**A Distribution-Free Control Chart for Retrospective
Location Analysis of Subgrouped Multivariate Data**

by

Richard C. Bell, Jr.

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2011

Keywords: phase I, preliminary, in-control reference sample,
robust, nonparametric, data depth

Approved by

Saeed Maghsoodloo, Co-chair, Professor of Industrial Engineering
L. Allison Jones-Farmer, Co-chair, Associate Professor of Management
Nedret Billor, Associate Professor of Mathematics and Statistics
Alice E. Smith, Professor of Industrial Engineering

Abstract

In multivariate quality control, a proper Phase I analysis is essential to the success of

Phase II monitoring. Even self-starting methods, which seek to minimize the Phase I process,

usually recommend a single retrospective analysis at some point in the control charting process.

This is true regardless of the underlying distribution of a process, which cannot often be assumed

to be multivariate normal. A literature review reveals no distribution-free Phase I multivariate

techniques in existence, so this research seeks to fill that gap by developing a distribution-free

method of establishing an in-control reference sample for subgrouped multivariate processes in

Phase I. The resulting multivariate sample, representing the in-control state of a process, can

then be used to estimate the appropriate parameters for the Phase II multivariate quality control

monitoring method of choice.

The proposed method, which assumes constant covariance within subgroups, uses data

depth in conjunction with robust estimators to detect both isolated and sustained shifts in

subgroup location. Using Monte Carlo simulation, the proposed method is compared to the

traditional Hotelling's T^2 chart with a Phase I upper control limit. Although Hotelling's T^2 chart

is preferred when data are multivariate normally distributed, the proposed method is shown to

perform significantly better than Hotelling's T^2 chart when a process distribution is heavy-tailed

or skewed.

Acknowledgements

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AP	alarm probability
ARL	average run length
BACON	blocked adaptive computationally efficient outlier nominators
CL	center line
CUSUM	cumulative sum
EAP	empirical alarm probability
EWMA	exponentially weighted moving average
FAP	false alarm probability
HT2	Hotelling's T^2
IC	in control
IS	isolated shift
LCL	lower control limit
LGN	lognormal
MA	moving average
MCD	minimum covariance determinant
MCUSUM	multivariate cumulative sum
MEWMA	multivariate exponentially weighted moving average
MHD	Mahalanobis depth
MMR	multivariate mean-rank

MSD   Mahalanobis spatial depth

MVE   minimum volume ellipsoid

OC   out of control

RBP   replacement breakdown point

RL   run length

RMCD   reweighted minimum covariance determinant

RMD   robust Mahalanobis depth

SD   simplicial depth

SPD   spatial depth

SS   sustained shift

UCL   upper control limit

Note: To avoid confusion, the reader should pay particular attention to the definitions provided

for OC, SD, and SS. These abbreviations are often used in statistical literature to stand for

operating characteristic, standard deviation, and sum of squares, respectively, but are defined

differently in this document.

1   Introduction and Literature Review

1.1   Background and Motivation

Multivariate statistical process control charts are necessary to simultaneously monitor two or more correlated variables representing quality characteristics of an industrial or other process. A multivariate control charting application usually involves a dimension reduction technique of converting multivariate observations to single dimensional control chart statistics which are then monitored using appropriate control limits. This approach accounts for the correlation structure in the data, whereas monitoring correlated variables using separate univariate control charts for each variable ignores the correlation among quality characteristics and can lead to erroneous conclusions about the state of a process. The first multivariate quality control chart is attributed to Hotelling (1947), who created the T^2 chart to monitor bombsight data during World War II.

Multivariate quality control charting has grown in both popularity and relevance since Hotelling's introduction. In a review of statistical process control research issues and ideas, Woodall and Montgomery (1999) pointed out the notable rise in multivariate quality control research due to increased measurement capability and computing power. Montgomery (2005, p. 489) noted that larger manufacturing databases have greatly increased the use of multivariate quality control methods in recent years. Bersimis, Psarakis, and Panaretos (2007) stated in a multivariate statistical process control overview that multivariate Shewhart-type charts are the

most common control charts in industry today, adding that more examination of this area is very

important. In particular, they pointed out the need for more research into robust design of

Hotelling's T^2 chart and nonparametric control charts.

As represented by Figure 1.1.1, the contribution of this research is the merger of three

separately researched but highly related fields (distribution-free Phase I quality control,

computational geometry, and robust parameter estimation) to provide a solution to the open

problem of establishing an outlier-free reference sample for a multivariate process without the

assumption of normality. Great strides have been made in each of the aforementioned research

**Distribution-Free
Phase I Quality
Control**

**Computational
Geometry**

**Robust
Parameter
Estimation**
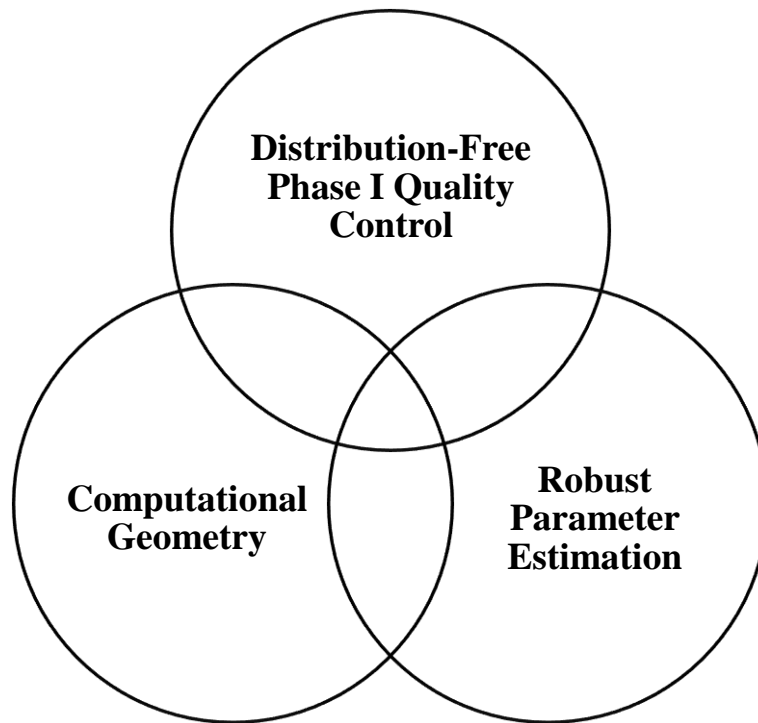
Figure 1.1.1 The Unification of Relevant Research Areas

areas in recent years, yet no one in the statistical quality control field has leveraged recent

developments in the manner accomplished by this research. The following chapters will detail

the multivariate extension of an existing univariate distribution-free control chart for subgroup

location, including the use of appropriate data depth functions for purposes of dimension reduction and the implementation of an effective robust parameter estimation technique, to provide a solution to this problem.

1.2   Differences Between Phase I and Phase II

A control charting application is typically divided into two distinct phases. In Phase I, also known as the preliminary analysis phase, when little is known about a process being studied, the objective is to identify an in-control (IC) reference sample. This involves retrospective analysis of a historical data set in order to eliminate any data points which do not accurately represent the routine operation of the process. The resulting data are described as IC because it is believed that all remaining variability in the process is inherent to the process itself and not due to assignable causes. Upon completion of Phase I, the IC reference sample is used to establish control limits for Phase II, the monitoring stage of a control charting application. In Phase II, newly observed data points are successively compared to the control limits to identify significant departures from the IC state. Should an observation fall outside the control limits, a search for an assignable cause is immediately undertaken. If the change in process behavior can be linked to special causes or external factors, the process is deemed out of control (OC) and remedial action is taken to correct the problem.

Prior to conducting any analysis in a control charting scenario, it is usually assumed that the unedited reference sample may contain OC points and the control limits are unknown. The challenging nature of a Phase I analysis under these conditions has been recognized since the earliest days of statistical process control. Shewhart (1939, p. 76) said, "In the majority of practical instances, the most difficult job of all is to choose the sample that is to be used as the

basis for establishing the tolerance range. If one chooses such a sample without respect to the assignable causes present, it is practically impossible to establish a tolerance range that is not subject to a huge error."

If a flawed Phase I analysis results in the erroneous inclusion of OC points in the IC reference sample, the control limits for Phase II monitoring will be too wide and OC situations will not be detected in a timely manner. This in turn will result in the production of poor quality goods or services for an unnecessarily protracted period of time. When the OC condition is finally detected, the substandard goods or services will have to be reworked, or scrapped and completely reproduced. This can cost the goods or services facility money in terms of labor and other operating expenses for rework or reproduction, additional materials necessary for reproduction, lost future production while previous work is redone, financial penalties for failure to meet contractual deadlines, and loss of customers due to dissatisfaction with faulty or untimely goods or services received.

On the other hand, if a faulty Phase I analysis results in the erroneous exclusion of IC points from the IC reference sample, the control limits for Phase II monitoring will be too narrow and false alarms will repeatedly occur. False alarms require work stoppages to search for assignable causes, potentially costing the goods or services facility money in terms of lower throughput, idle workers while OC signals are investigated, overtime for quality control personnel investigating OC signals, financial penalties for failure to meet contractual deadlines, and loss of customers due to goods or services not being received in a timely manner. Ultimately, whether the resulting control limits are too wide or too narrow, an incorrect Phase I analysis can also cause a lack of confidence by all in the quality control methodology in place, creating a challenging environment for managers.

4

Phase I control charts are designed with the goal of achieving a specified overall IC false alarm probability (FAP), defined as the probability of one or more observations plotting outside the control limits in the absence of assignable causes. Phase I usually involves iteratively comparing the reference sample to trial control limits (corresponding to the desired overall IC FAP) estimated from the sample. At each iteration of a Phase I analysis, an OC point is eliminated from the reference sample if an assignable cause is identified, and trial control limits are updated excluding the OC point. This iterative process continues until all points in the reference sample are IC.

Phase I analysis requires careful consideration when it involves methods which compute independent control chart statistics consisting of individual observations (e.g. the univariate X chart or the multivariate T^2 chart) or subgrouped observations (e.g. the univariate $\bar{X}$ chart or the multivariate T^2 chart). Provided the observations originate from random sampling, the control chart statistics are independent of one another. However, because the control limits are estimated from the reference sample itself in Phase I, the control limits are dependent on each sample point included in their calculation. Thus, *successive comparisons* of chart statistics to control limits are statistically dependent despite the control chart statistics themselves being independent. These dependencies often make it difficult to correctly determine the overall IC FAP for a Phase I analysis.

Phase II, on the other hand, consists of comparing *new* observations (in the form of a chart statistic) to the control limits previously established in Phase I. Because the control limits in Phase II are fixed through conditioning, successive comparisons of chart statistics to control limits are independent provided the chart statistics are independent of one another (e.g. the X, $\bar{X}$, and T^2 charts). This is in contrast to moving average (MA), exponentially weighted moving

5

average (EWMA), or cumulative sum (CUSUM) charts and their multivariate counterparts, whose chart statistics include past observations and are therefore naturally dependent.

Chart performance in Phase II is often measured using moments of the run length (RL) distribution. The RL is the number of observations until an OC signal is observed. If the comparisons of the chart statistics to the control limits are independent, the RL is a geometric random variable. The expected value of the IC RL is equal to $1/\alpha$, where α is equal to the probability that a single chart statistic plots outside the control limits in the absence of assignable causes. The expected value of the RL is known as the average run length (ARL) and is commonly used to describe control chart performance in Phase II.

The purpose of this research is to develop a Phase I procedure for subgrouped multivariate data that is distribution free when a process is IC. The procedure will be based on the use of data depth in conjunction with robust estimators of location and scale to reduce multivariate observations to univariate depth values, thus producing a center-outward ordering of the multivariate data. The corresponding ranks of the univariate depth values, in the form of a control statistic for each subgroup, will then be analyzed using a univariate chart. As the following literature review will demonstrate, this is an area in much need of additional research.

1.3   Phase II Multivariate Control Charting Methods

Existing Phase II multivariate control charting methods will be discussed first, beginning with parametric charts. This will be followed by an examination of distribution-free, nonparametric, and robust techniques, and will conclude with a synopsis of depth-based nonparametric procedures for use in Phase II. Before undertaking this discussion, however, it is

important to distinguish precisely what is meant by the terms distribution free, nonparametric, and robust.

Gibbons and Chakraborti (2003, p. 3) state, "In a *distribution-free* inference, whether for testing or estimation, the methods are based on functions of the sample observations whose corresponding random variable has a distribution which does not depend on the specific distribution of the population from which the sample was drawn." In other words, a "distribution-free" method uses a control chart statistic which follows the same distribution regardless of the underlying distribution of the process itself. Gibbons and Chakraborti (2003, p. 3) add, "On the other hand, strictly speaking, the term *nonparametric* test implies a test for a hypothesis which is not a statement about parameter values." This means that "nonparametric" control charting methods assess whether the distribution of a process, as opposed to specific parameters, has departed from the IC state. From this, it is clear that the terms distribution free and nonparametric are not synonymous, as a control charting method could be distribution free but not nonparametric and vice versa. Last but not least, the term "robust" will be used to refer to methods in which the distribution of the statistics are similar regardless of the distribution of the process data, but the methods may not be strictly distribution free. All characterizations of control charting methods as being distribution free, nonparametric, or robust refer to the IC state of a process only.

1.3.1   Phase II Multivariate Parametric Charts

Hotelling's T^2 control chart is the most familiar multivariate quality control chart in existence today [Montgomery (2005, p. 491)]. It is designed for detecting large shifts in the mean vector of a multivariate normally distributed process because it uses information only from

the current sample, and it can be applied during both Phase I and Phase II using appropriate control limits. Alternatively, authors such as Chenouri and Steiner (2009), Chenouri and Variyath (2011), and Mohammadi, Midi, Arasan, and Al-Talib (2011) have proposed bypassing Phase I by using the reweighted minimum covariance determinant (RMCD) method of Willems, Pison, Rousseeuw, and Van Alest (2002) to glean robust estimates of location and scatter from a reference sample, and implementing those estimates directly in a Phase II T^2 control chart. In all cases, however, the T^2 chart is reliant on the limiting assumption that the data follow a multivariate normal distribution. This chart's lack of robustness to nonnormality is well documented by distribution-free and nonparametric control chart authors such as Chou, Mason, and Young (2001), Liu, Singh, and Teng (2004), and Fricker and Chang (2009a) who evaluated their proposed methods in comparison to the traditional T^2 chart applied to nonnormal process data.

Crosier (1988) and Pignatiello and Runger (1990) proposed several multivariate cumulative sum (MCUSUM) charts which are more sensitive to small or gradual location shifts since they use past information in addition to the current sample, but these charts also rely on the assumption of multivariate normally distributed data. Jackson (1991) presented a T^2 chart using principal components scores, a control chart for principal components residuals, and a control chart for each independent principal component's scores, all based on the assumption of a multivariate normally distributed process. The multivariate exponential weighted moving average (MEWMA) control chart developed by Lowry, Woodall, Champ, and Rigdon (1992) is, like the MCUSUM chart, sensitive to small or gradual shifts but likewise based on the assumption of multivariate normally distributed data. It can be designed to be robust to nonnormality by using a small smoothing constant as noted by Stoumbos and Sullivan (2002),

Testik, Runger, and Borror (2003), and Testik and Borror (2004). However, the MEWMA chart assumes that the IC process mean vector and covariance matrix are known, which is unlikely to be the case in Phase I.

Numerous other parametric Phase II multivariate control charting methods, many of which are variants of the well-known T^2, MEWMA, and MCUSUM charts, have been proposed but will not be detailed here. For comprehensive reviews of such charts, see Wierda (1994), Lowry and Montgomery (1995), Mason, Champ, Tracy, Wierda, and Young (1997), Woodall and Montgomery (1999), and Bersimis et al. (2007).

1.3.2 Phase II Multivariate Distribution-Free, Nonparametric, and Robust Charts

Nonparametric, distribution-free, and robust multivariate control charting methods have been developed, yet they are usually designed for Phase II implementation. Hayter and Tsui (1994) proposed a nonparametric multivariate control chart to detect location changes in nonnormally distributed processes. This method is based on the empirical cumulative distribution function of a statistic formed from an IC reference sample of 500 or more observations, so it is strictly a Phase II method. Qiu and Hawkins (2001) developed a distribution-free, rank-based CUSUM procedure for detecting a location shift, but this method assumes knowledge of the IC mean vector. Chou et al. (2001) proposed a kernel smoothing technique for estimating the distribution of the T^2 control statistic and the upper control limit of the T^2 chart when the Phase II process data follow a multivariate exponential distribution. Qiu and Hawkins (2003) also introduced a nonparametric CUSUM procedure for detecting mean shifts in all directions. This method is based both on the order information among the measurement components as well as the order information between measurement components

and their IC means, but it assumes that the IC distribution of a process is known. Sun and Tsung (2003) developed a distribution-free multivariate control chart based on the distance between the "kernel centre" of the known IC sample and the new observation, using support vector methods to calculate the kernel distance. Thissen, Swierenga, de Weijer, Wehrens, Melssen, and Buydens (2005) used a combination of mixture modeling, which separates the data into Gaussian clusters, and statistical process control techniques to create a distribution-free multivariate control chart. This method requires an IC reference sample to estimate the moments of the Gaussian clusters and the fraction of observations in each cluster. Qiu (2008) proposed a distribution-free, log-linear modeling-based approach to estimating the IC multivariate distribution, as well as a distribution-free MCUSUM procedure for detecting location shifts in Phase II, but the availability of a set of IC data is assumed. Fricker and Chang (2009a) used a Kolmogorov-Smirnov test to compare the ranked kernel density estimates for a set of IC data and a set of the most recent data points. This method is nonparametric but again requires the existence of a multivariate reference sample.

1.3.3   Phase II Multivariate Rank-Based Charts

Nonparametric multivariate control charts have also been proposed using simplicial data depth, which was first introduced by Liu (1990), as a dimension reduction technique. The idea behind simplicial depth-based control charts is to use the simplicial depth of a given multivariate point x within the data cloud formed by a multivariate reference sample $\{X_1,...,X_n\}$ to produce a univariate center-outward ranking of the data points. A precise definition of simplicial depth in p dimensions will be presented in Chapter 2, but the simplicial depth of a bivariate point x is the

proportion of triangles formed by all possible triplets of points in $\{X_1, ..., X_n\}$ containing x.

Simplicial depth in higher dimensions follows the same logic.

Liu's (1995) suggested procedure is to calculate the simplicial depth of a given multivariate point, use the depth to create a control statistic reflecting the point's center-outward ranking relative to an IC reference sample, plot the control statistic on a univariate control chart, and finally compare the control statistic to control limits set to achieve a desired maximum IC FAP. The resulting control charts, called r, Q, and S charts, are essentially X, $\bar{X}$, and cumulative sum (CUSUM) charts respectively, using simplicial depth-based ranks instead of raw univariate data to compute control statistics. Liu (1995) describes these charts as completely nonparametric and able to simultaneously detect location and scale changes in a process. However, Stoumbos and Jones (2000) showed that the 500-observation IC reference sample recommended for Liu's (1995) charts was not large enough to achieve a satisfactory IC FAP for many process distributions, thus limiting the method's potential for widespread implementation. Liu et al. (2004) later introduced a simplicial data depth-based moving average (DDMA) control chart which is described as having better ability than the r and Q charts to detect changes in location while maintaining the same ability to detect changes in scale. The DDMA chart is also said to be completely nonparametric but as with most nonparametric methods, if the process data follow a multivariate normal distribution then a normal theory method (e.g. Hotelling's T^2 chart) is preferred. Notably, all results from this study are derived using an IC reference sample of 1000, yet again raising the question of how one is to obtain such a large IC data set.

Other data depth-based nonparametric approaches to the Phase II multivariate quality control problem have been developed, but they all assume a pre-existing IC reference sample. Zarate (2004) used principal components analysis to reduce the dimensionality of a process, and

11

then employed a nonparametric control chart based on data depth to monitor some of the principal components instead of the original variables. Beltran (2006) employed Liu's (1995) r chart using the simplicial depth ranks of the first and last set of principal components. Messaoud, Weihs, and Hering (2008) proposed a data depth-based, distribution-free EWMA control chart for multivariate observations. This procedure consists of computing the Mahalanobis or simplicial depth of a point with respect to the m most recent observations from a process, converting each depth to a sequential rank among the m most recent observations, and monitoring the standardized sequential ranks using the EWMA chart. The authors recommend an IC reference sample of 100 or more points to initiate this method. For multivariate data following an elliptical distribution, Hamurkaroglu, Mert, and Saykan (2004) developed a nonparametric control chart which consists of computing the Mahalanobis depth of each point, ranking each depth measurement with respect to a sample from an IC process, and then using r and Q charts proposed by Liu (1995) to monitor the ranks. Once more, the Phase I problem of identifying an IC reference sample must be solved before using any of these procedures.

1.4   Self-Starting Multivariate Control Charting Methods

Self-starting multivariate methods, in which successive observations are used to update parameter estimates and check for OC conditions, have been suggested as a substitute for solving the Phase I problem because they can be implemented at the very beginning of a process. These methods are designed to reduce reliance on large and potentially costly Phase I samples required by some multivariate control charting procedures. As noted by Sullivan and Jones (2002, p. 25), they can be especially advantageous when production is slow, early OC production is expensive, or there are insufficient samples available to estimate parameters.

One of the earliest attempts at a self-starting multivariate control chart is Quesenberry's (1997) *Q*-chart, in which the author proposed computing a control chart statistic based on the quadratic form of the deviation of the current observation vector from the estimated mean vector. The control chart statistic is then transformed to a $N(0, 1)$ scalar and monitored using a univariate Shewhart-type control chart. Schaffer (1998) employed the same basic methodology as Quesenberry (1997), but used a univariate EWMA scheme to monitor the resulting control chart statistic. Both methods assume multivariate normally distributed process data.

Sullivan and Jones (2002) introduced a self-starting MEWMA chart, showing that it is more effective than the methods of Quesenberry (1997) and Schaffer (1998) and has the added advantage of robustness to nonnormality with an appropriate choice of smoothing constant. Sullivan and Jones (2002) caution that because parameter estimates are updated with each new observation, changes occurring near the beginning of a process can be unknowingly absorbed into the parameter estimates, thus masking the shift. To guard against this, Sullivan and Jones (2002) recommend augmenting their self-starting chart with a single retrospective analysis at a suitable point in the process, with the exact timing dependent on the dimension as well as other factors.

Zamba and Hawkins (2006) developed a multivariate change-point model which claims to eliminate the requirement of a large Phase I sample. Their method analyzes standardized differences between potential preshift and postshift observations to identify the point at which the mean vector changes, but is only applicable to multivariate normal processes. Also, Zamba and Hawkins' (2006) chart assumes that the mean vector remains constant after a single shift occurs, so it is designed to detect a *sustained* shift of the mean only.

Hawkins and Maboudou-Tchao (2007) proposed a self-starting methodology which transforms multivariate normal observations with unknown parameters into multivariate standard normal observations which are then charted using the MEWMA chart or any other method requiring known parameters, thus bypassing the difficult task of parameter estimation. However, like most self-starting methods, this technique is susceptible to error resulting from early shifts in the process. Although the authors argue their method eliminates the need for a Phase I - Phase II distinction, they suggest that after the initial phase of data gathering, one should "start with the most recent process reading and successively add and chart the earlier readings back to the start of the sequence" [Hawkins and Maboudou-Tchao (2007, p. 206)] in order to diagnose undetected shifts occurring earlier in the process.

These self starting methods are certainly viable alternatives under certain conditions. Nevertheless, they have not diminished the need for a more universally applicable distribution-free Phase I multivariate control chart procedure.

1.5   Phase I Multivariate Control Charting Methods

There exist a number of control chart methods developed for use in Phase I, though they are mostly variations of Hotelling's T^2 control chart based on the assumption of a multivariate normally distributed process. In addition, the majority of them deal with *individual* as opposed to *subgrouped* data. Hotelling's T^2 control chart can be applied to individual data in Phase I using control limits outlined by Tracy, Young, and Mason (1992). However, Sullivan and Woodall (1996) showed that the usual practice of pooling all the individual observations to estimate the covariance matrix for a T^2 chart results in poor performance in detecting step (sudden) and ramp (gradual) shifts in the mean vector. They instead proposed using the vector

differences between successive individual observations to estimate the IC covariance matrix for the T^2 statistic, and demonstrated that this method works better in detecting mean shifts but not outliers.

For processes consisting of either individual or subgrouped observations, Sullivan and Woodall (1998) proposed modified MCUSUM and MEWMA charts using simulated control limits to account for the correlation among control statistics as well as a regression-based method with exact (not simulated) limits for detecting sustained shifts in the mean vector. Using simulation, they showed that each of their three proposed methods is better at detecting small shifts in the mean vector than Hotelling's T^2 chart. Nedumaran and Pignatiello (2000) addressed the issue of constructing T^2 control chart limits for retrospective testing when the parameters of a subgrouped multivariate normally distributed process are unknown. They described and compared a computationally intensive method of determining the exact control limit, Bonferroni adjustments to Alt's (1976) Phase I control limit, and Bonferroni adjustments to the standard χ^2 limit, ultimately recommending Bonferroni adjustments to Alt's (1976) Phase I limit as the best alternative.

Vargas (2003) proposed T^2 control charts for Phase I analysis of individual multivariate normally distributed data using robust estimators of location and dispersion instead of the usual sample mean vector and sample covariance matrix. A total of five different estimators were considered, including the minimum volume ellipsoid (MVE) estimators of Rousseeuw and Van Zomeren (1990), first introduced by Rousseeuw (1984), and the minimum covariance determinant (MCD) technique of Rousseeuw and Van Driessen (1999), also introduced by Rousseeuw (1984). The MVE method finds the ellipsoid of minimum volume that covers a specified minimum number of data points, and uses the geometrical center of the ellipsoid as the

location estimator and the matrix defining the ellipsoid itself (multiplied by a constant) as the covariance matrix estimator. The MCD method finds the subset of data that has the smallest covariance matrix determinant while covering a specified minimum number of points. It then uses the sample mean vector and the sample covariance matrix (also multiplied by a constant) of the points in the subset as estimators for location and dispersion. Vargas also considered a trimming approach which removes a proportion of extreme values based on Mahalanobis distance, Sullivan and Woodall's (1996) sample mean vector and covariance matrix estimated from differences of successive observations, and an outlier detection algorithm proposed by Sullivan and Woodall (1996). Based on simulation results, Vargas recommended using both a T^2 control chart based on MVE estimators for detecting multiple outliers and the T^2 control chart suggested by Sullivan and Woodall (1996) to detect sustained shifts in the mean vector in Phase I.

Jensen, Birch, and Woodall (2007) further detailed the advantages of using the MVE and MCD methods in conjunction with T^2 control charts for detecting outliers in individual multivariate normally distributed data during Phase I. They determined that the MVE estimator is best for smaller sample sizes and a smaller percentage of outliers, while the MCD estimator is preferred for larger sample sizes or a larger percentage of outliers. The authors also provided tables of simulated control limits for both estimators.

Other Phase I control charting efforts for multivariate normally distributed processes include Alfaro and Ortega's (2008) proposal to trim each variable to obtain robust estimates for the mean vector and covariance matrix, and then use those estimates in Hotelling's T^2 chart with Tracy et al.'s (1992) Phase I UCL to provide enhanced outlier detection. Jobe and Pokojovy (2009) created a computationally intensive two-step method of identifying the largest bulk of

16

similar data from a time-ordered sequence of individual multivariate normally distributed points, and used the estimated mean vector and covariance matrix from this bulk in the T^2 statistic with empirical control limits. The authors compared the performance of Hotelling's T^2 chart using their method, the classical method of parameter estimation, and the robust methods analyzed by Vargas (2003) and Jensen et al. (2007), showing that their method results in improved performance in detecting outliers as well as location shifts in Phase I. The authors attribute their success to the fact that their method considers the time order of the data, whereas other methods do not. Oyeyemi and Ipinyomi (2010) robustly estimated the covariance matrix for Hotelling's T^2 chart for individuals in Phase I by identifying a subset of data which meets specified optimality criteria, and then iteratively expanding the subset to a predetermined size. Their method was shown to outperform the MVE and MCD methods in a limited number of cases, but only bivariate normally distributed samples of size $m = 30$ were considered. Most recently, Yanez, Gonzalez, and Vargas (2010) proposed using biweight S estimators for location and scatter in a T^2 chart for individual multivariate normally distributed data with simulated limits, showing that it outperforms Hotelling's T^2 chart with MVE estimators for small samples.

Distribution-free and nonparametric Phase I methods, on the other hand, have received little attention in multivariate quality control literature. The only chart found is Dai, Zhou, and Wang's (2006a) unpublished halfspace (Tukey) data depth-based nonparametric MCUSUM chart.

1.6   Developing a Distribution-Free Phase I Procedure -- A Univariate Example

Although unanswered in the multivariate domain, the challenge of developing a distribution-free Phase I procedure has been addressed for the univariate case. The details of the

univariate Phase I solution are relevant to the multivariate Phase I problem because this research

will ultimately rely on a univariate chart to monitor control statistics resulting from dimension

reduction of a multivariate reference sample using data depth. The unique considerations

involved in developing a distribution-free Phase I procedure are best illustrated by an example.

Example 1.6.1

Consider a reference sample consisting of $m = 25$ independent subgroups, each

containing $n = 5$ observations from an unknown distribution. The widely used Shewhart $\bar{X}$ chart

with 3σ limits can be created using the procedure outlined in Montgomery (2005), under the

assumption that the distribution of subgroup averages is approximately normal due to the central

limit theorem. Since the IC parameters μ_o and σ_o are unknown, the lower control limit (LCL),

center line (CL), and upper control limit (UCL) are estimated using

$$LCL = \hat{\mu}_o - 3\frac{\hat{\sigma}_o}{\sqrt{n}} \tag{1.6.1}$$

$$CL = \hat{\mu}_o \tag{1.6.2}$$

$$UCL = \hat{\mu}_o + 3\frac{\hat{\sigma}_o}{\sqrt{n}}, \tag{1.6.3}$$

where $\hat{\mu}_o$ and $\hat{\sigma}_o$ are unbiased estimators for μ_o and σ_o. Montgomery (2005, pp. 196-198)

discusses several choices for $\hat{\mu}_o$ and $\hat{\sigma}_o$.

Using Equations (1.6.1), (1.6.2), and (1.6.3), the initial Phase I control chart for this

example is illustrated in the top panel of Figure 1.6.1. Suppose that investigation of the potential

OC point represented by subgroup average number 11 reveals an assignable cause, so the point is

deemed OC. The revised control limits in the bottom panel of Figure 1.6.1 are more narrow due

18

to the exclusion of subgroup 11, and all remaining subgroup averages now fall within the updated control limits. The IC reference sample has been established, and the most recent control limits can be used for Phase II monitoring.
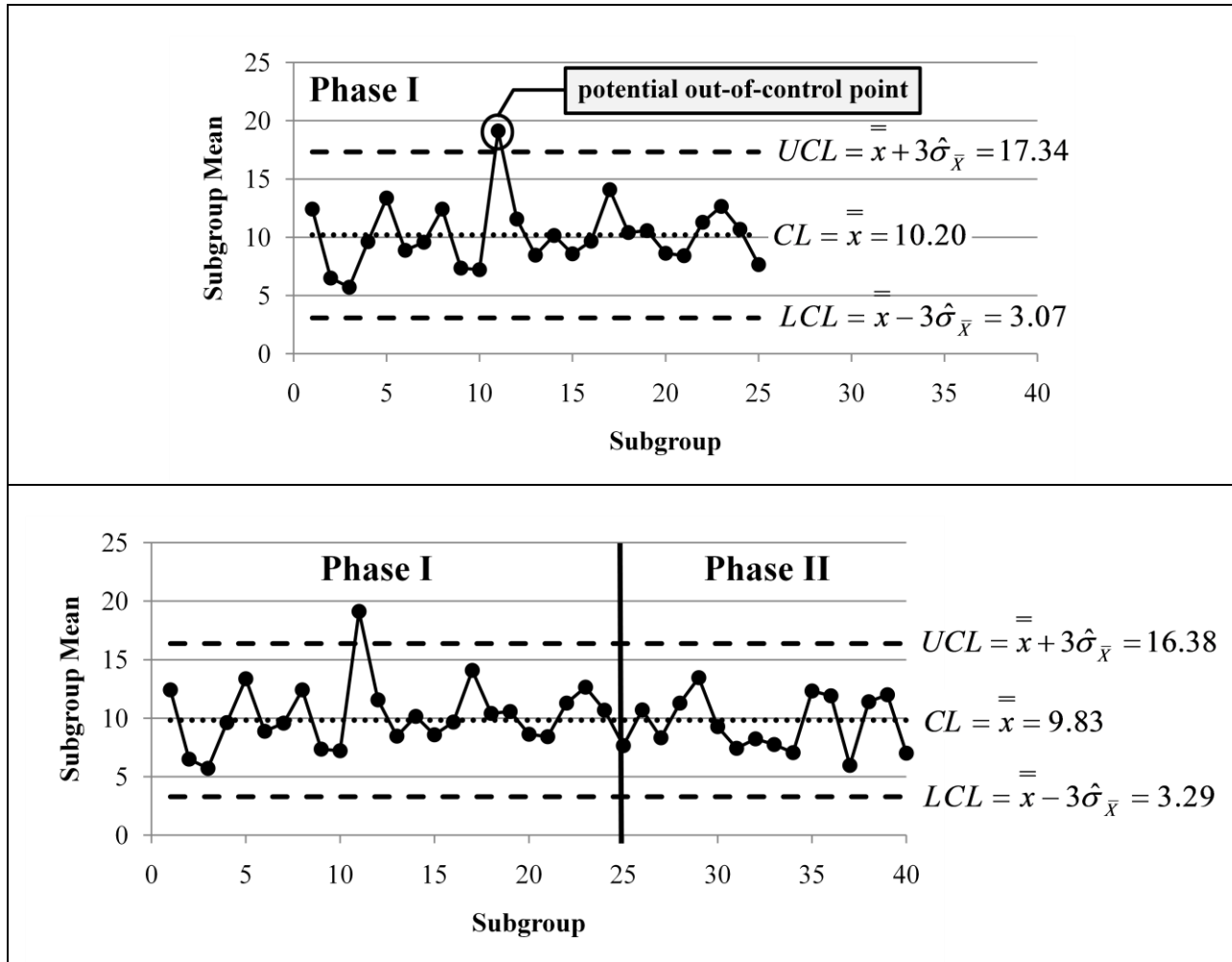
Figure 1.6.1 Initial (Top Panel) and Revised (Bottom Panel) Control Charts

Determination of the overall IC FAP for the control chart in Example 1.6.1 would be straightforward under conditions of normality of subgroup averages and known parameters. The overall IC FAP or P(at least one false alarm among all $m = 25$ comparisons) would be calculated

as follows: $(1 - (1 - 0.0027)^{25}) = 0.0654$. The overall IC FAP, while considerably higher than the individual FAP of 0.0027, could easily be lowered by using limits wider than 3σ.

If, on the other hand, the underlying distribution of the subgroup averages is not normal, the true overall IC FAP may be much larger. Suppose for example that the actual individual FAP is 0.01. Then the overall IC FAP equals $(1 - (1 - 0.01)^{25}) = 0.2222$. With only a slight increase in the individual IC FAP, the overall IC FAP increased dramatically. This could result in a large number of IC subgroups being erroneously excluded during Phase I.

Furthermore, when the parameters are unknown as in Example 1.6.1, successive comparisons of subgroup averages to control limits are dependent. Therefore, the overall IC FAP may *not* be determined using 1 minus the product of the complements of the $m = 25$ individual FAPs. Instead, control limits designed to achieve a specified overall IC FAP must be determined using the joint density function or the simulated empirical distribution of the subgroup averages.

Champ and Jones (2004) dealt with the case of a normally distributed process and unknown parameters by using the (joint) multivariate t distribution of the m control statistics to define control limits to achieve a desired overall IC FAP. For processes in which normality cannot be established and parameters are unknown, Jones-Farmer, Jordan, and Champ (2009) proposed a rank-based Phase I location chart which is essentially a Shewhart chart of standardized subgroup mean ranks. This method uses approximate multivariate normal theory control limits (for large subgroup sizes n) and simulated control limits (for smaller subgroup sizes n) to achieve a specified overall IC FAP.

The issues of data nonnormality and dependence among control statistics are problematic for any parametric Phase II control charting method used in Phase I, including multivariate

procedures. These are precisely the problems this research seeks to address by developing a distribution-free method of establishing an IC reference sample for a multivariate process consisting of subgrouped data.

1.7   Special Considerations in Multivariate Quality Control

There are two drawbacks to multivariate quality control that must be kept in mind in any research effort. The first is computational complexity. Multivariate control charting methods are inherently more computationally intensive than univariate methods. Despite advances in quality control software, complex methods can quickly become unmanageable as the dimension of the data increases. The development of methods that only work for two or three variables, or that are too complex to be used by practitioners, must be guarded against.

The second downside to multivariate quality control is the issue of interpretation. Multivariate control chart techniques do not directly identify which variable(s) caused an OC signal. As previously discussed, it is insufficient to simply separate and individually chart each variable belonging to an OC multivariate process, because correlated variables may behave differently alone than when in combination with each other. As a result, many useful approaches to interpreting OC signals in a multivariate setting have been proposed, and a summary of such works is provided by Bersimis et al. (2007) in an overview of multivariate statistical process control charts. While this problem will not be specifically addressed by this research, it should be considered when developing a new procedure.

1.8   Organization of Dissertation

The remainder of this document is dedicated to the detailed development and application of a data depth-based, distribution-free Phase I multivariate control charting method for detecting location changes in subgrouped data. In Chapter 2, data depth is explored as a distribution-free method of reducing multi-dimensional data to univariate ranks, and the advantages and disadvantages of several depth functions considered for implementation are discussed. Chapter 3 addresses the actual design of the data depth-based, distribution-free Phase I control chart for subgrouped multivariate data. In Chapter 4, the simulation-based performance assessment plan for the proposed method is discussed, and detailed algorithms for measuring performance under various location shifts in normal, heavy-tailed, and skewed distributions are provided. Chapter 5 contains the results of extensive simulation runs comparing the proposed data depth-based, distribution-free Phase I multivariate method to Hotelling's T^2 chart with Phase I UCL. Chapter 6 is dedicated to a comprehensive application of the proposed data depth-based, distribution-free Phase I multivariate method to a simulated historical data set containing several location shifts. This dissertation concludes in Chapter 7 with a synopsis of research conducted, recommendations for Phase I analysis when dealing with subgrouped multivariate data under conditions of normality and nonnormality, recommendations for subsequent Phase II monitoring, and discussion of areas in need of further investigation.

2   Measuring Centrality of Multivariate Data Using Data Depth

2.1   Fundamentals of Data Depth

A *data depth* measures how deep (or central) a point $x \in \mathbb{R}^p$ is with respect to a certain probability distribution F or a given data cloud $\mathrm{X}_n = \{X_1, ..., X_n\}$ in $\mathbb{R}^p$. A data depth is computed by applying one of many known data depth functions to a multivariate data point, thus reducing it from a p-vector to a univariate depth value. Assuming unimodality of the data, a large depth value indicates centrality and a low depth value suggests outlyingness of a given point. Depth values are usually normalized to have a range of [0, 1]. The point of maximal depth is considered the center of the data and is referred to as the multivariate median. A data depth function may be visualized in p-dimensional space as a series of nested contours around the multivariate median, where each contour represents the set of p-dimensional points with equal depth values. Some depth functions force contours of a particular geometric form (e.g. elliptical), whereas others allow contours to follow the actual geometric shape of the data.

Data depth facilitates the extension of order statistics to higher dimensions, because depth values can be ranked from largest to smallest to produce a center-outward ordering of the data. The ordered depth values can then be used to detect outliers, which are known in multivariate quality control literature as OC points. Data depth allows multivariate data from any distribution to be characterized by the relative position of the data points rather than parameters estimated

from the actual data values. This rank-based perspective makes data depth potentially very useful as a distribution-free method of multivariate analysis.

The concept of data depth dates as far back as Tukey (1975), but until recently its usefulness for statistical quality control has been limited by the tradeoff between statistical properties, robustness to nonnormality, and computational complexity. After a comprehensive review of numerous existing depth functions, this research will implement robust Mahalanobis depth and Mahalanobis spatial depth because they are computationally feasible in any dimension, sufficiently robust to outliers under the assumptions of this research, and satisfy the four desirable properties of data depth functions discussed by Zuo and Serfling (2000).

2.2 Desirable Properties of Data Depth Functions

For a depth function $D(x; F)$ to serve most effectively as an analytical tool, the following four properties are required [Liu (1990), Zuo and Serfling (2000)]. Denote the class of probability distributions on R^p by F.

- Property 1: Affine invariance. The depth of a point $x \in \mathrm{R}^p$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements. This ensures that a point classified as an outlier or nonoutlier in one coordinate system is similarly classified in another coordinate system resulting from an affine transformation. Formally stated, $D(Ax + b; F_{AX+b}) = D(x; F_X)$ holds for any random vector X in R^p, any p x p nonsingular matrix A, and any p-vector b.

- Property 2: Maximality at center. For a distribution having a uniquely defined "center" (e.g., the point of symmetry with respect to some notion of symmetry), the depth function should attain maximum value at this center. This supports an accurate center-outward

24

ordering of the data points. Formally stated, $D(\boldsymbol{\theta}; F) = \sup_{x \in \mathbb{R}^p} D(\boldsymbol{x}; F)$ holds for any

$F \in \mathbf{F}$ having center $\boldsymbol{\theta}$, where $\mathbf{F}$ is the class of distributions on the Borel sets of $\mathbb{R}^p$.

- Property 3: Monotonicity relative to deepest point. As a point $\boldsymbol{x} \in \mathbb{R}^p$ moves away from

  the "deepest point" (the point at which the depth function attains maximum value; in

  particular, for a symmetric distribution, the center) along any fixed ray through the

  center, the depth at $\boldsymbol{x}$ should decrease monotonically. This also supports an accurate

  center-outward ordering of the data points. Formally stated, for any $F \in \mathbf{F}$ having

  deepest point $\boldsymbol{\theta}$, $D(\boldsymbol{x}; F) \le D(\boldsymbol{\theta} + \alpha(\boldsymbol{x} - \boldsymbol{\theta}); F)$ holds for $\alpha \in [0, 1]$.

