MULTIVARIATE STATISTICAL PROCESS CONTROL FOR CORRELATION MATRICES

by

Mark F. Sindelar

B. S. in Electrical Engineering, University of Akron, 1995

M. B. A., John Carroll University, 1999

M. S. in Industrial Engineering, University of Pittsburgh, 2001

Submitted to the Graduate Faculty of the

School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

SCHOOL OF ENGINEERING

This dissertation was presented

by

Mark F. Sindelar

It was defended on

March 1, 2007

and approved by

Mainak Mazumdar, Professor Emeritus, Industrial Engineering

Larry J. Shuman, Professor & Associate Dean for Academic Affairs, School of Engineering

Pandu R. Tadikamalla, Professor, Business Administration

Harvey Wolfe, Professor Emeritus, Industrial Engineering

Dissertation Director: Mary E. Besterfield-Sacre, Associate Professor, Industrial Engineering

Copyright © by Mark F. Sindelar

2007

MULTIVARIATE STATISTICAL PROCESS CONTROL

FOR CORRELATION MATRICES

Mark F. Sindelar, Ph.D.

University of Pittsburgh, 2007

Measures of dispersion in the form of covariance control charts are the multivariate analog to the univariate R-chart, and are used in conjunction with multivariate location charts such as the Hotelling T^2 chart, much as the R-chart is the companion to the univariate X-bar chart. Significantly more research has been directed towards location measures, but three multivariate statistics ($|\boldsymbol{S}|$, W_i, and G) have been developed to measure dispersion. This research explores the correlation component of the covariance statistics and demonstrates that, in many cases, the contribution of the correlation component is less significant than originally believed, but also offers suggestions for how to implement a correlation control chart when this is the variable of primary interest.

This research mathematically analyzes the potential use of the three covariance statistics ($|\boldsymbol{S}|$, W_i, and G), modified for the special case of correlation. A simulation study is then performed to characterize the behavior of the two modified statistics that are found to be feasible. Parameters varied include the sample size (n), number of quality characteristics (p), the variance, and the number of correlation matrix entries that are perturbed. The performance and utility of the front-running correlation (modified W_i) statistic is then examined by comparison to similarly classed statistics and by trials with real and simulated data sets, respectively. Recommendations for the development of correlation control charts are presented along with a description of the

types of process to which they apply. An outgrowth of the research is the understanding that the correlation component often does not contribute as much as the scale factor component of the dispersion measure in many cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation is dedicated to Wilma Powell, for her encouragement, support, understanding and friendship.

My father, Kenneth F. Sindelar, was the first engineer in my life and the one that deserves credit for introducing me at an early age to the basic concepts of Engineering which would become the foundation of my schooling and profession; my mother, Mary Ann V. Sindelar deserves thanks for her patience, reassurance, and encouragement through my many decades of education.

United States Steel Corporation, and particularly my Engineering Manager, Raymond W. Boronyak, have made numerous accommodations to assist my pursuit of this degree. This cooperation has been very much appreciated.

I am indebted to my advisor and dissertation committee chair, Mary E. Besterfield-Sacre and am thankful for the contributions of committee members Mainak Mazumdar, Larry J. Shuman, Pandu R. Tadikamalla, and Harvey Wolfe.

I am also thankful for the assistance of Murat Caner Testik, Associate Professor and Department Chair, Industrial Engineering, of Hacettepe University, who reviewed the derivations of Chapter 3 and offered helpful suggestions.

Finally, I thank Robert S. Parker, Associate Professor of Chemical Engineering at the University of Pittsburgh, for use of the simulator software of Section 7.3, and the many fine people in the University Library System for their assistance obtaining resources.

1.0 INTRODUCTION

Multivariate quality control charts expand the options for process monitoring beyond the traditional univariate case by allowing for the simultaneous monitoring of several variables. In the multivariate setting, many developments have provided improved charts for monitoring location (*i.e.* the mean vector) while relatively little research has addressed measures of dispersion (Alt, 1985; Golnabi & Houshmand, 1996). However, it has been recognized that dispersion may have a significant impact on the location vector (Hayter & Tsui, 1994; Houshmand *et al.*, 1997).

Thus, while three statistics to measure dispersion exist ($|\boldsymbol{S}|$, W_i, and G), a comparison has not been made among them to determine their effectiveness for a variety of process conditions. The sample generalized variance, $|\boldsymbol{S}|$, is the oldest of the statistics, but several authors have cited possible disadvantages. The W_i statistic was introduced by Alt (1985) and appears to overcome many of the drawbacks reported with using the $|\boldsymbol{S}|$ chart. More recently, the G chart has been introduced by Levinson *et al.* (2002) and incorporates mean square successive differences in an attempt to increase sensitivity of the dispersion measure. Each of these statistics can be plotted in a traditional quality control manner against time or lot number to monitor a process as shown in Figure 1 where “CL” (center line) is the mean value and “UCL” and “LCL” are the upper and lower control limits, respectively. If the plotted point exceeds the control limits (as shown by the last point) the statistic is out-of-control, indicating a possible change in the process.

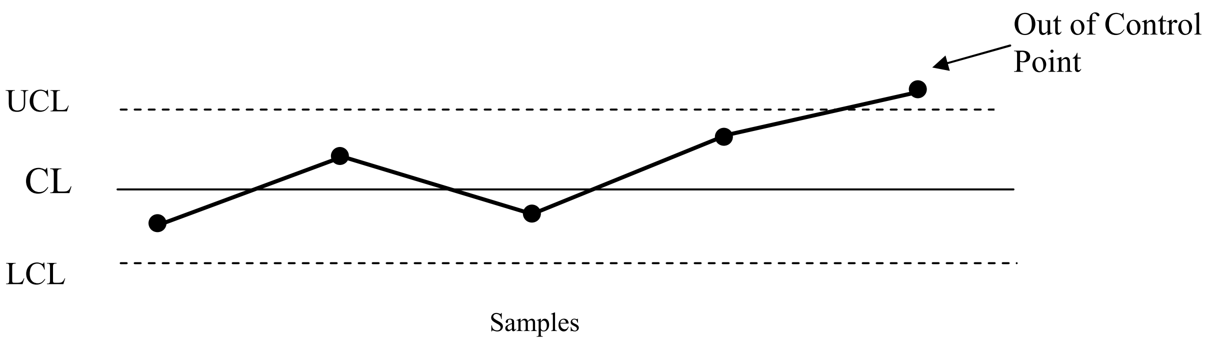


Figure 1 Control Chart

Multivariate dispersion charts are usually intended to supplement location charts, much as the univariate *R*-chart is used as a supplement to the *X*-bar chart. Despite this orientation, several authors have indicated that dispersion, in the form of the variance-covariance matrix, is often a component of location and, therefore, affects changes in the mean vector (Lowry and Montgomery, 1995; Golnabi and Houshmand, 1996). Understanding the behavior of covariance independent of the mean vector in a statistical quality control setting guides the direction of this research.

As an avenue to begin an investigation, consider instead the correlation matrix, a standardized variance-covariance matrix, for the same data set. The diagonal elements, each representing the correlation of a number with itself, are all equal to unity and provide an upper boundary condition. This bounded nature of the correlation matrix provides an alternative and initial approach to the comparison of dispersion statistics. Furthermore, this approach also separates the covariance matrix into two components: the correlation matrix and a scale factor. The contribution of the correlation component to the dispersion measure can, thus, also be evaluated in this manner and this is the focus of this dissertation.

The generation and use of correlation control charts rather than covariance charts may have some advantages in application as well as contribute to understanding dispersion in

Multivariate Statistical Process Control (MSPC). Similar to the variance-covariance case, changes in the correlation matrix could be used to help explain changes in the location vector that appear on a location chart such as the Hotelling T^2 chart (1947). A chart that monitors for changes in the correlation matrix for p quality characteristics could also be used to monitor for changes in the dispersion, specifically the correlation, matrix separate from the location vector.

Additionally, if the data for the in-control correlation matrix originates from a process where no changes are expected, then changes in the sampled correlation matrix may indicate that the underlying assumptions have changed and a new analysis for the current data would likely reveal the discrepancies. This is particularly applicable to process data such as that found in the chemicals industry where certain relationships are assumed to remain independent when the process is stable.

1.1 OBJECTIVES OF THE RESEARCH

This research investigates the modification of known statistics for covariance ($|\boldsymbol{S}|$, W_i, and G) to the special case of correlation to: (1) compare and evaluate the three statistics for a number of parameters; (2) provide guidance for the application of said statistics to correlation control charts; (3) investigate the application of correlation control charts; and (4) make assessments of the role correlation plays in dispersion.

1.2 RATIONALE FOR THE DEVELOPMENT OF CORRELATION CONTROL CHARTS

This research focuses on evaluation of the dispersion statistics used for development of multivariate control charts where the unique nature of correlation matrices would make the correlation control chart itself a tool in certain circumstances where the measures of linear association are of interest to an organization.

For example, in the chemicals industry a distillation column is expected to produce outputs which are correlated with one another because of their stoiciometric relationship. A change in the feedstock composition could change the correlation between outputs if this balance is upset. All the outputs of the column may be saleable but, if their relation changes due to a change in the input feedstock, then the column operator should be made aware.

This scenario also illustrates an important observation for interpretation of correlation control charts. Namely, it is the proposition that a shift in the correlation matrix does not necessarily denote an out-of-control condition in the traditional sense of the term. The shift may be attributable to a change in requirements, as opposed to requirements not being satisfied. Thus, the study of correlation, the standardized version of the variance-covariance matrix, provides additional value in its conceptual form.

1.3 ORGANIZATION OF THE WORK

To begin an exploration of correlation as a measure of dispersion, it is illustrative to start with the first MSPC chart and statistic. In Chapter 2, the Hotelling T^2 chart, which considered the location vector, is introduced. Since its inception in 1947, additional research has modified and improved upon Hotelling's approach to the monitoring of the location vector. Charts for dispersion then followed, albeit less in number. The Literature Review presented in Chapter 2 places particular emphasis on the dispersion statistics for covariance ($|\boldsymbol{S}|$, W_i, and G) that are considered in modified form in this research.

The evaluation of the three statistics in Chapter 3 begins with a mathematical analysis to determine the applicability of the covariance statistics modified to the correlation special case and investigates the associated control limits. From these analyses, the $|\boldsymbol{S}|$ and W_i statistics in modified form for correlation emerge as feasible possibilities for the construction of control charts. The bounded nature of the correlation matrix, while providing a framework for the analysis, also raises some concerns about the sensitivity of the control limits in certain situations.

Chapter 4 starts with a simulation study to evaluate the correlation case for various levels of quality characteristics (p) and sample size (n) for two common matrices. A terminating sequential simulation provides data with the performance measure being the Average Run Length, or *ARL*. The results of the simulation indicate that only a modified form of the W_i statistic (noted as the W_{R} statistic) is possible, under specified conditions, for the development of correlation control charts. An empirical equation developed from the simulation results is then used to characterize the behavior of this statistic.

A comparison with similarly developed MSPC statistics is presented in Chapter 5, along with a discussion of the effect that each parameter has on the statistic. Existing data sets are then employed in Chapter 6 to demonstrate possible applications with results that vary from encouraging to discouraging. The insight gained from attempting to control chart these data sets allows for a summarization of implementation issues and used to make conclusions about the utility of this branch of MSPC.

In Chapter 7, recommendations for associated future work and the direction thereof are then presented. A summary of the work is also provided, showing how the W_R statistic for correlation, developed from the W_i statistic for covariance, is really the only feasible alternative for control charting. While this statistic behaves similar to other statistics of its type, it also similarly has a number of issues with regard to practical implementation. However, the upshot of the conclusion is that the correlation component of the variance-covariance matrix is often of less influence than the scale factor component.

2.0 LITERATURE REVIEW & BACKGROUND

The literature review presented is partitioned into four areas. First an overview of multivariate statistical process control and specifically the most commonly used multivariate chart is discussed in Section 2.1. Next, in Section 2.2, a summary of the research associated with multivariate control charts is presented. In section 2.3, the three statistics of dispersion specific to this research are discussed in depth. Other dispersion measures are briefly discussed in Section 2.4. Section 2.5 provides the background and rationale for the use of the correlation component of the covariance matrix.

2.1 MULTIVARIATE STATISTICAL PROCESS CONTROL

Following the original univariate mean and range charts of Shewhart are a variety of multivariate control charts. Perhaps the best known is the Hotelling T^2 chart, the more common version of the χ^2 chart in which the covariance matrix and mean vector are not known and are estimated from the current sample (Montgomery, 1997). A single statistic is generated by the equation

$$T^2 = n(\overline{X} - \overline{\overline{x}})' S^{-1} (\overline{X} - \overline{\overline{x}}) \qquad \text{(2-1)},$$

where n is the sample size, $\bar{\bar{x}}$ is the estimate of the true mean vector, μ, and $\boldsymbol{S}$ is the estimate of the true covariance matrix, **Σ**. This statistic is plotted as shown in Figure 1, against control limits of the form

$$\begin{cases} UCL = \dfrac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\ LCL = 0 \end{cases} \qquad (2\text{-}2),$$

where p represents the number of quality characteristics, n is the sample size for each group, m represents the number of sample groups and the F-statistic comes from the distribution with the number of degrees of freedom as specified.

2.2 A SUMMARY OF LITERATURE ASSOCIATED WITH THE HOTELLING T^2 CONTROL CHART AND ITS DEVELOPMENTS

The concept of developing a correlation control chart, with an objective of characterizing covariance behavior, evolves naturally in the progression of research into multivariate charts that originated with the Hotelling T^2 chart. Figure 2 provides an overview of these research directions that are reviewed herein. Most research since the late 1950s involving the T^2 chart has concentrated on monitoring changes in the location or mean vector, as depicted on the left side of Figure 2. The developments for location (mean) can be subdivided into two general areas: determination and tolerance. Determination denotes research concerned with identifying univariate components responsible for generating an out-of-control point on the T^2 chart. Tolerance applies to research seeking the preferred method for assigning control limits to the T^2

chart. Similarly, as shown on the right side of Figure 2, there have also been developments investigating changes in the dispersion (variance), including two hybrid methods that also incorporate information related to the location vector. Figure 2 is explained in the Literature Review with the emphasis placed primarily on the dispersion measures.

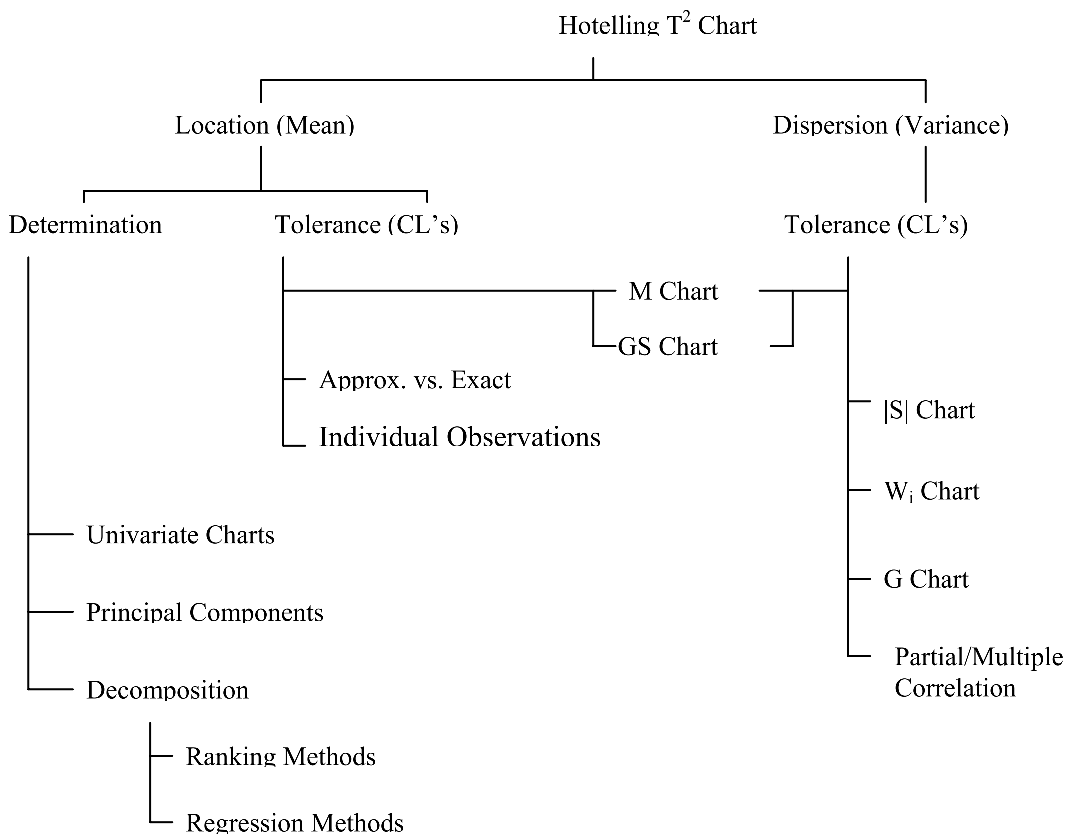


Figure 2 Research Directions Related to the Hotelling T^2 Chart

2.2.1 Research on Control Limits for Location

Research on the T^2 chart mentions several alternatives for setting control limits. Typically, Phase I limits are used to estimate process parameters from the sample when exact values are not known, and then Phase II limits are established for the process in control (Alt, 1985, Sullivan and Woodall, 1995). However, if the process mean and variance are known *a priori* it is generally accepted that Phase II limits can be applied. In general, Phase II control limits are established with an upper limit based on the chi-squared distribution with p degrees of freedom and the lower control limit at zero.

Much of the research investigates the differences between using asymptotic approximations and exact values in the setting of Phase I control limits and the asymptotic estimates are found to be acceptable for most cases (Alt, 1985). Tracy *et al.* (1996) have made a specific contribution in the area of MSPC addressing the problem of single observations plotted on the T^2 chart. That is, they address the situation where the sample size, n, equals one. Sullivan and Woodall (1996) review various approaches for single sample control charts, including the work of Tracy *et al.* related to the T^2 chart.

2.2.2 Research on the Explanation of Out-of-Control Points

A simple method for determination of out-of-control points is plotting a univariate control chart for each variable used in calculation of the T^2 statistic. The upper and lower control limits are calculated using a Bonferroni approach so that the tolerance on each univariate statistic is α/p, where p is the number of variables and α is the overall simultaneous tolerance requirement.

The earliest alternative approach to determining the out-of-control components involved decomposition by principal components analysis (PCA) and stems from the work of Jackson (1959, 1980, 1985). The chemical process industry readily uses this approach but more recent techniques that more accurately define out-of-control variables can be found in the literature. As noted by Hayter and Tsui (1994), the variables resulting from PCA are often quantities such as "(1/4) weight + (3/4) length" that do not have an analog in the process.

Two interrelated methods have recently been used in place of principal components analysis to determine those variables which cause an out-of-control signal on the T^2 chart. Doganaksoy *et al.* (1991) introduced a method that ranked the components of the observation vector based on a statistic compared to a *t*-distribution. Similarly, Hawkins (1991, 1993) has done extensive work to improve the capabilities of multivariate control charts by making regression adjustments to individual variables. Hawkins (1993) and Wade and Woodall (1993) have applied these regression adjustments to the T^2 chart to identify influential components. In 1995, Mason *et al.* demonstrated a comprehensive decomposition procedure by which the regression and ranking techniques were both shown to be subsets.

2.2.3 Research on Control Limits for Dispersion

Referring back to Figure 2, the other factor influencing T^2 charts concerns dispersion. It has been recognized that much work remains in this area, including a 1995 review of multivariate control charts by Lowry and Montgomery and an article by Golnabi and Houshmand (1996). Of particular note in the existing literature is the orientation, although it is not explicitly stated, that changes in the variance-covariance matrix (or correlation matrix at its most basic state) are considered nuisances that must be addressed in order to ensure that the T^2 chart is properly monitoring the process location. The *M* chart (Hayter and Tsui, 1994) and *GS* chart (Houshmand *et al.,* 1997) combine dispersion with location to address this issue. Research that directly attacks the dispersion question has resulted in the production of three statistics for control charting purposes: the $|\boldsymbol{S}|$, the W_i, and the G. In addition, the methods of partial and multiple correlation have been introduced. The latter methods, while providing the ability to indicate in- and out-of-control conditions, are algorithmic applications (Golnabi and Houshmand, 1996).

These methods were not designed with the intent of adaptation to control charts and do not produce a single, chartable statistic. Thus, they are only briefly summarized. The following section elaborates on the development of the three statistics ($|\boldsymbol{S}|$, W_i, and G) of interest to this research.

2.3 DISPERSION CONTROL CHARTS

As previously noted, three dispersion statistics ($|\boldsymbol{S}|$, W_i, and G) and associated control charts have been developed. Each is described below in the chronological order of their introduction—this facilitates understanding the motivation for the more recent developments.

2.3.1 The $|\boldsymbol{S}|$ Chart

The sample generalized variance, $|\boldsymbol{S}|$, is one of the most widely used measures of process dispersion (Alt, 1985). The foundation of the $|\boldsymbol{S}|$ chart is based on the assumption that the determinant of the covariance matrix is the multivariate analog to covariance (Wilks, 1932).

One attraction of the $|\boldsymbol{S}|$ chart is ease of calculation and the resultant scalar $|\boldsymbol{S}|$ that is plotted against the control limits. Alt (1985) notes that the ". . . $|S|$-control chart can be constructed using only the first two moments of $|S|$ and the property that most of the probability distribution of $|S|$ is contained in the interval $E(|S|) \pm 3\sqrt{V(|S|)}$" (p. 116).

The control limits are

$$\begin{cases} UCL = |\Sigma_0| \dfrac{\left(\chi^2_{2n-4,\alpha/2}\right)^2}{4(n-1)^2} \\ LCL = |\Sigma_0| \dfrac{\left(\chi^2_{2n-4,1-(\alpha/2)}\right)^2}{4(n-1)^2} \end{cases} \qquad (2\text{-}3),$$

where Σ_0 is the in-control covariance matrix, n is the sample size, and α is the specified tolerance. If $n < 6$, the lower control limit is replaced with zero. The control limits in equation 2-3 are more common, although others exist for specific purposes (Golnabi and Houshmand, 1996; Alt, 1985; Montgomery and Wadsworth, 1972).

Despite ease of calculation and intuitive plausibility, the |***S***| statistic has several potential drawbacks as a relatively simplistic, scalar representation of a complex multivariate structure (Lowry and Montgomery, 1997). