1;   % seed number to fix the data
    rand('seed',seed);
    randn('seed',seed);
    Z=randn(N,P);     % Generate the data from the standard normal
% Put the outlier in the data

Data1=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0];
    MEDIANDatamad1=median(Data1(1:N,:));  % Calcualte the median
    Delta1=zeros(N,1);
    MEDIANmadMatCov1=MadCov1(Data1(1:N,:));  % Calculate the
% covariance matrix of MADn
for j=1:N
    MEDIANminusmean1(j,:)= Data1(j,:)-MEDIANDatamad1;

Delta1(j)=MEDIANminusmean1(j,:)/MEDIANmadMatCov1*(MEDIANminusmean
1(j,:))';
end
%  Calculate the Mahalanobis distance
  RANKDELTA1=sort(Delta1);
  MAX1=zeros(K,1);
  MIN1=zeros(K,1);
for i=1:K
for j=1:N
if (Delta1(j)==RANKDELTA1(N-i+1))
          MAX1(i)=j;
end
if (Delta1(j)==RANKDELTA1(N-K-i+1))
          MIN1(i)=j;
```

```
end
end
Data1(MAX1(i),:)=Data1(MIN1(i),:);
end

% Calculate the trimmed mean
MEANTRIM1(r1,:)=(sum(Data1(1:N,:))-sum(Data1(MAX1(1:K),:)))/(N-K);
TRIMVARIANC1(:,:,r1)=(N-1)/(N-K-1)*cov((Data1(1:N,:)));  % Calculate the
% trimmed covariance matrix
end
% Phase II
%  Calculate the T-square for the vector of the observation
for r1=1:R1
   seed = 15391+r1;
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(1,P)*sigma0;
   Z1=Z+m;
   TRIMmeanminusmean1(r1,:)= Z-MEANTRIM1(r1,:);
   T41(r1,1)=TRIMmeanminusmean1(r1,:)/TRIMVARIANC1(:,:,r1)*
(TRIMmeanminusmean1(r1,:))';
   TRIMmeanminusmean2(r1,:)= Z1-MEANTRIM1(r1,:);
T42(r1,1)=TRIMmeanminusmean2(r1,:)/TRIMVARIANC1(:,:,r1)*
(TRIMmeanminusmean2(r1,:))';
end
Count1=0;
Count2=0;
for i=1:R1
if( T41(i)>UCL)
     Count1=Count1+1;
end
```

```
if( T42(i)>UCL)
      Count2=Count2+1;
end
end
 typeerror1=Count1/R1;  % Calculate false alarms rates
 probdetec=Count2/R1;   % Calculate the probability of detection
```

10) This program calculates the false alarms and the probability of detection outliers for the Hotelling T square of trimmed mean with the trimmed covariance matrix using the scale estimator *Sn.*

```
clear all; % to clear the previous calculations before the run
R=5000;       % number of iterations
R1=1000;
N=150;      %Total number of rows
P=2;        % Number of characteristics variables
pi=0.0;         % percentage of outliers
per=0.40;       % percentage of the trimming the outliers
K=floor(N*per); % the number of outliers points
TRIMmeanminusmean=zeros(1,P);
MEDIANDatasn=zeros(1,P);
TRIMVARIANC=zeros(P,P);
MEDIANsnMatCov1=zeros(P,P);
MEDIANsnMatCov=zeros(P,P);
MEDIANminusmean1=zeros(1,P);
MEDIANDatasn1=zeros(1,P);
Delta=zeros(N,1);
TRIMVARIANC1=zeros(P,P,R1);
TRIMSTANDARD1=zeros(P,P,R1);
TRIMmeanminusmean1=zeros(1,P);
TRIMmeanminusmean2=zeros(1,P);
MEANTRIM=zeros(R,P);
```

```
MEANTRIM1=zeros(R1,P);
ROUND=floor(pi*N);
m=[0 0];
sigma0=zeros(P,P);
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
        sigma0(i,j)=0.9;
end
end
end
T4=zeros(R,1);% location for the values of Hotelling T-square
for r=1:R
  seed = 3985+r;  % seed number to fix the data
  rand('seed',seed)
  randn('seed',seed)
  Z2=randn(N+1,P);  % generate the data from standard normal
  MEDIANDatasn=median(Z2(1:N,:));% Calculate the median for the data set
  Delta=zeros(N,1);
  MEDIANsnMatCov=SnCov11(Z2(1:N,:));% calculate the covariance
% matrix for the Sn
  MEDIANminusmean=zeros(N,P);
% Calculate the Mahalanobis distance for each row of the data matrix
for j=1:N
   MEDIANminusmean= Z2(j,:)-MEDIANDatasn;
   Delta(j)=MEDIANminusmean/MEDIANsnMatCov*(MEDIANminusmean)';
end
RANKDELTA=sort(Delta);  % sort the values of Mahalonobis distance
```

```
MAX=zeros(K,1);
MIN=zeros(K,1);
for i=1:K
for j=1:N
if (Delta(j)==RANKDELTA(N-i+1))
        MAX(i)=j;
end
if (Delta(j)==RANKDELTA(N-K-i+1))
        MIN(i)=j;
end
end
   Z2(MAX(i),:)=Z2(MIN(i),:);
end
MEANTRIM(r,:)=(sum(Z2(1:N,:))-sum(Z2(MAX(1:K),:)))/(N-K);
TRIMVARIANC(:,:,r)=(N-1)/(N-K-1)*cov((Z2(1:N,:)));
TRIMmeanminusmean(r,:)= Z2(N+1,:)-MEANTRIM(r,:);
T4(r,1)=TRIMmeanminusmean(r,:)/TRIMVARIANC(:,:,r)*(TRIMmeanminusmean(
r,:))';
end
UCL=prctile(T4,95);    %  Calculate the upper control limit
% Phase I
T42=zeros(R1,1);
T41=zeros(R1,1);
for r1=1:R1
   seed = 95395+r1; % seed number to fix the data
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(N,P); % Generate the data from the standard normal
% Put the outlier in the data

Data1=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0];
```

```
%Data1=[Z(1:ROUND,:)+repmat(m,ROUND,1);Z(ROUND+1:N,:)];
    MEDIANDatasn1=median(Data1);
    Delta1=zeros(N,1);
    MEDIANsnMatCov1=SnCov11(Data1);
%  Calculate the Mahalanobis distance
for j=1:N
    MEDIANminusmean1(j,:)= Data1(j,:)-MEDIANDatasn1;

Delta1(j)=MEDIANminusmean1(j,:)/MEDIANsnMatCov1*(MEDIANminusmean1(
j,:))';
end
  RANKDELTA1=sort(Delta1);
  MAX1=zeros(K,1);
  MIN1=zeros(K,1);
for i=1:K
for j=1:N
if (Delta1(j)==RANKDELTA1(N-i+1))
          MAX1(i)=j;
end
if (Delta1(j)==RANKDELTA1(N-K-i+1))
          MIN1(i)=j;
end
end
    Data1(MAX1(i),:)=Data1(MIN1(i),:);
end
% Calculate the trimmed mean
MEANTRIM1(r1,:)=(sum(Data1(1:N,:))-sum(Data1(MAX1(1:K),:)))/(N-K);
TRIMVARIANC1(:,:,r1)=(N-1)/(N-K-1)*cov((Data1(1:N,:)));% Calculate the
% trimmed covariance matrix
end
% Phase II
```

```
% Calculate the T-square for the vector of the observation
for r1=1:R1
    seed = 15391+r1;
    rand('seed',seed);
    randn('seed',seed); %#ok<*RAND>
    Z=randn(1,P)*sigma0;
    Z1=Z+m;
    TRIMmeanminusmean1(r1,:)= Z-MEANTRIM1(r1,:);
T41(r1,1)=TRIMmeanminusmean1(r1,:)/TRIMVARIANC1(:,:,r1)*(TRIMmeanmin
usmean1(r1,:))';
    TRIMmeanminusmean2(r1,:)= Z1-MEANTRIM1(r1,:);
T42(r1,1)=TRIMmeanminusmean2(r1,:)/TRIMVARIANC1(:,:,r1)*(TRIMmeanmin
usmean2(r1,:))';
end
Count1=0;
Count2=0;
for i=1:R1
if( T41(i)>UCL)
      Count1=Count1+1;
end
if( T42(i)>UCL)
      Count2=Count2+1;
end
end
typeerror1=Count1/R1;         % Calculate the false alarms rates
 probabdetection=Count2/R1;   % Calculate the probability of detection
```

11) Program calculates the false alarms and the probability of detection outliers for the Hotelling T square with trimmed mean and the trimmed covariance matrix using the scale estimator *Tn.*

```
clear all;      % to clear the previous calculations before the run
R=5000;         % number of iterations
R1=1000;
N=150;          %Total number of rows
P=2;            % Number of characteristics variables
pi=0.0;         % percentage of outliers
per=0.40;       % percentage of the trimming the outliers
K=floor(N*per); % the number of outliers points
TRIMmeanminusmean=zeros(1,P);
MEDIANDatatn=zeros(1,P);
TRIMVARIANC=zeros(P,P);
MEDIANtnMatCov1=zeros(P,P);
MEDIANtnMatCov=zeros(P,P);
MEDIANminusmean1=zeros(1,P);
MEDIANDatatn1=zeros(1,P);
Delta=zeros(N,1);
TRIMVARIANC1=zeros(P,P,R1);
TRIMSTANDARD1=zeros(P,P,R1);
TRIMmeanminusmean1=zeros(1,P);
TRIMmeanminusmean2=zeros(1,P);
MEANTRIM=zeros(R,P);
MEANTRIM1=zeros(R1,P);
ROUND=floor(pi*N);
sigma0=zeros(P,P);
m=[0 0];
for i=1:P
for j=1:P
if i==j
        sigma0(i,j)=1;
else
```

```
            sigma0(i,j)=0.9;
end
end
end
T4=zeros(R,1);    % location for the values of Hotelling T-square
for r=1:R
   seed = 3985+r;   % seed number to fix the data
   rand('seed',seed)
   randn('seed',seed)
   Z2=randn(N+1,P); % generate the data from standard normal

   MEDIANDatatn=median(Z2(1:N,:));% Calculate the median for the data set
   Delta=zeros(N,1);
   MEDIANtnMatCov=TnCov11(Z2(1:N,:));  % calculate the covariance matrix for
the Tn
   MEDIANminusmean=zeros(1,P);
% Calculate the Mahalanobis distance for each row of the data matrix
for j=1:N
    MEDIANminusmean= Z2(j,:)-MEDIANDatatn;
    Delta(j)=MEDIANminusmean/MEDIANtnMatCov*(MEDIANminusmean)';
end
RANKDELTA=sort(Delta);   % sort the values of Mahalanobis distance

MAX=zeros(K,1);
MIN=zeros(K,1);
for i=1:K
for j=1:N
if (Delta(j)==RANKDELTA(N-i+1))
          MAX(i)=j;
end
if (Delta(j)==RANKDELTA(N-K-i+1))
```

```
        MIN(i)=j;
end
end
%if (Delta(MAX(i))==Delta(MIN(i)));
    Z2(MAX(i),:)=Z2(MIN(i),:);
%end
end

MEANTRIM(r,:)=(sum(Z2(1:N,:))-sum(Z2(MAX(1:K),:)))/(N-K);
TRIMVARIANC(:,:,r)=(N-1)/(N-K-1)*cov((Z2(1:N,:)));
TRIMmeanminusmean(r,:)= Z2(N+1,:)-MEANTRIM(r,:);
T4(r,1)=TRIMmeanminusmean(r,:)/TRIMVARIANC(:,:,r)*(TRIMmeanminusmean(
r,:))';
end
UCL=prctile(T4,99);    %  Calculate the upper control limit
% Phase I
T41=zeros(R1,1);
T42=zeros(R1,1);
for r1=1:R1
   seed = 95395+r1;   % seed number to fix the data
   rand('seed',seed);
   randn('seed',seed);
   Z=randn(N,P); % Generate the data from the standard normal
% Put the outlier in the data

Data1=[Z(1:ROUND,:)*sigma0+repmat(m,ROUND,1);Z(ROUND+1:N,:)*sigma0];
%Data1=[Z(1:ROUND,:)+repmat(m,ROUND,1);Z(ROUND+1:N,:)];
   MEDIANDatatn1=median(Data1(1:N,:));
   Delta1=zeros(N,1);
   MEDIANtnMatCov1=Tncov11(Data1(1:N,:));
%  Calculate the Mahalanobis distance
```

```
for j=1:N
    MEDIANminusmean1(j,:)= Data1(j,:)-MEDIANDatatn1;

Delta1(j)=MEDIANminusmean1(j,:)/MEDIANtnMatCov1*(MEDIANminusmean1(j
,:))';
end
  RANKDELTA1=sort(Delta1);
  MAX1=zeros(K,1);
  MIN1=zeros(K,1);
for i=1:K
for j=1:N
if (Delta1(j)==RANKDELTA1(N-i+1))
          MAX1(i)=j;
end
if (Delta1(j)==RANKDELTA1(N-K-i+1))
          MIN1(i)=j;
end
end
       Data1(MAX1(i),:)=Data1(MIN1(i),:);
end

MEANTRIM1(r1,:)=(sum(Data1(1:N,:))-sum(Data1(MAX1(1:K),:)))/(N-K);% the
trimmed mean
TRIMVARIANC1(:,:,r1)=(N-1)/(N-K-1)*cov((Data1(1:N,:)));% Calculate the
% trimmed covariance matrix
end
% Phase II
%  Calculate the T-square for the vector of the new observation
for r1=1:R1
    seed = 15391+r1;
    rand('seed',seed);
```

```
    randn('seed',seed);
    Z=randn(1,P)*sigma0;
    Z1=Z+m;
    TRIMmeanminusmean1(r1,:)= Z-MEANTRIM1(r1,:);
T41(r1,1)=TRIMmeanminusmean1(r1,:)/TRIMVARIANC1(:,:,r1)*(TRIMmeanminusmean1(r1,:))';
TRIMmeanminusmean2(r1,:)= Z1-MEANTRIM1(r1,:);
T42(r1,1)=TRIMmeanminusmean2(r1,:)/TRIMVARIANC1(:,:,r1)*(TRIMmeanminusmean2(r1,:))';
end
Count1=0;
Count2=0;
for i=1:R1
if( T41(i)>UCL)
     Count1=Count1+1;
end
if( T42(i)>UCL)
     Count2=Count2+1;
end
end
typeerror1=Count1/R1;    % Calculate the false alarms
probdetection=Count2/R1;  % Calculate the probability of detection
```

Appendix B
Programs for The Winsorized Sample

1) Winsorized sample by using the criteria of one step M-estimator(*MOM*)

```
function Result=WMADn_sample(Y)
% This funcion compute MoM for data in matrix Y
% Y is suppose to be a data vector comes from main programm
[S1 S2]=size(Y);        % size of Y
if S2>1
   disp('error Only vectors not coulumns or Matrices');
return;
end
% the following part to trim the extreme values from two sides in the vectors

Med=median(Y);          % Median of the vector Y
Mad= MADn(Y);           % the law of MADnj=MADj/0.6745 for each subgroup
Low=-2.24*Mad;          % the criteria to trimm the lower extreme values in each
subgroup
High=2.24*Mad;          % the criteria to trimm the upper extreme values in each
subgroup
k=0;
for i=1:S1,
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
       k= k+1;
end
end
X = zeros(k,S2);
k=1;
for i=1:S1,
```

```
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
        X(k) = Y(i);
        k= k+1;
end
end
Max = max(X);
Min = min(X);
for i=1:S1,
if ((Y(i) - Med) < Low)
    Y(i) = Min;
end
if ((Y(i) - Med) > High)
    Y(i) = Max;
end
end
Result=Y; % the winsorized sample
```

2) Program constructs the Winsorized samples using the criteria of the *MOM* estimator with the scale estimator *Sn*

```
function Result=WSn_sample(Y)
% This funcion compute MoM for data in matrix Y
% Y is suppose to be a data vector comes from start programm
[S1 S2]=size(Y);        % size of Y
if S2>1
    disp('error Only vectors not coulumns or Matrices');
return;
end
% The following part of the program is the criteria that is used to trim the
% extreme values from two sides of data
```

```
Med=median(Y);          % Median of the vector Y
SN= Sn1(Y)/(1.1926*0.6745);    % calculate the value of Sn using Sn subroutine
Low=-2.24*SN;      % the criteria to trim the lower extreme values
High=2.24*SN;      % the criteria to trim the upper extreme values
k=0;
for i=1:S1,
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
        k= k+1;
end
end
X = zeros(k,S2);
k=1;
for i=1:S1,
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
        X(k) = Y(i);
        k= k+1;
end
end
Max = max(X);
Min = min(X);
for i=1:S1,
if ((Y(i) - Med) < Low)
    Y(i) = Min;
end
if ((Y(i) - Med) > High)
    Y(i) = Max;
end
end
Result=Y; % Winsorized sample
```

3) This program creates the Winsorized samples using the scale estimator *Tn*

```
function Result=WTn_sample(Y)
% This function compute MoM for data in matrix Y
% Y is suppose to be a data vector comes from start program
[S1 S2]=size(Y);        % size of Y
if S2>1
    disp('error Only vectors not columns or Matrices');
return;
end
% the following part is the programming of the criteria to trim the
% extreme values from two sides in the vectors in each data
Med=median(Y);          % Median of the vector Y
TN= Tn(Y)/(1.38*0.6745);     % calculate the value of Tn
Low=-2.24*TN;
High=2.24*TN;   % the criteria to trim the upper extreme values in each data
k=0;
for i=1:S1,
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
        k= k+1;
end
end
X = zeros(k,S2);
k=1;
for i=1:S1,
if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
        X(k) = Y(i);
        k= k+1;
end
```

```
end
Max = max(X);
Min = min(X);
for i=1:S1,
if ((Y(i) - Med) < Low)
    Y(i) = Min;
end
if ((Y(i) - Med) > High)
    Y(i) = Max;
end
end
Result=Y; % winsorized sample
```

Appendix C

Programs for The Location And Scale Estimators

1) Program calculates the value of *MADn*

```
function Result=MADn(X)
[s1 s2]=size(X);          % SIZE OF X
  Median=median(X);      % median of X
  Result= 1.4826*median(abs(X-repmat(Median,s1,s2)));
```

2) The following part of the program calculates the value of the robust scale estimator *Sn*

```
function Result=Sn1(X)  % X is a vector comes from the main program
  [s1 s2]=size(X);       % size of X s1 rows and s2 columns
  Mediandist=zeros(s1,s2);
for k=1:s1% loop for each number in the vector X
      dist=zeros(s1,s2);
      Count=0;  % counter for the number of distances
for i=1:s1   % loop for the following number which comes after the k
if k~=i
         Count=Count+1; % take another distances
         dist(Count,s2)= abs(X(k,s2)-X(i,s2));    % Calculate the distances
end
end
      Mediandist(k,s2)= median(dist(1:Count,s2));    % Calculate the median
end
   Result = median(Mediandist(1:s1,s2))*1.1926;   %the value of Sn for this vector
```

3) The following part of the program calculates the value of scale estimator *Tn*

```
function Result=Tn(X)  % x is a vector which comes from the main program
[s1 s2]=size(X);    % size of the vector x
MedianDist=zeros(s1,s2);  % size of the matrix which contains
% median of the distances the observation from the other.
for k=1:s1        % loop for the rows
dist=zeros(s1-1,s2); % size of the distances between
% any two number
%with the other
Count=0;
for i=1:s1
if k~=i
 Count=Count+1;
 dist(Count,s2)= abs(X(k,s2)-X(i,s2));
end
end
% the median of the distances each number with the other
  MedianDist(k,s2)= median(dist(1:s1-1,s2));
end
% sort the values of median distances
Sortmedian=sort(MedianDist);
% calculate the value of h = [n/2]+1
h= floor(s1/2)+1;
Result = mean(Sortmedian(1:h,s2))*1.38; % the value of Tn
```

4) Program calculates Hodges-Lehmann estimator

```
function Result=HL(X)
   [s1 s2]=size(X);             % size x
   Hlm=zeros((s1+1)*s1/2,s2);    % Location
```

```
   Count=0;
for k=1:s1
for i=k:s1
   Count=Count+1;
  Hlm(Count,s2)= (X(k,s2)+X(i,s2))/2;
end
end
% arrange these distances in ascending way
  Result=median(Hlm(1:Count,s2)); % Hodges-Lehmann estimator
```

Appendix D

Programs to Calculate the Covariance Matrices for the Scale Estimators

1) This program calculate the *MADn* covariance matrix

```
function xy = MadCov1(Data1)
 [N P]=size(Data1);% the size of Datamad with N rows and p columns
E=zeros(P,P);
for j=1:P          % loop to get the p characteristics variables
for k=1:P % loop to get other p characteristics variables
    Colx=Data1(1:N,j);       % the vector of i-variable
   Coly=Data1(1:N,k);       % the vector of k-variable
 MADnwis1=WMADn_sample(Colx);
 MADnwis2=WMADn_sample(Coly);
 E(j,k)=(sum(MADnwis1.*MADnwis2)-
                 N*mean(MADnwis1)*mean(MADnwis2))/(N-1);
end
end
xy=E;
```

2)Program calculate covariance matrix of *Sn*

```
function xy = SnCov1(Data1)
    [N P]=size(Data1);   % the size of Datasn with N rows and p columns
E=zeros(P,P);
for j=1:P         % loop to get the p characteristics variables
for k=1:P     % loop to get other  p characteristics variables

          Colx=Data1(1:N,j);       % the vector of i-variable
          Coly=Data1(1:N,k);       % the vector of k-variable
```

```
        snwis1=WSn_sample(Colx);
        snwis2=WSn_sample(Coly);
        E(j,k)=(sum(snwis1.*snwis2)-N*mean(snwis1)*mean(snwis2))/(N-1);
end
end
xy=E; % covariance matrix of Sn
```

3)Thisprogram calculate covariance matrix for the scale estimator *Tn*.

```
function xy = TnCov1(Datatn,N)
%E(j,k,Count)=zeros(j,k,Count)
[N P]=size(Datatn);     % the size of Datatn N rows and p columns
E=zeros(P,P);
%Count6=1;              % Counter for the subgroup
%for i=1:G:N          % loop to get G subgroups from N rows
for j=1:P        % loop to get the p characteristics variables
for k=1:P     % loop to get other  p characteristics variables
        Colx=Datatn(1:N,j);       % the vector of j-variable
        Coly=Datatn(1:N,k);       % the vector of k-variable
        tnwis1=WTn_sample(Colx);
        tnwis2=WTn_sample(Coly);
        E(j,k)=(sum(tnwis1.*tnwis2)-N*mean(tnwis1)*mean(tnwis2))/(N-1);
end
end
xy=E; % covariance matrix
```

4) This program calculate covariance matrix for *MADn*

```
function xy = MadCov11(Data1)
[N P]=size(Data1);% the size of Datamad with N rows and p columns
E=zeros(P,P);
for j=1:P         % loop to get the p characteristics variables
for k=1:P % loop to get other  p characteristics variables
  Colx=Data1(1:N,j);       % the vector of i-variable
 Coly=Data1(1:N,k);       % the vector of k-variable
 rankx=tiedrank((Colx));
ranky=tiedrank((Coly));
difranks=sum((rankx-ranky).^2);
corrsp=1-(6*difranks)/(N*(N^2-1));
 E(j,k)=MADn(Colx)*MADn(Coly)*corrsp;
end
end
xy=E;   % covariance matrix for MADn
```

6) Program calculates the covariance matrix of scale estimator *Sn*

```
function xy = SnCov11(Datasn)
%E(j,k,Count5)=zeros(j,k,Count5);
[N P]=size(Datasn);   % the size of Datasn with N rows and p columns
E=zeros(P,P);

for j=1:P        % loop to get the p characteristics variables
for k=1:P     % loop to get other  p characteristics variables
        Colx=Datasn(1:N,j);       % the vector of i-variable
        Coly=Datasn(1:N,k);       % the vector of k-variable
        rankx=tiedrank((Colx));
```

```
ranky=tiedrank((Coly));
difranks=sum((rankx-ranky).^2);
corrsp=1-(6*difranks)/(N*(N^2-1));
E(j,k)=Sn1(Colx)*Sn1(Coly)*corrsp;
end
end
xy=E;    % Covariance matrix of Sn
```

7) Program calculates covariance matrix for *MADn*

```
function xy = MadCov1(Data1)
[N P]=size(Data1);% the size of Datamad with N rows and p columns
E=zeros(P,P);
for j=1:P          % loop to get the p characteristics variables
for k=1:P % loop to get other  p characteristics variables
Colx=Data1(1:N,j);        % the vector of i-variable
Coly=Data1(1:N,k);        % the vector of k-variable
rankx=tiedrank((Colx));
ranky=tiedrank((Coly));
difranks=sum((rankx-ranky).^2);
corrsp=1-(6*difranks)/(N*(N^2-1));
E(j,k)=MADn(Colx)*MADn(Coly)*corrsp;
end
end
xy=E;   % Calculate the covariance matrixfor MADn
```

8) This program calculates covariance matrix of Sn estimator

```
function xy = SnCov11(Datasn)
[N P]=size(Datasn);   % the size of Datasn with N rows and p columns
E=zeros(P,P);
for j=1:P        % loop to get the p characteristics variables
for k=1:P    % loop to get other  p characteristics variables
 Colx=Datasn(1:N,j);       % the vector of i-variable
Coly=Datasn(1:N,k);       % the vector of k-variable
 rankx=tiedrank((Colx));
 ranky=tiedrank((Coly));
difranks=sum((rankx-ranky).^2);
corrsp=1-(6*difranks)/(N*(N^2-1));
 E(j,k)=Sn1(Colx)*Sn1(Coly)*corrsp;
end
end
  xy=E;  % the covariance matrix of Sn
```

9) This program calculates covariance matrix of *Tn*

```
 function xy = TnCov11(Datatn,N)
 [N P]=size(Datatn);    % the size of Datatn N rows and p columns
  E=zeros(P,P);
 for j=1:P        % loop to get the p characteristics variables
for k=1:P    % loop to get other  p characteristics variables
Colx=Datatn(1:N,j);       % the vector of j-variable
 Coly=Datatn(1:N,k);       % the vector of k-variable
  rankx=tiedrank(Colx);
 ranky=tiedrank(Coly);
  difranks=sum((rankx-ranky).^2);
```

```
 corrsp=1-(6*difranks)/(N*(N^2-1));
   E(j,k)=(Tn(Colx))*(Tn(Coly))*corrsp;% Tn variance
%covariance matrix
end
end
 xy=E;
```

Appendix E

Program for proving that the covariance matrices are positive definite

```
R=10000;
N=25;     % number of rows
P=10;     %number of columns
pi=0.1;   % percentage of outliers
j=3;
% shifted mean
%m=[3 3 3];
%m=[5 5];   %
%m=[3 3 3 3 3];
m=[3 3 3 3 3 3 3 3 3 3];
%m=[5 5 5 5 5 5 5 5 5 5];
ROUND=floor(pi*N);   % number of data that are generating from out of control
distribution
Totmatrix=zeros(P,3,R);
Pivotatrix=zeros(P,j);
Detmatrix=zeros(P,j);
Matrix1=zeros(P,P,R);
Matrix2=zeros(P,P,R);
Matrix3=zeros(P,P,R);
xy1=zeros(P,P,R);
xy2=zeros(P,P,R);
xy3=zeros(P,P,R);
rowindex=1;
C1=0;
C2=0;
C3=0;
count1=0;
count2=0;
```

```
count3=0;
newmatrix=zeros(P*R,j);
newmatrix1=zeros(P*R,j);
for r=1:R
seed = 3985+r;      % to fix all random data sets
rand('seed',seed);  % generate from uniform distribution
randn('seed',seed); %generate from standdard normal distribution

Z=randn(N,P);       % generate from standard normal distribution
Datatn=[Z(1:ROUND,:)+repmat(m,ROUND,1);Z(ROUND+1:N,:)]; % out outliers
by percent pi.
[N P]=size(Datatn); % the size of Datatn N rows and p columns
E1=zeros(P,P);      % location for the covariance matrix E1
E2=zeros(P,P);      % location for the covariance matrix E2
E3=zeros(P,P);      % location for the covariance matrix E3
% calculate the covariance by spearman correlation
for j=1:P         % loop to get the p characteristics variables
for k=1:P     % loop to get other  p characteristics variables
    Colx=Datatn(1:N,j);       % the vector of j-variable
    Coly=Datatn(1:N,k);       % the vector of k-variable
    rankx=tiedrank((Colx));
    ranky=tiedrank((Coly));
    difranks=sum((rankx-ranky).^2);
    corrsp=1-(6*difranks)/(N*(N^2-1));
    E1(j,k)=Tn(Colx)*Tn(Coly)*corrsp;
    E2(j,k)=MADn(Colx)*MADn(Coly)*corrsp;
    E3(j,k)=Sn1(Colx)*Sn1(Coly)*corrsp;
end
end
% matrices that are generated from the Tn,MADn,Sn
xy1(:,:,r)=E1;
```

```
xy2(:,:,r)=E2;
xy3(:,:,r)=E3;
% calculate the eigen values for three covariance matrices
[V1,D1]=eig(E1);
[V2,D2]=eig(E2);
[V3,D3]=eig(E3);
for i=1:P
if D1(i,i)<0
      C1=C1+1;
end
if D2(i,i)<0
      C2=C2+1;
end
if D3(i,i)<0
      C3=C3+1;
end
end
% by cholesky calculate the lower triangular
  [L1,p1] = chol(E1,'lower');
  [L2,p2] = chol(E2,'lower');
  [L3,p3] = chol(E3,'lower');
%cholesky decomposition
 Matrix1(:,:,r)= L1*L1';
 Matrix2(:,:,r)= L2*L2';
 Matrix3(:,:,r)= L3*L3';
% calculate the pivots
 A=zeros(P,1);B=zeros(P,1);C=zeros(P,1);
 d1=zeros(P,1);d2=zeros(P,1);d3=zeros(P,1);
for k=1:P
     A(k,1)=det(E1(1:k,1:k));
     B(k,1)=det(E2(1:k,1:k));
```

```
    C(k,1)=det(E3(1:k,1:k));
if k==1
        d1 (k,1)=A(k,1)/1;
        d2 (k,1)=B(k,1)/1;
        d3 (k,1)=C(k,1)/1;
else
        d1 (k,1)=A(k,1)/A(k-1,1);
        d2 (k,1)=B(k,1)/B(k-1,1);
        d3 (k,1)=C(k,1)/C(k-1,1);
end
end
for i=1:P
if A(i,1)<0
         count1=count1+1;
end
if B(i,1)<0
        count2=count2+1;
end
if C(i,1)<0
        count3=count3+1;
end
end

end
```