- Property 4: Vanishing at infinity. The depth of a point $\boldsymbol{x}$ should approach zero as $\|\boldsymbol{x}\|$

  approaches infinity, where $\|\boldsymbol{x}\|$ is the Euclidean norm of $\boldsymbol{x}$. This ensures the data depth

  function is both bounded and nonnegative. Formally stated, $D(\boldsymbol{x}; F) \to 0$ as $\|\boldsymbol{x}\| \to \infty$,

  for each $F \in \mathbf{F}$.

According to Zuo and Serfling (2000), depth functions which satisfy these four properties are

particularly well suited for nonparametric multivariate inference, so these properties will serve as

a useful basis for describing the data depth functions selected for implementation in this

research.

A depth function may be viewed as a location estimator, and as such may be

characterized by its finite-sample *replacement breakdown point* (RBP). First defined by Donoho

and Huber (1983), the RBP is the minimum fraction of a sample which must be replaced by

outliers in order to completely ruin an estimate, so a low RBP indicates nonrobustness and a high

RBP signifies robustness to outliers. When used to describe a depth function, the RBP is usually

stated in reference to the multivariate median estimated by a depth function. The RBP of the

multivariate median is important because if the center of the data (as determined by the multivariate median) is significantly affected by outliers, the subsequent center-outward ordering will likewise be affected and outliers may be masked.

Whether a depth function has a high or low RBP is often determined by the robustness of any location or scatter estimators used in its construction. The robustness of such location or scatter measures is also described using the RBP. Precise definitions of RBPs for both location and scatter estimators are adapted from Donoho and Huber (1983) and Lopuhaa and Rousseeuw (1991). Let $X_n = \{X_1, ..., X_n\}$ be a random sample of size n in R^p. The RBP of a location estimator T at X_n, or the smallest fraction k/n of outliers which can take the resulting estimate beyond any bound, is defined as

$$RBP(T; X_n) = \min\left\{ \frac{k}{n} : \sup_{X_{n,k}} \|T(X_n) - T(X_{n,k})\| = \infty \right\}, \tag{2.2.1}$$

where $X_{n,k}$ is a contaminated sample found by replacing k points of X_n with arbitrary values. The RBP of a scatter estimator C at X_n or the smallest fraction k/n of outliers which can drive the largest eigenvalue of the resulting estimate to infinity or the smallest eigenvalue of the resulting estimate to zero, is defined as

$$RBP(C; X_n) = \min\left\{ \frac{k}{n} : \sup_{X_{n,k}} M\left(C(X_n), C(X_{n,k}) \right) = \infty \right\}, \tag{2.2.2}$$

where $X_{n,k}$ is defined as before, $M(A, B) = \max\left\{ \left| \lambda_1(A) - \lambda_1(B) \right|, \left| \lambda_p(A)^{-1} - \lambda_p(B)^{-1} \right| \right\}$, and $\lambda_1(A) \geq \cdots \geq \lambda_p(A)$ are the ordered eigenvalues of the matrix A.

To illustrate the idea of an RBP, consider a sample of size n in R^1 and two common location estimators: the sample mean and the sample median. The sample mean has an RBP of

only $1/n$ because a single outlier could move the sample mean to infinity, so it is considered a nonrobust location estimator. In contrast, the sample median has the highest possible RBP of $1/2$ because $1/2$ of the sample would have to be contaminated with outliers in order to effect a corresponding shift in the sample median. Consequently, the sample median is the preferred location estimator in $\mathbf{R}^1$ from a robustness standpoint.

In addition to having a high RBP, any location or scatter estimator used in conjunction with a data depth function should also be *affine equivariant*. From Lopuhaa and Rousseeuw (1991), a location estimator T is affine equivariant if $T(\boldsymbol{AX}+\boldsymbol{b})=\boldsymbol{A}T(\boldsymbol{X})+\boldsymbol{b}$ for any p-vector $\boldsymbol{b}$ and any p x p nonsingular matrix $\boldsymbol{A}$, and a positive definite scatter estimator C is said to be affine equivariant if $C(\boldsymbol{AX}+\boldsymbol{b})=\boldsymbol{A}C(\boldsymbol{X})\boldsymbol{A}^T$ for any p-vector $\boldsymbol{b}$ and any p x p nonsingular matrix $\boldsymbol{A}$. Akin to the concept of affine invariance for data depth functions, affine equivariance means that an estimator does not depend on the location, scale, or orientation of the data. According to Lopuhaa and Rousseeuw (1991), finding affine equivariant estimators with high RBPs is a challenging problem. However, these properties are of paramount importance to any multivariate quality control application, so only estimators possessing these properties will be considered in this research.

2.3 Robust Mahalanobis Depth

The *Mahalanobis depth* (MHD) of a point $\boldsymbol{x}$ in $\mathbf{R}^p$ with respect to a distribution F in $\mathbf{R}^p$ is defined as

$$MHD(\boldsymbol{x};F)=\left[1+d^2_{\sum(F)}\left(\boldsymbol{x},\boldsymbol{\mu}(F)\right)\right]^{-1},\qquad\qquad(2.3.1)$$

27

where $\mu(F)$ and $\Sigma(F)$ are location and covariance measures defined on F and

$d_M^2(x, y) = (x - y)' M^{-1}(x - y)$ is the Mahalanobis distance [Mahalanobis (1936)] between two

points x and y in R^p with respect to a positive definite p x p matrix M. When the distribution F

is unknown and a random sample $X_n = \{X_1, ..., X_n\}$ is used to estimate $\mu(F)$ and $\Sigma(F)$, the

sample version of the depth function is annotated as $MHD(x; F_n)$, where F_n denotes the

empirical distribution function of the sample. MATLAB code for computing Mahalanobis

depth, based on a modification of S. Mazumder's (personal communication, July 7, 2010)

algorithm, is provided in Appendix A.

The Mahalanobis depth function satisfies the four desirable properties listed by Zuo and

Serfling (2000) and is relatively easy to compute, but assumes the underlying distribution F is

elliptical and therefore produces elliptical contours of equal depth. In addition, as noted by Zuo

and Serfling (2000), the RBP of the median determined by the Mahalanobis depth function is

completely dependent on the choice of location and covariance measures $\mu(F)$ and $\Sigma(F)$. If

the classical location and covariance estimators $\bar{X}_n$ and S_n are used, the Mahalanobis depth

function is nonrobust. The presence of even a single outlier can contaminate the estimators $\bar{X}_n$

and S_n, possibly masking the presence of outliers. In order to preclude this, Mahalanobis depth

should be used in conjunction with robust estimators.

Mahalanobis depth will be referred to as robust Mahalanobis depth (RMD) when used

with robust location and scatter estimators. There are numerous robust estimation methods from

which to choose. Dang and Serfling (2010) noted that the computationally complex MCD

method proposed by Rousseeuw (1984) or the more efficient Fast-MCD method of Rousseeuw

and Van Driessen (1999) could be used to produce affine equivariant, robust location and

28

covariance estimates. As discussed in Chapter 1, the MCD method finds the subset of data that has the smallest covariance matrix determinant while covering a user-specified number of points. It then uses the sample mean vector and the sample covariance matrix of the points in the subset as estimators for location and dispersion. According to Jensen et al. (2007), MCD estimators have a maximum RBP of $\lfloor (n-p+1)/2 \rfloor / n$, which is approximately 1/2 for reasonable values of n and p, when the number of points used is equal to the integer value of $(n+p+1)/2$. The Fast-MCD program is available in many statistical software packages such as R, S-PLUS, and SAS. In addition, a library of MATLAB codes for robust analysis including the Fast-MCD program may be obtained from the LIBRA website at http://wis.kuleuven.be/stat/robust/Libra.html.

Another alternative for finding robust estimators of location and scatter is the blocked adaptive computationally efficient outlier nominators (BACON) method of Billor, Hadi, and Velleman (2000). The BACON method is very computationally efficient, even for extremely large data sets. It begins with a small outlier-free subset of the data, and then allows this subset to grow rapidly until a stopping criteria is reached. Two versions of this iterative forward selection method are available: Version 2 which is nearly affine equivariant and has RBPs exceeding 40% for various combinations of dimension and sample size, and Version 1 which is completely affine equivariant with RBP of approximately 20%. The Type I error probability (α) for the BACON method can be set to any number between 0 and 1, but $\alpha = 0.05$ is suggested for most applications. MATLAB code for the BACON method is available from the authors.

After several rounds of experimentation, it was decided to use the BACON method (with $\alpha = 0.10$) to estimate the process mean vector and $\overline{S} = \dfrac{1}{m} \sum_{i=1}^{m} S_i$, the scatter estimator for

Hotelling's T^2 chart when data are divided into m subgroups, to estimate the process covariance matrix. The BACON method was chosen as the location estimator because of its excellent balance between computational efficiency and robustness. Although $\overline{S}$ is generally not considered a robust estimator, it was chosen as the scatter estimator because it is highly robust to location shifts (the focus of this research) when process data possess a common within-subgroup covariance structure. Details are provided in Chapter 5.

Example 2.3.1

To illustrate an application of the robust Mahalanobis depth function, consider the bivariate random sample $X_5 = \{X_1, ..., X_5\}$ from an unknown distribution, where each $X_i = [X_1 \quad X_2], i = 1, ..., 5,$ illustrated in Figure 2.3.1. The first step in computing RMD for this sample is to estimate the mean vector using the BACON method (with $\alpha = 0.10$) and the

i	X_i	
1	11.15	49.63
2	7.91	36.46
3	5.42	28.06
4	16.22	38.77
5	8.09	29.21
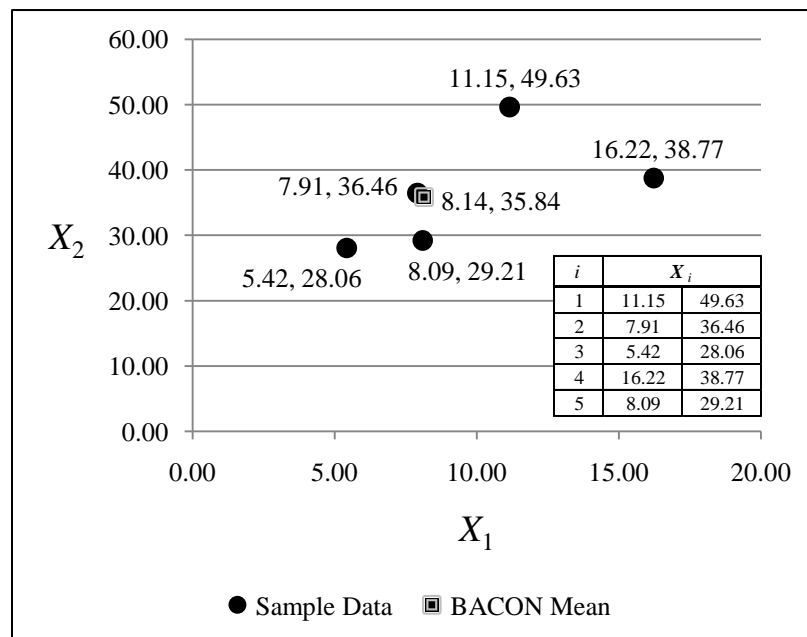
Figure 2.3.1 Bivariate Random Sample

30

covariance matrix using Hotelling's T^2 scatter estimator for subgrouped data. Because this

example involves individual as opposed to subgrouped observations, Hotelling's T^2 scatter

estimator for subgrouped data reduces to the classical nonrobust sample covariance matrix.

Under these conditions, the robust BACON scatter estimator may be a better choice, but

Hotelling's T^2 scatter estimator is used to maintain consistency with the methodology employed

throughout the remainder of this research. Estimates of location and scatter are determined to be:

$$\bar{X}_{BACON} = \begin{bmatrix} 8.14 & 35.84 \end{bmatrix}$$

$$S_{HT2} = \begin{bmatrix} 17.18 & 20.45 \\ 20.45 & 75.48 \end{bmatrix}$$

$$S_{HT2}^{-1} = \begin{bmatrix} 0.09 & -0.02 \\ -0.02 & 0.02 \end{bmatrix}.$$

Note that the BACON method excluded $X_4 = \begin{bmatrix} 16.22 & 38.77 \end{bmatrix}$ from the estimated mean vector

due to its outlyingness relative to the other points.

Using the RMD function, $RMD(x; F_n) = \left[1 + \left(x - \bar{X}_{robust} \right) S_{robust}^{-1} \left(x - \bar{X}_{robust} \right) \right]^{-1}$, and the

location and scatter estimates, the robust Mahalanobis depth for $X_1 = \begin{bmatrix} 11.15 & 49.63 \end{bmatrix}$ is computed

as follows:

$$RMD(X_1; F_5) = \left[1 + \left(X_1 - \bar{X}_{BACON} \right) S_{HT2}^{-1} \left(X_1 - \bar{X}_{BACON} \right) \right]^{-1}$$

$$= \left[1 + \left(\begin{bmatrix} 11.15 & 49.63 \end{bmatrix} - \begin{bmatrix} 8.14 & 35.84 \end{bmatrix} \right) \begin{bmatrix} 0.09 & -0.02 \\ -0.02 & 0.02 \end{bmatrix} \left(\begin{bmatrix} 11.15 & 49.63 \end{bmatrix} - \begin{bmatrix} 8.14 & 35.84 \end{bmatrix} \right)' \right]^{-1}$$

$$= \left[1 + \begin{bmatrix} 3.01 & 13.79 \end{bmatrix} \begin{bmatrix} 0.09 & -0.02 \\ -0.02 & 0.02 \end{bmatrix} \begin{bmatrix} 3.01 & 13.79 \end{bmatrix}' \right]^{-1}$$

$$= \begin{bmatrix} 1 + 2.56 \end{bmatrix}^{-1}$$

$$= 0.28.$$

RMD computations for the four remaining observations in the sample proceed in the same manner.

The final results, along with corresponding rankings, are provided in Table 2.3.1. As expected, X_2 attains the highest depth value since it is closest to the center of the data set (as defined by the BACON mean vector), and X_4 receives the lowest depth value since it is most outlying.

i	X_i		RMD$(X_i;F_5)$	rank
1	11.15	49.63	0.28	4
2	7.91	36.46	0.98	1
3	5.42	28.06	0.55	2
4	16.22	38.77	0.18	5
5	8.09	29.21	0.54	3

Table 2.3.1 Data Ranked According to RMD

2.4   Mahalanobis Spatial Depth

Mahalanobis spatial depth (MSD) [Dang and Serfling (2010)] is an attractive alternative to robust Mahalanobis depth because it is only slightly more difficult to compute yet is not restricted to elliptical distributions. This means that the contours of equal depth determined by the depth function conform to the geometric structure and shape of the data, as opposed to being constrained to an elliptical form. Mahalanobis spatial depth is based on the concept of *spatial depth* (SPD), defined by Vardi and Zhang (2000) for a point x in $\mathbb{R}^p$ with respect to a distribution F in $\mathbb{R}^p$ as

$$SPD(x;F) = 1 - \|ES(x-X)\|, \text{ where } S(x) = \begin{cases} \dfrac{x}{\|x\|} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases} \quad (2.4.1)$$

Intuitively, the spatial depth of a multivariate point x is equal to one minus the average of the unit vectors from x to all observations in the sample. Spatial depth is graphically illustrated in Figure 2.4.1.
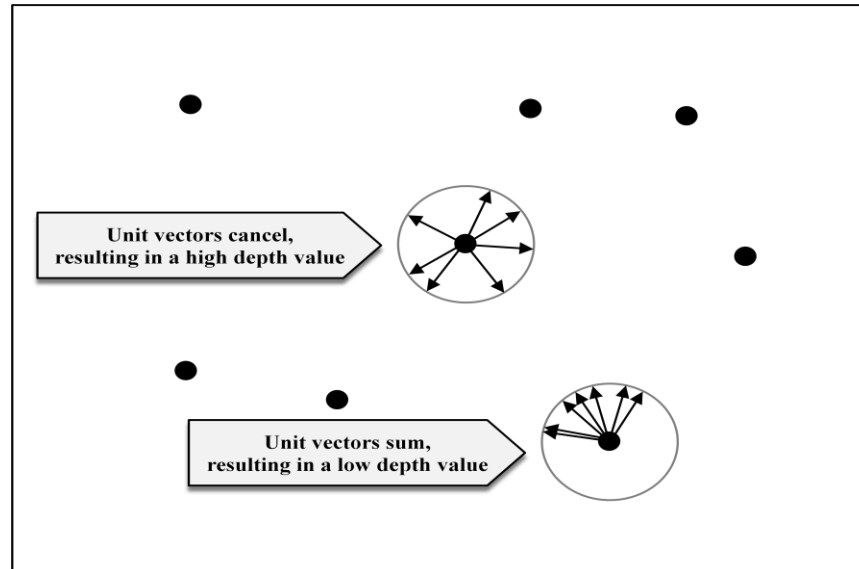
Figure 2.4.1 Illustration of Spatial Depth

The spatial depth function is quickly computable in any dimension, and its multivariate median has a very favorable RBP of 1/2 [Vardi and Zhang (2000)]. It also satisfies the properties of maximality at center (with some exceptions; see Zuo and Serfling (2000) for details), monotonicity relative to deepest point, and vanishing at infinity. However, it is not completely affine invariant. According to Serfling (2002), the spatial depth function is invariant with respect to shift, orthogonal, and homogeneous scale transformations of the data, but not heterogeneous scale transformations. This is sufficient if all variables share the same unit of measure, but this is not always the case in a multivariate quality control application so a modification of the spatial depth function is needed.

Serfling (2010) showed that a fully affine invariant modification of the spatial depth

function may be accomplished by standardizing the sample data using any *weak covariance*

functional, which is defined as follows [Serfling (2010, p. 9)]:

"A symmetric positive definite p x p matrix-valued functional $C(F)$ is called a

weak covariance functional if, for $Y = AX + b$ with any nonsingular p x p matrix

A and any vector b, $C(F_Y) = k_1 AC(F_X)A'$, with $k_1 = k_1(A, b, F_X)$ a positive

scalar function of A, b, and F_X. The sample version for a data set

$X_n = \{X_1, ..., X_n\}$ in R^p may be expressed, with $Y_n = AX_n + b$ and

$k_1 = k_1(A, b, X_n)$, as $C_n(Y_n) = k_1 AC_n(X_n)A'$."

Application of a weak covariance functional transformation leads to Serfling's (2010) formula

for computation of Mahalanobis spatial depth (MSD) for a point x in R^p with respect to a

distribution F in R^p :

$$MSD(x; F) = 1 - \left\| ES\left(C(F_X)^{-1/2}(x - X)\right)\right\|. \qquad (2.4.2)$$

The sample version for a point x with respect to a random sample $X_n = \{X_1, ..., X_n\}$ in

R^p is

$$MSD(x; F_n) = 1 - \left\| ES\left(C_n(F_n)^{-1/2}(x - X)\right)\right\|. \qquad (2.4.3)$$

There are a number of options available for determining the sample weak covariance functional

$C_n(X_n)$ but again $\overline{S} = \dfrac{1}{m}\sum_{i=1}^{m} S_i$, the scatter estimator for Hotelling's T^2 chart when data are

divided into m subgroups, will be used in this research because of its robustness to location shifts

under the assumption of constant within-subgroup covariance. MATLAB code for computing

Mahalanobis spatial depth, based on a modification of S. Mazumder's (personal communication,

July 7, 2010) algorithm, is provided in Appendix B.

Example 2.3.1 will be revisited to illustrate an application of the Mahalanobis spatial

depth function. Computing MSD begins by multiplying the data set by the negative square root

of the weak covariance functional $C_5(X_5) = S_{HT2}$ as follows:

$$X^* = X_5\left(C_5(X_5)\right)^{-1/2} = X_5 S_{HT2}^{-1/2} = \begin{bmatrix} 11.15 & 49.63 \\ 7.91 & 36.46 \\ 5.42 & 28.06 \\ 16.22 & 38.77 \\ 8.09 & 29.21 \end{bmatrix} \begin{bmatrix} 17.18 & 20.45 \\ 20.45 & 75.48 \end{bmatrix}^{-1/2} = \begin{bmatrix} 0.43 & 5.74 \\ 0.24 & 4.23 \\ -0.01 & 3.29 \\ 2.50 & 4.06 \\ 0.69 & 3.29 \end{bmatrix}.$$

Next, the spatial depth formula is applied to each observation in the transformed sample,

beginning with $X_1^* = \begin{bmatrix} 0.43 & 5.74 \end{bmatrix}$. The first step in this process is to determine the unit vectors

from X_1^* to every point in the sample:

$$\frac{X_1^* - X_1^*}{\left\|X_1^* - X_1^*\right\|} = \begin{bmatrix} 0.00 & 0.00 \end{bmatrix} \text{ by definition}$$

$$\frac{X_1^* - X_2^*}{\left\|X_1^* - X_2^*\right\|} = \frac{\begin{bmatrix} 0.20 & 1.51 \end{bmatrix}}{1.52} = \begin{bmatrix} 0.13 & 0.99 \end{bmatrix}$$

$$\frac{X_1^* - X_3^*}{\left\|X_1^* - X_3^*\right\|} = \frac{\begin{bmatrix} 0.44 & 2.44 \end{bmatrix}}{2.48} = \begin{bmatrix} 0.18 & 0.98 \end{bmatrix}$$

$$\frac{X_1^* - X_4^*}{\left\|X_1^* - X_4^*\right\|} = \frac{\begin{bmatrix} -2.07 & 1.68 \end{bmatrix}}{2.66} = \begin{bmatrix} -0.78 & 0.63 \end{bmatrix}$$

$$\frac{X_1^* - X_5^*}{\left\|X_1^* - X_5^*\right\|} = \frac{\begin{bmatrix} -0.26 & 2.45 \end{bmatrix}}{2.46} = \begin{bmatrix} -0.11 & 0.99 \end{bmatrix}.$$

Then, the average of the unit vectors from the point X_1^* to every point in the sample is computed:

$$\frac{\begin{bmatrix} 0.00 & 0.00 \end{bmatrix} + \begin{bmatrix} 0.13 & 0.99 \end{bmatrix} + \begin{bmatrix} 0.18 & 0.98 \end{bmatrix} + \begin{bmatrix} -0.78 & 0.63 \end{bmatrix} + \begin{bmatrix} -0.11 & 0.99 \end{bmatrix}}{5} = \begin{bmatrix} -0.12 & 0.72 \end{bmatrix}.$$

Finally, the Euclidean norm of the resulting vector is subtracted from one in order to arrive at the Mahalanobis spatial depth value of the point X_1^*:

$$\text{MSD}\left(X_1^*; F_5\right) = 1 - \sqrt{\left(-0.12\right)^2 + \left(0.72\right)^2} = 0.27.$$

Computations for the four remaining observations in the sample proceed in the same fashion.

The final results, along with corresponding rankings, are listed in Table 2.4.1. Rankings for X_1 and X_4 were assigned as indicated because $\text{MSD}\left(X_1^*; F_5\right) > \text{MSD}\left(X_4^*; F_5\right)$ when $\text{MSD}\left(X_1^*; F_5\right)$ and $\text{MSD}\left(X_4^*; F_5\right)$ are expanded to four significant digits. Depth values and rankings using the MSD function are somewhat different than those obtained using the RMD function. This is because RMD assumes the data are elliptically symmetric, whereas MSD makes no distributional assumptions about the data.

i	X_i		$\text{MSD}(X_i^*; F_5)$	rank
1	11.15	49.63	0.27	4
2	7.91	36.46	0.68	1
3	5.42	28.06	0.35	3
4	16.22	38.77	0.27	5
5	8.09	29.21	0.53	2

Table 2.4.1 Data Ranked According to MSD

2.5   Simplicial Depth

As discussed in Chapter 1, simplicial data depth played a prominent role in early depth-based nonparametric multivariate control charting efforts, so a justification for its exclusion from this research is necessary. Introduced by Liu (1990), the *simplicial depth* (SD) of a point x in $\mathbb{R}^p$ with respect to a distribution F in $\mathbb{R}^p$ is defined as the probability that x belongs to a random simplex in $\mathbb{R}^p$, formally stated as

$$SD(\pmb{x};F) = P_F\left(\pmb{x} \in S\left[\pmb{X}_1,...,\pmb{X}_{p+1}\right]\right), \tag{2.5.1}$$

where $\pmb{X}_1,...,\pmb{X}_{p+1}$ are independent observations from F and $S\left[\pmb{X}_1,...,\pmb{X}_{p+1}\right]$ denotes the p-

dimensional simplex with vertices $\pmb{X}_1,...,\pmb{X}_{p+1}$, or the set of all points in R^p that are convex

combinations of $\pmb{X}_1,...,\pmb{X}_{p+1}$.

For a random sample $X_n = \{\pmb{X}_1,...,\pmb{X}_n\}$ from F in R^p, the sample simplicial depth

function is derived from this definition to be

$$SD(\pmb{x};F_n) = \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{p+1} \le n} I\left(\pmb{x} \in S\left[\pmb{X}_{i_1},...,\pmb{X}_{i_{p+1}}\right]\right), \tag{2.5.2}$$

where I is the indicator function. $SD(\pmb{x};F_n)$ computes the fraction of the random sample

simplices containing the point $\pmb{x}$. In order to check whether a point $\pmb{x}$ in R^p is inside a simplex

$S\left[\pmb{X}_1,...,\pmb{X}_{p+1}\right]$, the following system of $p + 1$ equations with $p + 1$ unknowns must be solved:

$$\pmb{x} = a_1\pmb{x}_1 + a_2\pmb{x}_2 + ... + a_{p+1}\pmb{x}_{p+1} \tag{2.5.3}$$

$$a_1 + a_2 + ... + a_{p+1} = 1. \tag{2.5.4}$$

Equation (2.5.3) translates into p equations which check to see if the p-dimensional point $\pmb{x}$ can

be expressed as a linear combination of the $p + 1$ vertices forming a given simplex

$S\left[\pmb{X}_1,...,\pmb{X}_{p+1}\right]$. Equation (2.5.4) represents a constraint that the coefficients $a_1, a_2,..., a_{p+1}$ sum

to one. According to Liu (1990), if the simplex is nondegenerate, this system of equations has a

unique solution. Furthermore, the point $\pmb{x}$ is inside the simplex if and only if the coefficients

$a_1, a_2,..., a_{p+1}$ are all positive. For a given point $\pmb{x}$, this process must be repeated for each of the

$\binom{n}{p+1}$ possible p-dimensional simplices $S\left[X_1,...,X_{p+1}\right]$ formed by the sample

$X_n = \{X_1,...,X_n\}$.

In order to illustrate the simplicial depth function, a simple graphical example is provided. Consider a sample of size $n = 5$ from a continuous bivariate distribution F, and suppose the simplicial depth of a point θ is desired. There are a total of $\binom{5}{3} = \dfrac{5!}{3!(5-3)!} = 10$ possible triangles that can be formed from the sample, three of which contain the point θ as illustrated in Figure 2.5.1: $\left(X_1X_2X_4, X_1X_3X_4, X_1X_4X_5\right)$. Therefore, the simplicial depth of the point θ is $SD(\theta;F_5) = \dfrac{3}{10} = 0.30$.
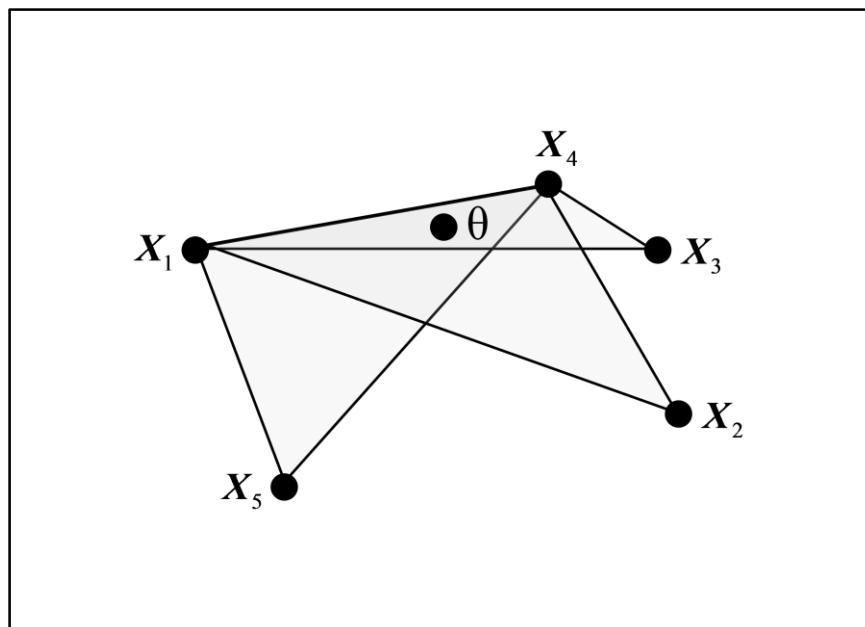
Figure 2.5.1 Illustration of Simplicial Depth

Liu (1990) showed that the simplicial depth function satisfies the affine invariance, vanishing at infinity, maximality at center, and monotonicity properties for continuous

38

distributions. However, as demonstrated by Zuo and Serfling (2000), the maximality and monotonicity properties fail for some discrete distributions, which could be problematic when dealing with a finite sample. As noted by Li and Liu (2004), the exact simplicial depth may be computed in any dimension by solving a system of linear equations, but more efficient algorithms are needed due to the increased computational complexity in higher dimensions. Rousseeuw and Ruts (1996) provided such an algorithm for the bivariate case, but for dimensions greater than two this remains an open problem. Since computational feasibility in higher dimensions is an important goal of this research, simplicial depth will not be implemented in the multivariate quality control charting method proposed in the following chapter.

3   The Multivariate Mean-Rank (MMR) Control Chart

3.1   Introduction

A multivariate quality control Phase I analysis begins with a p-dimensional reference sample, often from an unknown distribution, which may contain one or more OC points. Application of a data depth function to the multivariate reference sample reduces the dimension of the reference sample from p to one. Then a univariate control charting method, with control limits adjusted to account for the dependence among successive comparisons of control chart statistics to control limits, may be applied to the resulting depth values in order to identify and remove the OC points, thus producing an IC reference sample which will serve as a basis for Phase II monitoring.

Differences between Phase I and Phase II were explained in detail in Chapter 1, but will be briefly reiterated here as these differences directly impact the manner in which control limits are determined in a Phase I analysis. In Phase II, the monitoring stage of a control charting application, each new observation is compared (through a control chart statistic) to fixed control limits. With data depth-based methods such as those described by Liu (1995), control limits are often fixed by using an IC reference sample to approximate the univariate distribution of the control chart statistic. Knowledge of this distribution is used to set control limits designed to achieve a certain maximum IC FAP.

In Phase I, the retrospective analysis stage of a control charting application, a fixed number of m existing observations (or subgroups) from a reference sample are successively compared through control chart statistics to trial control limits which are constantly revised as OC points are identified and removed from the reference sample. This renders successive comparisons of control chart statistics to control limits dependent, so control limits must be determined by manipulation of the joint distribution of the control chart statistic, simulation of the empirical joint distribution of the control chart statistic, or other techniques which account for these dependencies.

Methods such as these will be necessary to design control limits for the data depth-based variation of the $\bar{X}$ chart used in this research. The $\bar{X}$ chart for subgrouped data was selected as the model for implementation because it is particularly well suited for use in a Phase I analysis. The $\bar{X}$ chart analyzes only the information from the most recent observation or subgroup. This makes it very effective at detecting single outliers or large shifts in a process which commonly occur in Phase I. According to Montgomery (2005, p. 385), Shewhart-type charts (such as $\bar{X}$ charts) "are extremely useful in Phase I implementation of statistical process control, where the process is likely to be OC and experiencing assignable causes that result in large shifts in the monitored parameters."

On the contrary, other methods such as cumulative sum (CUSUM), exponentially weighted moving average (EWMA), and moving average (MA) charts use more information from a sample and are therefore typically preferred for Phase II monitoring. A CUSUM chart is used to plot the cumulative sum of deviations of sample values from a specified target value [Montgomery (2005, p. 388)]. An EWMA control chart statistic is a weighted average of all previous sample means, with the weights declining geometrically [Montgomery (2005, p. 406)].

The control chart statistic of an MA chart is a simple unweighted average of a specified number of the most recent observations [Montgomery (2005, p. 417)]. Because they accumulate information over time, CUSUM, EWMA, and MA charts detect small shifts in a process more effectively than X and $\bar{X}$ charts, but are slower to respond to large shifts and have less ability to detect single outliers. Furthermore, these charts are based on an implicit assumption that the most recent observations are the most important. This assumption may not be reasonable in Phase I when the sample size is fixed and new observations are not being added. Consistent with this perspective, Montgomery (2005, p. 386) characterizes CUSUM and EWMA control charts as "excellent alternatives to the Shewhart control chart for Phase II process monitoring situations."

3.2 Design of the MMR Chart

The chart implemented in this research is the multivariate analog of Jones-Farmer et al.'s (2009) Phase I mean-rank chart, which was designed as a distribution-free method of identifying an IC reference sample for a *univariate* process with subgrouped data. The mean-rank chart is similar in construct to the $\bar{X}$ chart for univariate subgrouped data, but it uses the standardized average subgroup rank rather than the average of raw subgroup data values as a control statistic. The use of ranks rather than actual data values renders the method distribution free, since the distribution of ranks is the same regardless of the underlying distribution of a univariate process. The mean-rank chart's IC and OC performance was shown to be comparable to the traditional $\bar{X}$ chart when a univariate process is normally distributed, and better than the $\bar{X}$ chart in many scenarios when a univariate process follows a heavy-tailed or skewed distribution.

It will be shown that the mean-rank chart of Jones-Farmer et al. (2009) performs similarly well when adapted for use with ranked data depth values corresponding to a multivariate process. The mean-rank chart modified for use with data depth values from a multivariate process will be hereafter referred to as the multivariate mean-rank (MMR) chart. Like the mean-rank chart, the MMR chart will monitor standardized average subgroup ranks which follow the same distribution regardless of the underlying distribution of a multivariate process, so it too will be distribution free when a process is IC.

In general, any continuous process consisting of two or more correlated variables, usually but not always representing quality characteristics, in which data are subgrouped by design or can be rationally subgrouped, could potentially benefit from the MMR chart proposed by this research. Since most existing multivariate Phase I methods rely on the assumption of a multivariate normally distributed process, the MMR chart will be particularly useful when the process under study is clearly nonnormal or lacks sufficient history to verify an assumption of normality. In addition, because the MMR chart is computationally inexpensive, it will be especially useful for processes consisting of a large number of variables. Example applications of the MMR chart include, but are not limited to industrial (e.g. chemical, power, mining, steel, petroleum, pharmaceutical, electronics, textile, polymer, and automotive), healthcare (e.g. clinical trials and patient satisfaction), military (e.g. weapons development, combat operations, and soldier performance), and service organizations (e.g. finance, marketing, and customer support).

An example military application of the MMR chart, and the one which inspired this author's interest in quality control, is charting the progress of combat operations in Iraq. This problem rose to the forefront of the military operations research community in early 2007, when

43

the President of the United States ordered the deployment of approximately 40,000 additional American troops (known as "The Surge") to reverse a trend of escalating violence in Iraq. Because the troop increase was politically polarizing and therefore closely scrutinized by the United States Congress, it was imperative that an accurate method of assessing its effectiveness be emplaced. Military analysts thus faced a two-fold problem -- determining a historical data set reflecting "normal" violence levels in Iraq and implementing an appropriate method of prospectively monitoring future violence levels during "The Surge."

In hindsight, the difficult problem of determining a historical data set would have been a prime opportunity for application of the MMR chart. First of all, the overall level of violence in Iraq was measured by several correlated variables related to the performance of the US-led coalition and Iraqi security forces, the terrorist actions of various insurgent groups in Iraq, and the safety of the Iraqi civilian populace. In addition, early data on violence levels was extremely volatile and highly skewed due to Iraq's troubled history as well as immature and often inaccurate reporting procedures. Furthermore, data were collected daily but aggregated into weekly subgroups to account for differences in the pace of combat operations on different days of the week. In this situation, the MMR chart would have been a useful tool to establish an IC reference sample against which future weekly violence levels during "The Surge" could have been compared using a Phase II multivariate control chart.

An all-inclusive list of potential applications for the MMR chart is not possible, but it is the opinion of this author that it has the potential to serve as a valuable analytical tool for a wide range of organizations in diverse settings. Its ease of execution and flexibility in solving the distribution-free Phase I multivariate quality control charting problem for subgrouped data fills

in many of the existing gaps in current literature, thus providing a useful methodology for researchers and practitioners alike.

3.2.1 The MMR Control Chart Statistic

Consider a reference sample consisting of m subgroups of size n from a p-dimensional multivariate process in which all variables are continuous. Let the random vector $\boldsymbol{X}_{ij}$ represent the 1 x p row vector containing the jth observation from the ith subgroup. Treating the observations from the m mutually independent samples of size n as a single sample of size $N = n$ x m as described by Jones-Farmer et al. (2009) and attributed to Kruskal and Wallis (1952), a data depth function is applied to each $\boldsymbol{X}_{ij}$, resulting in a corresponding depth value $D\left(\boldsymbol{X}_{ij};F_{N}\right)$, where F_{N} denotes the empirical distribution function of the pooled reference sample. Next, integer ranks $R_{ij} = 1, 2,..., N$ are assigned to each $D\left(\boldsymbol{X}_{ij};F_{N}\right)$ in the pooled sample of size N, beginning with the largest $D\left(\boldsymbol{X}_{ij};F_{N}\right)$ and continuing in descending order. In other words, R_{ij} denotes the rank of $D\left(\boldsymbol{X}_{ij};F_{N}\right)$ when compared to all other depth values in the pooled sample of size N, with the largest $D\left(\boldsymbol{X}_{ij};F_{N}\right)$ receiving rank 1 and the smallest receiving rank N. When the process is IC, the mean of the random variable R_{ij} is $E\left(R_{ij}\right) = \dfrac{N+1}{2}$ and the variance is $Var\left(R_{ij}\right) = \dfrac{(N-1)(N+1)}{12}$ [Jones-Farmer et al. (2009, p. 306)].

In the event of a tie, the midrank method is used as a correction without affecting the mean and variance of the random variable R_{ij} [Jones-Farmer et al. (2009, p. 306)]. According to the midrank method, each tied depth value receives the average of the ranks they would receive

45

if the ties were broken [Lehman (2006, p. 18)]. For example, suppose the four depth values

{0.93, 0.67, 0.67, 0.22} are to be ranked in *descending* order. It is clear that the largest depth

value (0.93) should be assigned rank 1 and the smallest depth value (0.22) rank 4, but the

assignment of ranks 2 and 3 to the equivalent depth values (0.67, 0.67) is ambiguous. In order to

preserve the equality of these two depth values in terms of their ranks, they will both be assigned

the average of the middle two ranks. In this example, the duplicate depth values will both be

assigned rank = (2+3)/2 = 2.5. Thus, the set of ranks corresponding to the four depth values is

{1, 2.5, 2.5, 4}.

Now consider the average of the ranks in each subgroup i, denoted by

$$\bar{R}_i = \frac{\sum_{j=1}^{n} R_{ij}}{n}. \tag{3.2.1}$$

If a process is IC, the ranks should be distributed evenly throughout the m subgroups, resulting in

approximately equal $\bar{R}_i$ for each subgroup. For an IC process, the mean and variance of $\bar{R}_i$ are,

respectively [Bakir (1989, pp. 764-765)]:

$$E(\bar{R}_i) = \frac{N+1}{2} \tag{3.2.2}$$

$$Var(\bar{R}_i) = \frac{(N-n)(N+1)}{12n}. \tag{3.2.3}$$

Invoking the central limit theorem, the random variable representing the standardized subgroup

mean rank,

$$Z_i = \frac{\bar{R}_i - E(\bar{R}_i)}{\sqrt{Var(\bar{R}_i)}}, \tag{3.2.4}$$

follows an approximate standard normal distribution when n is sufficiently large [Jones-Farmer et al. (2009, p. 306)], although small subgroup sizes (e.g. $n = 4, 5,$ or 6) are more likely in most quality control applications [Montgomery (2005, p. 196)]. To create the MMR control chart for use in Phase I, the control statistic Z_i in Equation (3.2.4) is plotted for each of the m subgroups.

3.2.2   Empirical Control Limits for the MMR Chart

As opposed to both lower and upper control limits required for the univariate mean-rank chart of Jones-Farmer et al. (2009), the MMR chart has only an upper control limit. This is because with the MMR chart, observations are ranked based on data depth values rather than raw data values. An extremely negative control chart statistic Z_i occurs when a subgroup consists of observations having extremely high depth values and correspondingly low ranks. This indicates near-perfect centrality with respect to the p-dimensional data cloud, and is therefore no cause for concern. Conversely, an extremely positive control chart statistic Z_i is realized when a subgroup of observations is located far away from the center of the p-dimensional data cloud, resulting in extremely low depth values and correspondingly high ranks. Such a subgroup indicates a potential OC condition which requires further investigation.

For each m, n combination of interest, Monte Carlo simulation of the empirical joint distribution of the standardized subgroup mean rank was used to determine the MMR chart upper control limits in Table 3.2.1. Recall that the joint distribution is required because successive comparisons of control chart statistics to control limits are dependent in Phase I. Limits are tabled for a maximum overall IC FAP of 0.10, where the FAP is the probability that the Phase I chart with m subgroups of size n signals at least once when the process is IC. Due to the discrete nature of the mean-rank distribution as well as simulation noise, simulated FAP values do not

47

m	n	Desired FAP = 0.10	
		UCL	Simulated FAP
20	5	2.476	0.0941
50	5	2.702	0.0983
100	5	2.854	0.0982
150	5	2.932	0.0983
200	5	2.985	0.0981

Table 3.2.1 Empirical Control Limits for the MMR Chart

precisely match the desired FAP values. Conservative limits were chosen in order to ensure the

simulated FAP came as close as possible to the desired FAP without exceeding it. A more

comprehensive table of limits for various combinations of m, n, and FAP is provided in

Appendix C, and MATLAB code for simulating additional limits is provided in Appendix D.

The general construct of the simulation algorithm is as follows:

1) Establish a trial UCL to attain the desired overall IC FAP for a given (m, n)

   combination.

2) Simulate $N = m$ x n random numbers from a Uniform(0, 1) distribution. Assign each

   number a rank from largest (rank $= 1$) to smallest (rank $= N$). Divide the resulting

   ranks into m subgroups of size n.

3) Compute the average rank $\bar{R}_i$ for each subgroup. Determine the corresponding

   standardized subgroup mean rank Z_i.

4) Compare each of the m standardized subgroup mean ranks, Z_i, $i = 1,...,m$, to the trial

   UCL. Increment a counter by one if any Z_i exceeds the UCL.

5) Repeat steps 2 - 4 a total of 100,000 times.

6) Determine the empirical FAP = (final counter value)/100,000.

7) If the empirical FAP exceeds the desired FAP, increase the UCL. If the empirical

FAP is lower than the desired FAP, decrease the LCL.

8) Reset the counter to zero.

9) Repeat steps 2 - 8 until the desired overall IC FAP is achieved.

10) Record m, n, the desired FAP, the empirical FAP, and the UCL.

3.2.3 Analytical Control Limits for the MMR Chart

Prior to simulating empirical limits for the MMR chart, analytical control limits were

attempted using the joint distribution of the standardized mean ranks. As reported by Jones-

Farmer et al. (2009), the central limit theorem suggests that the individual standardized mean

ranks follow a standard normal distribution for sufficiently large subgroup size n. From Bakir

(1989), the joint distribution of the standardized mean ranks is asymptotically multivariate

normal with correlation matrix $R_{m \times m} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{bmatrix}$, where $\rho_{ij} = \dfrac{-1}{(m-1)}$ when

subgroup sizes are equal. Using a zero mean vector and the correlation structure given by $R_{m \times m}$,

asymptotic control limits for the MMR chart were numerically determined through a

modification of Genz' (2011) MATLAB algorithm for evaluating the multivariate normal

distribution. Control limits were computed to achieve a maximum IC FAP of 0.10.

Next, the IC performance of the multivariate normal theory control limits was evaluated

by simulating 10,000 applications of the MMR chart using robust Mahalanobis depth to IC

bivariate normally distributed data with zero mean vector and identity covariance matrix, without

loss of generality. Multivariate normal theory control limits and corresponding empirical IC

FAPs for m = 20, 50(50)200 subgroups of size n = 5(5)20 are recorded in Table 3.2.2.

m	n	Desired FAP = 0.10	
		MVN UCL	Simulated FAP
20	5	2.565	0.0752
20	10	2.565	0.0822
20	15	2.565	0.0881
20	20	2.565	0.0873
50	5	2.865	0.0485
50	10	2.865	0.0766
50	15	2.865	0.0889
50	20	2.865	0.0861
100	5	3.077	0.0296
100	10	3.077	0.0692
100	15	3.077	0.0831
100	20	3.077	0.0837
150	5	3.195	0.0199
150	10	3.195	0.0602
150	15	3.195	0.0744
150	20	3.195	0.0789
200	5	3.277	0.0131
200	10	3.277	0.0567
200	15	3.277	0.0725
200	20	3.277	0.0751

Table 3.2.2 Simulated IC FAPs Using Normal Theory Limits

Multivariate normal theory control limits produced empirical IC FAPs which are close to

the desired IC FAP of 0.10 for large n but unacceptably low for small n. This is because small

subgroup sizes n are insufficient to ensure the individual standardized mean ranks Z_i follow a

standard normal distribution in accordance with the central limit theorem, thus preventing the

joint distribution of the standardized mean ranks from achieving asymptotic multivariate

normality. This can be seen graphically in Figure 3.2.1 depicting Q-Q plots of simulated

standardized mean ranks for $m = 50$ and $n = 5(5)20$. The individual Q-Q plots show a clear departure from normality when $m = 50$ and $n = 5$ (top left), and increasing normality as n is raised to 20 (bottom right).
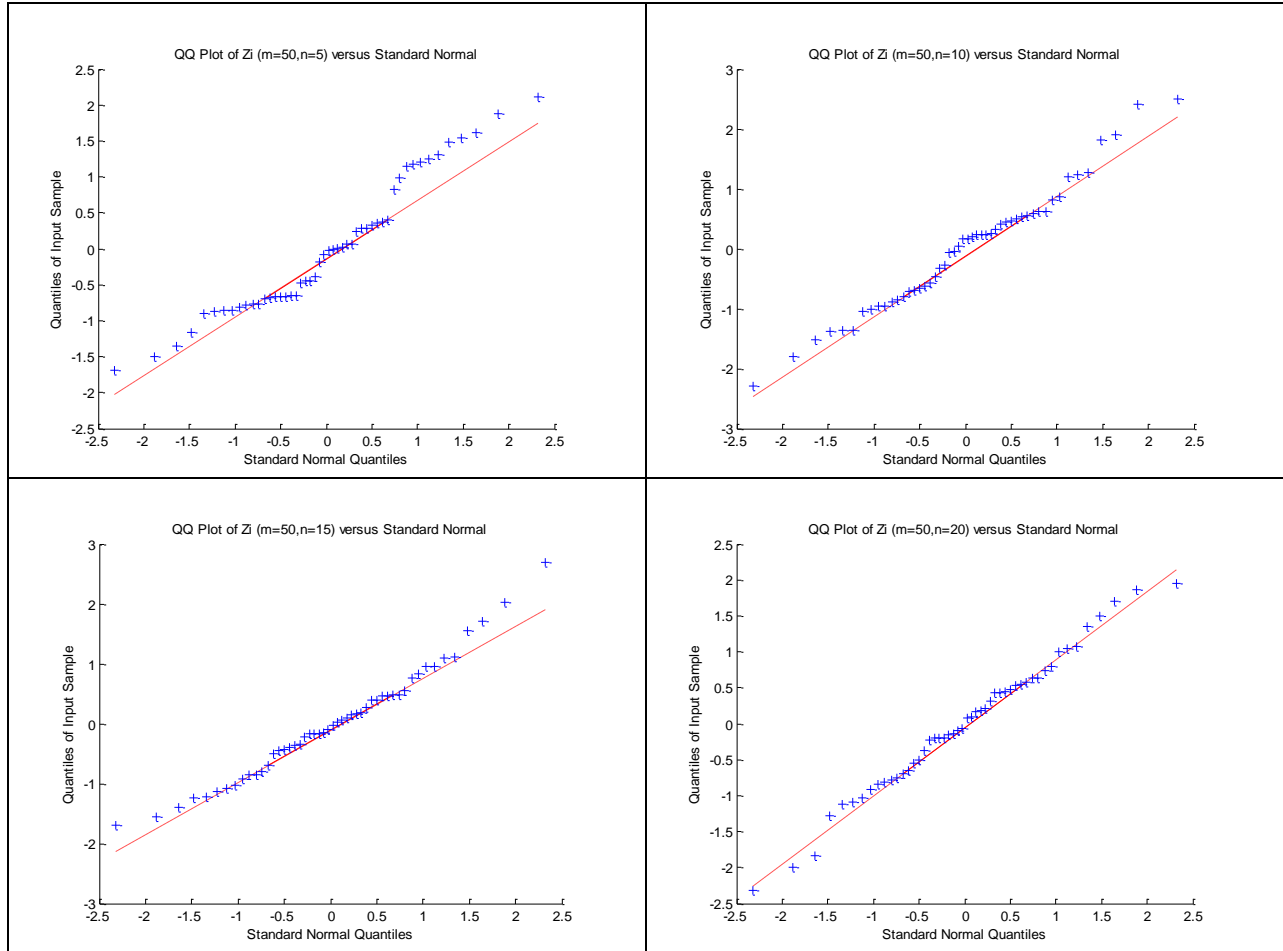
Figure 3.2.1 *Q-Q Plots of Z_i for $m = 50$, $n = 5(5)20$*

Table 3.2.2 also illustrates that MMR chart performance using multivariate normal theory control limits worsens with increasing m. This is easily understood if a Phase I analysis is viewed as the partitioning of a desired overall IC FAP among m simultaneous individual comparisons of control chart statistics to an UCL. A larger m means that a smaller portion of the overall IC FAP is allocated to each of the m individual comparisons. This can be visualized as

51

the UCL being pushed progressively farther into the upper tail of the standard normal distribution of each individual control chart statistic. As this happens, the effects of any departures of the distribution of the individual control chart statistic from standard normality will be exacerbated. This in turn will lead to undesired empirical FAPs for the MMR chart using multivariate normal theory control limits.

Multivariate normal theory control limits could be used to provide conservative limits for a very small number of subgroups or very large subgroup sizes, but empirical control limits are much more consistent in maintaining the desired IC FAP for the range of m and n considered in this research. An alternative to the "one size fits all" multivariate normal theory control limits for the MMR chart is to enumerate the distribution of the standardized mean rank for each combination of number of subgroups m and subgroup size n, and use this information to derive the corresponding joint distribution of the standardized mean ranks. However, this method is clearly impractical for the number of subgroups considered in this research, again supporting the use of empirically determined control limits for the MMR chart.

3.3 Example Application of the MMR Chart

In order to fully understand the workings of an MMR chart, a simple example is provided. Consider the first subgroup of a bivariate process consisting of $m = 50$ subgroups of size $n = 5$ from an unknown distribution F. Let the random vector X_{ij} represent the 1 x 2 row vector containing the jth observation from the ith subgroup, where $i = 1$ and $j = 1$ - 5. The data, along with corresponding robust Mahalanobis depth values and ranks, are listed in Table 3.3.1.

i	j	X_{ij}		RMD$(X_{ij};F_{250})$	R_{ij}
1	1	5.1880	2.4570	0.3311	197
1	2	0.7332	4.7681	0.2904	218
1	3	3.3695	4.3434	0.4533	127
1	4	4.5465	4.7078	0.3258	201
1	5	3.0102	3.8656	0.5677	61

Table 3.3.1 MMR Chart Data for the First Subgroup of a Bivariate Process

Note that R_{ij} reflects rankings with respect to the pooled reference sample of size $N =$

250. The average of the ranks in the first subgroup is $\bar{R}_1 = \dfrac{(197+218+127+201+61)}{5} = 160.80.$

Using Equations (3.2.2) and (3.2.3), $E(\bar{R}_i) = \dfrac{N+1}{2} = \dfrac{250+1}{2} = 125.50$ and

$Var(\bar{R}_i) = \dfrac{(N-n)(N+1)}{12n} = \dfrac{(250-5)(250+1)}{12(5)} = 1024.92.$ Using Equation (3.2.4), the

standardized mean rank for the first subgroup is $Z_1 = \dfrac{\bar{R}_1 - E(\bar{R}_i)}{\sqrt{Var(\bar{R}_i)}} = \dfrac{160.80 - 125.50}{\sqrt{1024.92}} = 1.103.$

Given a desired IC FAP of 0.10, the MMR chart UCL for $m = 50$, $n = 5$ is found from

Table 3.2.1 to be 2.702. Since Z_1 is less than 2.702, it is concluded that the first subgroup is IC.

In order to complete the MMR chart, this process is repeated for subgroups $i = 2$ - 50. Any Z_i

exceeding the UCL will have its corresponding subgroup $X_{i,}$ removed from the sample if no

assignable cause is found, thus establishing the IC reference sample for use in Phase II.

Using the control limits in Table 3.2.1, the next step is to compare the performance of the

MMR chart using both robust Mahalanobis depth and Mahalanobis spatial depth to the best

multivariate parametric Phase I alternative. All control charts will be tested on normal, heavy-

tailed, and skewed multivariate data, with both isolated and sustained shifts of the mean. Details

concerning the testing and evaluation process are provided in Chapter 4.

4   MMR Chart Performance Assessment Methodology

4.1   Introduction

To assess the effectiveness of the MMR chart as a distribution-free method of establishing an IC reference sample, its performance will be compared to an equivalent Phase I *parametric* multivariate method. If there were any other multivariate nonparametric or distribution-free Phase I methods in existence, they would also yield useful comparisons. However, the MMR chart appears to be the first in this class of control charts.

Because the MMR chart is a Shewhart-type chart, it must naturally be compared to another Shewhart-type chart for subgrouped multivariate data. From the literature review in Chapter 1, there is no clear consensus on the preferred Phase I parametric method. Because the original Hotelling's T^2 chart is the most common baseline performance measure for subsequently developed Phase I parametric multivariate methods, it will likewise be used as a basis of comparison for the distribution-free MMR chart.

4.2   Establishing Baseline Performance Using Hotelling's T^2 Chart

Constructing Hotelling's T^2 chart for a reference sample consisting of m subgroups of size n from a p-dimensional multivariate process requires first calculating unbiased estimates of the mean vector and covariance matrix. From Montgomery (2005, p. 495) the classical estimators

are

$$\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_i \tag{4.2.1}$$

and

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m}S_i, \tag{4.2.2}$$

where $\overline{\overline{X}}$ represents the average of the m subgroup mean vectors and $\overline{S}$ represents the average of the m subgroup covariance matrices. Using these estimated parameters, the control statistic is computed as

$$T_i^2 = n\left(\overline{X}_i - \overline{\overline{X}}\right)' \overline{S}^{-1}\left(\overline{X}_i - \overline{\overline{X}}\right). \tag{4.2.3}$$

The control statistic for each subgroup is compared to the Phase I UCL given by Alt's (1976) formula:

$$UCL_{T^2} = C(m,n,p)F_{\alpha,p,mn-m-p+1}, \text{ where } C(m,n,p) = \frac{p(m-1)(n-1)}{(mn-m-p+1)}. \tag{4.2.4}$$

In Equation (4.2.4) above, $F_{\alpha,p,mn-m-p+1}$ represents the $(1 - \alpha)$th percentile of the F distribution with p and $(mn - m - p + 1)$ degrees of freedom, and α is the desired IC FAP for each *individual* subgroup. In order to achieve a desired *overall* IC FAP for all m subgroups in a reference data set, α must be set as follows:

$$\alpha = 1 - \left(1 - \alpha_{overall}\right)^{1/m}, \tag{4.2.5}$$

where $\alpha_{overall}$ is the desired overall IC FAP. For example, for a reference sample consisting of m = 50 subgroups and a desired *overall* IC FAP of 0.05, $\alpha = 1 - (1 - 0.05)^{1/50} = 0.001025$ would be used in Equation (4.2.4) to determine the Phase I UCL.

Alt's (1976) formula given in Equation (4.2.4) was derived using the IC distribution of the T^2 statistic given in Equation (4.2.3) under the assumption of multivariate normally distributed data. Therefore, it is not appropriate for use when the distribution of the data is nonnormal because it will not result in the desired IC FAP. Having a common baseline level of performance is essential to a valid comparison of OC performance among all charts considered, so control limits for Hotelling's T^2 chart must be empirically adjusted when the data under study are nonnormally distributed. This will be accomplished using an algorithm similar to the one for determining MMR empirical control limits detailed in Chapter 3. Hotelling's T^2 empirical control limits used in this research are provided in Appendix E, and the MATLAB code used to determine them is provided in Appendix F.

4.3   Simulating Symmetric and Skewed Process Distributions

The MMR and Hotelling's T^2 charts will be tested on IC as well as mean-shifted data from normal, heavy-tailed, and skewed distributions with dimensions $p = 2$, 5, and 10. Due to affine equivariance of the mean vector and covariance matrix, multivariate normal data will be generated without loss of generality from the standard multivariate normal distribution, $N_p(\mathbf{0}, \mathbf{I})$, where $\mathbf{0}$ is a p-dimensional mean vector of all zeros and $\mathbf{I}$ is a p x p identity matrix. Heavy-tailed data will be represented by the multivariate t distribution, also using $\mathbf{I}_{p \, x \, p}$ as the covariance matrix. Variations of the multivariate t distribution will include both 10 and 3 degrees of freedom corresponding to increasingly fatter tails. Finally, skewed data will come from a

multivariate lognormal distribution, standardized to have zero mean vector and identity

covariance matrix. The data will be simulated using MATLAB code from the MathWorks

Statistics Toolbox at http://www.mathworks.com/help/toolbox/stats/. A summary of all planned

experiments is illustrated in Table 4.3.1.

Control Chart	Shift Type	Number/Size of Subgroups: m = 20, 50(50)200; n = 5											
		Process Distribution (in p = 2, 5, or 10 Dimensions):											
		normal			t (10)			t (3)			lognormal		
Hotelling's T^2 Chart	no shift	2	5	10	2			2	5	10	2	5	
	isolated shift	2	5	10	2			2	5	10	2	5	
	5/15/30% sustained shifts	2		10				2		10	2	5	
MMR-RMD Chart	no shift	2	5	10	2			2	5	10	2		
	isolated shift	2	5	10	2			2	5	10	2		
	5/15/30% sustained shifts	2		10				2		10	2		
MMR-MSD Chart	no shift	2			2			2	5		2	5	
	isolated shift	2			2			2	5		2	5	
	5/15/30% sustained shifts							2			2	5	

Table 4.3.1 Summary of Planned Experiments

4.4   Evaluating In-Control Performance

The MMR and Hotelling's T^2 charts will first be evaluated based on their ability to

maintain a desired IC FAP for subgrouped data from multivariate normal, multivariate t, and

multivariate lognormal distributions. It is expected that only the MMR chart, because it is

distribution-free, will be able to maintain the desired IC FAP across all combinations of sample

and subgroup sizes. Furthermore, IC performance of the MMR chart should be invariant to the

choice of depth function used.

The algorithm for these simulations, which will be performed in MATLAB, is as follows:

1) Simulate m subgroups of size n from a p-dimensional normal, t, or lognormal

distribution.

2) Establish the UCL for the MMR or Hotelling's T^2 chart.

3) Compute control chart statistics for each subgroup and compare to the UCL. If at least one control chart statistic exceeds the UCL, increment a counter by one.

4) Repeat steps 1 - 3 a total of 10,000 times.

5) Estimate the overall IC FAP = (final counter value)/10,000.

This process will be repeated for all desired combinations of m, n, p, process distribution, and control chart. MATLAB algorithms for simulating IC performance for the MMR and Hotelling's T^2 charts are provided in Appendix G and Appendix H, respectively.

4.5 Evaluating Out-of-Control Performance

Next, the MMR and Hotelling's T charts will be evaluated in terms of their ability to detect isolated and sustained shifts of the mean. An isolated shift of the mean is defined as a location shift occurring in a single subgroup of size n. Because the probability of detection is independent of the location of a shift within a data set, isolated shifts will take place in the first subgroup of each simulated data set without loss of generality. A sustained shift of the mean is defined as a location shift occurring in a certain percentage of the pooled sample of size N. Sustained shift percentages tested will include 5%, 15%, and 30%, and will take place at the end of each data set. Sustained shifts could be induced anywhere in the data set without loss of generality, but being at the end is most logical since it is unlikely that a process would go from an OC state back to an IC state without outside intervention.

The magnitude of the various shifts imposed will vary depending on the scenario being evaluated. This is because both the dimension of the data and the type of shift have a direct impact on the probability of a shift being detected. In general, all shifts are easier to detect in

lower dimensions than in higher dimensions, and sustained shifts are easier to detect than isolated shifts. The magnitude of a shift will be measured by the noncentrality parameter

$$\lambda = \sqrt{\delta \Sigma^{-1} \delta'},$$
(4.5.1)

where the process mean vector shifts from μ_o to $\mu_o + \delta$ and Σ is the process covariance matrix. Because the direction of a shift does not affect control chart performance with elliptically symmetric distributions, shifts will be fixed in the direction of $e_1 = [1, 0, ..., 0]$ without loss of generality [Stoumbos and Sullivan (2002), p. 265]. Shift directions for skewed distributions will be discussed in Section 4.6.

OC performance for a control chart will be quantified in terms of the empirical alarm probability (EAP), where EAP is defined as the estimated probability of a chart signaling at least once in an OC situation. Ideally, a control chart's EAP should be 100% for all scenarios involving induced location shifts. It is hoped that the MMR chart's performance will match that of Hotelling's T^2 chart for normally distributed data and surpass the T^2 chart's performance for nonnormally distributed data.

The algorithm for simulating OC performance is slightly different than the IC case, and is detailed as follows:

1) Simulate m subgroups of size n from a p-dimensional normal, t, or lognormal distribution.

2) Add isolated or sustained location shifts to the desired subgroups.

3) Establish the UCL for the MMR or Hotelling's T^2 chart.

4) Compute control chart statistics for each subgroup and compare to the UCL. If at least one control chart statistic exceeds the UCL, increment a counter by one.

5) Repeat steps 1 - 4 a total of 10,000 times.

59

6) Estimate the EAP = (final counter value)/10,000.

This process will be repeated for all combinations of *m*, *n*, *p*, process distribution, shift type, and control chart. MATLAB algorithms for simulating OC performance for the MMR and Hotelling's T^2 charts are also provided in Appendix G and Appendix H, respectively.

4.6   Evaluating Out-of-Control Performance with Skewed Data

Control chart performance with skewed distributions will be assessed using multivariate lognormally distributed data, simulated using the transformational relationship between the multivariate normal and the multivariate lognormal distributions. A *p*-dimensional multivariate lognormal random vector X can be represented as $X = \left(e^{Y_1}, e^{Y_2}, ..., e^{Y_p} \right)$, where Y is multivariate normal $N_p \left(\mu_Y, \Sigma_Y \right)$ [Law and Kelton (2000), p. 382]. Applying this transformation using a multivariate normal random vector Y with mean vector $\mu_Y = \left(\mu_1, \mu_2, ..., \mu_p \right)$ and covariance matrix Σ_Y with σ_{ij} the (*i*,*j*)th entry, the resulting multivariate lognormal random vector X has the following properties [Law and Kelton (2000), p. 382]:

$$E(X_i) = e^{(\mu_i + \sigma_{ii}/2)} \tag{4.6.1}$$

$$V(X_i) = e^{(2\mu_i + \sigma_{ii})} \left(e^{\sigma_{ii}} - 1 \right) \tag{4.6.2}$$

$$Cov(X_i, X_j) = \left(e^{\sigma_{ij}} - 1 \right) e^{\left(\mu_i + \mu_j + \frac{\sigma_{ii} + \sigma_{jj}}{2} \right)}. \tag{4.6.3}$$

Simulating multivariate lognormal observations is therefore simply a matter of generating $Y = \left(Y_1, Y_2, ..., Y_p \right) \sim N_p \left(\mu_Y, \Sigma_Y \right)$ and then evaluating $X = \left(e^{Y_1}, e^{Y_2}, ..., e^{Y_p} \right)$. Without loss of

generality, this research will use $Y \sim N_p(\mathbf{0}, \mathbf{I})$ to create multivariate lognormal data X having the following properties:

$$E(X_i) = e^{1/2} = 1.6487 \qquad (4.6.4)$$

$$V(X_i) = e(e-1) = 4.6708 \qquad (4.6.5)$$

$$Cov(X_i, X_j) = 0. \qquad (4.6.6)$$

In order to maintain consistency with other simulated distributions used in this research, the multivariate lognormal data X will be standardized using $X_i = (X_i - \mu_X)\Sigma_X^{-1/2}$, where μ_X is a 1 x p mean vector with all entries equal to 1.6487 and Σ_X is a p x p covariance matrix with diagonal entries equal to 4.6708 and zeros everywhere else, resulting in $X \sim$ lognormal $(\mathbf{0}, \mathbf{I})$.

Once the multivariate lognormal data are simulated, isolated and sustained shifts will be induced to evaluate OC performance. As noted by Stoumbos and Sullivan (2002, p. 265), while the direction of a shift has no effect on control chart performance with elliptically symmetric distributions, it can substantially affect a control chart's detection power with skewed distributions. One method of handling this is to focus on the shift direction having the most dramatic effect on control chart performance, but this is a difficult task because there are an infinite number of shift directions from which to choose in a multivariate setting [Stoumbos and Sullivan (2000), p. 267]. Even if the most impactful shift direction could be determined, its odds of occurring in practice are unknown. As pointed out by J. Sullivan (personal communication, February 2, 2011), there is no guidance found in the literature regarding the likelihood of certain shift directions occurring, so a better approach is to assume that all shift directions are equally probable. Under this assumption, as done by Stoumbos and Sullivan (2000), the effects of shift directions randomly generated over a uniform distribution will be averaged.

The shift directions will be generated using an algorithm proposed by Johnson (1987, p. 127), who stated that a p-dimensional shift can be created by first generating p independent standard normal random variates $Z_1, Z_2, ..., Z_p$. Next, the shift vector δ which follows a uniform distribution on the p-sphere is computed using

$$\delta_i = \frac{Z_i}{\left(Z_1^2 + ... + Z_p^2\right)}, \quad i = 1, 2, ..., p. \tag{4.6.7}$$

A different δ will be generated for each of the 10,000 iterations of the simulation, and the results will be averaged at the conclusion of the simulation. In two dimensions, this method of creating shift vectors is analogous to randomly generating a series of unit vectors which emanate from the origin and terminate along the boundary of the unit sphere.

As with elliptically symmetric distributions, the magnitude of the various shifts imposed will be measured by the noncentrality parameter given in Equation (4.5.1), where the multivariate lognormal process mean vector shifts from μ_o to $\mu_o + \delta$ and Σ is the asymptotic covariance matrix of the multivariate lognormal process. With δ as defined by Equation (4.6.7) and Σ equal to the identity matrix, λ always equals one. In order to induce shifts corresponding to $\lambda \neq 1$, the shift vector δ resulting from Equation (4.6.7) must be multiplied by the desired λ, thus shortening or lengthening the unit vector to achieve the desired λ.

For example, suppose it is desired to induce a shift of size $\lambda = 3$ into a bivariate lognormal process with identity covariance matrix. Using Equation (4.6.7), a possible shift vector is $\delta = [-0.7468, 0.6651]$. If this shift vector is applied directly to the process without any scaling, the magnitude of the resulting shift is $\lambda = \sqrt{\delta \Sigma^{-1} \delta'} = 1$. However, using

$3\delta = [-2.2404, 1.9953]$ produces the desired result of $\lambda = \sqrt{3\delta \Sigma^{-1} 3\delta'} = 3$. This methodology

will be employed for all simulations involving OC conditions in multivariate lognormally

distributed data. Once all simulations have been completed and results analyzed,

recommendations will be provided on how best to proceed in a Phase I multivariate quality

control scenario when a process distribution is normal, heavy-tailed, or skewed.

5   MMR Chart Performance Comparisons

5.1   Introduction

MMR chart performance comparisons to Hotelling's T^2 (HT2) chart were focused primarily on $m = 20, 50(50)200$ subgroups of size $n = 5$. The number of subgroups was chosen to be relatively small because a Phase I analysis often occurs early in the life of a process when very little historical data is available. A subgroup size of five was chosen because Jones-Farmer et al. (2009) showed that this is the minimum subgroup size necessary for reliable univariate mean-rank chart performance, and further testing using the MMR chart confirmed this to be true in the multivariate case as well. Limited experimentation was conducted using subgroup sizes $n = 5(5)20$ in order to demonstrate the enhancing effect of larger subgroup sizes on MMR chart performance. In all simulations, the desired IC FAP was set to 0.10, but the results can be generalized to other common IC FAPs such as 0.05.

5.2   MMR Chart Performance with Symmetric Distributions

Symmetric distributions tested include the multivariate normal, $t(10)$, and $t(3)$ distributions. When evaluating IC performance of Hotelling's T^2 chart, Alt's (1976) Phase I UCL was used for all process distributions. For OC assessments, Alt's (1976) Phase I UCL was used for the multivariate normal case only, and empirically adjusted UCLs were used for the $t(10)$ and $t(3)$ cases. RMD was the primary depth function used in the MMR chart because it is well-suited

for elliptically symmetric distributions and one of the simplest depth functions to compute, but MSD was implemented in a few cases for comparison purposes. Simulation results show that when data are normally or nearly normally distributed, a normal-theory method such as Hotelling's T^2 chart is preferred. However, when data are heavy-tailed as with the $t(3)$ distribution, the distribution-free MMR chart is usually a superior alternative.

5.2.1   In-Control Performance with Symmetric Distributions

The fundamental advantage of a distribution-free control chart is its ability to maintain a desired IC FAP for any process distribution. Accordingly, the MMR chart using both RMD and MSD was first compared to Hotelling's T^2 chart using IC bivariate normal, $t(10)$, and $t(3)$ processes with a desired IC FAP of 0.10. For these comparisons, Hotelling's T^2 chart was constructed using only Alt's (1976) Phase I UCL given by Equation (4.2.4), adjusted for the number of subgroups using Equation (4.2.5), in order to demonstrate the effects of applying a normal-theory method to both normally and nonnormally distributed data.

As indicated in Figure 5.2.1, Hotelling's T^2 chart maintains the desired IC FAP for the bivariate normal process, but becomes progressively worse as the distribution deviates from normality and the number of subgroups is increased. For a bivariate $t(3)$ process, the IC FAP for Hotelling's T^2 chart using Alt's (1976) Phase I UCL ranges from approximately 30% when $m = 20$ to over 90% when $m = 200$. This is why, for OC assessments with nonnormally distributed data, the UCL for Hotelling's T^2 chart must be empirically tailored to achieve the desired IC FAP of 0.10 for each (m, n) combination and process distribution studied. Although this is impracticable outside of a simulation environment because it requires knowing the exact process distribution, it is necessary in order to ensure a common basis of comparison for all charts

included in OC performance comparisons. The MMR chart, on the other hand, consistently

maintains the desired IC FAP for all process distributions and any number of subgroups. This

holds true regardless of the data depth measure used, so no adjustments to the MMR chart UCLs

given in Table 3.2.1 are necessary.
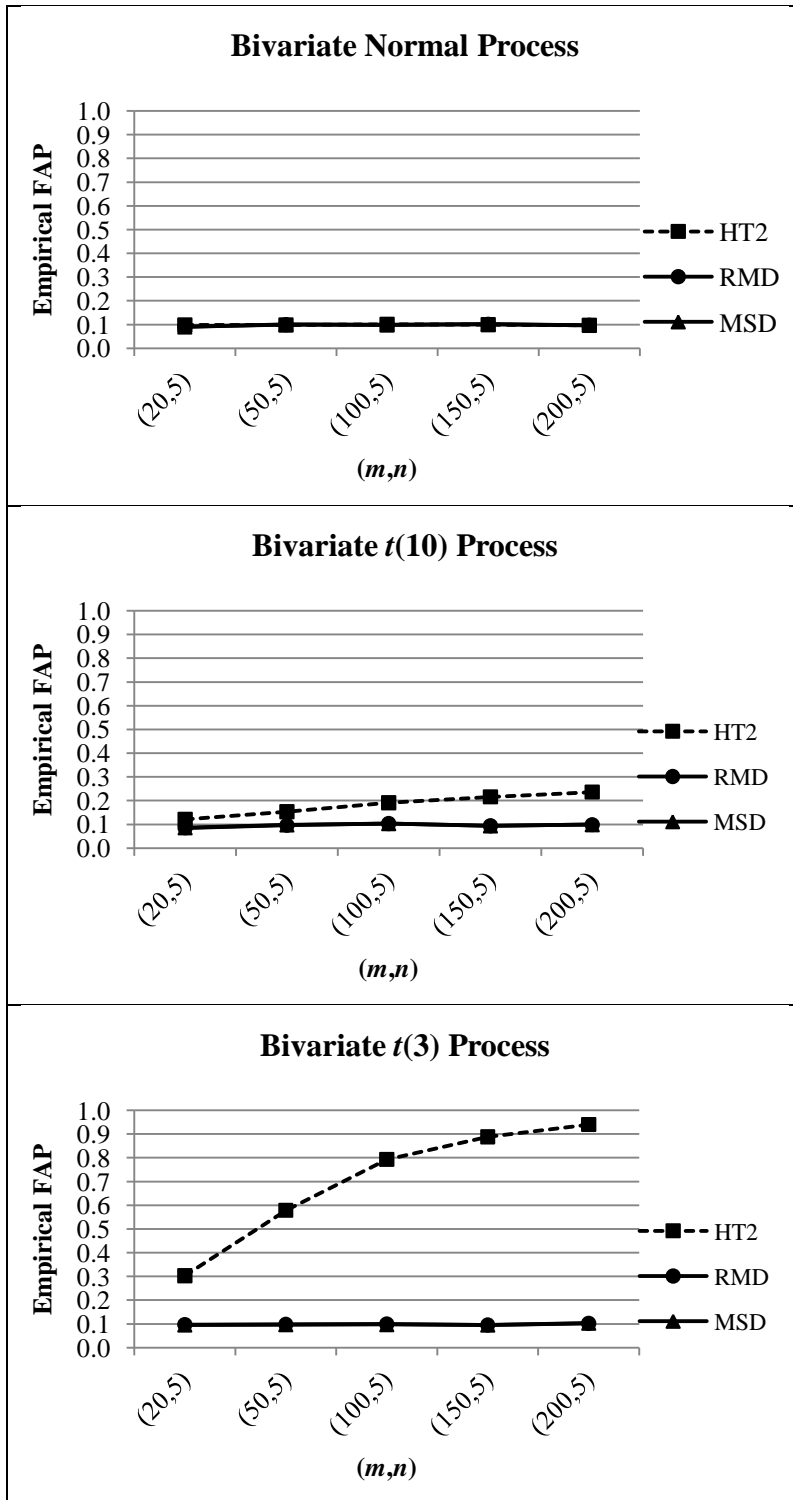
Figure 5.2.1 Empirical IC FAPs for Symmetric Bivariate Distributions

Figure 5.2.2 shows the effects of dimensionality on control chart performance using a $t(3)$ process. Again, the MMR chart consistently maintains the desired IC FAP for any number of subgroups m and any dimension p. Hotelling's T^2 chart becomes distinctly worse in higher dimensions, reaching empirical IC FAPs near 100% for all but the smallest number of subgroups considered when $p = 10$. These results show that the MMR chart is distribution free in any dimension when applied to elliptically symmetric data using RMD, MSD, or presumably any other depth function with similar statistical properties. A complete table of IC performance data for symmetric distributions is provided in Appendix I.
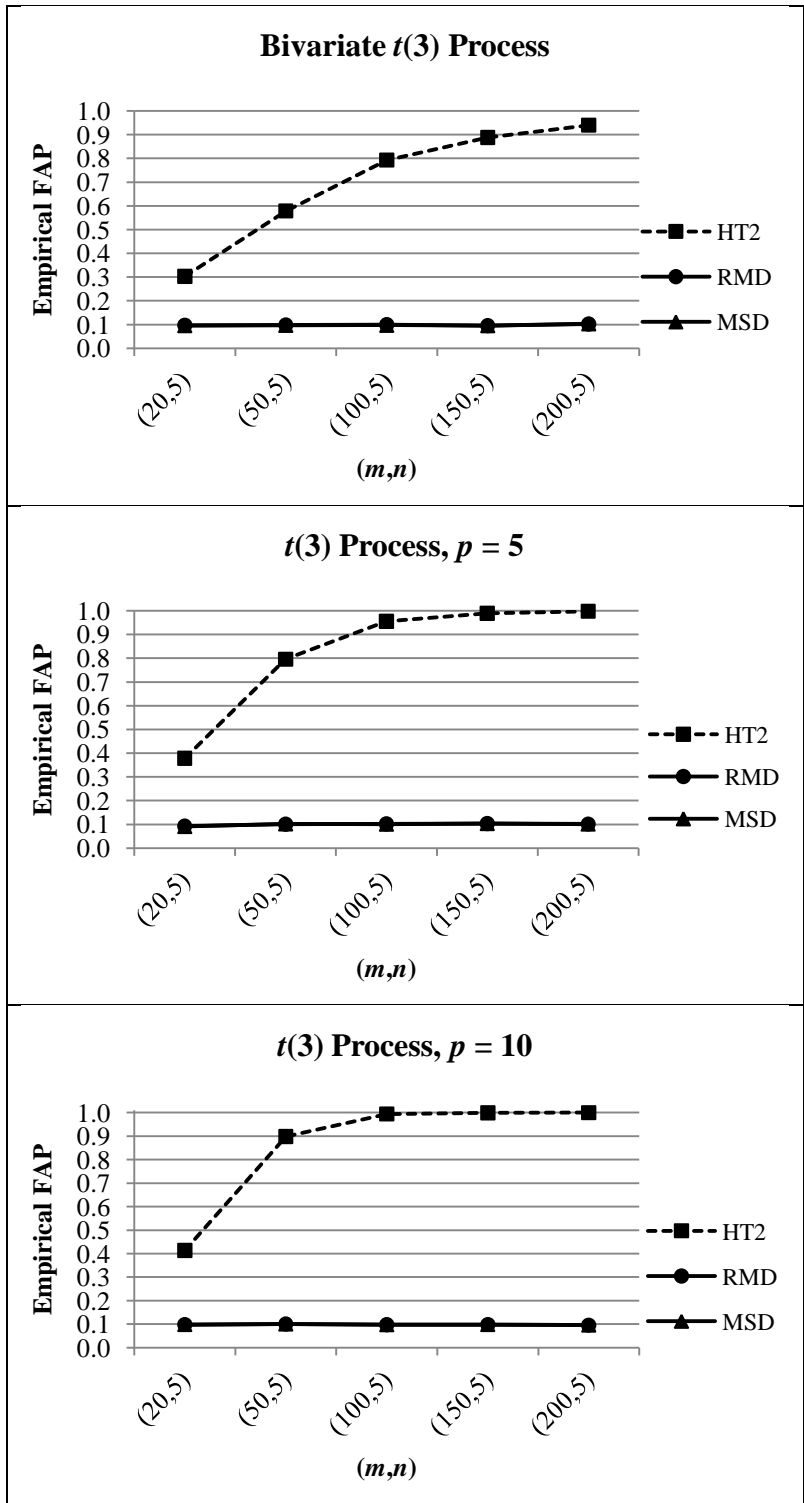
Figure 5.2.2 Empirical IC FAPs for *t*(3) Processes in Higher Dimensions

5.2.2 Isolated Shifts of the Mean with Symmetric Distributions

MMR-RMD and Hotelling's T^2 chart performance for isolated shifts in two dimensions for (m, n) combinations (20, 5), (100, 5), and (200, 5) are shown in Figure 5.2.3. Hotelling's T^2 chart using Alt's (1976) Phase I limits is superior in the case of bivariate normally distributed data, as expected. For slightly nonnormal data following a bivariate $t(10)$ distribution, Hotelling's T^2 chart with empirically adjusted UCL maintains a smaller but still notable advantage over the MMR-RMD chart. For heavy-tailed process data following a bivariate $t(3)$ distribution depicted in the bottom panel of Figure 5.2.3, however, the MMR-RMD chart is both significantly better and much more consistent than Hotelling's T^2 chart in terms of EAP. The two control charts are roughly equivalent when $m = 20$, but the performance of Hotelling's T^2 chart declines dramatically as m is increased to 200, whereas MMR chart performance is far less affected when the number of subgroups is increased. For example, in the case of bivariate $t(3)$ data with an isolated shift of magnitude $\lambda = 6$, EAPs for $m = 20$, 100, and 200 are approximately 100% using the MMR-RMD chart as compared to 100%, 92%, and 46%, respectively, for Hotelling's T^2 chart with empirical UCL.
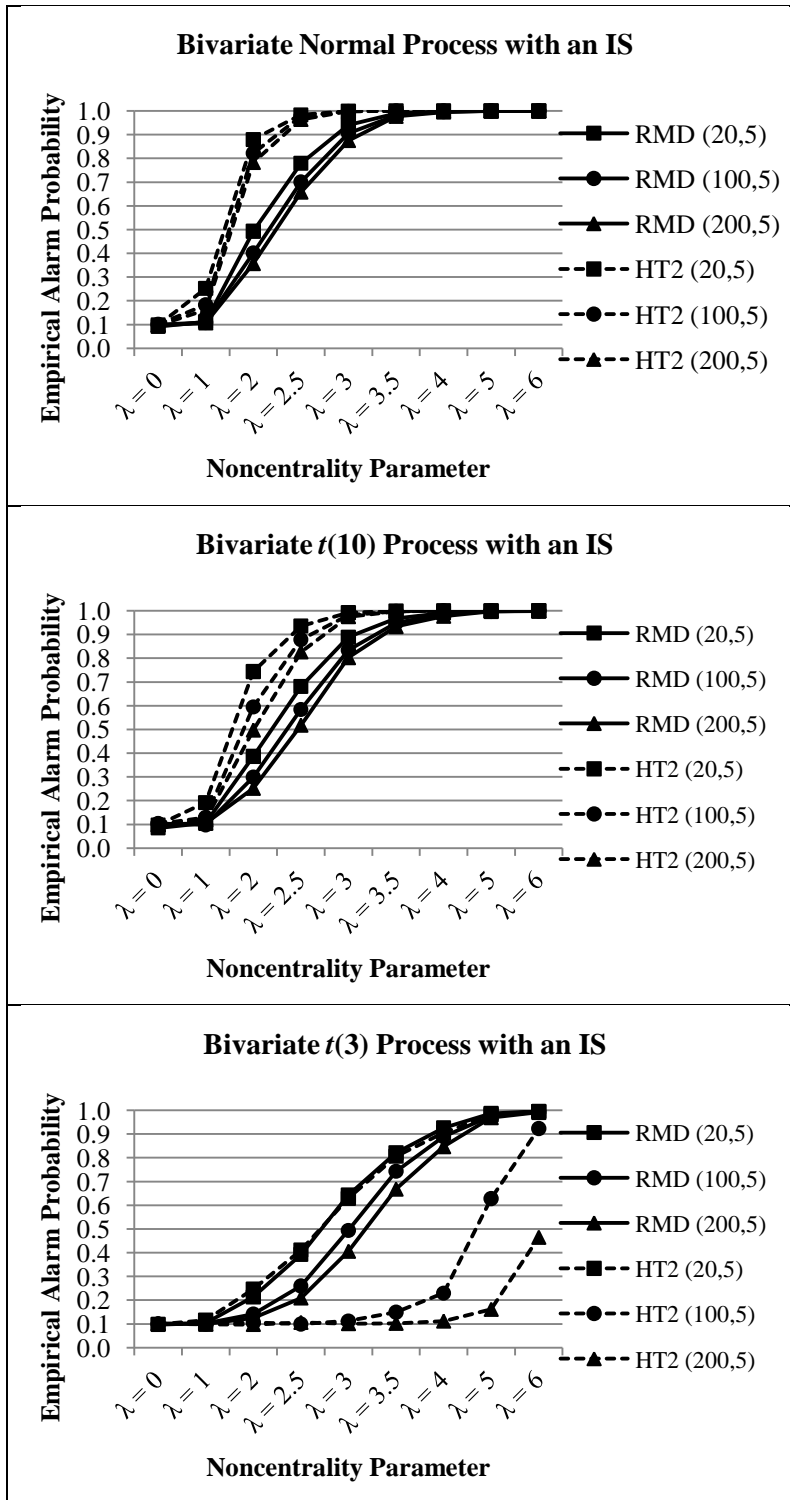
Figure 5.2.3 Control Chart Performance on Symmetric Bivariate Data with an IS

MMR chart performance for isolated shifts is relatively invariant to the choice of depth function. Figure 5.2.4 shows the application of an MMR chart using both RMD and MSD to a bivariate $t(3)$ process with (m, n) combinations $(20, 5)$ and $(200, 5)$. The MMR-RMD chart has a slight advantage over the MMR-MSD chart for $m = 20$ subgroups, but the two charts are nearly identical in terms of EAP when $m = 200$. Repeating this analysis using other symmetric distributions yielded similar results in both two and five dimensions.
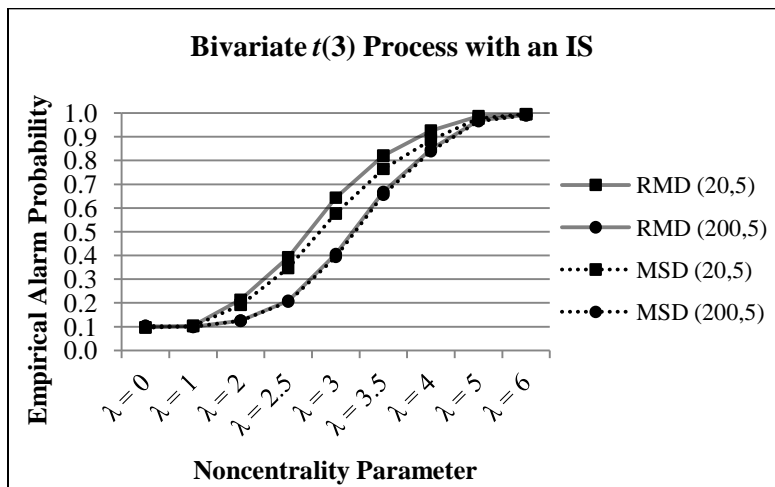
Figure 5.2.4 MMR-RMD/MSD Chart Performance on $t(3)$ Data with an IS

The MMR chart loses some power to detect isolated shifts of the mean as the dimension of the data is increased, but it retains clear superiority over Hotelling's T^2 chart in most scenarios considered. When applied to a heavy-tailed process represented by the $t(3)$ distribution, the MMR-RMD chart is a substantially better alternative for $m \geq 50$ in five dimensions and for $m \geq 100$ in ten dimensions. This is illustrated in Figure 5.2.5.
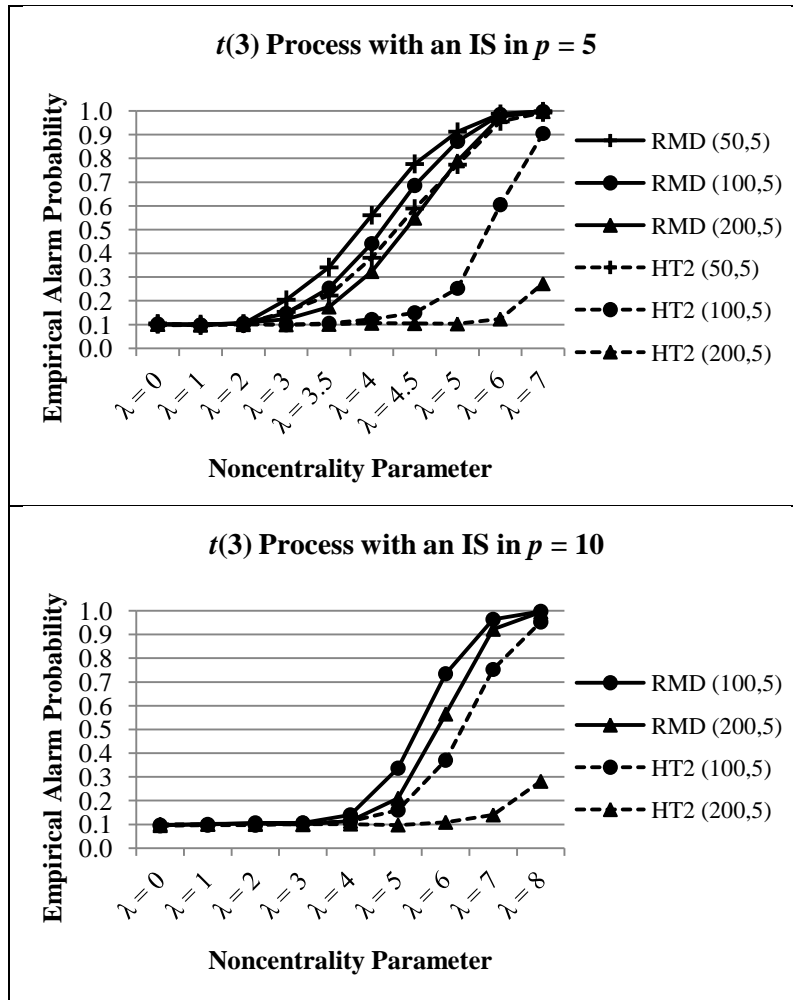
Figure 5.2.5 Control Chart Performance on *t*(3) Data with an IS in Higher Dimensions

Complete tables of results for all simulations performed using symmetric distributions with isolated shifts of the mean are provided in Appendices J - L.

5.2.3   Sustained Shifts of the Mean with Symmetric Distributions

The MMR-RMD chart is generally superior to Hotelling's T^2 chart in detecting sustained shifts of the mean in a bivariate *t*(3) process, although some loss of power is observed as the level of contamination in the sample is increased. Figure 5.2.6 depicts control chart performance for sustained mean shifts composing 5%, 15%, and 30% of the total data sets. For a 5%

contamination level, MMR-RMD chart performance matches Hotelling's T^2 chart performance for $m = 20$ and surpasses it by an increasing margin as m is increased from 50 to 200. Similar trends are observed for a 15% level of contamination, but $m \geq 50$ subgroups are necessary for MMR-RMD chart performance to exceed that of Hotelling's T^2 chart. When the level of contamination is raised to 30%, $m \geq 150$ subgroups are necessary for the MMR-RMD chart to consistently outperform Hotelling's T^2 chart.

For each sustained shift scenario considered, MMR-RMD chart performance is remarkably consistent when at least 50 subgroups are present. For example, in the 15% sustained shift scenario depicted in the middle panel of Figure 5.2.6, the lines representing MMR-RMD chart performance for $m = 50$, 100, and 200 subgroups are nearly coincident. On the other hand, Hotelling's T^2 chart performance declines rapidly as the number of subgroups is increased. However, the fact that the overall detection power of the MMR chart declines as the level of contamination is raised from 5% to 30% is counterintuitive, as one would expect the opposite to hold true. This is shown in Section 5.5 to be an unavoidable consequence of a rank-based control charting method.
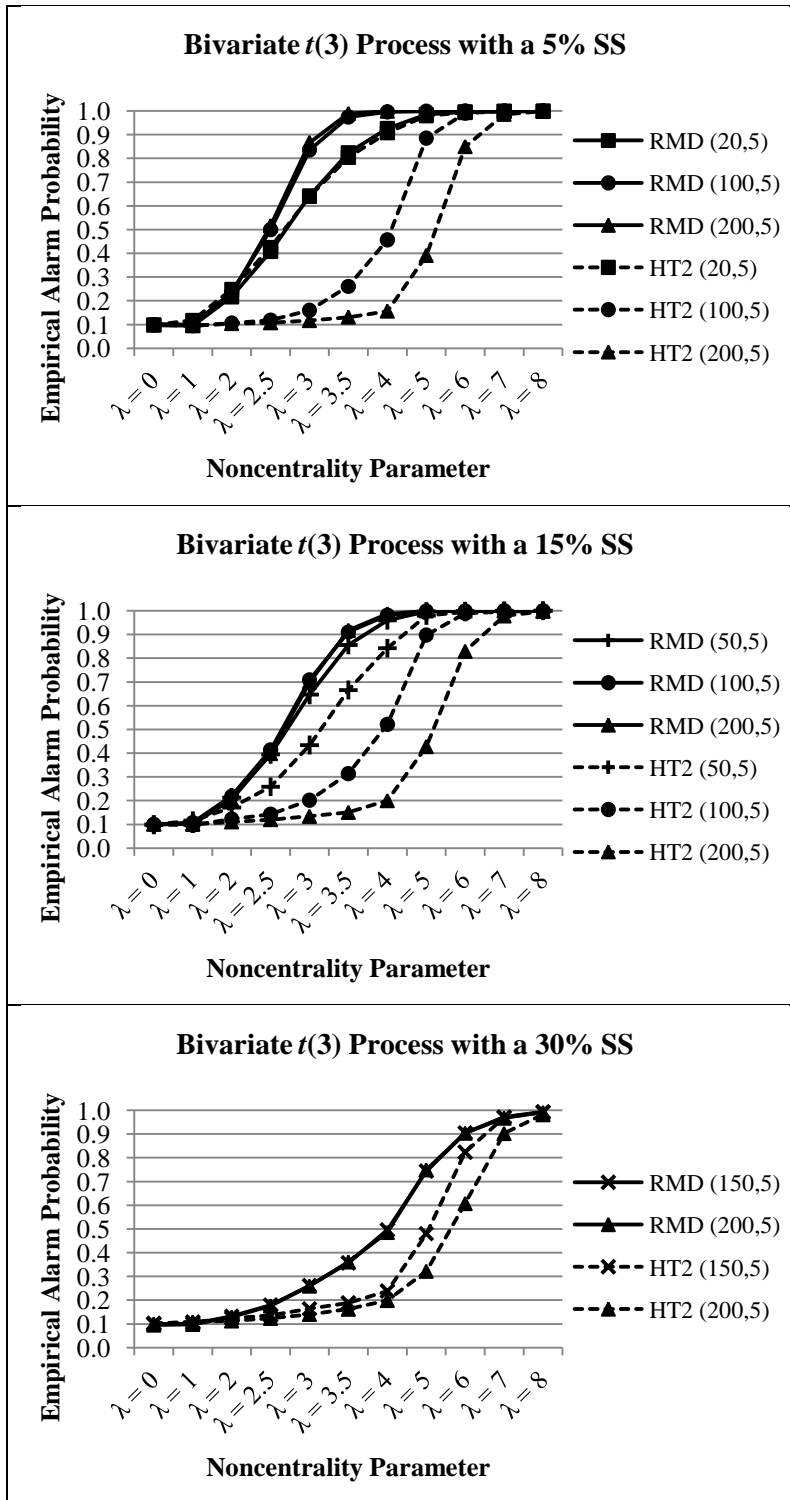
Figure 5.2.6 Control Chart Performance on Increasingly Contaminated Bivariate *t*(3) Data

For the MMR chart, RMD is a more effective depth measure than MSD in the presence of a sustained mean shift in a bivariate $t(3)$ process. MMR-MSD chart detection power lags only slightly behind MMR-RMD chart performance under a 5% contamination level, but falls farther behind when the contamination is increased to 15% and becomes unacceptably low at the 30% contamination level. This effect is illustrated in Figure 5.2.7. Based on these results, RMD is clearly the preferred depth measure for the MMR chart when data are elliptically symmetric.
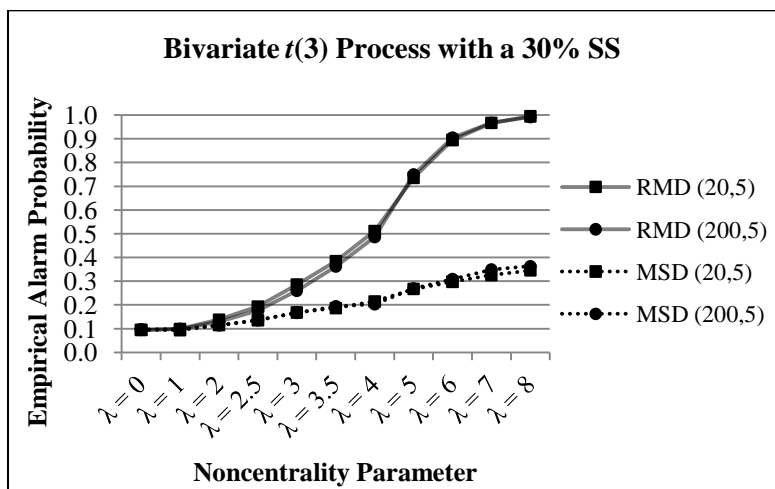
Figure 5.2.7 MMR-RMD/MSD Chart Performance on Bivariate $t(3)$ Data with a 30% SS

As with isolated shifts of the mean, the MMR chart's ability to detect sustained shifts of the mean is somewhat degraded as the dimension of the data is increased. In the ten-dimensional $t(3)$ process with a 15% sustained mean shift shown in Figure 5.2.8, the MMR-RMD chart matches or exceeds Hotelling's T^2 chart performance for $m \geq 100$, in contrast to $m \geq 50$ required for the bivariate $t(3)$ case depicted in the middle panel of Figure 5.2.6. Similar results are seen with 5% and 30% sustained shifts of the mean imposed upon a ten-dimensional $t(3)$ process.
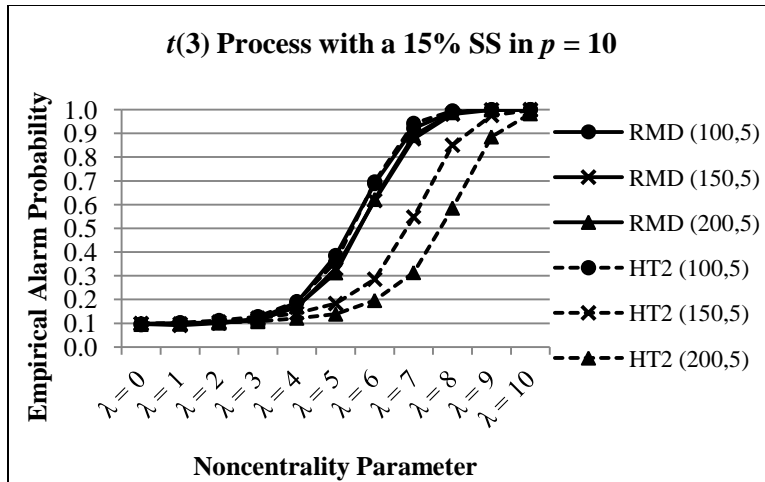
Figure 5.2.8 Control Chart Performance on $t(3)$ Data with a 15% SS in $p = 10$

Complete tables of results for all simulations performed using symmetric distributions with sustained shifts of the mean are provided in Appendices M - R. In addition, a matrix of recommended control chart usage with heavy-tailed multivariate data under both isolated and sustained shifts of the mean is provided in Table 5.2.1. Although Hotelling's T^2 chart outperforms the MMR-RMD chart for all scenarios in which $m \leq 50$, $n = 5$, and $p = 10$, even $m = 50$ subgroups of size $n = 5$ would be considered an exceptionally small reference sample for a ten-dimensional process. The MMR-RMD chart is a better alternative than Hotelling's T^2 chart for most of the more realistic scenarios considered in ten dimensions. Furthermore, for any scenario in which Hotelling's T^2 chart outperforms the distribution-free MMR chart, it should be reiterated that its implementation requires empirically adjusted UCLs based on the exact distribution of the process under study. Since the process distribution is unlikely to be known in practice, another control charting technique must be sought for scenarios in Table 5.2.1 labeled "HT2."

p	Contamination Level	(m,n)				
		(20,5)	(50,5)	(100,5)	(150,5)	(200,5)
2	IS			MMR-RMD		
	5% SS					
	15% SS					
	30% SS	HT2				
10	IS		HT2		MMR-RMD	
	5% SS					
	15% SS					
	30% SS					

Table 5.2.1 Recommended Phase I Control Chart Usage for Heavy-Tailed Data

5.3 MMR Chart Performance with Skewed Data

The multivariate lognormal distribution was the lone skewed distribution tested. As with the symmetric distributions evaluated, Hotelling's T^2 chart was used in conjunction with Alt's (1976) Phase I UCL for the IC case and with empirically adjusted UCLs for the OC scenarios. Most MMR charts were created using MSD, since MSD was expected to outperform RMD on skewed process data. Performance comparisons were focused on $m = 20$, 100, and 200 subgroups and dimensions $p = 2$ and 5 because MSD, although quickly computable for a single data set, is considerably more time consuming than RMD when performing 10,000 replications. Simulation results show that when data are skewed, the distribution-free MMR chart almost always represents the best available control charting methodology.

5.3.1 In-Control Performance with Skewed Data

In order to validate its performance as a distribution-free method when process data are skewed, MMR and Hotelling's T^2 charts were first applied to IC lognormal processes in both two and five dimensions using a desired IC FAP of 0.10. As with the symmetric distributions tested,

Hotelling's T^2 chart was constructed using Alt's (1976) Phase I UCL given by Equation (4.2.4), adjusted for the number of subgroups using Equation (4.2.5), in order to demonstrate the negative consequences of applying a normal-theory method to skewed data. The results of the IC performance analysis are illustrated in Figure 5.3.1.
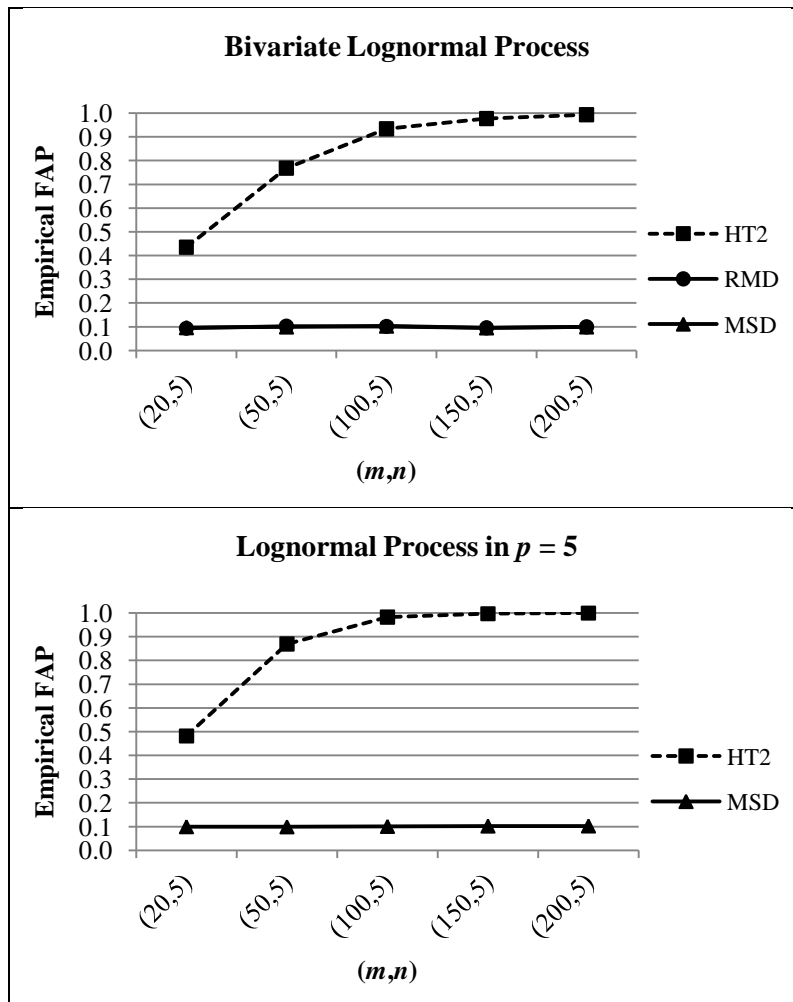
Figure 5.3.1 Empirical IC FAPs for Lognormal Processes in $p = 2$ and $p = 5$

Hotelling's T^2 using Alt's (1976) Phase I UCL chart categorically fails to maintain the desired IC FAP for multivariate lognormal processes. The IC FAP for Hotelling's T^2 chart ranges from approximately 43% to 99% in two dimensions and from approximately 48% to

100% in five dimensions. In contrast, the MMR charts using RMD and MSD with the UCLs

from Table 3.2.1 consistently maintain the desired IC FAP of 0.10 for all (m, n) combinations

considered, solidifying its characterization as a distribution-free method. A complete table of IC

performance data for skewed data is provided in Appendix S.

5.3.2   Isolated Shifts of the Mean with Skewed Data

The performance of MMR-MSD and Hotelling's T^2 charts under isolated shifts of the

mean in bivariate lognormally distributed data is displayed in Figure 5.3.2. Even with UCLs

empirically adjusted to achieve an IC FAP of 0.10, Hotelling's T^2 chart performance deteriorates

rapidly for $m > 20$. The MMR chart not only outperforms Hotelling's T^2 chart by a wide margin,

but its performance is extremely consistent for all m.
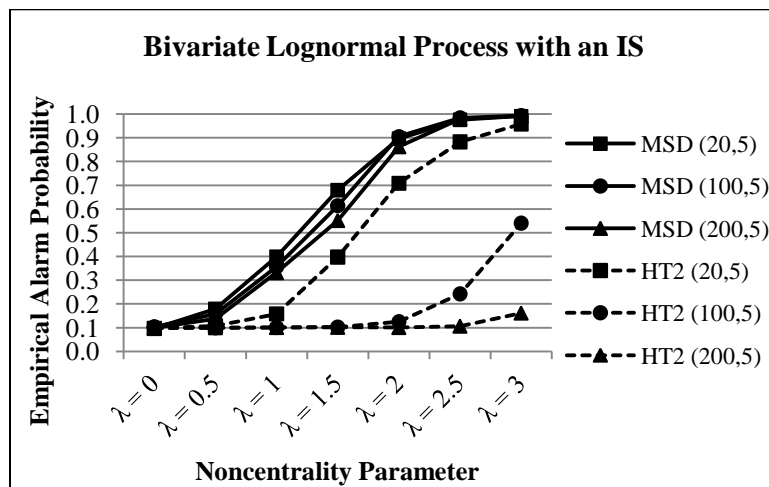
Figure 5.3.2 Control Chart Performance on Bivariate Lognormal Data with an IS

In Figure 5.3.3, the previous scenario is repeated for $m = 100$, $n = 5$ using the MMR-

RMD chart in order to compare the performance of MSD and RMD as depth functions. As

originally hypothesized, it is seen that the MMR-MSD chart detects smaller isolated shifts with

higher probability than the MMR-RMD chart, and offers equivalent performance in the case of larger shifts.
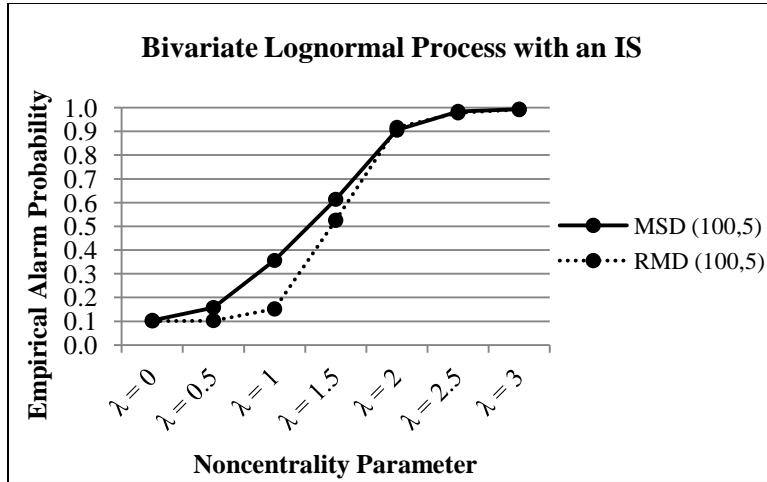
Figure 5.3.3 MMR-MSD/RMD Chart Performance on Bivariate LGN Data with an IS

Although the MMR chart's gradual loss in power with symmetric distributions in higher dimensions is also observed with skewed data, it remains notably better than Hotelling's T^2 chart. In the five-dimensional scenario depicted in Figure 5.3.4, the MMR-MSD chart matches Hotelling's T^2 chart performance for $m = 20$ and dominates for $m > 20$, making it clearly the best alternative for detecting isolated shifts occurring in skewed process data when $p \leq 5$.

Lognormal Process with an IS in *p*=5
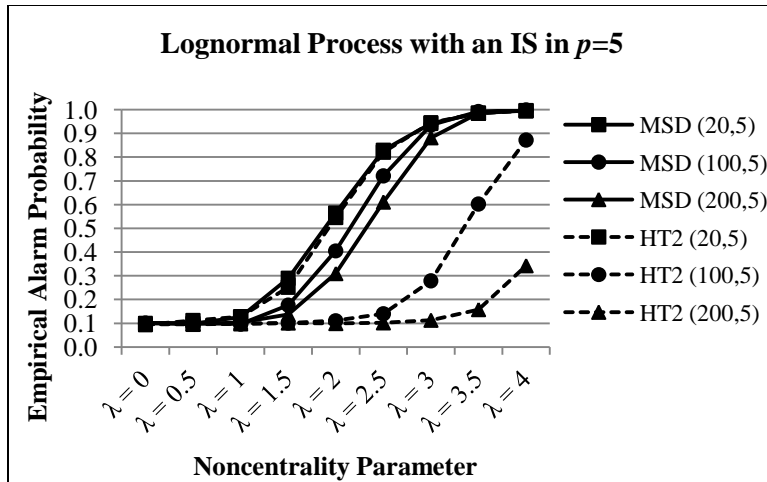
Figure 5.3.4 Control Chart Performance on LGN Data with an IS in *p* = 5

Complete tables of results for all simulations performed using the multivariate lognormal distribution with isolated shifts of the mean are provided in Appendices T and U.

5.3.3   Sustained Shifts of the Mean with Skewed Data

MMR-MSD chart performance in detecting sustained shifts of the mean in skewed bivariate data varies greatly with the percentage of data shifted. As shown in Figure 5.3.5, the MMR-MSD chart is universally more powerful than Hotelling's T^2 chart in detecting 5% and 15% sustained shifts in a bivariate lognormal process and demonstrates very consistent performance across the range of *m* considered. However, a different story is seen with a 30% sustained shift of the mean, as MMR-MSD chart performance falls to unacceptable levels.
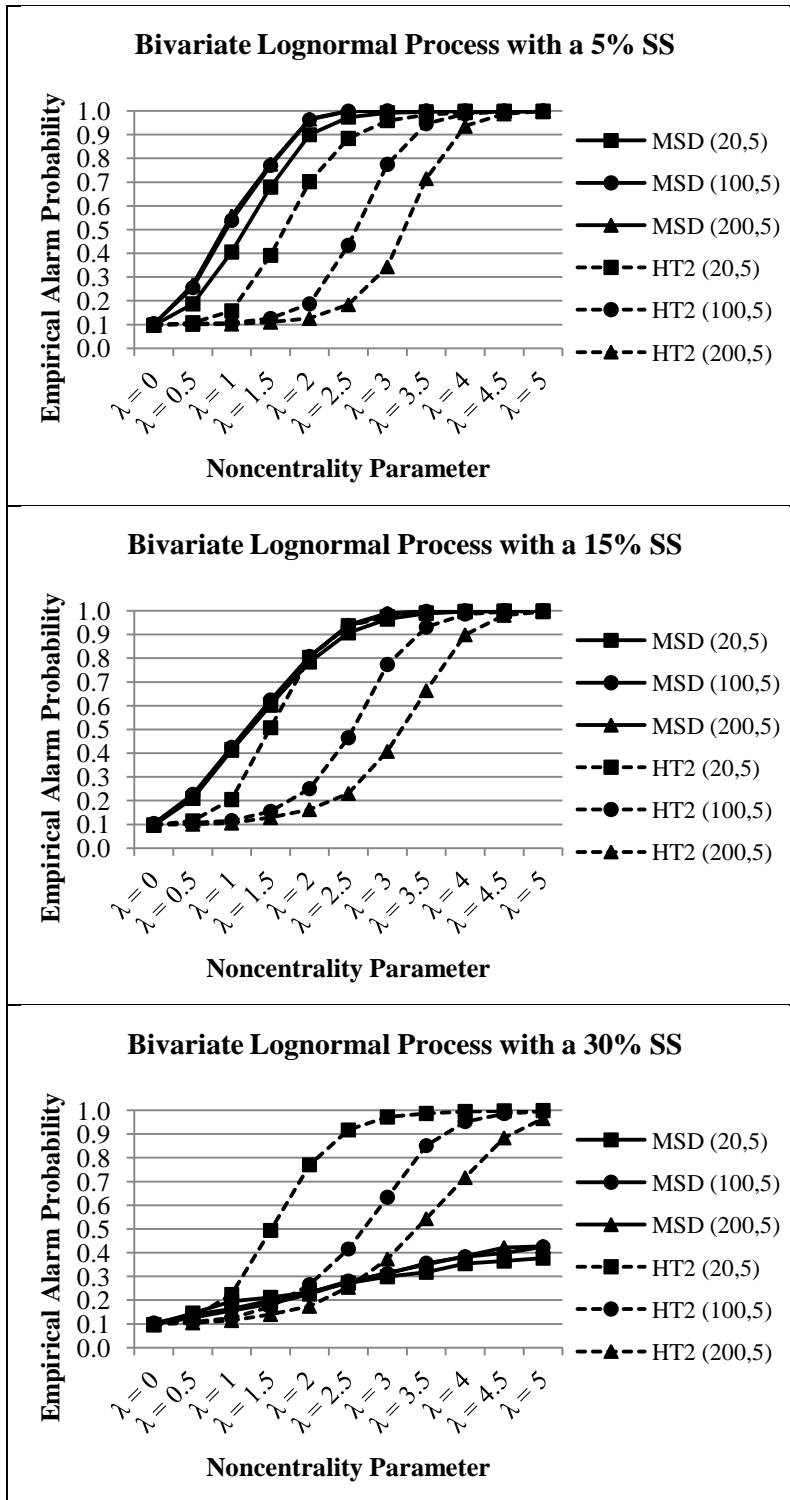
Figure 5.3.5 Control Chart Performance on Increasingly Contaminated LGN Data in $p = 2$

Further testing revealed that the MMR-MSD chart is robust to sustained shifts of the mean in skewed bivariate data with contamination levels up to approximately 20%. MMR-MSD chart performance for $m = 100$, $n = 5$ and contamination levels 5(5)30% is illustrated in Figure 5.3.6.
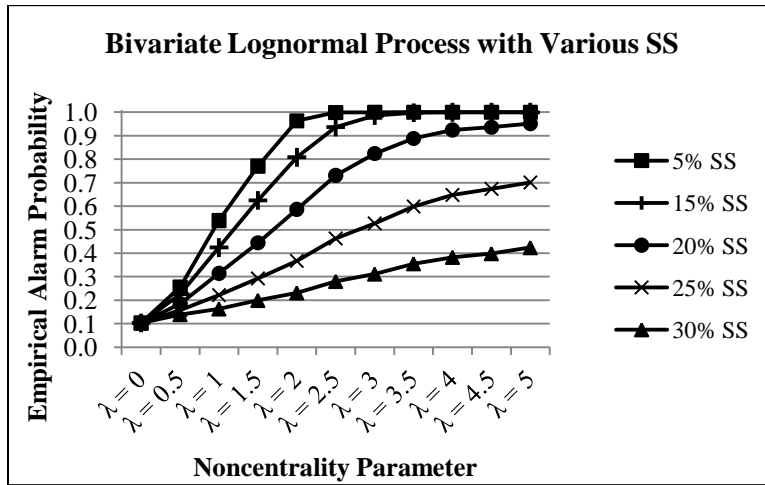
Figure 5.3.6 MMR-MSD Chart Performance on Increasingly Contaminated LGN Data

Surprisingly, this happens despite the fact that the multivariate median determined by the MSD function has RBP equal to 1/2, indicating a high degree of robustness to outliers. However, as noted by R. Serfling (personal communication, June 6, 2011), the RBPs of other quantiles determined by MSD decrease from the median outward. This means that when a high percentage of a data set is shifted, even though the center of the data is well estimated by MSD, the overall center-outward ordering may be adversely affected by outlying points.

In order to demonstrate this, a simple example using bivariate lognormal data with a 30% randomly directed, sustained mean shift with $\lambda = 4$ is presented in Figure 5.3.7. For illustrative purposes, only 20 individual observations are simulated, so 14 points are IC and 6 are OC. In Figure 5.3.7, each bivariate data point is labeled with two numbers representing ranks

84

determined by the MSD function and the RMD function, respectively. Ideally, the IC points should be labeled with the highest ranks from 1 (most central) to 14, and the OC points should be labeled with the lowest ranks from 15 to 20 (most outlying).
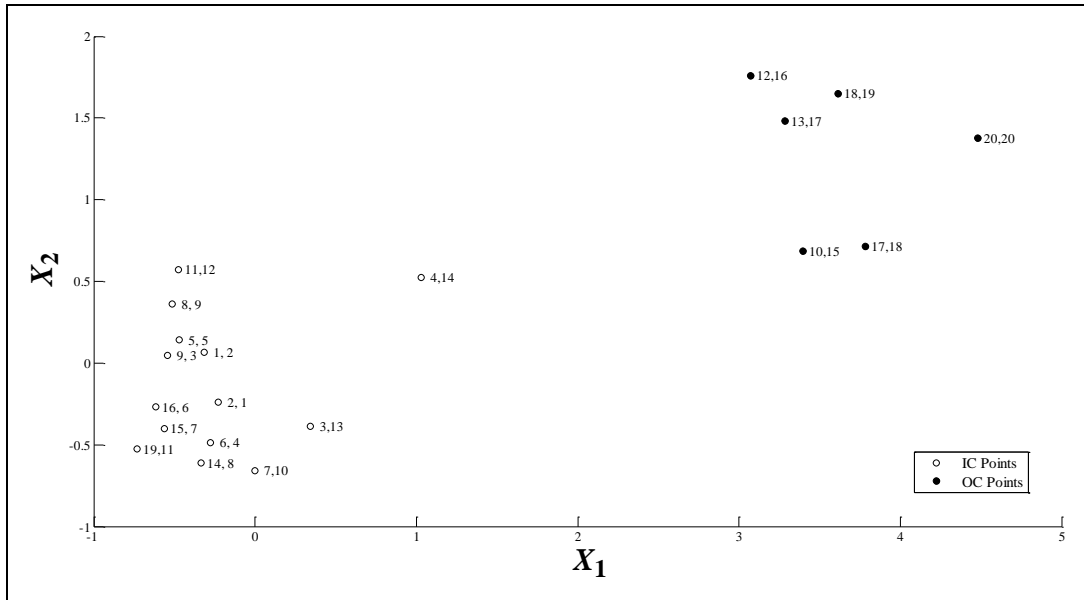
Figure 5.3.7 MSD and RMD Rankings for Bivariate LGN Data with a 30% SS

This holds true for the ranks determined by the RMD function, but is clearly not the case with MSD. Close scrutiny of Figure 5.3.7 reveals that the center determined by the MSD function is relatively close to the center determined by the RMD function, but the similarities end there. The MSD function assigns the lowest ranks (indicating outlyingness) to points along the outer limits of entire data cloud consisting of both IC and OC points. This disrupts the entire ranking scheme, resulting in several IC points being assigned ranks which suggest outlyingness, and several OC points receiving ranks which indicate centrality. For example, consider the point located at approximate coordinates (-0.75, -0.50). This point is assigned a rank of 19 by the MSD function, which indicates a high degree outlyingness, and a distinctly different rank of 11 by the RMD function, which strongly suggests that the point belongs to the IC cluster.

85

Further analysis was performed by simulating $m = 200$ subgroups of size $n = 5$ of

bivariate lognormal data with both 5% and 30% randomly directed, sustained mean shifts with λ

$= 4$. For each scenario, a scatterplot was constructed of ranks determined by the MSD function

versus ranks determined by the RMD function. A straight line was drawn to represent the path

the plotted ranks would follow if both depth functions generated equivalent rankings for each

observation. Results are provided in Figure 5.3.8.
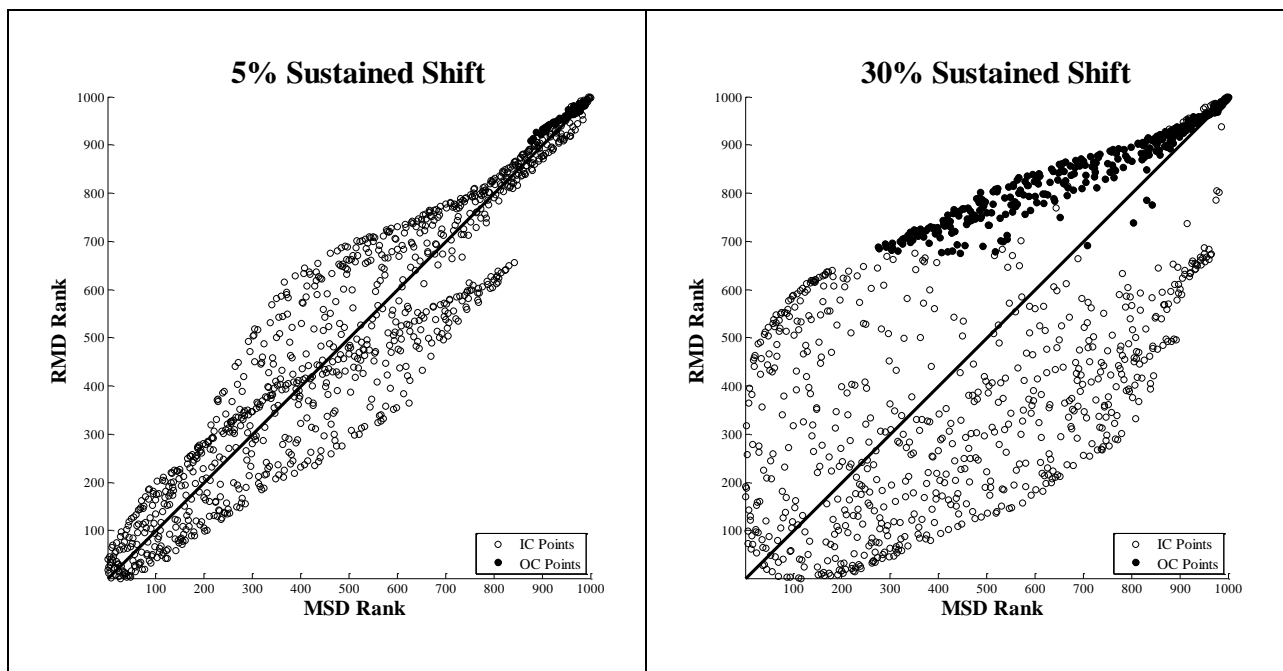
Figure 5.3.8 Scatterplots of MSD vs. RMD Ranks for Shifted Bivariate LGN Data

In the case of a 5% sustained shift, MSD and RMD rankings are in general agreement for

the lowest and highest rankings, and there is a moderate amount of variation in the middle.

However, with a 30% sustained shift, the differences in depth functions become more apparent.

There is much more variability overall, but the rankings assigned to the OC points are especially

notable. The MSD function consistently assigns higher rankings to the OC points than does the

RMD function, as evidenced by the fact that most of the OC points fall well above the diagonal

line in the right panel of Figure 5.3.8. In other words, at the 30% contamination level, the MSD

function usually classifies OC points as more central than they truly are. Because of this, many

of the IC points correspondingly receive lower rankings (incorrectly suggesting outlyingness)

from the MSD function than from the RMD function. The rankings determined by the MSD and

RMD functions are similar only for the most extreme OC points (rankings near 1000). Although

simulation results vary with other randomly generated shift directions, the general conclusion

remains the same -- a more robust depth function is needed for skewed data with contamination

levels exceeding 15%.

Since the MMR chart using RMD did not break down at the 30% contamination level

with symmetric distributions, it was decided to rerun the skewed distribution scenarios depicted

in Figure 5.3.5 using RMD instead of MSD as the depth function. The results are displayed in

Figure 5.3.9.
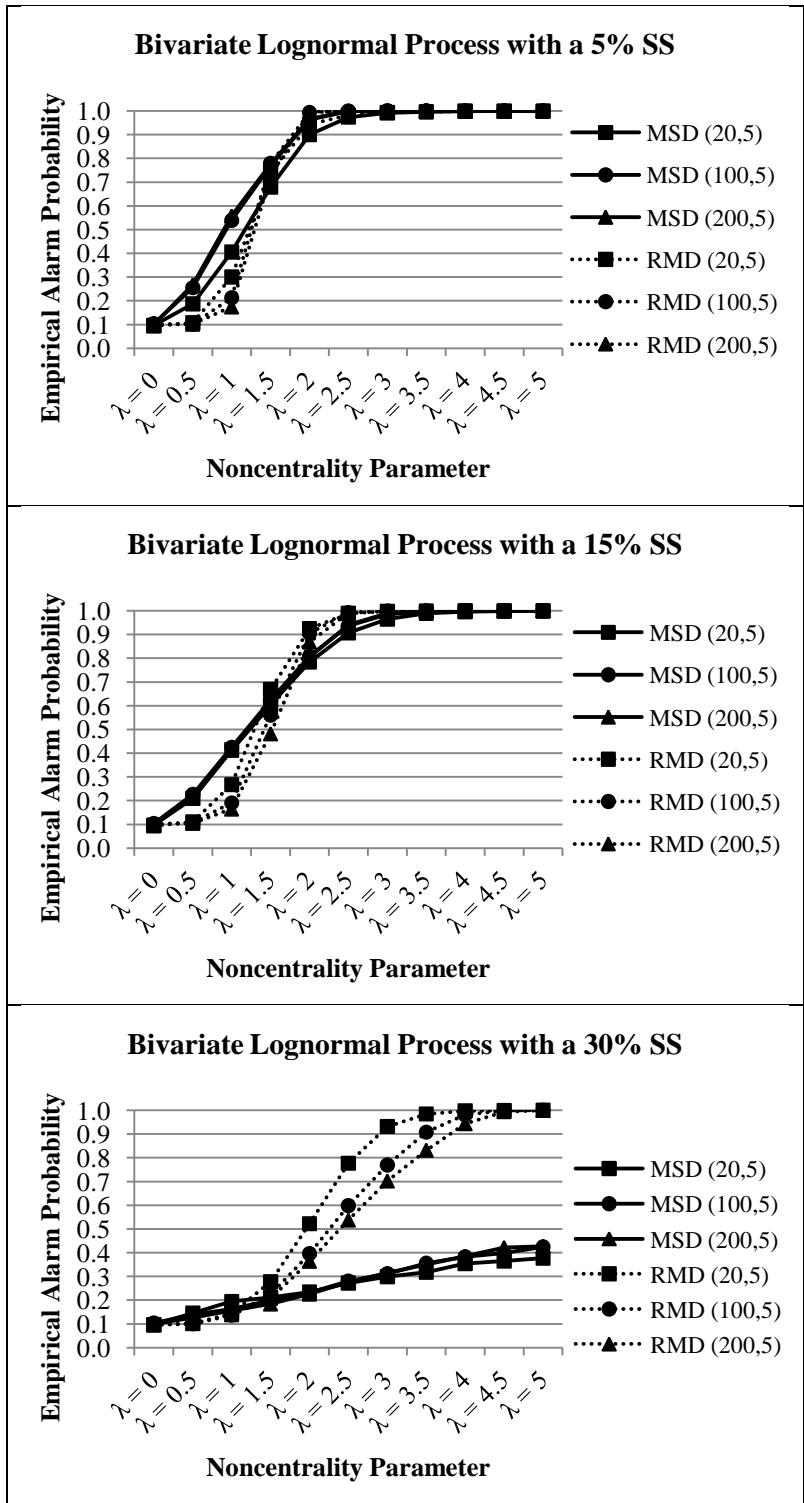
Figure 5.3.9 MMR-MSD/RMD Chart Performance on Increasingly Shifted LGN Data

For a 5% sustained shift of the mean, the MMR-RMD chart is less effective than the

MMR-MSD chart in detecting shifts of magnitude $\lambda = 0.5 - 1$ and marginally better in detecting

shifts of magnitude $\lambda = 1.5 - 2$. The same is true for a 15% level of contamination, but the

differences in chart performance are slightly magnified. When 30% of the data are shifted,

however, the MMR-RMD chart is clearly the better alternative because it does not break down in

the presence of severe contamination levels. The MMR-RMD chart's performance as compared

to Hotelling's T^2 chart with a 30% sustained shift of the mean is illustrated in Figure 5.3.10. The

MMR-RMD chart clearly outperforms Hotelling's T^2 chart for $m \geq 100$, but more importantly

offers reasonable distribution-free performance for all m even in the presence of severe

contamination levels.
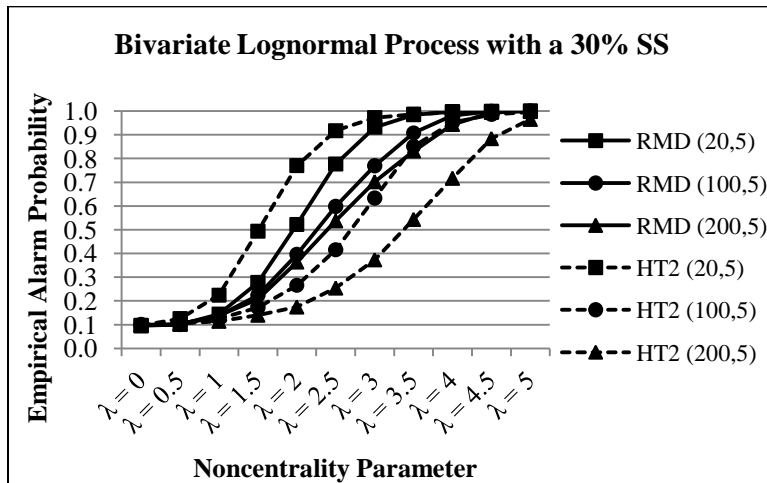
Figure 5.3.10 MMR-RMD Chart Performance on Bivariate LGN Data with a 30% SS

When the dimension is increased to five, the same trends in MMR-MSD chart

performance under sustained shifts of the mean in skewed data are observed, along with the

slight loss in power which accompanies increased dimensionality. Figure 5.3.11 shows the

results of applying the MMR-MSD and Hotelling's T^2 charts to a five-dimensional lognormally

distributed process with a 15% sustained shift of the mean. At least 100 subgroups, as opposed

to $m \geq 20$ in the bivariate case, are required for MMR-MSD chart performance to surpass

Hotelling's T^2 chart performance.
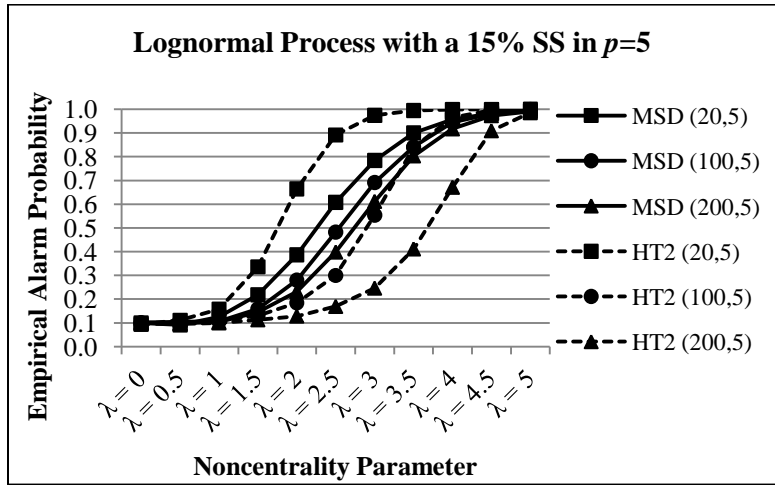
Figure 5.3.11 Control Chart Performance on LGN Data with a 15% SS in $p = 5$

Based on these results, it is concluded that when dealing with skewed data containing

sustained mean shifts, the MMR-MSD chart is preferred for contamination levels up to 15%, and

the MMR-RMD chart is the best option if the contamination level is suspected to exceed 15%.

Alternatively, both MMR-MSD and MMR-RMD charts could be run on the same data set in

order to provide maximum detection capability for all possible contamination levels.

Complete tables of results for all simulations performed using the multivariate lognormal

distribution with sustained shifts of the mean are provided in Appendices V - Y. In addition, a

matrix of recommended control chart usage with skewed multivariate data under both isolated

and sustained shifts of the mean is provided in Table 5.3.1. The MMR-MSD chart is almost

always preferred for contamination levels of 15% or less, and the MMR-RMD chart can be used

for higher contamination levels when the number of subgroups is sufficiently large. For the few

cases in which Hotelling's T^2 chart outperforms the MMR chart, more research is necessary

because implementation of Hotelling's T^2 chart with empirical UCL is only possible if the exact

process distribution is known.

p	Contamination Level	(m,n)		
		(20,5)	(100,5)	(200,5)
2	IS	MMR-MSD		
	5% SS			
	15% SS			
	30% SS	HT2	RMD	
5	IS	MMR-MSD		
	5% SS			
	15% SS			
	30% SS	HT2		MMR-RMD

Table 5.3.1 Recommended Phase I Control Chart Usage for Skewed Multivariate Data

5.4   MMR Chart Performance with Larger Subgroup Sizes

In order to assess the effects of larger subgroup sizes on MMR chart performance, a

targeted analysis using $m = 100$ and $n = 5(5)20$ was undertaken. The MMR-RMD chart was

evaluated under both isolated and 15% sustained shifts of the mean in five-dimensional $t(3)$ and

lognormally distributed processes. Simulation results reveal that increasing the subgroup size for

a given m enhances the performance of both the MMR and Hotelling's T^2 charts, but the MMR

chart reigns supreme in all cases considered.

As exhibited in Figure 5.4.1, the empirical probability of an MMR-RMD or Hotelling's

T^2 chart detecting an isolated shift in five dimensions is raised substantially by increasing the

subgroup size from 5 to 20. The difference in performance between the MMR-RMD and

Hotelling's T^2 charts is smallest when $n = 20$, but the MMR-RMD chart remains the superior alternative throughout the range of subgroup sizes evaluated. The overall trends for detection of isolated shifts in heavy-tailed and skewed processes are very similar, although shifts of smaller magnitude are detected more rapidly in a skewed process.
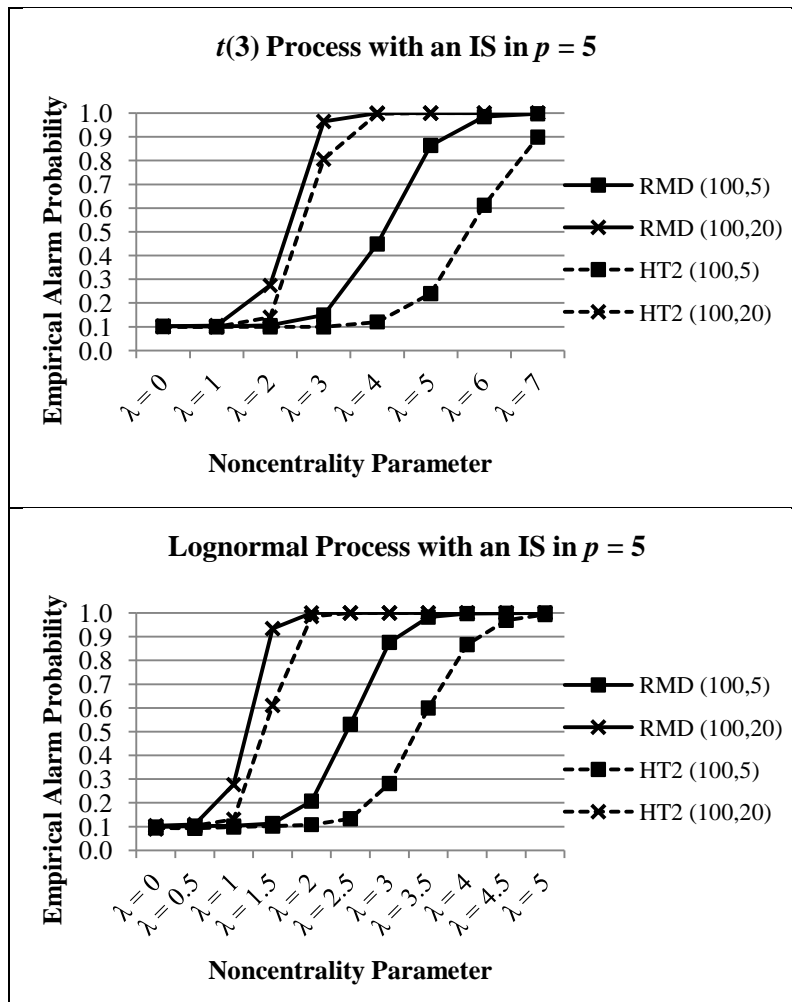
Figure 5.4.1 Effects of Subgroup Size on Control Chart Performance Under an IS in $p = 5$

A comparable pattern of performance is witnessed in detection of 15% sustained shifts of the mean by the MMR-RMD and Hotelling's T^2 charts. Figure 5.4.2 shows that increasing the

subgroup size raises the EAP for both charts considerably, but the MMR-RMD chart always

performs better than Hotelling's T^2 chart with empirically adjusted UCL.
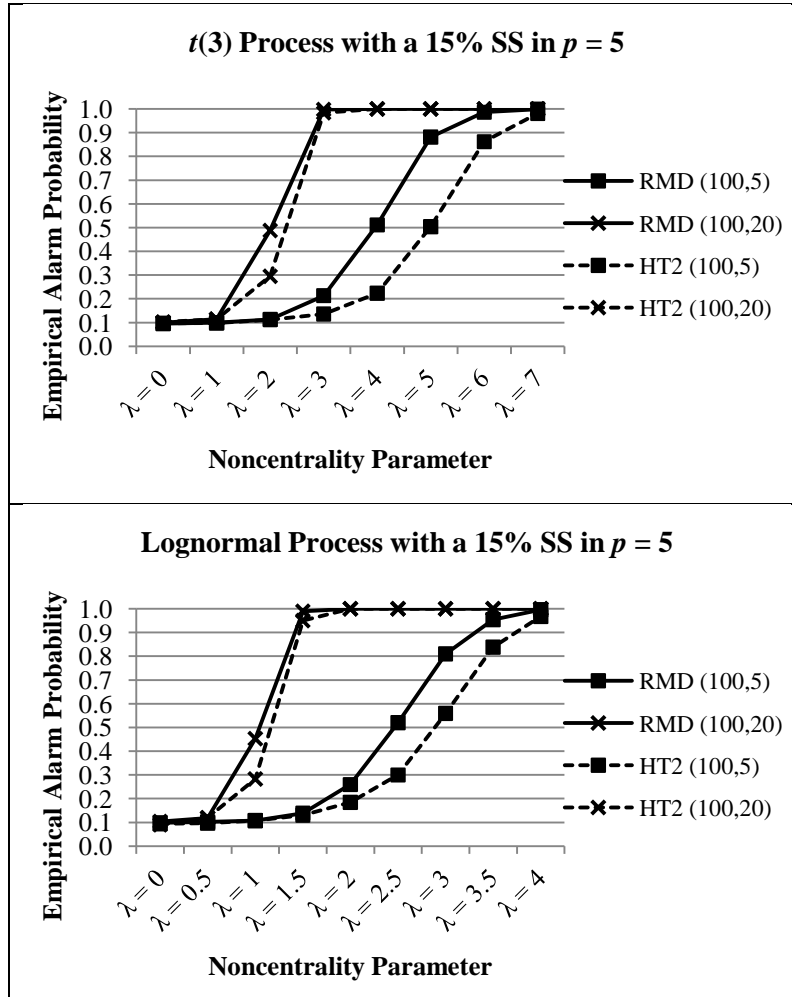
Figure 5.4.2 Effects of Subgroup Size on Chart Performance Under a 15% SS in $p = 5$

These results are somewhat surprising, as one might think that increasing the subgroup

size to $n = 20$ would result in approximate normality of subgroup averages. This in turn would

make a normal-theory method such as Hotelling's T^2 chart a better option than the MMR chart.

Although normality of subgroup averages will eventually be achieved for sufficiently large n due

to the central limit theorem, it is unlikely that subgroup sizes $n > 20$ will be observed in reality.

For more practical subgroup sizes such as $5 \leq n \leq 20$, the distribution-free MMR chart is clearly the best alternative. Complete tables of results for all subgroup size analyses performed are provided in Appendices Z - BB.

5.5   Robust Estimators of Location and Scatter for the MMR Chart

It was originally decided to use the BACON method of Billor et al. (2000) to robustly estimate both the mean vector and covariance matrix for use with the MMR chart. However, it was later determined that using the BACON location estimator with Type I error probability $\alpha = 0.10$ and Hotelling's T^2 scatter estimator $\overline{S}$ would result in significantly enhanced MMR chart performance. This choice of robust estimators was briefly addressed in Chapter 2, and will be discussed in detail here.

In early test runs, the MMR-RMD chart using strictly BACON estimators was compared to Hotelling's T^2 chart with empirically adjusted UCL using a bivariate $t(3)$ process with a sustained shift of the mean. The BACON method of estimation with $\alpha = 0.05$ performed nearly perfectly in detecting large process shifts ($\lambda \geq 8$) and subsequently excluding OC points from the resulting location and scatter estimates. With smaller shifts ($\lambda < 8$) however, the BACON method did not consistently identify outlying points, often resulting in estimated mean vectors and covariance matrices which were approximately equivalent to the classical nonrobust estimates. The contamination in the estimated parameters resulted in degraded performance of the MMR chart and as indicated in Figure 5.5.1, this effect was magnified as the level of contamination in the data set was raised from 15% to 30%. Limited testing of the MCD method to determine robust location and scatter estimates yielded similar results at the cost of a significantly higher computational burden.

94
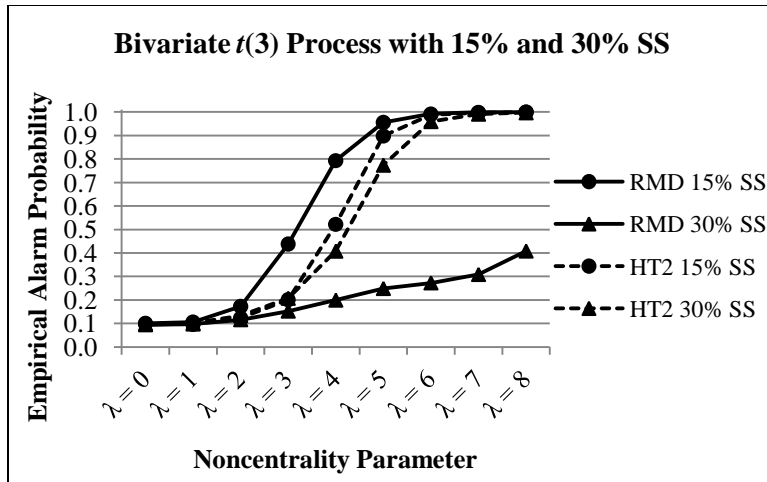
Figure 5.5.1 Comparison of MMR-RMD (Using BACON Estimators) and HT2 Charts

Figure 5.5.2 shows why it is so difficult for even robust methods to distinguish IC from OC data when λ is small. The univariate $t(3)$ plots represent probability density functions for various unshifted and shifted $t(3)$ distributions. The bivariate graphs were created by randomly generating 500 observations from a bivariate $t(3)$ distribution and inducing a location shift upon 15% of the data. In the first row of Figure 5.5.2, a one unit shift is barely distinguishable. In the second row, a four unit shift is more noticeable but still results in significant overlap between unshifted and shifted data. It takes an eight unit shift, as depicted in the third row of Figure 5.5.2, to clearly separate shifted data from unshifted data.
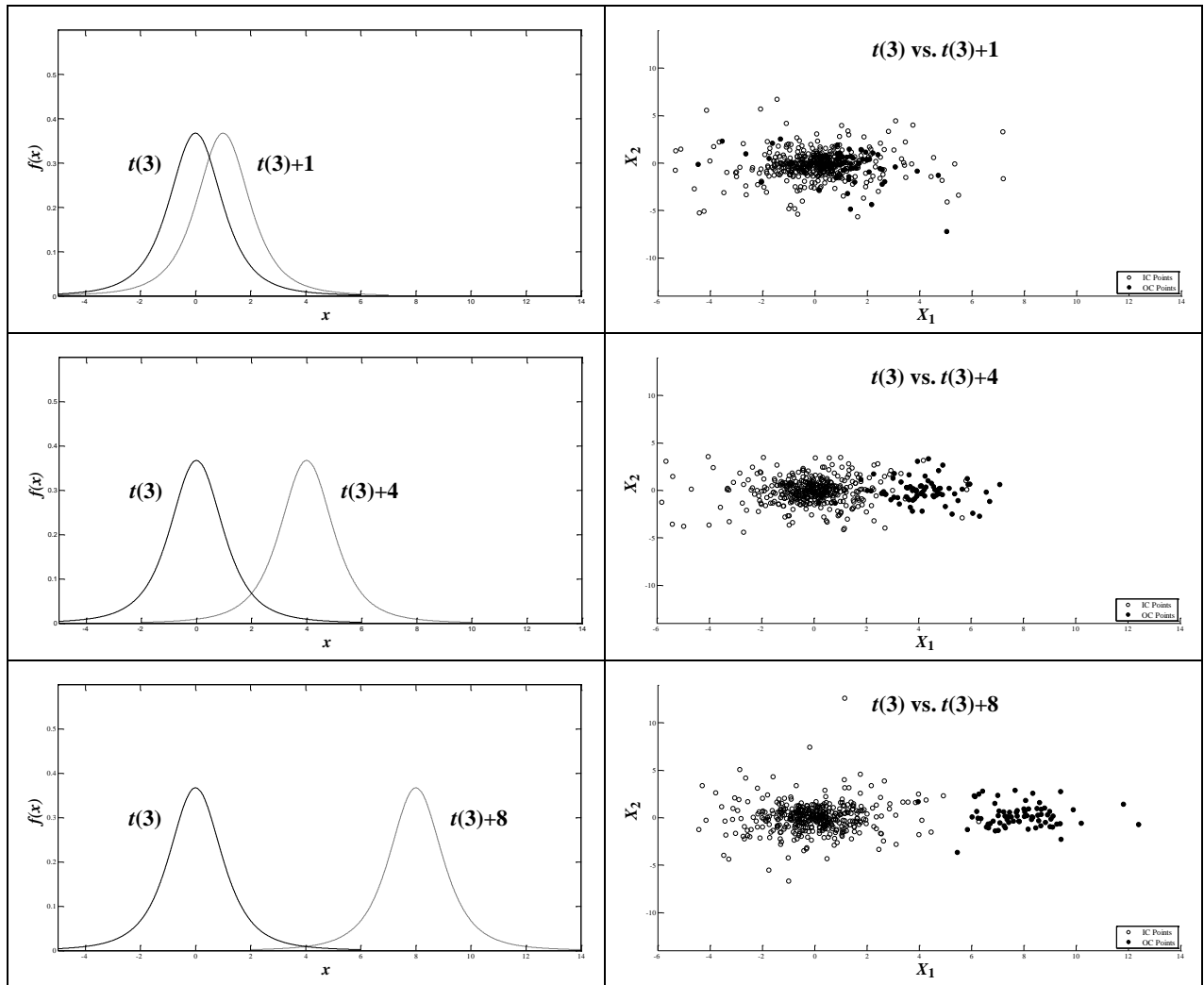
Figure 5.5.2 The Effects of Increasing Shift Sizes on Univariate and Bivariate $t(3)$ Data

Also illustrated in Figure 5.5.1, Hotelling's T^2 control chart with empirical UCL is

substantially less affected by higher contamination levels than the MMR-RMD chart. This is

because Hotelling's scatter estimator for data consisting of m subgroups of size n, $\overline{S} = \dfrac{1}{m}\sum_{i=1}^{m} S_i$,

where S_i has (k,l)th entry $\dfrac{1}{n-1}\sum_{j=1}^{n}\left(X_{jk} - \bar{X}_i\right)\left(X_{jl} - \bar{X}_i\right)'$, is robust to location shifts under the

assumption that shifted subgroups possess the same covariance structure as unshifted subgroups.

$\overline{S}$ represents the average of the m subgroup covariance matrices, each of which is computed

with respect to its subgroup mean $\bar{X}_i$ rather than the mean of the entire data set $\bar{\bar{X}}$ as with the

classical covariance estimator S, which has (k,l)th entry $\dfrac{1}{N-1}\displaystyle\sum_{j=1}^{N}\left(X_{jk}-\bar{\bar{X}}\right)\left(X_{jl}-\bar{\bar{X}}\right)'$.

Accordingly, $\bar{S}$ is not inflated by OC subgroups as are classical methods which consider the data set as a whole or robust methods which fail to exclude outliers.

This result is true only for subgrouped data. When individual data are encountered in a control charting application, Hotelling's T^2 scatter estimator becomes the nonrobust classical covariance matrix. Under those circumstances, robust parameter estimation methods such as BACON may be preferred because they exclude OC points corresponding to shifts with large λ.

Based on these findings, it was decided to substitute Hotelling's T^2 scatter estimator $\bar{S}$ for the BACON scatter estimator in the MMR chart. To achieve a more robust location estimate for the MMR chart, the BACON method was implemented with a higher Type I error probability. Experimentation with the BACON method using $\alpha = 0.05, 0.10, 0.20,$ and 0.35 showed that $\alpha = 0.10$ provides the best compromise between Type I and Type II error. As indicated in Figure 5.5.3, implementation of the MMR-RMD chart using the new estimators results in significantly enhanced performance over the MMR-RMD chart using strictly BACON estimators, especially when the contamination level is high.
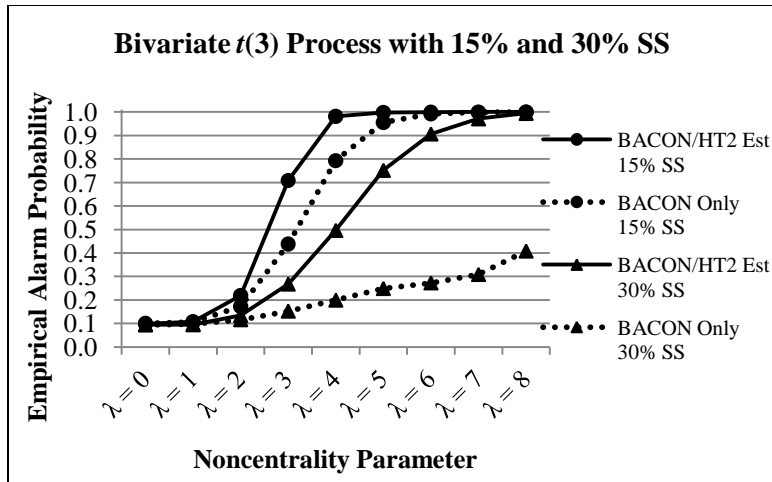
Figure 5.5.3 Improvement in MMR-RMD Chart Performance with New Estimators

Surprisingly, even with the new estimators, 30% sustained shifts of the mean are detected by both charts with lower probability than 15% sustained shifts. In the case of Hotelling's T^2 chart, this occurs because Hotelling's T^2 scatter estimator is naturally robust but Hotelling's T^2 location estimator $\overline{\overline{X}} = \dfrac{1}{m}\sum_{i=1}^{m} \overline{X}_i$ is equivalent to the classical mean vector and is therefore nonrobust. To verify this, the charts in Figure 5.5.1 were repeated using a known mean vector of all zeros. As expected, Figure 5.5.4 shows that 30% sustained shifts are detected by Hotelling's T^2 chart with higher probability than 15% sustained shifts when the mean vector is known, yet the same does not hold true for the MMR-RMD chart. Additional experimentation revealed that this occurs because of the redistribution of the ranks assigned to depth values during the MMR control charting process, and is simply an unavoidable consequence of rank-based control charting.
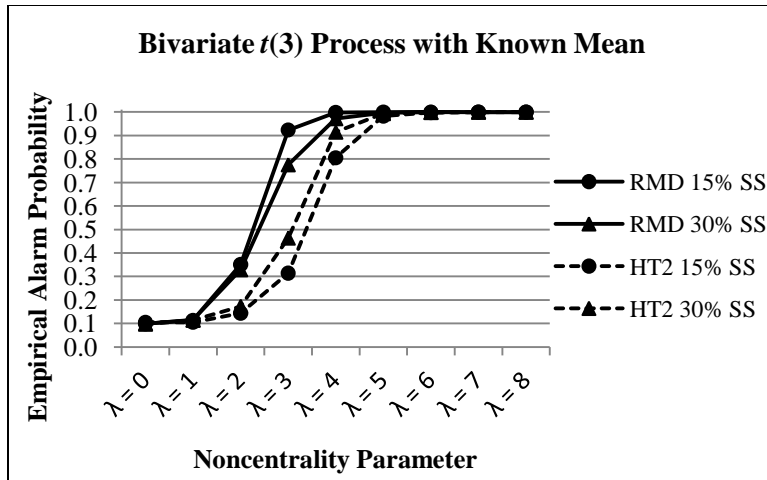
Figure 5.5.4 Change in Chart Performance When the Mean is Known

To illustrate the redistribution of ranks in conjunction with higher contamination levels, 100 observations consisting of $m = 20$ subgroups of size $n = 5$ from an in-control bivariate standard normal process were simulated. Depth values for each point were computed using RMD with BACON location estimator ($\alpha = 0.10$) and Hotelling's T^2 scatter estimator, and ranks were assigned to each point from nearest (rank = 1) to farthest (rank = 100) from the center. Next, 5% of the data were shifted by three units to the right, RMD values were recomputed, and new ranks were recorded. Finally, this process was repeated using a 30% contamination level.

For both 5% and 30% shifts, Figure 5.5.5 illustrates scatterplots and rank charts of IC and OC data before and after the shifts. If the rankings of IC points were unaffected by the shifts as one might expect, they would follow a straight line on the before versus after rank charts. For the 5% shift in column one this is nearly the case, but the rank chart for the 30% shift in column two shows that the IC rankings are significantly affected by the induced shift. This is true because as previously illustrated in Figure 5.5.2, a shift of magnitude three is not large enough to clearly separate the IC points from the OC points. Rather, many of the IC and OC points are comingled, thus distorting the rankings and making it more difficult for the MMR chart to

distinguish between them. As the level of contamination in the data is raised, the level of

distortion in the rankings increases and MMR chart performance decreases accordingly.
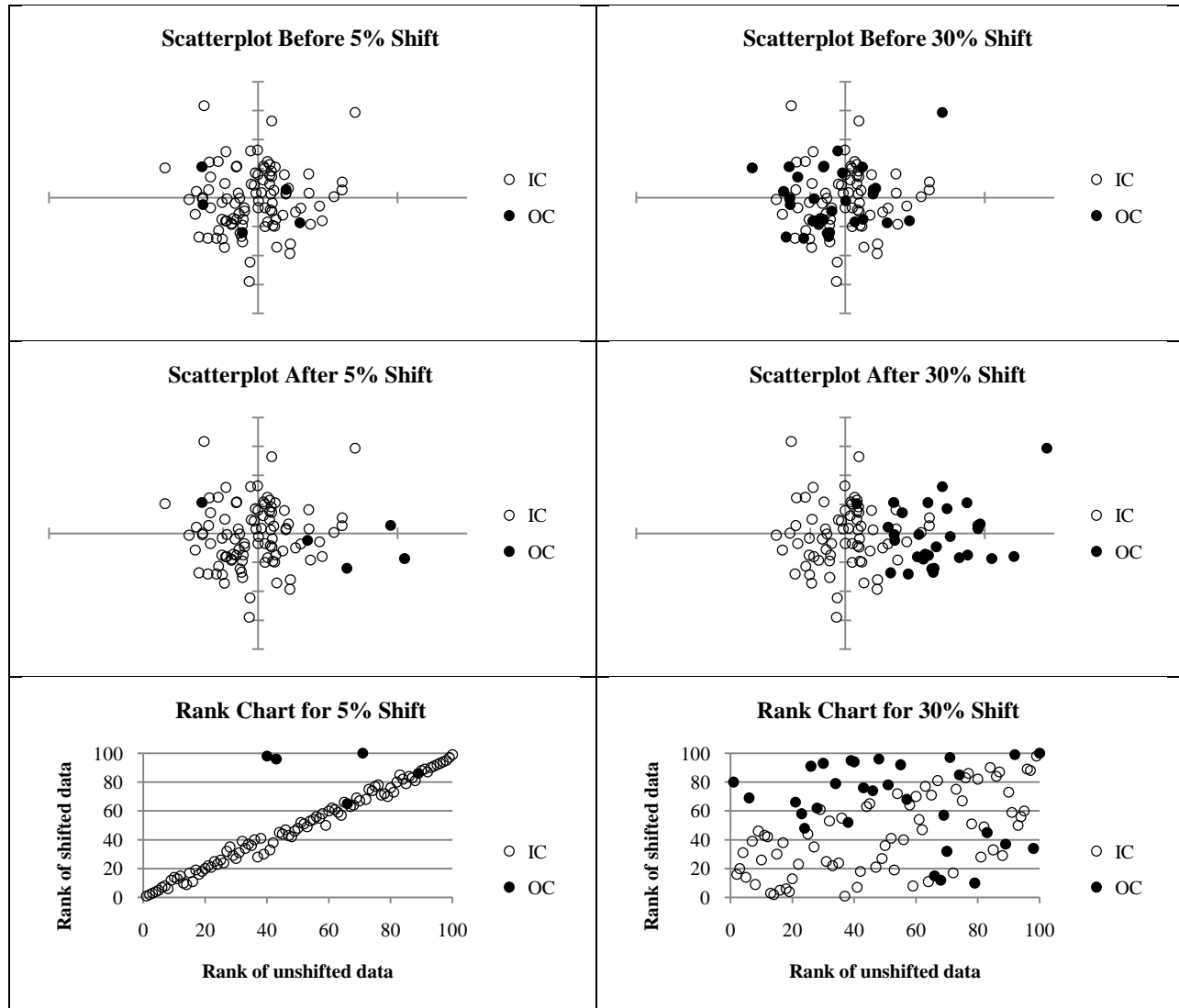
Figure 5.5.5 Redistribution of Ranks Under 5% and 30% Sustained Shifts

Despite this effect, the MMR chart using data depth with BACON location estimator (α

= 0.10) and Hotelling's T^2 scatter estimator is much more effective than Hotelling's T^2 chart in

detecting isolated and sustained shifts of the mean in both symmetric and skewed multivariate

data. This was illustrated throughout Chapter 5 for various combinations of m = 20, 50(50)200,

$n = 5(5)20$, and $p = 2, 5$, and 10. The MMR chart has the added advantage of being distribution free, unlike Hotelling's T^2 chart which has to be tailored to the specific process distribution under study. In order to illustrate a complete application of the MMR chart, an example is offered in the following chapter.

6   An Example Phase I Analysis Using the MMR Chart

6.1   Simulating the Contaminated Reference Sample

In order to demonstrate an application of the MMR-MSD chart from start to finish, a simulated example involving a five-dimensional, lognormally distributed reference sample with $m = 100$, $n = 5$, and three isolated shifts of the mean is presented. Data and shift directions were generated in accordance with the procedures outlined in Chapter 4. Isolated shifts of increasing magnitude were applied to single subgroups as follows: $\lambda = 3$ at subgroup 4, $\lambda = 5$ at subgroup 41, and $\lambda = 200$ at subgroup 91. The shift of magnitude $\lambda = 3$ represents the smallest shift for which the MMR-MSD chart was shown in Chapter 5 to have a nearly perfect detection ability, and the shift of magnitude $\lambda = 200$ is designed to illustrate the sensitivity of robust and nonrobust estimators to extreme outliers. Using a desired IC FAP of 0.05, the MMR-MSD chart using UCLs from Table 6.1.1 was compared to Hotelling's T^2 chart with Alt's (1976) Phase I UCL.

m	n	Desired FAP = 0.05	
		UCL	Simulated FAP
100	5	2.992	0.0483
99	5	2.99	0.0484
98	5	2.987	0.0482
97	5	2.986	0.0485

Table 6.1.1 MMR Chart UCLs for Chapter 6 Example

6.2   Removing Outliers from the Sample

MMR-MSD and Hotelling's T^2 charts applied to the unedited reference sample are

pictured in Figure 6.2.1. Each chart contains a superimposed table of potential OC subgroups.

MMR-MSD Chart

i	Z_i	UCL
4	3.156	2.992
41	3.598	2.992
91	3.850	2.992

Hotelling's T^2 Chart

i	T_i^2	UCL
4	151.25	22.59
41	238.79	22.59
91	268,471	22.59
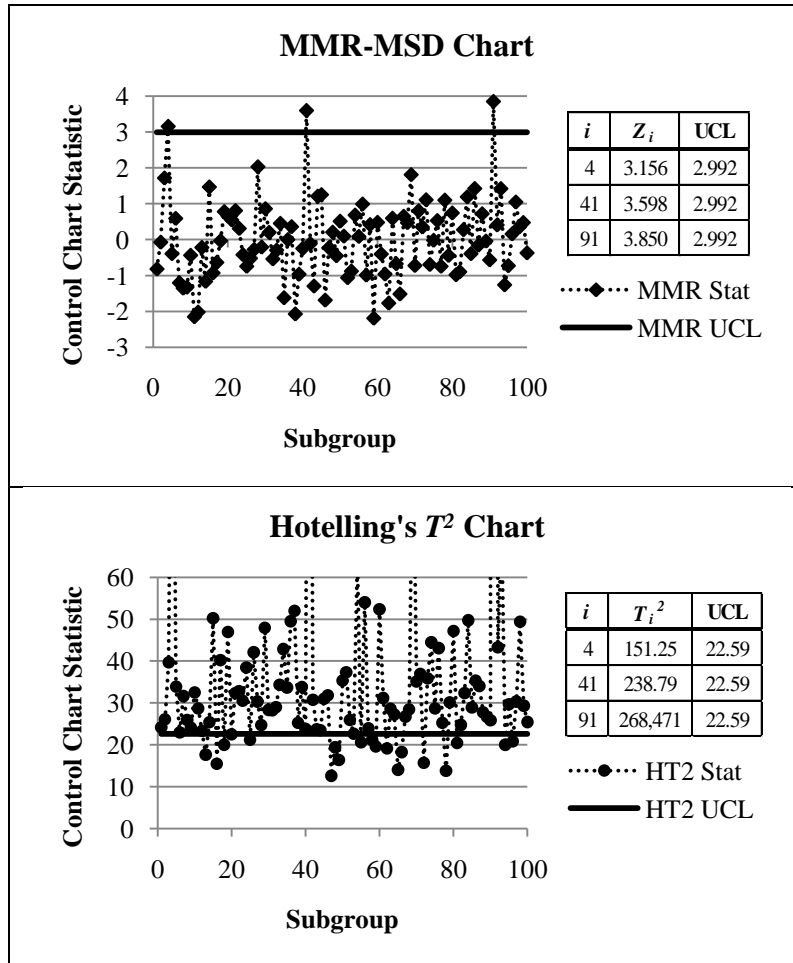
Figure 6.2.1 Initial Application of Phase I Control Charts to the Lognormal Sample

The three shifted subgroups are readily apparent on the MMR-MSD chart, as they all fall above

the initial UCL for $m = 100$, $n = 5$. The extreme outlier represented by subgroup 91 does not

look considerably different than the other two outliers for several reasons. First of all, its

extreme outlyingness is mitigated by the rank-based nature of the MMR-MSD control chart

statistic. A rank from 1 to N assigned to a point in a reference sample represents only the

position of that point with respect to the other $N = m \times n$ points in the sample as determined by a depth function -- the degree of outlyingness is not reflected in the ranking. The most outlying point in a data set will receive a rank of N, regardless of whether the point is only marginally more outlying than all others or a significant distance away from the rest of the p-dimensional data cloud. Also, computation of MSD does not involve estimation of a location vector, hence its robustness to isolated shifts in location no matter how extreme. As was shown in Chapter 5, very high contamination levels can redistribute the ranks in such a manner that the MMR-MSD chart becomes ineffective at detecting sustained shifts, but extreme isolated shifts are detected with ease. Even if the RMD function (which does require an estimated mean vector to compute) was used in this scenario, the BACON location estimator would exclude the extreme outlier represented by subgroup 91 and the resulting MMR-RMD control chart would be very similar to the MMR-MSD chart. As a result of these properties, the MMR chart is well insulated against the effects of a single extreme outlier in a given reference sample.

With Hotelling's T^2 chart, however, the extreme outlier has a dramatic effect on the T^2 statistic for each subgroup, as evidenced by the fact that the majority of the control chart statistics fall above the initial UCL. This occurs because the grand mean $\bar{\bar{X}}$ used in computing

$$T_i^2 = n\left(\bar{X}_i - \bar{\bar{X}}\right)' \bar{\bar{S}}^{-1}\left(\bar{X}_i - \bar{\bar{X}}\right)$$ is not robust to outliers. A more robust estimator for the mean vector such as BACON could prevent this from occurring, but is beyond the scope of this research.

The next step in a Phase I analysis is to investigate each potential OC subgroup for an assignable cause. In this example, it is assumed that all potential OC subgroups have assignable causes and therefore warrant removal from the data set. Some control chart authors advocate

removing all OC subgroups at once, and then recalculating control limits. Others believe that OC subgroups should be removed one at a time beginning with the most outlying subgroup, with the control limits being recalculated at each iteration. This example will take the latter approach.

The most extreme OC subgroup for both the MMR-MSD and Hotelling's T^2 control charts is subgroup 91, so it will be removed first. Once an OC subgroup is removed from the data set, both control charts are reconstructed using control limits appropriate for the reduced number of subgroups. The control charts for $m = 99$, $n = 5$ after removal of the first OC subgroup are depicted in Figure 6.2.2.
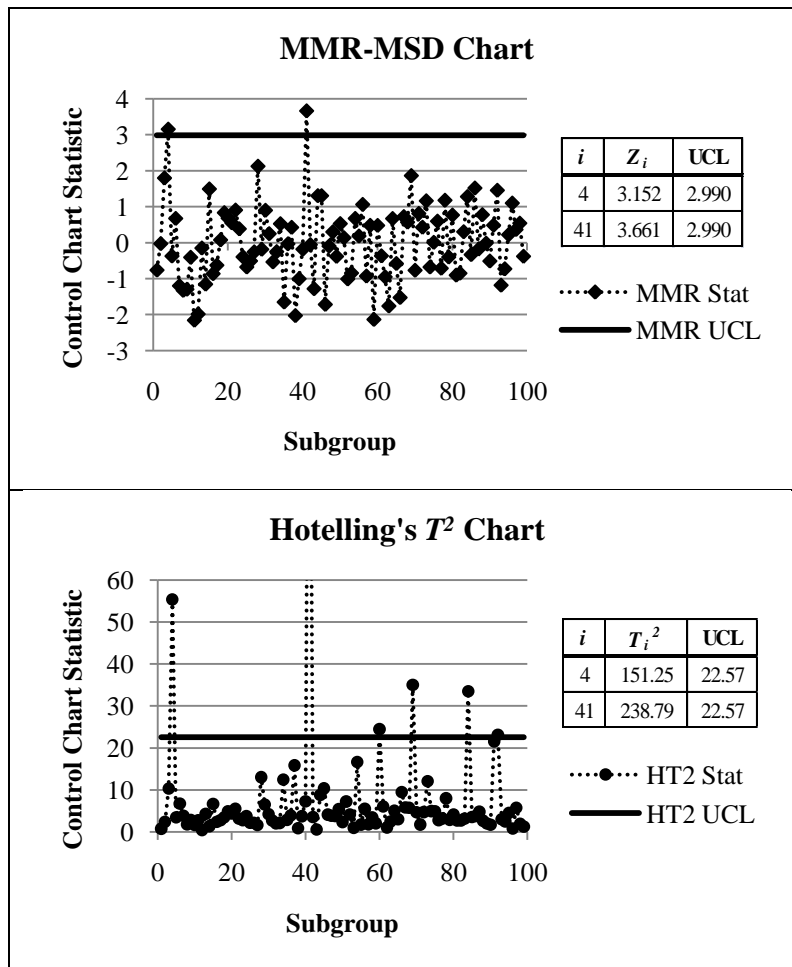
Figure 6.2.2 Second Iteration of the MMR-MSD Control Chart

105

After removing the extreme outlier, control chart statistics for the remaining two planted

outliers still exceed the UCLs for both the MMR-MSD and Hotelling's T^2 control charts. Next,

the outlier represented by subgroup 41 is removed and both control charts are recalculated using

$m = 98$, $n = 5$. Finally, the outlier represented by subgroup 4 is eliminated and the control charts

are recomputed using $m = 97$, $n = 5$. The final MMR-MSD and Hotelling's T^2 control charts

after sequentially removing all planted outliers are illustrated in Figure 6.2.3.
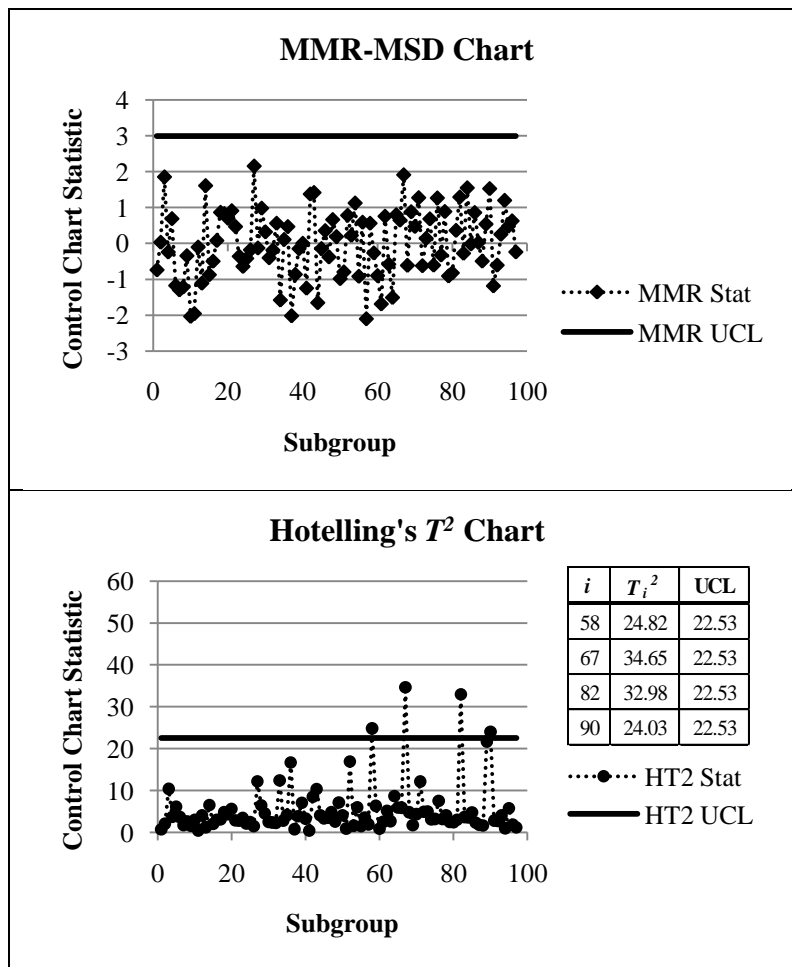
Figure 6.2.3 Final Control Charts After Four Iterations of Phase I Analysis

At this point, all control chart statistics for the MMR-MSD chart fall under the UCL, so

the remaining reference sample consisting of $m = 97$ subgroups is correctly declared to be IC.

Hotelling's T^2 chart, despite following the same outlier removal process as the MMR-MSD chart, still identifies four potential OC subgroups after the third iteration of a Phase I analysis. Further iterations could result in the identification of even more potential OC subgroups because the Phase I UCL for Hotelling's T^2 chart is adjusted downward as the number of subgroups is decreased with each iteration.

6.3   Analyzing the Results

Hotelling's T^2 chart falsely identifies multiple potential OC subgroups because normal-theory Phase I UCLs were applied to skewed data, illustrating the danger of applying a normal-theory method without regard to the underlying distribution of a process. Using UCLs empirically tailored to a five-dimensional multivariate lognormal distribution would solve the problem of multiple false alarms, but would also result in a loss in detection power as only the first two OC subgroups would be identified and removed. In addition, the exact process distribution would not be known in anything but a simulation example such as the one presented here, so empirical UCLs for Hotelling's T^2 chart are not practical for widespread implementation. The MMR chart is clearly a superior alternative because it offers accurate, distribution-free performance with a low computational burden.

7   Conclusion

7.1   Synopsis of Findings

The MMR chart for detecting location shifts in subgrouped data represents the first known distribution-free Phase I multivariate control chart. This work represents the culmination of extensive research to synthesize appropriate statistical process control techniques, data depth functions, and robust parameter estimation methods to create a distribution-free, computationally feasible, and accurate Phase I multivariate control charting methodology. The MMR chart has been shown to be extremely effective in detecting isolated and sustained shifts of the mean in both heavy-tailed and skewed multivariate data.

7.2   Summary of Research Conducted

The MMR chart was created as a multivariate extension of Jones-Farmer et al.'s (2009) univariate distribution-free Phase I mean-rank chart for subgroup location. Given an unedited p-dimensional reference sample consisting of m subgroups of size n, data depth functions in conjunction with robust estimators were used to reduce multivariate data to univariate depth values. The robust Mahalanobis depth function for elliptically symmetric data was implemented using the BACON location estimator and Hotelling's T^2 scatter estimator for subgrouped data. The Mahalanobis spatial depth function, which is not reliant on distributional assumptions and does not require a location estimator, was employed using Hotelling's T^2 scatter estimator for

subgrouped data. Depth values resulting from these functions were ranked and converted into

MMR control chart statistics for each subgroup, which were then compared to empirical UCLs

determined through simulation of the joint distribution of the MMR control chart statistic.

Hotelling's T^2 control chart with Alt's (1976) Phase I UCLs for normally distributed data

and empirically adjusted UCLs for nonnormally distributed data was used to establish a baseline

level of Phase I performance. Performance comparisons of the MMR chart to Hotelling's T^2

chart included scenarios involving simulated multivariate normally distributed data, heavy-tailed

data represented by the multivariate $t(3)$ distribution, and skewed data represented by the

multivariate lognormal distribution for $m = 20, 50(50)200$ subgroups of size $n = 5$ and

dimensions $p = 2, 5$, and 10. All data were standardized, without loss of generality, to have zero

mean vector and identity covariance matrix. IC performance was measured by each chart's

ability to maintain the desired FAP using simulated IC data. OC performance was measured by

each chart's EAP under isolated as well as 5%, 15%, and 30% sustained shifts of the mean

assuming constant within-subgroup covariance. Shifts were fixed in a specific direction with

elliptically symmetric distributions without loss of generality, and averaged over a uniform

distribution of shift directions with skewed distributions. Limited analysis was performed on the

effect of increased subgroup sizes on control chart performance in Phase I.

7.3   Recommendations for Phase I Analysis

A comprehensive simulation study shows that when normality of Phase I multivariate

process data can be established, Hotelling's T^2 chart with Alt's (1976) Phase I UCL is preferred

for detecting isolated or sustained shifts of the mean. This is not unexpected, as one would

expect a normal-theory method to outperform a distribution-free method when a process is

multivariate normally distributed, and the original intent of the MMR chart was to provide a distribution-free control charting methodology for processes demonstrating clear departures from normality.

When Phase I process data are heavy-tailed or skewed, the MMR chart usually outperforms Hotelling's T^2 chart in detecting isolated or sustained shifts of the mean. More importantly, the MMR chart offers truly distribution-free performance because the UCL for a given application depends only on the number of subgroups, the size of each subgroup, and the desired IC FAP without regard to the form of the underlying process distribution. UCLs for Hotelling's T^2 chart, on the contrary, must be empirically tailored to the exact distribution of a nonnormally distributed process to achieve the desired IC FAP, something which is only possible in a simulation environment. An added benefit of the MMR chart is that, for a given OC scenario involving nonnormally distributed data, its performance is far more invariant to the size of m than Hotelling's T^2 chart with empirical UCL, thus making it even more attractive as a distribution-free alternative.

As indicated in Table 5.2.1, the MMR-RMD chart is recommended for most situations involving heavy-tailed data as long as the required minimum number of subgroups is present. As shown in Table 5.3.1, when process data are skewed, the MMR-MSD chart is almost always recommended if the contamination level is less than 15%, and the MMR-RMD chart is preferred for contamination levels above 15% if the number of subgroups is sufficiently large. In all cases tested, as the dimension of the data or the level of contamination is raised, the minimum number of subgroups required for the MMR chart to achieve superiority over Hotelling's T^2 chart with empirical limits correspondingly increases but remains within reasonable bounds. These general conclusions are based on a subgroup size of at least $n = 5$. Larger subgroup sizes reduce the

minimum number of subgroups required for MMR chart performance to surpass that of Hotelling's T^2 chart.

7.4   Recommendations for Phase II Monitoring

Once an IC reference sample has been determined through a successful Phase I analysis using the MMR chart, it can be used in conjunction with an appropriate Phase II method to monitor future observations for any departures from the IC state. As noted by C. Champ (personal communication, May 12, 2011), since more is known about a process at the conclusion of a Phase I analysis, the form of a Phase II control chart does not necessarily have to match the form of a Phase I control chart. Although this research will assume nonnormally distributed data throughout the retrospective analysis and monitoring phases, this means that even though the Phase I MMR chart is specifically designed for multivariate data collected in subgroups, the search for the most suitable Phase II complement to the Phase I MMR chart need not be limited to methods requiring subgrouped multivariate data.

After an extensive literature review, the MEWMA chart proposed by Lowry et al. (1992), with small smoothing parameter as recommended by Stoumbos and Sullivan (2002), is recommended for Phase II monitoring because it is easy to understand and implement, well documented in statistical process control literature, and robust to the underlying process distribution. The MEWMA control chart statistic represents a weighted average of all Phase II observations, with the most recent observation assigned a weight equal to the smoothing constant r and all previous observations assigned weights which geometrically decrease according to their age. Stoumbos and Sullivan (2002) showed that the MEWMA chart can be successfully applied to nonnormally distributed individual or subgrouped multivariate data if a sufficiently small

111

smoothing constant is chosen. Based on the results of a comprehensive simulation exercise, the authors recommend a smoothing constant of $r \in [0.02, 0.05]$ for five or less dimensions and $r \le 0.02$ for more than five dimensions for reliable detection of sustained location shifts in heavy-tailed or skewed multivariate data. A subsequent study by Testik et al. (2003) mirrored the findings of Stoumbos and Sullivan (2002) regarding use of the MEWMA chart as a robust Phase II method.

It should be noted that all three aforementioned MEWMA chart studies are based on the assumption that the IC mean vector and covariance matrix are known. If the MEWMA chart is employed following a Phase I analysis using the MMR chart, the mean vector and covariance matrix are not known but rather estimated from an IC reference sample. If the IC reference sample is too small, using estimated as opposed to known parameters can lead to more frequent false alarms and a lower probability of detecting of OC conditions, especially when the smoothing constant is small. For the univariate EWMA chart, this effect was detailed by Jones, Champ, and Rigdon (2001), and design strategies to alleviate this problem were offered by Jones (2002). For the MEWMA chart, Champ and Jones-Farmer (2007) showed that widening the control limits through simulation to account for the additional variability introduced by the use of estimated parameters results in nearly the same performance as the known parameter case. An analytical method of determining control limits for the MEWMA chart with estimated parameters as well as the minimum sample size required for estimated parameter performance to equal known parameter performance are topics for future research. Despite these open issues, the MEWMA chart represents the most broadly applicable control charting methodology for Phase II monitoring of nonnormally distributed multivariate data.

A potential criticism of the MEWMA chart is that using a small smoothing parameter to improve robustness to nonnormality decreases control chart sensitivity to large sustained mean shifts and isolated outlying observations, but it can be argued that this is not a significant disadvantage in a Phase II control charting scenario. As previously noted in Chapter 3 of this document, Montgomery (2005, p. 386) characterizes control charts which accumulate information from sequences of points (e.g. CUSUM, EWMA, and their multivariate counterparts) as being ideally suited for Phase II monitoring because they are more sensitive to small process shifts than Shewhart type charts which use information only from the most recent observation. According to Montgomery (2005, p. 386), sensitivity to small shifts is desirable for a Phase II control chart because in contrast to Phase I, "assignable causes do not typically result in large process upsets or disturbances" in Phase II. If greater control chart sensitivity to large sustained mean shifts or individual outliers is desired, the reader is directed to the Chapter 1 discussion of Phase II nonparametric, distribution-free, and robust control charts. Although a few such methods could potentially supplement the MEWMA control chart in certain scenarios, none have proven as effective as the MEWMA chart with small smoothing constant on a wide range of nonnormally distributed data in higher dimensions.

7.5 Future Research Directions

The MMR chart fills a notable gap in current multivariate quality control literature, yet much work remains to be done in the field of distribution-free Phase I multivariate quality control. Although it is believed that the fundamental structure of the MMR chart is sound, potential refinements include further exploration of the BACON method to determine optimal input parameters (e.g. Type I error probability) for maximum robustness to shifts of all

magnitudes, implementation of other location and scatter estimators to improve robustness to higher contamination levels, and experimentation with alternative data depth functions which may enhance MMR chart performance. Additionally, since the MMR chart is designed to detect location changes in subgrouped multivariate data during Phase I, an equivalent distribution-free chart for detecting scale changes is needed for Phase I scenarios in which the assumption of constant within-subgroup covariance is not appropriate. Finally, Phase I distribution-free charts for detecting both location and scale changes in small subgroups ($n < 5$) and individual multivariate observations ($n = 1$) should be sought as well. It is the hope of this author that the success of the MMR chart as the first proposed distribution-free Phase I multivariate method will serve as the catalyst for some or all of this additional research.

References

Alfaro, J.L., & Ortega, J.F. (2008). A Robust Alternative to Hotelling's T^2 Control Chart Using Trimmed Estimators. *Quality and Reliability Engineering International,* 24, 601-611.

Aloupis, G. (2005, August). *Geometric and Combinatorial Issues in Data Depth*. Presented at the Franco-Canadian Workshop on Combinatorial Algorithms, Hamilton, Ontario.

Aloupis, G. (2006). Geometric Measures of Data Depth. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (Vol. 72, pp. 147-158). Providence, RI: American Mathematical Society.

Alt, F.B. (1976). Small Sample Probability Limits for the Mean of a Multivariate Normal Process. *ASQC Technical Conference Transactions*, pp. 170-176.

Bakir, S.T. (1989). Analysis of Means Using Ranks. *Communications in Statistics – Simulation and Computation*, 18(2), 757-776.

Beltran, L.A. (2006). *Nonparametric Multivariate Statistical Process Control Using Principal Component Analysis and Simplicial Depth*. Dissertation, University of Central Florida.

Bersimis, S., Panaretos, J., & Psarakis, S. (2005). Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry. *Proceedings of the 7th Hellenic European Conference on Computer Mathematics and Its Applications*.

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International,* 23, 517-543.

Billor, N., Hadi, A.S., & Velleman, P.F. (2000). BACON: Blocked Adaptive Computationally Efficient Outlier Nominators. *Computational Statistics & Data Analysis*, 34, 279-298.

Chakraborty, B., & Chaudhuri, P. (1999). A Note on the Robustness of Multivariate Medians. *Statistics & Probability Letters*, 45, 269-276.

Champ, C.W., & Jones, L.A. (2004). Designing Phase I $\bar{X}$ Charts with Small Sample Sizes. *Quality and Reliability Engineering International,* 20, 497-510.

Champ, C.W., & Jones-Farmer, L.A. (2007). Properties of Multivariate Control Charts with Estimated Parameters. *Sequential Analysis,* 26, 153-169.

Chatterjee, S. & Qiu, P. (2009). Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits. *The Annals of Applied Statistics*, 3(1), 349-369.

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

Chenouri, S., & Variyath, A.M. (2011). A Comparative Study of Phase II Robust Multivariate Control Charts for Individual Observations. *Quality and Reliability Engineering International*, 27(3) [Electronic version].

Chou, Y.M., Mason, R.L., & Young, J.C.(2001). The Control Chart for Individual Observations from a Multivariate Non-Normal Distribution. *Communications in Statistics – Theory and Methods*, 30(8), 1937-1949.

Crosier, R.B. (1988). Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics*, 30(3), 291-303.

Dai, Y., Zhou, C., & Wang, Z. (2006a). Multivariate CUSUM Control Charts Based on Data Depth for Preliminary Analysis (Working paper).

Dang, X., & Serfling, R. (2010). Nonparametric Depth-Based Multivariate Outlier Identifiers, and Masking Robustness Properties. *Journal of Statistical Planning and Inference*, 140, 198-213.

Donoho, D.L. & Huber, P.J. (1983). The Notion of a Breakdown Point. In P.J. Bickel, K.A. Doksum and J.L. Hodges, Jr. (Eds.), *A Festschrift for Eric L. Lehmann* (pp. 157-184). Belmont, CA: Wadsworth.

Fricker, R.D., & Chang, J.T. (2009a). The Repeated Two-Sample Rank (RTR) Procedure: A Nonparametric Multivariate Individuals Control Chart (Working paper).

Gao, Yonghong (2003). Data Depth Based on Spatial Rank. *Statistics and Probability Letters,* 65(3), 217-225.

Genz, A. (2011). *QSIMVNV*. Retrieved April 22, 2011, from http://www.math.wsu.edu/faculty/genz/software/software.html.

Gibbons, J.D., & Chakraborti, S. (2003). *Nonparametric Statistical Inference* (4th ed.). New York: Marcel Dekker.

Hamurkaroglu, C., Mert, M., & Saykan, Y. (2004). Nonparametric Control Charts Based on Mahalanobis Depth. *Hacettepe Journal of Mathematics and Statistics*, 33, 57-67.

Hawkins, D.M., & Maboudou-Tchao, E.M. (2007). Self-Starting Multivariate Exponentially Weighted Moving Average Control Charting. *Technometrics,* 49(2), 199-209.

Hayter, A.J., & Tsui, K. (1994). Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Control,* 26, 197-208.

Hotelling, H. (1947). Multivariate Quality Control – Illustrated By the Air Testing of Sample Bombsights. In C. Eisenhart, M.W. Hastay, & W.A. Wallis (Eds.), *Techniques of Statistical Analysis* (pp. 111-184). New York: McGraw-Hill.

Hugg, J., Rafalin, E., Seyboth, K., & Souvaine, D. (2006, January). *An Experimental Study of Old and New Depth Measures*. Paper presented at the Workshop on Algorithm Engineering and Experiments, Miami, FL.

Hugg, J., Rafalin, E., & Souvaine, D. (2006, July). *Depth Explorer – A Software Tool for the Analyis of Depth Measures*. Presented at the International Conference on Robust Statistics, Lisbon, Portugal.

Jackson, J.E. (1991). *A User Guide to Principal Components*. New York: Wiley.

Jensen, W.A., Jones-Farmer, L.A., Champ, C.W., & Woodall, W.H. (2006). Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology,* 38(4), 349-364.

Jensen, W.A., Birch, J.B., & Woodall, W.H. (2007). High Breakdown Estimation Methods for Phase I Multivariate Control Charts. *Quality and Reliability Engineering International,* 23(5), 615-629.

Jobe, J.M., & Pokojovy, M. (2009). A Multistep, Cluster-Based Multivariate Chart for Retrospective Monitoring of Individuals. *Journal of Quality Technology,* 41(4), 323-339.

Johnson, M.E. (1987). *Multivariate Statistical Simulation*. New York: Wiley.

Jones, L.A. (2002). The Statistical Design of EWMA Control Charts with Estimated Parameters. *Journal of Quality Technology*, 34(3), 277-288.

Jones, L.A., & Woodall, W.H. (1998). The Performance of Bootstrap Control Charts. *Journal of Quality Technology*, 30(4), 362-375.

Jones, L.A., Champ, C.W., & Rigdon, S.E. (2001). The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters. *Technometrics*, 43(2), 156-167.

Jones-Farmer, L.A., Jordan, V., & Champ, C.W. (2009). Distribution-Free Phase I Control Charts for Subgroup Location. *Journal of Quality Technology,* 41(3), 304-316.

Kruskal, W.H., & Wallis, W.A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583-621.

Law, A.M., & Kelton, W.D. (2000). *Simulation Modeling and Analysis* (3rd ed.). Boston: McGraw-Hill.

Lehmann, E.L. (2006). *Nonparametrics – Statistical Methods Based on Ranks* (revised 1st ed.). New York: Springer Science+Business Media, LLC.

Li, J., & Liu, R. (2004). New Nonparametric Tests of Multivariate Locations and Scales Using Data Depth. *Statistical Science*, 19(4), 686-696.

Liu, R.Y. (1990). On a Notion of Data Depth Based on Random Simplices. *The Annals of Statistics,* 18, 405-414.

Liu, R.Y. (1995). Control Charts for Multivariate Processes. *Journal of the American Statistical Association,* 90, 1380-1388.

Liu, R.Y., & Singh, K. (1993). A Quality Index Based on Data Depth and Multivariate Rank Tests. *Journal of the American Statistical Association*, 88, 252-260.

Liu, R.Y., Singh, K., & Teng, J.H. (2004). DDMA-Charts: Nonparametric Multivariate Moving Average Control Charts Based on Data Depth. *Allgemeines Statistisches,* 88, 235-258.

Lowry, C.A., Woodall W.H., Champ, C.W., & Rigdon, S.E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics,* 34, 46-53.

Lowry, C.A., & Montgomery, D.C. (1995). A Review of Multivariate Control Charts. *IIE Transactions*, 27, 800-810.

Mahalanobis, P. C. (1936). On the Generalized Distance in Statistics. *Proceedings of the National Institute of Science of India*, 12, 49-55.

Mason, R.L., Champ, C.W., Tracy, N.D., Wierda, S.J., & Young, J.C. (1997). Assessment of Multivariate Process Control Techniques. *Journal of Quality Technology*, 29(2), 140-143.

Mason, R.L., Chou, Y.M., & Young, J.C. (2001). Applying Hotelling's T^2 Statistic to Batch Processes. *Journal of Quality Technology*, 33(4), 466-479.

Mason, R.L., & Young, J.C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Alexandria, VA: American Statistical Association; Philadelphia, PA: Society for Industrial and Applied Mathematics.

Messaoud, A., Weihs, C., & Hering, F. (2008). Detection of Chatter Vibration in a Drilling Process Using Multivariate Control Charts. *Computational Statistics & Data Analysis*, 52(6), 3208-3219.

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

Montgomery, D.C. (2005). *Introduction to Statistical Quality Control* (5th ed.). Hoboken, NJ: Wiley.

Nedumaran, G., & Pignatiello, J.J. (2000). On Constructing T^2 Control Charts for Retrospective Examination. *Communications in Statistics – Simulation and Computation*, 29(2), 621-632.

Nedumaran, G., & Pignatiello, J.J. (2005). On Constructing Retrospective $\bar{X}$ Control Chart Limits. *Quality and Reliability Engineering International*, 21, 81-89.

Oyeyemi, G.M., & Ipinyomi, R.A. (2010). A Robust Method of Estimating Covariance Matrix in Multivariate Data Analysis. African Journal of Mathematics and Computer Science Research, 3(1), 1-18.

Pignatiello, J.J., & Runger, G.C. (1990). Comparison of Multivariate CUSUM Charts. *Journal of Quality Technology*, 22(3), 173-186.

Polansky, A.M. (2005). A General Framework for Constructing Control Charts. *Quality and Reliability Engineering International*, 21, 633-653.

Qiu, P. (2008). Distribution-Free Multivariate Process Control Based on Log-Linear Modeling. *IIE Transactions,* 40(7), 664-691.

Qiu, P., & Hawkins, D. (2001). A Rank-Based Multivariate CUSUM Procedure. *Technometrics* 43(2), 120-132.

Qiu, P., & Hawkins, D. (2003). A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts In All Directions. *The Statistician,* 52(2), 151-164.

Quesenberry, C.P. (1997). *SPC Methods for Quality Improvement*. New York: Wiley.

Rafalin, E.K. (2005). *Algorithms and Analysis of Depth Functions Using Computational Geometry.* Dissertation, Tufts University.

Rousseeuw, P.J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P.J., & Ruts, I. (1996). Algorithm AS 307: Bivariate Location Depth. *Applied Statistics*, 45, 516-526.

Rousseeuw, P.J., & Van Driessen, K. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, 41, 212-223.

Rousseeuw, P.J., & Van Zomeren, B.C. (1990). Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association*, 85, 633-651.

Schaffer, J.R. (1998, August). *A Multivariate Application of the Q Chart*. Paper presented at the 1998 Joint Statistical Meetings, Dallas, TX.

Serfling, R. (2002). A Depth Function and a Scale Curve Based on Spatial Quantiles. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L1-Norm and Related Methods* (pp. 25-28). Berlin, Germany: Birkhäuser.

Serfling, R. (2006). Depth Functions in Nonparametric Multivariate Inference. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (Vol. 72, pp. 1-16). Providence, RI: American Mathematical Society.

Serfling, R. (2010). Equivariance and Invariance Properties of Multivariate Quantile and Related Functions, and the Role of Standardization. *Journal of Nonparametric Statistics*, 22, 915-936.

Serfling, R., & Zuo, Y. (2010). Discussion. *The Annals of Statistics*, 38(2), 676-684.

Shewhart, W.A. (1939). *Statistical Method from the Viewpoint of Quality Control*. New York: Dover Publications.

Stoumbos, Z. G., & Jones, L. A. (2000). On the Properties and Design of Individuals Control Charts Based on Simplicial Depth. *Nonlinear Studies,* 7(2), 147-178.

Stoumbos, Z. G., & Sullivan, J.H. (2002). Robustness to Non-Normality of the Multivariate EWMA Control Chart. *Journal of Quality Technology*, 34(3), 260-276.

Sullivan, J.H., & Woodall, W.H. (1996). A Comparison of Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 28(4), 398-408.

Sullivan, J.H., & Woodall, W.H. (1998). Adapting Control Charts for the Preliminary Analysis of Multivariate Observations. *Communications in Statistics – Simulation and Computation*, 27(4), pp. 953-979.

Sullivan, J.H., & Jones, L.A. (2002). A Self-Starting Control Chart for Multivariate Individual Observations. *Technometrics*, 44(1), 24-33.

Sun, R., & Tsung, F. (2003). A Kernel-Distance-Based Multivariate Control Chart Using Support Vector Methods. *International Journal of Production Research*, 41(13), 2975-2989.

Teng, H.C. (2000). *New Methodology in Regression and Multivariate Quality Control Via Data Depth*. Dissertation, Rutgers University.

Testik, M.C., Runger, G.C., & Borror, C.M. (2003). Robustness Properties of Multivariate EWMA Control Charts. *Quality and Reliability Engineering International*, 19, 31-38.

Testik, M.C., & Borror, C.M. (2004). Design Strategies for the Multivariate Exponentially Weighted Moving Average Control Chart. *Quality and Reliability Engineering International*, 20, 571-577.

Thissen, U., Swierenga, H., de Weijer, A.P., Wehrens, R., Melssen, W.J., & Buydens, L.M.C. (2005). Multivariate Statistical Process Control Using Mixture Modelling [*sic*]. *Journal of Chemometrics*, 19, 23-31.

Tracy, N.D., Young, J.C., & Mason, R.L. (1992). Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 24, 88-95.

Tukey, J. W. (1975). Mathematics and Picturing Data. In R. James (Ed.), *Proceedings of the 1974 International Congress of Mathematicians* (Vol. 2, pp. 523-531). Vancouver, BC.

Vardi, Y., & Zhang, C. (2000). The Multivariate L_1-Median and Associated Data Depth. *Proceedings of the National Academy of Sciences of the USA*, 97(4), 1423-1426.

Vargas, J.A. (2003). Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology,* 35(4), 367-376.

Wierda, S.J. (1994). Multivariate Statistical Process Control – Recent Results and Directions for Future Research. *Statistica Neerlandica*, 48, 147-168.

Willems, G., Pison, G., Rousseeuw, P.J., & Van Aelst, S. (2002). A Robust Hotelling Test. *Metrika*, 55, 125-138.

Wood, M., Kaye, M., & Capon, N. (1999). The Use of Resampling for Estimating Control Chart Limits. *Journal of the Operational Research Society*, 50, 651-659.

Woodall, W.H., & Montgomery, D.C. (1999). Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology,* 31(4), 376-386.

Yanez, S., Gonzalez, N., & Vargas, J.A. (2010). Hotelling's T^2 Control Charts Based on Robust Estimators. *Dyna*, 163, 239-247.

Zamba, K.D., & Hawkins, D.M. (2006). A Multivariate Change-Point Model for Statistical Process Control. *Technometrics*, 48(4), 539-549.

Zarate, P.B. (2004). *Design of Nonparametric Control Chart for Monitoring Multivariate Processes Using Principal Components Analysis and Data Depth*. Dissertation, University of South Florida.

Zuo, Y. (2003). Projection-Based Depth Functions and Associated Medians. *The Annals of Statistics*, 31(5), 1460-1490.

Zuo, Y., & He, X. (2006). On the Limiting Distributions of Multivariate Depth-Based Rank Sum Statistics and Related Tests. *The Annals of Statistics*, 34(6), 2879-2896.

Zuo, Y., & Serfling, R. (2000). General Notions of Statistical Depth Functions. *The Annals of Statistics,* 28(2), 461-482.

Appendices

Appendix A: MATLAB Code for Computing Robust Mahalanobis Depth

Appendix B: MATLAB Code for Computing Mahalanobis Spatial Depth

Appendix C: Expanded Table of Empirical UCLs for the MMR Chart

Appendix D: MATLAB Code for Finding Empirical UCLs for the MMR Chart

Appendix E: Empirical UCLs for Hotelling's T^2 Chart

Appendix F: MATLAB Code for Finding Empirical UCLs for Hotelling's T^2 Chart

Appendix G: MATLAB Code for Assessing MMR Chart Performance

Appendix H: MATLAB Code for Assessing Hotelling's T^2 Chart Performance

Appendix I: Simulation Results Using In-Control Symmetric Data

Appendix J: Simulation Results Using Symmetric Data with an IS in $p = 2$

Appendix K: Simulation Results Using Symmetric Data with an IS in $p = 5$

Appendix L: Simulation Results Using Symmetric Data with an IS in $p = 10$

Appendix M: Simulation Results Using Symmetric Data with a 5% SS in $p = 2$

Appendix N: Simulation Results Using Symmetric Data with a 15% SS in $p = 2$

Appendix O: Simulation Results Using Symmetric Data with a 30% SS in $p = 2$

Appendix P: Simulation Results Using Symmetric Data with a 5% SS in $p = 10$

Appendix Q: Simulation Results Using Symmetric Data with a 15% SS in $p = 10$

Appendix R: Simulation Results Using Symmetric Data with a 30% SS in $p = 10$

Appendix S: Simulation Results Using In-Control Skewed Data

Appendix A: MATLAB Code for Computing Robust Mahalanobis Depth

```matlab
function depth=computeRMDv1(X,Xbar_robust,S_robust)

% Computes the Robust Mahalanobis Depth (RMD) of each point in a multivariate
data set.
% Adapted by Richard Bell on 20100928 from code provided by Satyaki Mazumder
on 20100707.
% X is the multivariate reference data set.
% Xbar_robust is the robust location estimate.
% S_robust is the robust scatter estimate.
% Version 2 uses the square root in the Mahalanobis distance computation,
whereas Version 1 does not.

rows=length(X(:,1));  % identify the number of rows in the sample data set

depth=zeros(rows,1);  % initialize the (rows x 1) vector of depth values for
speed

for i=1:rows
    depth(i)=1/(1+((X(i,:)-Xbar_robust)/S_robust*(X(i,:)-Xbar_robust)'));
% compute the RMD for each observation in the sample; don't use the "mahal"
function in MATLAB because it uses the (nonrobust) sample mean vector and
covariance matrix
end
```

125

Appendix B: MATLAB Code for Computing Mahalanobis Spatial Depth

```matlab
function depth=computeMSDfast(X,S_robust)

% Computes the Mahalanobis Spatial Depth of each point in a multivariate data
set.
% Adapted by Richard Bell on 20100928 from code provided by Satyaki Mazumder
on 20100707.
% X is the (N x p) multivariate reference data set.
% S_robust is the (p x p) robust scatter matrix, raised to the -1/2 power and
used as the transformation-retransformation functional.

Xtr=X/(sqrtm(S_robust));  % transform the data using the TR functional
[rows,cols]=size(Xtr);  % store the dimensions of the transformed data set
depth=zeros(rows,1);  % initialize the vector of depth values for speed

% implementation of the Mahalanobis Spatial Depth function
for i=1:rows  % perform the outer loop for each x
    e=zeros(rows,cols);  % initialize the matrix of unit vectors from x to
all Xi's in the sample
    for j=1:rows  % perform the inner loop to compare each x to all Xi's in
the sample (including itself)
        Euclid=norm(Xtr(i,:)-Xtr(j,:));  % compute the Euclidean distance
between the current x and all Xi's
        if (Euclid~=0)
            e(j,:)=(Xtr(i,:)-Xtr(j,:))/Euclid;  % if the Euclidean distance
is nonzero, use it to normalize the distance between the current x and all
other Xi's in the sample
        else
            e(j,:)=0;  % if the Euclidean distance is zero, x is being
compared to itself so the normalized distance is zero
        end
    end  % end of inner loop
    depth(i)=1-norm(mean(e));  % compute Mahalanobis Spatial Depth of the
point x as one minus the average of the unit vectors from x to all Xi's in
the sample
end  % end of outer loop
```

Appendix C: Expanded Table of Empirical UCLs for the MMR Chart

m	*n*	Desired FAP = 0.10		Desired FAP = 0.05	
		UCL	Simulated FAP	UCL	Simulated FAP
20	5	2.476	0.0941	2.650	0.0486
20	10	2.519	0.0984	2.737	0.0476
30	5	2.581	0.0964	2.749	0.0471
30	10	2.642	0.0975	2.849	0.0477
40	5	2.650	0.0984	2.815	0.0488
40	10	2.724	0.0982	2.925	0.0484
50	5	2.702	0.0983	2.861	0.0487
50	10	2.787	0.0981	2.980	0.0480
60	5	2.743	0.0974	2.895	0.0485
60	10	2.840	0.0983	3.030	0.0486
70	5	2.776	0.0972	2.924	0.0472
70	10	2.881	0.0982	3.065	0.0485
80	5	2.810	0.0967	2.949	0.0487
80	10	2.917	0.0982	3.100	0.0487
90	5	2.831	0.0961	2.969	0.0485
90	10	2.946	0.0983	3.127	0.0489
100	5	2.854	0.0982	2.992	0.0483
100	10	2.972	0.0974	3.150	0.0488
110	5	2.872	0.0980	3.008	0.0482
110	10	2.998	0.0967	3.176	0.0488
120	5	2.890	0.0974	3.019	0.0489
120	10	3.022	0.0980	3.198	0.0478
130	5	2.904	0.0975	3.038	0.0479
130	10	3.042	0.0984	3.214	0.0480
140	5	2.919	0.0984	3.048	0.0486
140	10	3.060	0.0969	3.226	0.0489
150	5	2.932	0.0983	3.057	0.0486
150	10	3.076	0.0971	3.244	0.0486
160	5	2.945	0.0980	3.067	0.0488
160	10	3.088	0.0984	3.262	0.0484
170	5	2.953	0.0983	3.082	0.0477
170	10	3.104	0.0976	3.274	0.0488
180	5	2.964	0.0982	3.089	0.0488
180	10	3.119	0.0977	3.285	0.0485
190	5	2.977	0.0961	3.098	0.0483
190	10	3.134	0.0983	3.300	0.0486
200	5	2.985	0.0981	3.104	0.0485
200	10	3.144	0.0985	3.310	0.0482

Appendix D: MATLAB Code for Finding Empirical UCLs for the MMR Chart

```matlab
%=========================================================================%
%            FINDING EMPIRICAL CONTROL LIMITS FOR THE MMR CHART           %
%=========================================================================%
%  -Created by Richard Bell on 3/1/2011; last updated on 3/22/2011.       %
%  -Variables named for robust Mahalanobis depth (RMD) are used here,     %
%   although this file is not reliant on any particular depth measure.    %
%=========================================================================%


%>>>>> INSTRUCTIONS: Start with 10k iterations to get a ballpark estimate,
then fine-tune with 100k iterations.


clear all  % clear all objects in the MATLAB workspace
clc  % clear the output screen


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%% INPUT SIMULATION PARAMETERS %%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


% AUTOMATED INPUTS (for simulating multiple scenarios using an input file)


% read in m, n, UCL, shift size, and p from an Excel file
iterations=100000;  % number of simulation iterations to be performed
input=xlsread('c:\Users\Rich\Documents\InputFile.xlsx','Sheet1','A1:C50');
inputRows=length(input(:,1));  % determine the number of rows of data in the
input file
APtable=zeros(inputRows,3);  % initialize the array of estimated alarm
probability (AP) values for speed


for row=1:inputRows  % perform the simulation below for each m, n, p, UCL,
and shift size combination in the input file
m=input(row,1);  % read in the desired value for sample size (m)
n=input(row,2);  % read in the desired value for subgroup size (n)
UCL=input(row,3);  % read in the upper control limit
N=m*n;  % determine the pooled sample size (=m in the case of individual
observations)
AP=1;  % initialize the AP to 1 so at least one repetition of the UCL search
will be performed
reps=0;  % initialize the counter for the number of repetitions required to
find the optimal UCL


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%% GENERATE DATA AND COMPUTE ROBUST ESTIMATES %%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


while AP > 0.0985  % set the threshold AP based on the lower limit of an
upper 95% CI for a proportion


UCL=UCL+0.001;  % set the desired increment for each iteration of the UCL
search; use 0.10 first, then 0.01 and 0.001 to refine
reps=reps+1;  % count the number of repetitions required to find the optimal
UCL
```

```matlab
count=0;  % initialize the counter for the number of iterations performed
alarmCount=0;  % initialize the alarm counter

while count < iterations  % run the entire loop for a set number of
iterations

%=====> SIMULATE UNIFORM(0,1) NUMBERS REPRESENTING DEPTH VALUES FROM 0 TO 1

X=unifrnd(0,1,[N,1]);

%=====> PARTITION DATA INTO SUBGROUPS

% assign a subgroup identifier to each simulated data point

i=1;  % start with the first observation in the data set
assigned=0;  % initialize the total number of observations which have been
assigned subgroups
ID=1;  % initialize the subgroup identifier for the first subgroup
subgroup=zeros(N,1);  % initialize the N x 1 vector of subgroup identifiers
for speed

while assigned <= N-n  % perform loop until all observations in the data set
have been assigned subgroup identifiers
size=0;  % initialize the number of observations contained in each subgroup
while size < n  % perform loop until each subgroup reaches size n
subgroup(i)=ID;  % assign the subgroup identifier "ID" to an observation
size=size+1;  % increment the number of observations in the current subgroup
i=i+1;  % move to the next observation
end
ID=ID+1;  % increment the subgroup identifier
assigned=assigned+n;  % increment the total number of observations which have
been assigned subgroups
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%% RANK DATA AND COMPUTE SUBGROUP MEAN RANKS %%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% rank each uniform random number generated

RMDrank=tiedrank(X);  % use the midrank method in the event of a tie; MATLAB
default is to rank from smallest (rank=1) to largest (rank=N)

% compute subgroup mean ranks

subgroup(N+1)=0;  % create a fictitious subgroup identifier for the
nonexistent (N+1)st rank so the following while loop doesn't cause an error
at the Nth rank in the data set
RMDtotal=0;  % initialize the total RMD rank for the first subgroup to 0
i=1;  % initialize the index for the N x 1 vector of ranks resulting from the
depth function
k=1;  % initialize the index for the m x 1 vector of subgroup mean ranks to
be computed
alarm=0;  % initialize the number of RMD alarms to 0
```

```
RMDsubgrpAvg=zeros(m,1);  % initialize the m x 1 vector of RMD subgroup mean
ranks for speed

while i <= N  % perform loop for all N ranks resulting from application of
the depth function
j=i;  % initialize the rank identifier to point to the first observation in
each subgroup
RMDtotal=RMDrank(j);  % initialize the total RMD rank for each subgroup to be
the first rank in the subgroup
while subgroup(j)==subgroup(j+1)  % perform loop until the subgroup
identifier changes
RMDtotal=RMDtotal+RMDrank(j+1);  % add the next RMD rank in the current
subgroup to the total
j=j+1;  % increment the rank identifier by 1
end
RMDsubgrpAvg(k)=RMDtotal/n;  % compute the average subgroup RMD rank for the
current subgroup
k=k+1;  % increment the index for the vector of subgroup mean ranks
i=i+n;  % count the number of ranks for which subgroup averages have been
computed in order to regulate the while loop
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% COMPARE STANDARDIZED SUBGROUP MEAN RANKS TO CONTROL LIMITS %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% compute the theoretical mean and variance of subgroup mean ranks

ExpRbar=(N+1)/2;  % compute the expected value of the subgroup mean rank
VarRbar=((N-n)*(N+1))/(12*n);  % compute the variance of the subgroup mean
rank
Z_RMD=zeros(m,1);  % initialize the m x 1 vector of standardized subgroup RMD
mean ranks

% standardize subgroup mean ranks resulting from the RMD function and compare
to the UCL

for i = 1:m  % perform loop for all m subgroup mean ranks
if alarm==0  % continue loop as long as no alarms occur
Z_RMD(i)=(RMDsubgrpAvg(i)-ExpRbar)/sqrt(VarRbar);  % standardize each
subgroup mean rank
if Z_RMD(i)>UCL  % compare each standardized subgroup mean rank statistic to
the UCL
alarm=1;  % signal if a standardized subgroup mean rank falls above the UCL
end
end
end

if alarm==1
alarmCount=alarmCount+1;  % if a control chart issues an alarm, increment the
counter representing total alarms for all iterations
end
count=count+1;  % increment counter for total number of iterations performed
end
```

```matlab
AP=alarmCount/iterations;  % estimate the alarm probability (AP) for the
current scenario

APtable(row,1)=reps;  % record the results of each UCL evaluation in a table
APtable(row,2)=UCL;
APtable(row,3)=AP;

disp(APtable);  % display AP for the current scenario

end

% send the results to an Excel file
xlswrite('c:\Users\Rich\Documents\OutputFile.xlsx',APtable,'Sheet1','A1');

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%% END OF PROGRAM %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix E: Empirical UCLs for Hotelling's T^2 Chart

Process Distribution	p	m	n	Desired FAP = 0.10	
				UCL	Simulated FAP
t (10)	2	20	5	11.51	0.0970
		50	5	14.01	0.0963
		100	5	15.88	0.0976
		150	5	17.06	0.0971
		200	5	17.94	0.0977
t (3)	2	20	5	15.66	0.0954
		50	5	27.47	0.0973
		100	5	43.79	0.0967
		150	5	58.02	0.0965
		200	5	70.40	0.0976
t (3)	5	20	5	25.79	0.0967
		50	5	42.34	0.0967
		100	5	68.76	0.0958
		150	5	92.55	0.0967
		200	5	114.51	0.0968
t (3)	5	100	5	68.76	0.0958
		100	10	61.43	0.0980
		100	15	57.42	0.0963
		100	20	54.69	0.0970
t (3)	10	20	5	42.41	0.0970
		50	5	59.65	0.0976
		100	5	91.52	0.0971
		150	5	123.51	0.0970
		200	5	154.03	0.0959
lognormal	2	20	5	19.05	0.0977
		50	5	33.05	0.0971
		100	5	50.32	0.0964
		150	5	64.26	0.0975
		200	5	75.56	0.0963
lognormal	5	20	5	29.73	0.0978
		50	5	45.79	0.0976
		100	5	68.01	0.0975
		150	5	86.07	0.0973
		200	5	102.01	0.0977
lognormal	5	100	5	68.01	0.0975
		100	10	56.94	0.0976
		100	15	51.12	0.0974
		100	20	47.25	0.0972

```matlab
%==========================================================================%
%   FINDING EMPIRICAL CONTROL LIMITS FOR HOTELLING'S T^2 CONTROL CHART    %
%==========================================================================%
%  -Created by Richard Bell on 9/15/2010; last updated on 4/26/2011.      %
%  -Based on Hotelling's T2 control chart with Alt's (1976) Phase I UCL    %
%   adjusted for the number of subgroups.                                 %
%  -File is set up to run multiple scenarios; before using, undesired     %
%   sections must be commented out using "%".                             %
%==========================================================================%

%>>>>> INSTRUCTIONS: Start with 10k iterations to get a ballpark estimate,
then fine-tune with 50k iterations.

clear all  % clear all objects in the MATLAB workspace
clc  % clear the output screen

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%% INPUT SIMULATION PARAMETERS %%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% AUTOMATED INPUTS (for simulating multiple scenarios using an input file)

% read in m, n, UCL, shift size, and p from an Excel file
iterations=50000;  % number of simulation iterations to be performed
input=xlsread('c:\Users\Rich\Documents\InputFile.xlsx','Sheet1','A1:E50');
inputRows=length(input(:,1));  % determine the number of rows of data in the
input file
APtable=zeros(inputRows,3);  % initialize the array of estimated alarm
probability (AP) values for speed

for row=1:inputRows  % perform the simulation below for each m, n, p, UCL,
and shift size combination in the input file
m=input(row,1);  % read in the desired value for sample size (m)
n=input(row,2);  % read in the desired value for subgroup size (n)
UCL=input(row,3);  % read in the upper control limit
shiftSize=input(row,4); % read in the desired shift size
p=input(row,5);  % read in the number of variables
N=m*n;  % determine the pooled sample size (=m in the case of individual
observations)
AP=1;  % initialize the AP to 1 so at least one repetition of the UCL search
will be performed
reps=0;  % initialize the counter for the number of repetitions required to
find the optimal UCL

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%% GENERATE DATA AND CONSTRUCT HOTELLING'S T2 CHART %%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

while AP > 0.0978  % set the threshold AP based on the lower limit of an
upper 95% CI for a proportion
```

```
UCL=UCL+0.01;  % set the desired increment for each iteration of the UCL
search; use 1.0 first, then 0.25 and 0.01 to refine
reps=reps+1;  % count number of repetitions required to find the optimal UCL
count=0;  % initialize the counter for the number of iterations performed
alarmCount=0;  % initialize the alarm counter

while count < iterations  % run the entire loop for a set number of
iterations

%=====> SIMULATE MULTIVARIATE NORMAL AND MULTIVARIATE T DATA (ELLIPTICAL)

% OPTION 1: Simulate in-control data.

% multivariate normal distribution
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvnrnd(mu,sigma,N);  % generate multivariate normal data

% multivariate t distribution
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvtrnd(sigma,df,N);  % generate multivariate t data with specified degrees
of freedom

% OPTION 2: Simulate out-of-control data with isolated or sustained shifts of
the mean.

% multivariate normal -- isolated shift of the mean during the first subgroup
only
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu+shift,sigma,n);  % generate the shifted subgroup
Xb=mvnrnd(mu,sigma,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate t -- isolated shift of the mean during the first subgroup only
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
```

```matlab
Xa=mvtrnd(sigma,df,n)+repmat(shift,n,1);  % generate the first subgroup and
add the shift
Xb=mvtrnd(sigma,df,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate normal -- sustained shift of the mean during the last
"percentOC" % of the sample (irrespective of subgroups)
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
percentOC=0.15;  % designate the percentage of out-of-control points
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu,sigma,N-numberOC);  % generate the in-control points
Xb=mvnrnd(mu+shift,sigma,numberOC);  % generate the out-of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate t -- sustained shift of the mean during the last "percentOC" %
of the sample (irrespective of subgroups)
percentOC=0.15;  % designate the percentage of out-of-control points
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvtrnd(sigma,df,N-numberOC);  % generate the in-control points
Xb=mvtrnd(sigma,df,numberOC)+repmat(shift,numberOC,1);  % generate the out-
of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

%=====> SIMULATE MULTIVARIATE LOGNORMAL DATA (SKEWED)

% STEP 1: Simulate uniformly distributed vector of shift directions using
algorithm by Johnson (1987), page 127.

StdNorm=zeros(1,p);  % initialize vector of standard normal random numbers
Unif=zeros(1,p);     % initialize vector of shift directions

for i = 1:p
StdNorm(1,i)=normrnd(0,1);  % generate p independent standard normal variates
end

for i = 1:p
Unif(1,i)=StdNorm(1,i)/sqrt(sum(StdNorm.^2));  % create vector of shift
directions IAW Johnson (1987), page 127
end
```

135

```matlab
% STEP 2: Simulate the sample data set and standardize.

mu_Y=zeros(1,p);  % create a mean vector of all zeros
sigma_Y=eye(p);  % set the covariance matrix equal to the identity matrix
Y=mvnrnd(mu_Y,sigma_Y,N);  % simulate N multivariate normal observations
X=exp(Y);  % transform multivariate normal observations to multivariate
lognormal observations

% NOTE: THE FOLLOWING RESULTS ONLY APPLY TO MULTIVARIATE LOGNORMAL DATA
CREATED USING MULTIVARIATE NORMAL DATA WITH ZERO MEAN VECTOR AND IDENTITY
COVARIANCE MATRIX!

ExpX=exp(1/2);  % compute theoretical expected value of X

sigma_X=zeros(p,p);  % initialize covariance matrix to all zeros
for i=1:p  % fill in diagonals of covariance matrix
for j=1:p
if i==j
sigma_X(i,j)=exp(1)*(exp(1)-1);  % from Law and Kelton (2000), page 382
end
end
end

X=(X-ExpX)/sqrtm(sigma_X);  % standardize multivariate lognormal observations
to have zero mean vector and identity covariance matrix

% STEP 3: Scale the vector of shift directions to achieve a specified
noncentrality parameter.

sigma_X=eye(p);  % specify theoretical covariance matrix of standardized data
Unif=shiftSize*Unif;  % scale the directional shift vector
NCP=sqrt(Unif/sigma_X*Unif');  % check the noncentrality parameter to ensure
it equals the desired value

% STEP 4: Induce isolated or sustained shifts of the mean.

% isolated shift of the mean during the first subgroup only

Xa=X(1:n,:)+repmat(Unif,n,1);  % replicate the shift vector n times and add
it to the first subgroup
Xb=X(n+1:N,:);  % identify the remaining (unshifted) observations in the data
set
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% <OR> sustained shift of the mean during the last "percentOC" % of the
sample (irrespective of subgroups)

percentOC=0.15;  % designate the percentage of out-of-control points
numberOC=round(percentOC*N);  % determine the number of in-control points,
rounded to the nearest integer
Xa=X(1:(N-numberOC),:);  % identify unshifted observations in the data set
Xb=X(N-numberOC+1:N,:)+repmat(Unif,numberOC,1);  % replicate the shift vector
and add it to the remaining observations
X=vertcat(Xa,Xb);  % combine shifted and unshifted data
```

```matlab
%=====> PARTITION DATA INTO SUBGROUPS

% assign a subgroup identifier to each simulated data point

i=1;  % start with the first observation in the data set
assigned=0;  % initialize the total number of observations which have been
assigned subgroups
ID=1;  % initialize the subgroup identifier for the first subgroup
subgroup=zeros(N,1);  % initialize the N x 1 vector of subgroup identifiers
for speed

while assigned <= N-n  % perform loop until all observations in the data set
have been assigned subgroup identifiers
size=0;  % initialize the number of observations contained in each subgroup
while size < n  % perform loop until each subgroup reaches size n
subgroup(i)=ID;  % assign the subgroup identifier "ID" to an observation
size=size+1;  % increment the number of observations in the current subgroup
i=i+1;  % move to the next observation
end
ID=ID+1;  % increment the subgroup identifier
assigned=assigned+n;  % increment the total number of observations which have
been assigned subgroups
end

%=====> COMPUTE ROBUST ESTIMATES OF LOCATION AND SCATTER

subgroupMeans=zeros(m,p);  % initialize the matrix of individual subgroup
mean vectors
totalMeans=zeros(1,p);  % initialize the total of all subgroup mean vectors
totalCovs=zeros(p,p);  % initialize the total of all subgroup covariance
matrices
subgroup(N+1)=0;  % create a fictitious subgroup for the nonexistent (N+1)st
observation so the following while loop doesn't cause an error at the Nth
observation
i=1;  % initialize the index for the N x p vector of observations

while i <= N  % perform loop for all N observations
currentSubgroup=X(i,:);  % start with first observation in the data set
j=i;  % initialize the subgroup index to point to the first observation in
each subgroup
while subgroup(j)==subgroup(j+1)  % perform loop until the subgroup
identifier changes (this is where the fake subgroup is needed)
currentSubgroup=cat(1,currentSubgroup,X(j+1,:));  % combine individual
observations into their respective subgroups
j=j+1;  % increment the subgroup index by 1
end
subgroupMeans(j/n,:)=mean(currentSubgroup);  % store individual subgroup
means in a vector
totalMeans=totalMeans+subgroupMeans(j/n,:);  % keep a running total of all
subgroup mean vectors
totalCovs=totalCovs+cov(currentSubgroup);  % keep a running total of all
subgroup covariance matrices
i=i+n;  % count the number of observations for which subgroup averages have
been computed in order to regulate the while loop
end
```

```matlab
Xbar_robust=totalMeans/m;  % compute average of subgroup means; serves as
unbaised estimate of mean vector
S_robust=totalCovs/m;  % compute average of subgroup variances; serves as
unbiased estimate of covariance matrix


%=====> COMPUTE HOTELLING'S T2 STATISTICS AND COMPARE TO UCL


alarm=0;  % initialize indicator variable representing an alarm (=1) or no
alarm (=0)
T2vector=zeros(m,1);  % initialize vector of T2 statistics


for i=1:m
if alarm==0  % continue loop as long as no false alarms occur
T2stat=n*(subgroupMeans(i,:)-Xbar_robust)/S_robust*(subgroupMeans(i,:)-
Xbar_robust)';  % compute T2 control statistic
T2vector(i)=T2stat;  % store T2 control statistics in a vector
if T2stat > UCL
alarm=1;  % issue a false alarm if the T2 control statistic exceeds the UCL
end
end
end


if alarm==1
alarmCount=alarmCount+1;  % if a control chart issues a false alarm,
increment the counter representing total false alarms for all iterations
end


count=count+1;  % increment the counter for the total number of iterations
performed

end


AP=alarmCount/iterations;  % estimate the alarm probability (AP) for the
current scenario
APtable(row,1)=reps;  % record the results of each UCL evaluation in a table
APtable(row,2)=UCL;
APtable(row,3)=AP;
disp(APtable);  % display AP table for Hotelling's T2 chart on screen, if
desired


end


% send the results to an Excel file
xlswrite('c:\Users\Rich\Documents\OutputFile.xlsx',APtable,'Sheet1','A1');


end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%% END OF PROGRAM %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix G: MATLAB Code for Assessing MMR Chart Performance

```matlab
%=========================================================================%
%          MULTIVARIATE MEAN-RANK (MMR) CONTROL CHART PROGRAM FILE         %
%=========================================================================%
%  -Created by Richard Bell on 9/18/2010; last updated on 3/1/2011.        %
%  -Can be modified to find empirical APs for specified scenarios,         %
%   determine empirical UCLs for specific distributions, or construct      %
%   control charts for preliminary data sets.                             %
%  -File is set up to run multiple scenarios; before using, undesired      %
%   sections must be commented out using "%".                             %
%=========================================================================%


clear all  % clear all objects in the MATLAB workspace
clc  % clear the output screen


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%% INPUT SIMULATION PARAMETERS %%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


% AUTOMATED INPUTS (for simulating multiple scenarios using an input file)

% read in m, n, control limits, shift size, and p from an Excel file
iterations=10000;  % number of simulation iterations to be performed
input=xlsread('c:\Users\Rich\Documents\InputFile.xlsx','Sheet1','A1:E50');
inputRows=length(input(:,1));  % determine the number of rows of data in the
input file
RMD_APtable=zeros(inputRows,1);  % initialize the array of estimated alarm
probability (AP) values for the MMR chart using RMD
MSD_APtable=zeros(inputRows,1);  % initialize the array of estimated alarm
probability (AP) values for the MMR chart using MSD

for row=1:inputRows  % perform the simulation below for each m, n, UCL, shift
size, and p combination in the input file
m=input(row,1);  % read in the desired value for sample size (m)
n=input(row,2);  % read in the desired value for subgroup size (n)
UCL=input(row,3);  % read in the upper control limit (UCL) corresponding to
the m,n combination
shiftSize=input(row,4);  % read in the size of the desired shift
p=input(row,5);  % read in the number of variables

N=m*n;  % determine the pooled sample size (=m in the case of individual
observations)
count=0;  % initialize the counter for the number of iterations performed
RMDalarmCount=0;  % initialize the alarm counter for the RMD function
MSDalarmCount=0;  % initialize the alarm counter for the MSD function


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%% GENERATE DATA AND COMPUTE ROBUST ESTIMATES %%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


while count < iterations  % run the entire loop for a set number of
iterations
```

```
%=====> SIMULATE MULTIVARIATE NORMAL AND MULTIVARIATE T DATA (ELLIPTICAL)

% OPTION 1: Simulate in-control data.

% multivariate normal distribution
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvnrnd(mu,sigma,N);  % generate multivariate normal data

% multivariate t distribution
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvtrnd(sigma,df,N);  % generate multivariate t data with specified degrees
of freedom

% OPTION 2: Simulate out-of-control data with isolated or sustained shifts of
the mean.

% multivariate normal -- isolated shift of the mean during the first subgroup
only
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu+shift,sigma,n);  % generate the shifted subgroup
Xb=mvnrnd(mu,sigma,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate t -- isolated shift of the mean during the first subgroup only
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvtrnd(sigma,df,n)+repmat(shift,n,1);  % generate the first subgroup and
add the shift
Xb=mvtrnd(sigma,df,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate normal -- sustained shift of the mean during the last
"percentOC" % of the sample (irrespective of subgroups)
percentOC=0.15;  % designate the percentage of out-of-control points
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu,sigma,N-numberOC);  % generate the in-control points
Xb=mvnrnd(mu+shift,sigma,numberOC);  % generate the out-of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data
```

```matlab
% multivariate t -- sustained shift of the mean during the last "percentOC" %
of the sample (irrespective of subgroups)
percentOC=0.15;  % designate the percentage of out-of-control points
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvtrnd(sigma,df,N-numberOC);  % generate the in-control points
Xb=mvtrnd(sigma,df,numberOC)+repmat(shift,numberOC,1);  % generate the out-
of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data


%=====> SIMULATE MULTIVARIATE LOGNORMAL DATA (SKEWED)


% STEP 1: Simulate uniformly distributed vector of shift directions using
algorithm by Johnson (1987), page 127.


StdNorm=zeros(1,p);  % initialize vector of standard normal random numbers
Unif=zeros(1,p);     % initialize vector of shift directions


for i = 1:p
StdNorm(1,i)=normrnd(0,1);  % generate p independent standard normal variates
end


for i = 1:p
Unif(1,i)=StdNorm(1,i)/sqrt(sum(StdNorm.^2));  % create vector of shift
directions IAW Johnson (1987), page 127
end


% STEP 2: Simulate the sample data set and standardize.


mu_Y=zeros(1,p);  % create a mean vector of all zeros
sigma_Y=eye(p);  % set the covariance matrix equal to the identity matrix
Y=mvnrnd(mu_Y,sigma_Y,N);  % simulate N multivariate normal observations
X=exp(Y);  % transform multivariate normal observations to multivariate
lognormal observations


% NOTE: THE FOLLOWING RESULTS ONLY APPLY TO MULTIVARIATE LOGNORMAL DATA
CREATED USING MULTIVARIATE NORMAL DATA WITH ZERO MEAN VECTOR AND IDENTITY
COVARIANCE MATRIX!


ExpX=exp(1/2);  % compute theoretical expected value of X
VarX=exp(1)*(exp(1)-1);  % compute theoretical variance of X
X=(X-ExpX)/sqrt(VarX);  % standardize multivariate lognormal observations to
have zero mean vector and identity covariance matrix (can use this method
since the observations are independent)
```

```
% STEP 3: Scale the vector of shift directions to achieve a specified
noncentrality parameter.

sigma_X=eye(p);   % specify theoretical covariance matrix of standardized data
Unif=shiftSize*Unif;   % scale the directional shift vector
NCP=sqrt(Unif/sigma_X*Unif');   % check the noncentrality parameter to ensure
it equals the desired value


% STEP 4: Induce isolated or sustained shifts of the mean.

% isolated shift of the mean during the first subgroup only

Xa=X(1:n,:)+repmat(Unif,n,1);   % replicate the shift vector n times and add
it to the first subgroup
Xb=X(n+1:N,:);   % identify the remaining (unshifted) observations in the data
set
X=vertcat(Xa,Xb);   % combine shifted and unshifted data

% <OR> sustained shift of the mean during the last "percentOC" % of the
sample (irrespective of subgroups)

percentOC=0.15;   % designate the percentage of out-of-control points
numberOC=round(percentOC*N);   % determine the number of in-control points,
rounded to the nearest integer
Xa=X(1:(N-numberOC),:);   % identify the unshifted observations in the data
set
Xb=X(N-numberOC+1:N,:)+repmat(Unif,numberOC,1);   % replicate the shift vector
and add it to the remaining observations
X=vertcat(Xa,Xb);   % combine shifted and unshifted data

%=====> PARTITION DATA INTO SUBGROUPS

% assign a subgroup identifier to each simulated data point

i=1;   % start with the first observation in the data set
assigned=0;   % initialize the total number of observations which have been
assigned subgroups
ID=1;   % initialize the subgroup identifier for the first subgroup
subgroup=zeros(N,1);   % initialize the N x 1 vector of subgroup identifiers
for speed

while assigned <= N-n   % perform loop until all observations in the data set
have been assigned subgroup identifiers
size=0;   % initialize the number of observations contained in each subgroup
while size < n   % perform loop until each subgroup reaches size n
subgroup(i)=ID;   % assign the subgroup identifier "ID" to an observation
size=size+1;   % increment the number of observations in the current subgroup
i=i+1;   % move to the next observation
end
ID=ID+1;   % increment the subgroup identifier
assigned=assigned+n;   % increment the total number of observations which have
been assigned subgroups
end
```

```matlab
%=====> COMPUTE ROBUST ESTIMATES USING HOTELLING'S T^2 OR BACON METHODS

% OPTION 1:  Hotelling's T^2 Method

totalMeans=zeros(1,p);  % initialize the total of all subgroup mean vectors
totalCovs=zeros(p,p);  % initialize the total of all subgroup covariance
matrices
subgroup(N+1)=0;  % create a fictitious subgroup for the nonexistent (N+1)st
observation so the following while loop doesn't cause an error at the Nth
observation
i=1;  % initialize the index for the N x p vector of observations

while i <= N  % perform loop for all N observations
currentSubgroup=X(i,:);  % start with first observation in the data set
j=i;  % initialize the subgroup identifier to point to the first observation
in each subgroup
while subgroup(j)==subgroup(j+1)  % perform loop until the subgroup
identifier changes
currentSubgroup=cat(1,currentSubgroup,X(j+1,:));  % combine individual
observations into their respective subgroups
j=j+1;  % increment the subgroup identifier by 1
end
totalMeans=totalMeans+mean(currentSubgroup);  % keep a running total of all
subgroup mean vectors
totalCovs=totalCovs+cov(currentSubgroup);  % keep a running total of all
subgroup covariance matrices
i=i+n;  % count the number of observations for which subgroup averages have
been computed in order to regulate the while loop
end

Xbar_robust=totalMeans/m;  % compute average of subgroup means; serves as
unbaised estimate of mean vector
S_robust=totalCovs/m;  % compute average of subgroup variances; serves as
unbiased estimate of covariance matrix

% OPTION 2:  BACON method for estimating mean vector and covariance matrix

out=baconV(X,1,.10,4);  % compute BACON estimates for location and scatter
using Mahalanobis distance, alpha=0.05, and c=4; use version 2 (Euclidean
distance) if expected contamination exceeds 20 percent

Xbar_robust=out.center3;  % BACON estimate for mean vector
S_robust=out.cov3;  % BACON estimate for covariance matrix

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%% RANK DATA USING DATA DEPTH %%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% NOTE: The following code simultaneously applies both the robust Mahalanobis
depth (RMD) and Mahalanobis spatial depth (MSD) functions to the same data
set.

[RMD]=computeRMDv1(X,Xbar_robust,S_robust);  % compute the Robust Mahalanobis
Depth of each point in the sample
```

```matlab
RMDrank_interim=tiedrank(RMD);  % rank each depth value; use the midrank
method in the event of a tie; MATLAB default is to rank from smallest
(rank=1) to largest (rank=N)

RMDrank=N-RMDrank_interim+1;  % following data depth convention, adjust the
ranks to go from largest depth value (rank=1) to smallest depth value
(rank=N)

[MSD]=computeMSDfast(X,S_robust);  % compute the Mahalanobis Spatial Depth of
each point in the sample

MSDrank_interim=tiedrank(MSD);  % rank each depth value; use the midrank
method in the event of a tie; MATLAB default is to rank from smallest
(rank=1) to largest (rank=N)

MSDrank=N-MSDrank_interim+1;  % following data depth convention, adjust the
ranks to go from largest depth value (rank=1) to smallest depth value
(rank=N)


% compute subgroup mean ranks

subgroup(N+1)=0;  % create a fictitious subgroup identifier for the
nonexistent (N+1)st rank so the following while loop doesn't cause an error
at the Nth rank in the data set
RMDtotal=0;  % initialize the total RMD rank for the first subgroup to 0
MSDtotal=0; % initialize the total MSD rank for the first subgroup to 0
i=1;  % initialize the index for the N x 1 vector of ranks resulting from the
depth function
k=1;  % initialize the index for the m x 1 vector of subgroup mean ranks to
be computed
RMDalarm=0;  % initialize the number of RMD alarms to 0
MSDalarm=0;  % initialize the number of MSD alarms to 0
RMDsubgrpAvg=zeros(m,1);  % initialize the m x 1 vector of RMD subgroup mean
ranks for speed
MSDsubgrpAvg=zeros(m,1);  % initialize the m x 1 vector of MSD subgroup mean
ranks for speed


while i <= N  % perform loop for all N ranks resulting from application of
the depth function
j=i;  % initialize the rank identifier to point to the first observation in
each subgroup
RMDtotal=RMDrank(j);  % initialize the total RMD rank for each subgroup to be
the first rank in the subgroup
MSDtotal=MSDrank(j);  % initialize the total MSD rank for each subgroup to be
the first rank in the subgroup
while subgroup(j)==subgroup(j+1)  % perform loop until the subgroup
identifier changes
RMDtotal=RMDtotal+RMDrank(j+1);  % add the next RMD rank in the current
subgroup to the total
MSDtotal=MSDtotal+MSDrank(j+1);  % add the next MSD rank in the current
subgroup to the total
j=j+1;  % increment the rank identifier by 1
end
RMDsubgrpAvg(k)=RMDtotal/n;  % compute the average subgroup RMD rank for the
current subgroup
```

144

```
MSDsubgrpAvg(k)=MSDtotal/n;  % compute the average subgroup MSD rank for the
current subgroup
k=k+1;  % increment the index for the vector of subgroup mean ranks
i=i+n;  % count the number of ranks for which subgroup averages have been
computed in order to regulate the while loop
end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% COMPARE STANDARDIZED SUBGROUP MEAN RANKS TO CONTROL LIMITS %%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% compute the theoretical mean and variance of subgroup mean ranks

ExpRbar=(N+1)/2;  % compute the expected value of the subgroup mean rank
VarRbar=((N-n)*(N+1))/(12*n);  % compute the variance of the subgroup mean
rank
Z_RMD=zeros(m,1);  % initialize the m x 1 vector of standardized subgroup RMD
mean ranks
Z_MSD=zeros(m,1);  % initialize the m x 1 vector of standardized subgroup MSD
mean ranks

% standardize subgroup mean ranks resulting from the RMD function and compare
to the UCL

for i = 1:m  % perform loop for all m subgroup mean ranks
if RMDalarm==0  % continue loop as long as no alarms occur; no need to
perform further computations once an alarm occurs (for example, recall that
FAP is the probability of observing ONE OR MORE signals from a control chart
when the process is in control, so the total number of false alarms on a
single chart is irrelevant; same concept applies to EAP in out-of-control
scenarios)
Z_RMD(i)=(RMDsubgrpAvg(i)-ExpRbar)/sqrt(VarRbar);  % standardize each
subgroup mean rank
if Z_RMD(i)>UCL  % compare each standardized subgroup mean rank statistic to
the UCL
RMDalarm=1;  % signal if a standardized subgroup mean rank falls above UCL
end
end
end

if RMDalarm==1
RMDalarmCount=RMDalarmCount+1;  % if a control chart issues an alarm,
increment the counter representing total alarms for all iterations
end

% standardize subgroup mean ranks resulting from the MSD function and compare
to the UCL
for i = 1:m
if MSDalarm==0
Z_MSD(i)=(MSDsubgrpAvg(i)-ExpRbar)/sqrt(VarRbar);
if Z_MSD(i)>UCL
MSDalarm=1;
end
end
end
```

145

```matlab
if MSDalarm==1
MSDalarmCount=MSDalarmCount+1;
end

count=count+1;   % increment the counter for the total number of iterations
performed
end
% record results for both RMD and MSD methods

RMD_AP=RMDalarmCount/iterations;   % estimate the RMD alarm probability (AP)
for the current scenario and store in an array
RMD_APtable(row,1)=RMD_AP;

MSD_AP=MSDalarmCount/iterations;   % estimate the MSD AP for the current
scenario and store in an array
MSD_APtable(row,1)=MSD_AP;

disp('EAP Table for MMR-RMD');
disp(RMD_APtable);   % display AP table for MMR chart using RMD on screen, if
desired
disp('EAP Table for MMR-MSD');
disp(MSD_APtable);   % display AP table for MMR chart using MSD on screen, if
desired

% send the estimated APs to an Excel file
xlswrite('c:\Users\Rich\Documents\OutputFile.xlsx',APtable,'Sheet1','A1');

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%% END OF PROGRAM %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix H: MATLAB Code for Assessing Hotelling's T^2 Chart Performance

```matlab
%============================================================================%
%                   HOTELLING'S T^2 CONTROL CHART PROGRAM FILE                %
%============================================================================%
%  -Created by Richard Bell on 9/15/2010; last updated on 3/22/2011.         %
%  -Based on Hotelling's T2 control chart with Alt's (1976) Phase I UCL       %
%   adjusted for the number of subgroups.                                     %
%  -Can be modified to find empirical APs for specified scenarios,            %
%   determine empirical UCLs for specific distributions, or construct         %
%   control charts for preliminary data sets.                                 %
%  -File is set up to run multiple scenarios; before using, undesired         %
%   sections must be commented out using "%".                                 %
%============================================================================%

clear all  % clear all objects in the MATLAB workspace
clc  % clear the output screen


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%% INPUT SIMULATION PARAMETERS %%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


% AUTOMATED INPUTS (for simulating multiple scenarios using an input file)

% read in m, n, UCL, shift size, and p from an Excel file
iterations=10000;  % number of simulation iterations to be performed
input=xlsread('c:\Users\Rich\Documents\InputFile.xlsx','Sheet1','A1:E50');
inputRows=length(input(:,1));  % determine the number of rows of data in the
input file
APtable=zeros(inputRows,1);  % initialize the array of estimated alarm
probability (AP) values for speed

for row=1:inputRows  % perform the simulation below for each m, n, p, UCL,
and shift size combination in the input file
m=input(row,1);  % read in the desired value for sample size (m)
n=input(row,2);  % read in the desired value for subgroup size (n)
UCL=input(row,3);  % read in the upper control limit
shiftSize=input(row,4); % read in the number of variables
p=input(row,5);  % read in the number of variables

N=m*n;  % determine the pooled sample size (=m in the case of individual
observations)
count=0;  % initialize the counter for the number of iterations performed
alarmCount=0;  % initialize the alarm counter

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%% GENERATE DATA AND CONSTRUCT HOTELLING'S T2 CHART %%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

while count < iterations  % run the entire loop for a set number of
iterations
```

```
%=====> SIMULATE MULTIVARIATE NORMAL AND MULTIVARIATE T DATA (ELLIPTICAL)

% OPTION 1: Simulate in-control data.

% multivariate normal distribution
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvnrnd(mu,sigma,N);  % generate multivariate normal data

% multivariate t distribution
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
X=mvtrnd(sigma,df,N);  % generate multivariate t data with specified degrees
of freedom

% OPTION 2: Simulate out-of-control data with isolated or sustained shifts of
the mean.

% multivariate normal -- isolated shift of the mean during the first subgroup
only
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu+shift,sigma,n);  % generate the shifted subgroup
Xb=mvnrnd(mu,sigma,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate t -- isolated shift of the mean during the first subgroup only
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvtrnd(sigma,df,n)+repmat(shift,n,1);  % generate the first subgroup and
add the shift
Xb=mvtrnd(sigma,df,N-n);  % generate the rest of the (unshifted) sample
X=vertcat(Xa,Xb);  % combine shifted and unshifted data

% multivariate normal -- sustained shift of the mean during the last
"percentOC" % of the sample (irrespective of subgroups)
alpha=.10;  % desired overall false alarm probability (FAP) for the chart
alphaAdjusted=1-(1-alpha)^(1/m);  % desired FAP for each individual
comparison
```

```
UCL=((p*(m-1)*(n-1))/(m*n-m-p+1))*finv(1-alphaAdjusted,p,m*n-m-p+1);  % Alt's
Phase I upper control limit for Hotelling's T2 chart
percentOC=0.15;  % designate the percentage of out-of-control points
mu=zeros(1,p);  % set the mean vector to all zeros
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvnrnd(mu,sigma,N-numberOC);  % generate the in-control points
Xb=mvnrnd(mu+shift,sigma,numberOC);  % generate the out-of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data


% multivariate t -- sustained shift of the mean during the last "percentOC" %
of the sample (irrespective of subgroups)
percentOC=0.30;  % designate the percentage of out-of-control points
df=3;  % degrees of freedom for multivariate t distribution
sigma=eye(p);  % set the covariance matrix equal to the identity matrix
numberOC=round(percentOC*N);  % determine the number of out-of-control
points, rounded to the nearest integer
shift=zeros(1,p);  % initialize the shift vector
shift(1)=shiftSize;  % place the desired shift in the first position of the
shift vector
Xa=mvtrnd(sigma,df,N-numberOC);  % generate the in-control points
Xb=mvtrnd(sigma,df,numberOC)+repmat(shift,numberOC,1);  % generate the out-
of-control points
X=vertcat(Xa,Xb);  % combine shifted and unshifted data


%=====> SIMULATE MULTIVARIATE LOGNORMAL DATA (SKEWED)

% STEP 1: Simulate uniformly distributed vector of shift directions using
algorithm by Johnson (1987), page 127.

StdNorm=zeros(1,p);  % initialize vector of standard normal random numbers
Unif=zeros(1,p);      % initialize vector of shift directions

for i = 1:p
StdNorm(1,i)=normrnd(0,1);  % generate p independent standard normal variates
end

for i = 1:p
Unif(1,i)=StdNorm(1,i)/sqrt(sum(StdNorm.^2));  % create vector of shift
directions IAW Johnson (1987), page 127
end

% STEP 2: Simulate the sample data set and standardize.

mu_Y=zeros(1,p);  % create a mean vector of all zeros
sigma_Y=eye(p);  % set the covariance matrix equal to the identity matrix
Y=mvnrnd(mu_Y,sigma_Y,N);  % simulate N multivariate normal observations
X=exp(Y);  % transform multivariate normal observations to multivariate
lognormal observations
```

```matlab
% NOTE: THE FOLLOWING RESULTS ONLY APPLY TO MULTIVARIATE LOGNORMAL DATA
CREATED USING MULTIVARIATE NORMAL DATA WITH ZERO MEAN VECTOR AND IDENTITY
COVARIANCE MATRIX!

ExpX=exp(1/2);   % compute theoretical expected value of X
sigma_X=zeros(p,p);   % initialize covariance matrix to all zeros
for i=1:p   % fill in diagonals of covariance matrix
for j=1:p
if i==j
sigma_X(i,j)=exp(1)*(exp(1)-1);   % from Law and Kelton (2000), page 382
end
end
end


X=(X-ExpX)/sqrtm(sigma_X);   % standardize multivariate lognormal observations
to have zero mean vector and identity covariance matrix

% STEP 3: Scale the vector of shift directions to achieve a specified
noncentrality parameter.

sigma_X=eye(p);   % specify theoretical covariance matrix of standardized data
Unif=shiftSize*Unif;   % scale the directional shift vector
NCP=sqrt(Unif/sigma_X*Unif');   % check the noncentrality parameter to ensure
it equals the desired value

if abs(NCP-shiftSize)>0.00001   % display error message if calculated NCP does
not equal the shift size (they should be equal since the theoretical
covariance matrix of X is I)
disp('ERROR in NCP!')
end

% STEP 4: Induce isolated or sustained shifts of the mean.

% isolated shift of the mean during the first subgroup only

Xa=X(1:n,:)+repmat(Unif,n,1);   % replicate the shift vector n times and add
it to the first subgroup
Xb=X(n+1:N,:);   % identify the remaining (unshifted) observations in the data
set
X=vertcat(Xa,Xb);   % combine shifted and unshifted data

% <OR> sustained shift of the mean during the last "percentOC" % of the
sample (irrespective of subgroups)

percentOC=0.15;   % designate the percentage of out-of-control points
numberOC=round(percentOC*N);   % determine the number of in-control points,
rounded to the nearest integer
Xa=X(1:(N-numberOC),:);   % identify the unshifted observations in the data
set
Xb=X(N-numberOC+1:N,:)+repmat(Unif,numberOC,1);   % replicate the shift vector
and add it to the remaining observations
X=vertcat(Xa,Xb);   % combine shifted and unshifted data
```

```
%=====> PARTITION DATA INTO SUBGROUPS

% assign a subgroup identifier to each simulated data point

i=1;  % start with the first observation in the data set
assigned=0;  % initialize the total number of observations which have been
assigned subgroups
ID=1;  % initialize the subgroup identifier for the first subgroup
subgroup=zeros(N,1);  % initialize the N x 1 vector of subgroup identifiers
for speed

while assigned <= N-n  % perform loop until all observations in the data set
have been assigned subgroup identifiers
size=0;  % initialize the number of observations contained in each subgroup
while size < n  % perform loop until each subgroup reaches size n
subgroup(i)=ID;  % assign the subgroup identifier "ID" to an observation
size=size+1;  % increment the number of observations in the current subgroup
i=i+1;  % move to the next observation
end
ID=ID+1;  % increment the subgroup identifier
assigned=assigned+n;  % increment the total number of observations which have
been assigned subgroups
end

%=====> COMPUTE ROBUST ESTIMATES OF LOCATION AND SCATTER

subgroupMeans=zeros(m,p);  % initialize the matrix of individual subgroup
mean vectors
totalMeans=zeros(1,p);  % initialize the total of all subgroup mean vectors
totalCovs=zeros(p,p);  % initialize the total of all subgroup covariance
matrices
subgroup(N+1)=0;  % create a fictitious subgroup for the nonexistent (N+1)st
observation so the following while loop doesn't cause an error at the Nth
observation
i=1;  % initialize the index for the N x p vector of observations

while i <= N  % perform loop for all N observations
currentSubgroup=X(i,:);  % start with first observation in the data set
j=i;  % initialize the subgroup index to point to the first observation in
each subgroup
while subgroup(j)==subgroup(j+1)  % perform loop until the subgroup
identifier changes (this is where the fake subgroup is needed)
currentSubgroup=cat(1,currentSubgroup,X(j+1,:));  % combine individual
observations into their respective subgroups
j=j+1;  % increment the subgroup index by 1
end
subgroupMeans(j/n,:)=mean(currentSubgroup);  % store individual subgroup
means in a vector
totalMeans=totalMeans+subgroupMeans(j/n,:);  % keep a running total of all
subgroup mean vectors
totalCovs=totalCovs+cov(currentSubgroup);  % keep a running total of all
subgroup covariance matrices
i=i+n;  % count the number of observations for which subgroup averages have
been computed in order to regulate the while loop
end
```

151

```matlab
Xbar_robust=totalMeans/m;  % compute average of subgroup means; serves as
unbaised estimate of mean vector
S_robust=totalCovs/m;  % compute average of subgroup variances; serves as
unbaised estimate of covariance matrix

%=====> COMPUTE HOTELLING'S T2 STATISTICS AND COMPARE TO UCL

alarm=0;  % initialize indicator variable representing an alarm (=1) or no
alarm (=0)
T2vector=zeros(m,1);  % initialize vector of T2 statistics

for i=1:m
if alarm==0  % continue loop as long as no false alarms occur
T2stat=n*(subgroupMeans(i,:)-Xbar_robust)/S_robust*(subgroupMeans(i,:)-
Xbar_robust)';  % compute T2 control statistic
T2vector(i)=T2stat;  % store T2 control statistics in a vector
if T2stat > UCL
alarm=1;  % issue a false alarm if the T2 control statistic exceeds the UCL
end
end
end

if alarm==1
alarmCount=alarmCount+1;  % if a control chart issues a false alarm,
increment the counter representing total false alarms for all iterations
end

count=count+1;  % increment the counter for the total number of iterations
performed

end

AP=alarmCount/iterations;  % estimate the alarm probability (AP) for the
current scenario and store in an array
APtable(row,1)=AP;

disp('AP Table for Hotellings T2 Chart');
disp(APtable);  % display AP table for Hotelling's T2 chart on screen, if
desired

% send the estimated APs to an Excel file
xlswrite('c:\Users\Rich\Documents\OutputFile.xlsx',APtable,'Sheet1','A1');

end

%%%%%%%%%%%%%%%%%%%%%%%%%%% END OF PROGRAM %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix I: Simulation Results Using In-Control Symmetric Data

Process Distribution	m	n	Empirical FAP for Hotelling's T^2 Chart			Empirical FAP for MMR-RMD Chart			Empirical FAP for MMR-MSD Chart		
			$p = 2$	$p = 5$	$p = 10$	$p = 2$	$p = 5$	$p = 10$	$p = 2$	$p = 5$	$p = 10$
normal	20	5	0.0984	0.0992	0.0995	0.0919	0.0967	0.0947	0.0902		0.0947
	50	5	0.0979	0.0990	0.1019	0.0996	0.1020	0.0975	0.1006		0.0975
	100	5	0.1004	0.1006	0.0979	0.0998	0.0985	0.0972	0.0983		0.0972
	150	5	0.0997	0.1040	0.0995	0.1009	0.1019	0.0990	0.1019		0.0990
	200	5	0.0972	0.1005	0.1016	0.0973	0.0929	0.0973	0.0967		0.0973
$t\,(10)$	20	5	0.1205			0.0851			0.0862		
	50	5	0.1634			0.0961			0.0981		
	100	5	0.1907			0.1031			0.1035		
	150	5	0.2203			0.0939			0.0940		
	200	5	0.2317			0.0988			0.0988		
$t\,(3)$	20	5	0.3040	0.3843	0.4266	0.0973	0.0930	0.0981	0.0959	0.0912	0.0981
	50	5	0.5892	0.7876	0.9055	0.0978	0.1010	0.1004	0.0973	0.1021	0.1004
	100	5	0.7876	0.9591	0.9932	0.0994	0.1022	0.0974	0.0974	0.1009	0.0974
	150	5	0.8864	0.9895	0.9998	0.0950	0.1037	0.0981	0.0957	0.1035	0.0981
	200	5	0.9348	0.9971	1.0000	0.1019	0.1013	0.0957	0.1035	0.1015	0.0957

Appendix J: Simulation Results Using Symmetric Data with an IS in $p = 2$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart								
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
normal	2	20	5	0.0984	0.2523	0.8795	0.9835	0.9990	1.0000	1.0000	1.0000	1.0000
		50	5	0.0979	0.2091	0.8497	0.9847	0.9991	1.0000	1.0000	1.0000	1.0000
		100	5	0.1004	0.1826	0.8222	0.9770	0.9983	0.9999	1.0000	1.0000	1.0000
		150	5	0.0997	0.1703	0.7876	0.9712	0.9979	1.0000	1.0000	1.0000	1.0000
		200	5	0.0972	0.1610	0.7827	0.9641	0.9985	1.0000	1.0000	1.0000	1.0000
$t(10)$	2	20	5	0.0974	0.1918	0.7438	0.9360	0.9917	0.9993	1.0000	1.0000	1.0000
		50	5	0.0934	0.1505	0.6629	0.9110	0.9894	0.9991	0.9998	1.0000	1.0000
		100	5	0.1019	0.1295	0.5951	0.8790	0.9815	0.9978	0.9999	1.0000	1.0000
		150	5	0.0970	0.1231	0.5417	0.8442	0.9733	0.9978	0.9998	1.0000	1.0000
		200	5	0.0997	0.1196	0.4977	0.8268	0.9751	0.9976	0.9999	1.0000	1.0000
$t(3)$	2	20	5	0.0985	0.1154	0.2468	0.4115	0.6287	0.8065	0.9060	0.9811	0.9952
		50	5	0.0987	0.1061	0.1170	0.1496	0.2546	0.4322	0.6503	0.9242	0.9852
		100	5	0.0990	0.0988	0.1058	0.0991	0.1122	0.1488	0.2296	0.6279	0.9233
		150	5	0.1025	0.1005	0.0957	0.0984	0.1055	0.1013	0.1257	0.3033	0.7200
		200	5	0.1007	0.0983	0.0973	0.1060	0.1008	0.1020	0.1115	0.1610	0.4645

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart								
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
normal	2	20	5	0.0919	0.1120	0.4927	0.7795	0.9408	0.9894	0.9984	1.0000	1.0000
		50	5	0.0996	0.1090	0.4482	0.7518	0.9268	0.9861	0.9980	1.0000	1.0000
		100	5	0.0998	0.1092	0.4013	0.7012	0.9063	0.9807	0.9969	0.9998	1.0000
		150	5	0.1009	0.1091	0.3694	0.6740	0.8936	0.9760	0.9959	0.9999	1.0000
		200	5	0.0973	0.1065	0.3565	0.6577	0.8752	0.9771	0.9953	0.9997	1.0000
$t(10)$	2	20	5	0.0851	0.1067	0.3870	0.6811	0.8891	0.9674	0.9912	0.9994	0.9999
		50	5	0.0961	0.1037	0.3447	0.6374	0.8679	0.9622	0.9891	0.9986	0.9998
		100	5	0.1031	0.0977	0.2989	0.5842	0.8339	0.9488	0.9862	0.9974	0.9996
		150	5	0.0939	0.1062	0.2754	0.5473	0.8055	0.9406	0.9804	0.9975	0.9995
		200	5	0.0988	0.1041	0.2518	0.5180	0.8026	0.9335	0.9773	0.9973	0.9992
$t(3)$	2	20	5	0.0973	0.1033	0.2135	0.3926	0.6432	0.8208	0.9263	0.9872	0.9956
		50	5	0.0978	0.1025	0.1774	0.3318	0.5795	0.7957	0.9095	0.9817	0.9939
		100	5	0.0994	0.1014	0.1411	0.2598	0.4940	0.7432	0.8892	0.9765	0.9933
		150	5	0.0950	0.1010	0.1265	0.2344	0.4493	0.6983	0.8667	0.9714	0.9912
		200	5	0.1019	0.0988	0.1251	0.2091	0.4061	0.6681	0.8471	0.9694	0.9914

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-MSD Chart								
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
normal	2	20	5	0.0902	0.1064	0.4549	0.7387	0.9219	0.9836	0.9976	1.0000	1.0000
		50	5	0.1006	0.1087	0.4302	0.7318	0.9168	0.9837	0.9978	1.0000	1.0000
		100	5	0.0983	0.1074	0.3899	0.6902	0.8987	0.9794	0.9968	0.9999	0.9999
		150	5	0.1019	0.1082	0.3631	0.6653	0.8902	0.9755	0.9958	0.9999	1.0000
		200	5	0.0967	0.1059	0.3515	0.6506	0.8717	0.9758	0.9956	0.9999	1.0000
$t(10)$	2	20	5	0.0862	0.1046	0.3508	0.6351	0.8559	0.9543	0.9856	0.9996	0.9997
		50	5	0.0981	0.1047	0.3302	0.6157	0.8513	0.9555	0.9872	0.9989	0.9996
		100	5	0.1035	0.0970	0.2907	0.5702	0.8237	0.9435	0.9817	0.9989	0.9995
		150	5	0.0940	0.1071	0.2679	0.5372	0.7977	0.9382	0.9781	0.9980	0.9995
		200	5	0.0988	0.1041	0.2480	0.5088	0.7948	0.9312	0.9762	0.9979	0.9992
$t(3)$	2	20	5	0.0959	0.1024	0.1914	0.3475	0.5767	0.7645	0.8873	0.9766	0.9956
		50	5	0.0973	0.1010	0.1679	0.3065	0.5462	0.7683	0.8948	0.9793	0.9932
		100	5	0.0974	0.1007	0.1388	0.2502	0.4760	0.7233	0.8790	0.9760	0.9924
		150	5	0.0957	0.0995	0.1253	0.2305	0.4371	0.6847	0.8611	0.9715	0.9902
		200	5	0.1035	0.0992	0.1249	0.2049	0.3936	0.6559	0.8397	0.9657	0.9902

Appendix K: Simulation Results Using Symmetric Data with an IS in $p = 5$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart									
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
normal	5	20	5	0.0992	0.1791	0.7391	0.9939	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0990	0.1532	0.7034	0.9944	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.1006	0.1408	0.6658	0.9935	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.1040	0.1323	0.6440	0.9939	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.1005	0.1212	0.6137	0.9892	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	5	20	5	0.0987	0.1039	0.1697	0.4600	0.6703	0.8194	0.9184	0.9660	0.9930	0.9984
		50	5	0.1045	0.0960	0.1037	0.1538	0.2215	0.3817	0.5886	0.7732	0.9563	0.9934
		100	5	0.0973	0.0986	0.1024	0.0997	0.1047	0.1218	0.1495	0.2523	0.6051	0.9049
		150	5	0.0936	0.0990	0.0978	0.1055	0.1009	0.0960	0.1089	0.1176	0.2296	0.5849
		200	5	0.1008	0.0997	0.1001	0.0992	0.1010	0.1055	0.1043	0.1030	0.1235	0.2728

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart									
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
normal	5	20	5	0.0967	0.1044	0.3214	0.8282	0.9539	0.9911	0.9993	0.9999	1.0000	1.0000
		50	5	0.1020	0.1060	0.2838	0.7967	0.9455	0.9891	0.9988	0.9998	1.0000	1.0000
		100	5	0.0985	0.1069	0.2295	0.7485	0.9173	0.9838	0.9980	1.0000	1.0000	1.0000
		150	5	0.1019	0.1068	0.2223	0.7105	0.9090	0.9799	0.9964	0.9991	1.0000	1.0000
		200	5	0.0929	0.0956	0.2025	0.6996	0.8992	0.9765	0.9965	0.9994	1.0000	1.0000
t (3)	5	20	5	0.0930	0.0980	0.1184	0.2897	0.4630	0.6691	0.8206	0.9200	0.9873	0.9981
		50	5	0.1010	0.1014	0.1032	0.2045	0.3414	0.5607	0.7768	0.9126	0.9881	0.9986
		100	5	0.1022	0.0971	0.1044	0.1474	0.2535	0.4412	0.6859	0.8720	0.9849	0.9973
		150	5	0.1037	0.0998	0.0998	0.1357	0.2105	0.3736	0.6098	0.8190	0.9810	0.9975
		200	5	0.1013	0.0977	0.1074	0.1222	0.1740	0.3234	0.5481	0.7898	0.9761	0.9971

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-MSD Chart									
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
t (3)	5	20	5	0.0912	0.0980	0.1141	0.2520	0.3997	0.5950	0.7547	0.8708	0.9716	0.9943
		50	5	0.1021	0.0998	0.1041	0.1899	0.3110	0.5200	0.7321	0.8848	0.9847	0.9975
		100	5	0.1009	0.0970	0.1033	0.1431	0.2394	0.4147	0.6537	0.8424	0.9792	0.9978
		150	5	0.1035	0.0989	0.0997	0.1336	0.2046	0.3549	0.5846	0.8132	0.9779	0.9968
		200	5	0.1015	0.0988	0.1076	0.1199	0.1706	0.3137	0.5309	0.7691	0.9725	0.9964

Appendix L: Simulation Results Using Symmetric Data with an IS in $p = 10$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart								
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	10	20	5	0.0995	0.1387	0.5438	0.9670	0.9998	1.0000	1.0000	1.0000	1.0000
		50	5	0.1019	0.1268	0.5379	0.9764	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0979	0.1199	0.5038	0.9736	0.9999	1.0000	1.0000	1.0000	1.0000
		150	5	0.0995	0.1197	0.4692	0.9677	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.1016	0.1157	0.4553	0.9667	0.9999	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0987	0.1038	0.1395	0.3195	0.6983	0.9302	0.9894	0.9991	0.9996
		50	5	0.1030	0.1035	0.1068	0.1284	0.2556	0.5971	0.9037	0.9843	0.9981
		100	5	0.0943	0.0971	0.0964	0.1045	0.1131	0.1594	0.3705	0.7529	0.9531
		150	5	0.0979	0.0972	0.0949	0.0985	0.0995	0.1029	0.1342	0.2857	0.6659
		200	5	0.1005	0.0994	0.0993	0.1006	0.1011	0.0973	0.1085	0.1396	0.2816

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart								
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	10	20	5	0.0947	0.1079	0.2509	0.6957	0.9668	0.9992	1.0000	1.0000	1.0000
		50	5	0.0975	0.0993	0.1889	0.6289	0.9546	0.9985	0.9999	1.0000	1.0000
		100	5	0.0972	0.1015	0.1644	0.5607	0.9379	0.9985	0.9999	1.0000	1.0000
		150	5	0.0990	0.1044	0.1518	0.5260	0.9225	0.9964	0.9999	1.0000	1.0000
		200	5	0.0973	0.1026	0.1494	0.5009	0.9162	0.9952	0.9999	1.0000	1.0000
t (3)	10	20	5	0.0981	0.0979	0.1070	0.1528	0.3362	0.6653	0.8966	0.9823	0.9968
		50	5	0.1004	0.1029	0.0966	0.1146	0.2004	0.4930	0.8380	0.9778	0.9973
		100	5	0.0974	0.0999	0.1059	0.1064	0.1402	0.3368	0.7346	0.9640	0.9972
		150	5	0.0981	0.0975	0.0985	0.1005	0.1273	0.2522	0.6426	0.9393	0.9956
		200	5	0.0957	0.1026	0.1034	0.0995	0.1139	0.2106	0.5652	0.9221	0.9947

Appendix M: Simulation Results Using Symmetric Data with a 5% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0956	0.2576	0.8755	0.9835	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.1012	0.2906	0.9625	0.9987	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0990	0.4052	0.9991	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.1031	0.4354	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0993	0.4841	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	2	20	5	0.0985	0.1176	0.2473	0.4248	0.6404	0.8041	0.9079	0.9793	0.9936	0.9985	0.9991
		50	5	0.0987	0.1022	0.1313	0.1891	0.3091	0.5196	0.7544	0.9594	0.9941	0.9989	1.0000
		100	5	0.0990	0.0946	0.1058	0.1186	0.1597	0.2605	0.4569	0.8859	0.9902	0.9992	0.9997
		150	5	0.1025	0.1007	0.1122	0.1134	0.1186	0.1428	0.2103	0.5985	0.9344	0.9944	0.9990
		200	5	0.1007	0.0959	0.1038	0.1071	0.1163	0.1310	0.1561	0.3909	0.8497	0.9856	0.9986

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0919	0.1157	0.4920	0.7743	0.9414	0.9910	0.9989	1.0000	1.0000	1.0000	1.0000
		50	5	0.0996	0.1194	0.5842	0.8872	0.9892	0.9993	0.9999	1.0000	1.0000	1.0000	1.0000
		100	5	0.0998	0.1345	0.7941	0.9872	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.1009	0.1417	0.8360	0.9951	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0973	0.1503	0.8985	0.9987	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	2	20	5	0.0973	0.0953	0.2167	0.4078	0.6423	0.8240	0.9264	0.9848	0.9963	0.9983	0.9993
		50	5	0.0978	0.0984	0.2060	0.4160	0.6855	0.8878	0.9728	0.9983	0.9994	0.9995	0.9998
		100	5	0.0994	0.0993	0.2359	0.4995	0.8355	0.9743	0.9960	0.9994	0.9998	0.9999	1.0000
		150	5	0.0950	0.0981	0.2209	0.4957	0.8335	0.9812	0.9980	0.9998	1.0000	1.0000	1.0000
		200	5	0.1019	0.0986	0.2272	0.5167	0.8672	0.9890	0.9988	0.9997	0.9999	0.9999	1.0000

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
t (3)	2	20	5	0.0959	0.0949	0.1953	0.3621	0.5784	0.7676	0.8914	0.9768	0.9939	0.9976	0.9990
		50	5	0.0973	0.0965	0.1868	0.3507	0.5913	0.8139	0.9349	0.9933	0.9994	0.9995	0.9999
		100	5	0.0974	0.0998	0.2100	0.4191	0.7379	0.9316	0.9879	0.9995	0.9997	0.9996	1.0000
		150	5	0.0957	0.0999	0.2005	0.4121	0.7169	0.9336	0.9934	0.9996	0.9995	0.9999	0.9999
		200	5	0.1035	0.1009	0.2029	0.4271	0.7623	0.9554	0.9957	0.9996	0.9994	1.0000	0.9999

Appendix N: Simulation Results Using Symmetric Data with a 15% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0901	0.3966	0.9854	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.1026	0.4790	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0990	0.6014	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.1009	0.6486	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0941	0.6901	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	2	20	5	0.0985	0.1369	0.3595	0.5747	0.7921	0.9088	0.9617	0.9941	0.9976	0.9992	0.9995
		50	5	0.0987	0.1150	0.1713	0.2584	0.4344	0.6663	0.8420	0.9795	0.9950	0.9988	0.9998
		100	5	0.0990	0.0965	0.1229	0.1425	0.2015	0.3140	0.5217	0.8979	0.9888	0.9964	0.9995
		150	5	0.1025	0.1043	0.1056	0.1240	0.1388	0.1940	0.2585	0.6328	0.9326	0.9915	0.9980
		200	5	0.1007	0.1004	0.1106	0.1202	0.1339	0.1506	0.2000	0.4280	0.8296	0.9790	0.9981

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0919	0.1266	0.5698	0.8621	0.9820	0.9984	0.9998	1.0000	0.9999	0.9997	1.0000
		50	5	0.0996	0.1402	0.6561	0.9375	0.9973	1.0000	1.0000	0.9999	0.9999	1.0000	1.0000
		100	5	0.0998	0.1519	0.7984	0.9876	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.1009	0.1541	0.8408	0.9945	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0973	0.1550	0.8786	0.9978	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	2	20	5	0.0973	0.0996	0.2185	0.3975	0.6356	0.8208	0.9317	0.9935	0.9985	0.9996	1.0000
		50	5	0.0978	0.1042	0.2132	0.3951	0.6463	0.8559	0.9593	0.9974	0.9989	0.9999	0.9999
		100	5	0.0994	0.1071	0.2188	0.4136	0.7085	0.9106	0.9813	0.9985	0.9996	1.0000	1.0000
		150	5	0.0950	0.1041	0.2084	0.4129	0.6911	0.9072	0.9854	0.9993	0.9994	0.9999	0.9999
		200	5	0.1019	0.1021	0.2042	0.3984	0.6999	0.9176	0.9870	0.9993	0.9997	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
t (3)	2	20	5	0.0959	0.0991	0.1814	0.2853	0.4414	0.5971	0.7399	0.9126	0.9722	0.9919	0.9976
		50	5	0.0973	0.1036	0.1751	0.2743	0.4180	0.5841	0.7152	0.9040	0.9677	0.9889	0.9966
		100	5	0.0974	0.1062	0.1811	0.2913	0.4575	0.6264	0.7664	0.9362	0.9844	0.9966	0.9993
		150	5	0.0957	0.1041	0.1771	0.2845	0.4415	0.6041	0.7576	0.9320	0.9835	0.9957	0.9987
		200	5	0.1035	0.1021	0.1740	0.2804	0.4408	0.6185	0.7657	0.9406	0.9859	0.9970	0.9991

Appendix O: Simulation Results Using Symmetric Data with a 30% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0970	0.4486	0.9877	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0971	0.5508	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.1012	0.6391	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0974	0.6695	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0965	0.7107	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	2	20	5	0.0985	0.1478	0.3740	0.5731	0.7725	0.8920	0.9542	0.9875	0.9961	0.9988	0.9993
		50	5	0.0987	0.1181	0.1860	0.2727	0.4284	0.6168	0.8144	0.9666	0.9934	0.9970	0.9986
		100	5	0.0990	0.1020	0.1335	0.1577	0.2076	0.2829	0.4083	0.7743	0.9599	0.9915	0.9973
		150	5	0.1025	0.1073	0.1236	0.1357	0.1635	0.1884	0.2384	0.4802	0.8231	0.9701	0.9917
		200	5	0.1007	0.1094	0.1127	0.1236	0.1404	0.1622	0.1990	0.3221	0.6076	0.9028	0.9820

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
normal	2	20	5	0.0952	0.1097	0.2626	0.4342	0.6330	0.7909	0.9085	0.9922	0.9998	0.9998	1.0000
		50	5	0.0985	0.1159	0.3212	0.5213	0.7209	0.8661	0.9517	0.9974	1.0000	1.0000	1.0000
		100	5	0.0969	0.1245	0.3571	0.5759	0.7708	0.9073	0.9694	0.9988	1.0000	1.0000	1.0000
		150	5	0.0973	0.1185	0.3755	0.6052	0.8109	0.9276	0.9794	0.9989	1.0000	1.0000	1.0000
		200	5	0.0996	0.1233	0.4044	0.6288	0.8314	0.9364	0.9811	0.9994	0.9999	1.0000	1.0000
t (3)	2	20	5	0.0942	0.0987	0.1379	0.1935	0.2853	0.3839	0.5116	0.7353	0.8950	0.9669	0.9947
		50	5	0.0945	0.0997	0.1406	0.1919	0.2743	0.3875	0.5054	0.7428	0.8997	0.9725	0.9957
		100	5	0.0948	0.0948	0.1345	0.1852	0.2681	0.3769	0.4966	0.7524	0.9073	0.9717	0.9947
		150	5	0.0992	0.1012	0.1316	0.1795	0.2609	0.3566	0.4955	0.7442	0.9049	0.9712	0.9933
		200	5	0.0954	0.0981	0.1288	0.1761	0.2598	0.3615	0.4859	0.7498	0.9043	0.9675	0.9915

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$
t (3)	2	20	5	0.0948	0.0941	0.1145	0.1354	0.1679	0.1864	0.2148	0.2668	0.2974	0.3257	0.3451
		50	5	0.0948	0.0992	0.1210	0.1412	0.1702	0.1945	0.2151	0.2586	0.3087	0.3292	0.3553
		100	5	0.0945	0.0972	0.1193	0.1379	0.1687	0.1921	0.2164	0.2722	0.3126	0.3422	0.3705
		150	5	0.0991	0.1021	0.1200	0.1363	0.1662	0.1845	0.2196	0.2712	0.3001	0.3332	0.3621
		200	5	0.0968	0.0981	0.1144	0.1353	0.1668	0.1944	0.2032	0.2691	0.3084	0.3494	0.3631

Appendix P: Simulation Results Using Symmetric Data with a 5% SS in $p = 10$

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.1010	0.1417	0.5407	0.9664	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.1029	0.1569	0.7098	0.9978	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0977	0.1790	0.9170	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0936	0.1894	0.9526	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.1005	0.2041	0.9808	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0987	0.1050	0.1396	0.3116	0.6846	0.9251	0.9896	0.9990	0.9999	0.9999	1.0000
		50	5	0.1030	0.0996	0.1086	0.1491	0.3094	0.6567	0.9241	0.9894	0.9986	0.9999	1.0000
		100	5	0.0943	0.1013	0.1070	0.1153	0.1551	0.2771	0.6562	0.9387	0.9938	0.9992	0.9999
		150	5	0.0979	0.0985	0.1032	0.1086	0.1128	0.1446	0.2346	0.5196	0.8724	0.9851	0.9972
		200	5	0.1005	0.0981	0.0951	0.1016	0.1060	0.1215	0.1591	0.2767	0.6207	0.9207	0.9897

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.0947	0.1107	0.2460	0.6923	0.9681	0.9994	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0975	0.1095	0.2431	0.7505	0.9876	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0972	0.1073	0.3075	0.9262	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0990	0.1029	0.3264	0.9447	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0973	0.1034	0.3605	0.9756	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0981	0.0988	0.1094	0.1587	0.3444	0.6559	0.8976	0.9802	0.9969	0.9994	0.9999
		50	5	0.1004	0.1001	0.1022	0.1186	0.2185	0.4938	0.8180	0.9693	0.9979	0.9994	0.9999
		100	5	0.0974	0.0953	0.1037	0.1167	0.2178	0.5369	0.9032	0.9907	0.9996	1.0000	1.0000
		150	5	0.0981	0.1015	0.0995	0.1096	0.1747	0.4674	0.8701	0.9894	0.9998	1.0000	1.0000
		200	5	0.0957	0.1013	0.1030	0.1146	0.1757	0.4738	0.8888	0.9952	0.9996	1.0000	1.0000

Appendix Q: Simulation Results Using Symmetric Data with a 15% SS in $p = 10$

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.0995	0.1831	0.7445	0.9974	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0950	0.2261	0.9049	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0968	0.2464	0.9848	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0973	0.2713	0.9949	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0982	0.2946	0.9992	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0987	0.1121	0.1772	0.4380	0.8039	0.9680	0.9962	0.9997	0.9999	1.0000	1.0000
		50	5	0.1030	0.1065	0.1329	0.1984	0.4178	0.7694	0.9578	0.9949	0.9994	0.9998	1.0000
		100	5	0.0943	0.1026	0.1113	0.1282	0.1909	0.3578	0.6962	0.9421	0.9944	0.9988	0.9998
		150	5	0.0979	0.0964	0.1045	0.1205	0.1428	0.1828	0.2854	0.5475	0.8508	0.9762	0.9976
		200	5	0.1005	0.0980	0.1040	0.1071	0.1216	0.1386	0.1961	0.3136	0.5851	0.8851	0.9825

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.0947	0.1094	0.2859	0.7599	0.9839	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0975	0.1075	0.2744	0.7892	0.9951	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.0972	0.1088	0.3130	0.9012	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0990	0.1082	0.3314	0.9240	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0973	0.1121	0.3573	0.9519	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0981	0.1004	0.1051	0.1649	0.3047	0.5756	0.8267	0.9511	0.9920	0.9993	1.0000
		50	5	0.1004	0.0959	0.1057	0.1272	0.2051	0.3934	0.6600	0.8851	0.9816	0.9993	1.0000
		100	5	0.0974	0.1009	0.1029	0.1186	0.1881	0.3850	0.6887	0.9214	0.9905	0.9998	1.0000
		150	5	0.0981	0.0908	0.1002	0.1147	0.1675	0.3294	0.6172	0.8768	0.9814	0.9993	1.0000
		200	5	0.0957	0.0973	0.1017	0.1126	0.1703	0.3129	0.6219	0.8908	0.9871	0.9996	1.0000

Appendix R: Simulation Results Using Symmetric Data with a 30% SS in $p = 10$

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.0957	0.2040	0.7375	0.9952	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		50	5	0.0969	0.2504	0.9008	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	5	0.1030	0.2719	0.9682	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		150	5	0.0978	0.2935	0.9860	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		200	5	0.0962	0.3092	0.9934	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0987	0.1122	0.1859	0.4187	0.7523	0.9482	0.9931	0.9994	0.9999	0.9999	1.0000
		50	5	0.1030	0.1094	0.1439	0.2264	0.4202	0.7460	0.9434	0.9938	0.9992	0.9998	1.0000
		100	5	0.0943	0.1066	0.1196	0.1510	0.1979	0.3169	0.5278	0.8218	0.9639	0.9954	0.9994
		150	5	0.0979	0.1004	0.1093	0.1254	0.1489	0.1960	0.2900	0.4356	0.6899	0.9171	0.9870
		200	5	0.1005	0.1012	0.1094	0.1200	0.1337	0.1589	0.2058	0.2651	0.4123	0.6195	0.8686

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
normal	10	20	5	0.1013	0.1136	0.1759	0.4025	0.7251	0.9344	0.9944	0.9993	1.0000	1.0000	1.0000
		50	5	0.1024	0.1121	0.1818	0.4155	0.7651	0.9613	0.9976	1.0000	1.0000	1.0000	1.0000
		100	5	0.0969	0.1032	0.1736	0.4387	0.8137	0.9713	0.9989	1.0000	1.0000	1.0000	1.0000
		150	5	0.0967	0.1001	0.1769	0.4457	0.8223	0.9807	0.9996	1.0000	1.0000	1.0000	1.0000
		200	5	0.1021	0.1057	0.1835	0.4711	0.8435	0.9845	0.9990	1.0000	1.0000	1.0000	1.0000
t (3)	10	20	5	0.0946	0.0996	0.1112	0.1237	0.1710	0.2612	0.4224	0.6628	0.8747	0.9701	0.9923
		50	5	0.0967	0.1015	0.1052	0.1094	0.1406	0.2009	0.2903	0.4543	0.7347	0.9553	0.9973
		100	5	0.0985	0.0997	0.0996	0.1068	0.1324	0.1619	0.2527	0.3618	0.5623	0.9034	0.9960
		150	5	0.0995	0.0993	0.0999	0.1017	0.1253	0.1551	0.2254	0.3298	0.4823	0.8402	0.9928
		200	5	0.0961	0.0969	0.0991	0.1064	0.1221	0.1476	0.2184	0.3138	0.4491	0.7572	0.9903

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
t (3)	10	20	5	0.0909	0.0948	0.1041	0.1093	0.1278	0.1605	0.2134	0.2560	0.3182	0.3837	0.4417
		50	5	0.0960	0.0997	0.1029	0.1060	0.1176	0.1445	0.1661	0.2075	0.2530	0.3155	0.3456
		100	5	0.0979	0.0979	0.0980	0.1037	0.1211	0.1266	0.1586	0.1930	0.2341	0.2755	0.3192
		150	5	0.0992	0.0999	0.0997	0.0996	0.1131	0.1283	0.1493	0.1832	0.2119	0.2544	0.2983
		200	5	0.0965	0.0967	0.0994	0.1033	0.1141	0.1202	0.1459	0.1779	0.2087	0.2390	0.2912

Appendix S: Simulation Results Using In-Control Skewed Data

Process Distribution	m	n	Empirical FAP for Hotelling's T^2 Chart		Empirical FAP for MMR-RMD Chart		Empirical FAP for MMR-MSD Chart	
			$p = 2$	$p = 5$	$p = 2$	$p = 5$	$p = 2$	$p = 5$
lognormal	20	5	0.4414	0.4803	0.0935		0.0965	0.0991
	50	5	0.7618	0.8676	0.1019		0.0994	0.0987
	100	5	0.9353	0.9827	0.1012		0.1030	0.1005
	150	5	0.9779	0.9972	0.0949		0.0957	0.1020
	200	5	0.9915	0.9998	0.0996		0.0997	0.1023

Appendix T: Simulation Results Using Skewed Data with an IS in $p = 2$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart						
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
normal	2	20	5	0.0984	0.1234	0.2527	0.5734	0.8771	0.9857	0.9995
		50	5	0.0979	0.1138	0.2139	0.5128	0.8494	0.9812	0.9994
		100	5	0.1004	0.1041	0.1846	0.4545	0.8181	0.9777	0.9994
		150	5	0.0997	0.1071	0.1717	0.4171	0.7983	0.9682	0.9986
		200	5	0.0972	0.1036	0.1617	0.4075	0.7840	0.9668	0.9978
lognormal	2	20	5	0.0967	0.1071	0.1579	0.3973	0.7084	0.8826	0.9586
		50	5	0.0956						
		100	5	0.1009	0.0991	0.1003	0.1016	0.1254	0.2423	0.5404
		150	5	0.1001						
		200	5	0.0979	0.0999	0.1008	0.1017	0.1008	0.1066	0.1618

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart						
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
normal	2	20	5	0.0919	0.0899	0.1240	0.2274	0.4873	0.7792	0.9401
		50	5	0.0996	0.0962	0.1140	0.2012	0.4509	0.7458	0.9348
		100	5	0.0998	0.0977	0.1085	0.1858	0.4046	0.7065	0.9075
		150	5	0.1009	0.1001	0.1067	0.1725	0.3713	0.6672	0.8929
		200	5	0.0973	0.0995	0.1030	0.1555	0.3491	0.6587	0.8768
lognormal	2	20	5	0.0935	0.1048	0.2979	0.7470	0.9430	0.9861	0.9918
		50	5	0.1019						
		100	5	0.1012	0.1027	0.1518	0.5256	0.9168	0.9788	0.9918
		150	5	0.0949						
		200	5	0.0996	0.0956	0.1173	0.3561	0.8533	0.9684	0.9891

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-MSD Chart						
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
normal	2	20	5	0.0902	0.0887	0.1181	0.2117	0.4489	0.7387	0.9216
		50	5	0.1006	0.0948	0.1107	0.1910	0.4327	0.7244	0.9163
		100	5	0.0983	0.0969	0.1087	0.1809	0.3958	0.6934	0.9005
		150	5	0.1019	0.1003	0.1066	0.1682	0.3618	0.6640	0.8891
		200	5	0.0967	0.0990	0.1033	0.1549	0.3465	0.6517	0.8745
lognormal	2	20	5	0.0965	0.1785	0.3991	0.6790	0.8966	0.9759	0.9914
		50	5	0.0994						
		100	5	0.1030	0.1579	0.3564	0.6137	0.9052	0.9836	0.9938
		150	5	0.0957						
		200	5	0.0997	0.1366	0.3322	0.5517	0.8622	0.9805	0.9918

Appendix U: Simulation Results Using Skewed Data with an IS in $p = 5$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart								
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$
lognormal	5	20	5	0.0934	0.1119	0.1279	0.2515	0.5461	0.8207	0.9455	0.9851	0.9945
		50	5	0.0963								
		100	5	0.0958	0.0968	0.0975	0.1001	0.1110	0.1400	0.2783	0.6024	0.8719
		150	5	0.0968								
		200	5	0.0967	0.0953	0.0976	0.1012	0.0989	0.1018	0.1128	0.1577	0.3426

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-MSD Chart								
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$
lognormal	5	20	5	0.0991	0.0977	0.1258	0.2897	0.5647	0.8285	0.9420	0.9855	0.9962
		50	5	0.0987								
		100	5	0.1005	0.1001	0.0964	0.1760	0.4048	0.7213	0.9350	0.9926	0.9987
		150	5	0.1020								
		200	5	0.1023	0.0969	0.1038	0.1370	0.3094	0.6104	0.8813	0.9850	0.9986

Appendix V: Simulation Results Using Skewed Data with a 5% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0967	0.1072	0.1577	0.3914	0.7022	0.8842	0.9589	0.9839	0.9938	0.9967	0.9974
		50	5	0.0956										
		100	5	0.1009	0.1028	0.1055	0.1258	0.1868	0.4334	0.7751	0.9465	0.9876	0.9978	0.9996
		150	5	0.1001										
		200	5	0.0979	0.1012	0.1024	0.1102	0.1258	0.1842	0.3438	0.7156	0.9361	0.9885	0.9983

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0935	0.1072	0.3005	0.7387	0.9436	0.9821	0.9934	0.9966	0.9985	0.9990	0.9993
		50	5	0.1019										
		100	5	0.1012	0.1024	0.2126	0.7795	0.9933	0.9995	0.9999	1.0000	1.0000	1.0000	1.0000
		150	5	0.0949										
		200	5	0.0996	0.1020	0.1746	0.7258	0.9955	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0965	0.1868	0.4053	0.6796	0.8998	0.9737	0.9921	0.9963	0.9984	0.9995	0.9993
		50	5	0.0994										
		100	5	0.1030	0.2552	0.5387	0.7700	0.9632	0.9987	0.9999	1.0000	1.0000	1.0000	1.0000
		150	5	0.0957										
		200	5	0.0997	0.2678	0.5578	0.7762	0.9664	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000

Appendix W: Simulation Results Using Skewed Data with a 15% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0967	0.1150	0.2044	0.5074	0.8041	0.9374	0.9746	0.9904	0.9972	0.9984	0.9991
		50	5	0.0956										
		100	5	0.1009	0.1058	0.1160	0.1553	0.2504	0.4650	0.7737	0.9310	0.9858	0.9960	0.9982
		150	5	0.1001										
		200	5	0.0979	0.1011	0.1056	0.1287	0.1626	0.2307	0.4078	0.6638	0.8988	0.9806	0.9957

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0935	0.1099	0.2674	0.6685	0.9247	0.9895	0.9981	0.9993	0.9999	0.9998	0.9997
		50	5	0.1019										
		100	5	0.1012	0.1075	0.1904	0.5607	0.9053	0.9927	0.9992	1.0000	1.0000	1.0000	1.0000
		150	5	0.0949										
		200	5	0.0996	0.1050	0.1649	0.4819	0.8685	0.9907	0.9993	1.0000	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0965	0.2099	0.4151	0.6046	0.7839	0.9072	0.9648	0.9889	0.9955	0.9983	0.9993
		50	5	0.0994										
		100	5	0.1030	0.2263	0.4239	0.6243	0.8080	0.9372	0.9857	0.9976	0.9997	0.9998	0.9999
		150	5	0.0957										
		200	5	0.0997	0.2103	0.4125	0.6021	0.8056	0.9404	0.9894	0.9978	0.9989	0.9997	0.9996

Appendix X: Simulation Results Using Skewed Data with a 30% SS in $p = 2$

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0967	0.1256	0.2241	0.4939	0.7701	0.9173	0.9714	0.9868	0.9949	0.9973	0.9989
		50	5	0.0956										
		100	5	0.1009	0.1073	0.1247	0.1722	0.2658	0.4155	0.6337	0.8513	0.9522	0.9859	0.9962
		150	5	0.1001										
		200	5	0.0979	0.1048	0.1144	0.1400	0.1746	0.2546	0.3735	0.5437	0.7171	0.8843	0.9654

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0955	0.1013	0.1440	0.2777	0.5222	0.7769	0.9319	0.9846	0.9972	0.9993	1.0000
		50	5	0.0983										
		100	5	0.0982	0.0999	0.1368	0.2239	0.3961	0.5983	0.7691	0.9078	0.9808	0.9985	0.9997
		150	5	0.1008										
		200	5	0.0992	0.1019	0.1386	0.2079	0.3640	0.5372	0.7020	0.8314	0.9433	0.9940	0.9997

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	2	20	5	0.0965	0.1458	0.1946	0.2119	0.2345	0.2724	0.2992	0.3175	0.3537	0.3645	0.3771
		50	5	0.0994										
		100	5	0.1030	0.1390	0.1630	0.1993	0.2314	0.2801	0.3116	0.3556	0.3829	0.3986	0.4240
		150	5	0.0957										
		200	5	0.0997	0.1258	0.1562	0.1843	0.2258	0.2756	0.3145	0.3499	0.3852	0.4220	0.4270

Appendix Y: Simulation Results Using Skewed Data with a SS in $p = 5$

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0934	0.1031	0.1236	0.2490	0.5473	0.8198	0.9456	0.9857	0.9970	0.9989	0.9998
		50	5	0.0963										
		100	5	0.0958	0.0980	0.1020	0.1112	0.1419	0.2260	0.5187	0.8406	0.9713	0.9965	0.9994
		150	5	0.0968										
		200	5	0.0967	0.0947	0.0973	0.1025	0.1097	0.1356	0.1916	0.3595	0.6965	0.9435	0.9919

Process Distribution	p	m	n	Empirical AP for a 5% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0991	0.0920	0.1296	0.2995	0.5772	0.8201	0.9460	0.9865	0.9974	0.9989	0.9998
		50	5	0.0987										
		100	5	0.1005	0.1050	0.1067	0.2296	0.5250	0.8371	0.9689	0.9974	0.9994	0.9999	1.0000
		150	5	0.1020										
		200	5	0.1023	0.0998	0.1054	0.1922	0.4607	0.7969	0.9658	0.9989	0.9999	0.9999	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0934	0.1109	0.1574	0.3359	0.6643	0.8916	0.9749	0.9945	0.9989	0.9996	0.9999
		50	5	0.0963										
		100	5	0.0958	0.1013	0.1085	0.1309	0.1843	0.2998	0.5539	0.8396	0.9653	0.9960	0.9988
		150	5	0.0968										
		200	5	0.0967	0.0958	0.0993	0.1123	0.1282	0.1688	0.2459	0.4111	0.6712	0.9096	0.9856

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0991	0.0907	0.1251	0.2176	0.3862	0.6074	0.7847	0.9006	0.9572	0.9829	0.9931
		50	5	0.0987										
		100	5	0.1005	0.1046	0.1055	0.1588	0.2801	0.4819	0.6916	0.8418	0.9446	0.9816	0.9932
		150	5	0.1020										
		200	5	0.1023	0.0978	0.1034	0.1453	0.2316	0.3983	0.6119	0.8036	0.9173	0.9730	0.9907

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0934	0.1137	0.1713	0.3298	0.6137	0.8508	0.9568	0.9896	0.9973	0.9984	0.9997
		50	5	0.0963										
		100	5	0.0958	0.1014	0.1143	0.1431	0.1914	0.2863	0.4627	0.6808	0.8868	0.9739	0.9942
		150	5	0.0968										
		200	5	0.0967	0.0967	0.1020	0.1200	0.1401	0.1793	0.2447	0.3458	0.5092	0.7166	0.8836

Process Distribution	p	m	n	Empirical AP for a 30% Sustained Shift Using the MMR-MSD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	20	5	0.0991	0.0878	0.1065	0.1140	0.1477	0.1793	0.2306	0.2725	0.3227	0.3643	0.4048
		50	5	0.0987										
		100	5	0.1005	0.1028	0.0992	0.1102	0.1316	0.1546	0.2014	0.2350	0.2736	0.3244	0.3615
		150	5	0.1020										
		200	5	0.1023	0.0975	0.0980	0.1150	0.1283	0.1491	0.1870	0.2183	0.2597	0.2960	0.3410

Appendix Z: Subgroup Size Analysis Using In-Control Data

Process Distribution	p	m	n	Empirical FAP for Hotelling's T^2 Chart	Empirical FAP for MMR-RMD Chart
t (3)	5	100	5	0.9541	0.0964
		100	10	0.9054	0.1050
		100	15	0.8708	0.1020
		100	20	0.8332	0.1047
lognormal	5	100	5	0.9833	0.0986
		100	10	0.9437	0.1042
		100	15	0.9029	0.0949
		100	20	0.8602	0.1030

Appendix AA: Subgroup Size Analysis Using Data with an IS in $p = 5$

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart							
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
$t\,(3)$	5	100	5	0.0999	0.0991	0.0990	0.0990	0.1201	0.2394	0.6120	0.8989
		100	10	0.0942	0.0999	0.0974	0.1434	0.5595	0.9532	0.9973	0.9998
		100	15	0.0984	0.1000	0.1075	0.4214	0.9614	0.9990	1.0000	0.9999
		100	20	0.1004	0.0994	0.1402	0.8056	0.9971	0.9999	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart							
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
$t\,(3)$	5	100	5	0.1022	0.1013	0.1055	0.1498	0.4484	0.8643	0.9851	0.9975
		100	10	0.0986	0.1038	0.1395	0.5409	0.9767	0.9998	1.0000	1.0000
		100	15	0.0964	0.1030	0.1915	0.8414	0.9995	1.0000	1.0000	0.9999
		100	20	0.1021	0.1050	0.2741	0.9650	0.9999	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using Hotelling's T^2 Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	100	5	0.0933	0.0925	0.0977	0.1019	0.1080	0.1333	0.2812	0.5998	0.8664	0.9692	0.9929
		100	10	0.1016	0.0951	0.0952	0.1190	0.2992	0.7866	0.9787	0.9987	0.9997	1.0000	1.0000
		100	15	0.0977	0.0998	0.1133	0.2610	0.8405	0.9947	0.9998	0.9999	1.0000	1.0000	1.0000
		100	20	0.0898	0.1038	0.1310	0.6108	0.9852	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for an Isolated Shift Using the MMR-RMD Chart										
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5$
lognormal	5	100	5	0.0986	0.1016	0.1028	0.1130	0.2076	0.5308	0.8765	0.9826	0.9976	0.9994	1.0000
		100	10	0.1042	0.1012	0.1173	0.3757	0.9080	0.9981	1.0000	1.0000	1.0000	1.0000	1.0000
		100	15	0.0949	0.1093	0.1767	0.7495	0.9972	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		100	20	0.1030	0.1085	0.2768	0.9339	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Appendix BB: Subgroup Size Analysis Using Data with a 15% SS in $p = 5$

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T2 Chart							
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
$t(3)$	5	100	5	0.0963	0.0980	0.1120	0.1361	0.2236	0.5038	0.8621	0.9806
		100	10	0.0917	0.1067	0.1270	0.3049	0.8734	0.9970	1.0000	1.0000
		100	15	0.1003	0.1015	0.1805	0.7814	0.9979	1.0000	1.0000	1.0000
		100	20	0.1021	0.1144	0.2951	0.9834	0.9999	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-RMD Chart							
				$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
$t(3)$	5	100	5	0.0950	0.0996	0.1130	0.2132	0.5115	0.8817	0.9861	0.9988
		100	10	0.0989	0.1058	0.1997	0.7429	0.9961	0.9997	0.9999	1.0000
		100	15	0.0998	0.0985	0.3297	0.9696	0.9996	1.0000	1.0000	1.0000
		100	20	0.0996	0.1147	0.4881	0.9977	1.0000	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using Hotelling's T^2 Chart							
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$
lognormal	5	100	5	0.0933	0.0966	0.1075	0.1304	0.1843	0.2998	0.5591	0.8385	0.9673
		100	10	0.1016	0.1014	0.1256	0.2416	0.6156	0.9663	0.9992	0.9998	1.0000
		100	15	0.0977	0.1111	0.1791	0.5726	0.9905	0.9999	1.0000	1.0000	1.0000
		100	20	0.0898	0.1202	0.2824	0.9500	0.9999	1.0000	1.0000	1.0000	1.0000

Process Distribution	p	m	n	Empirical AP for a 15% Sustained Shift Using the MMR-RMD Chart							
				$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$
lognormal	5	100	5	0.0986	0.1026	0.1078	0.1390	0.2592	0.5200	0.8104	0.9541	0.9957
		100	10	0.1042	0.1035	0.1601	0.5318	0.9633	0.9994	1.0000	1.0000	1.0000
		100	15	0.0949	0.1157	0.2799	0.8985	0.9995	1.0000	1.0000	1.0000	1.0000
		100	20	0.1030	0.1189	0.4525	0.9899	1.0000	1.0000	1.0000	1.0000	1.0000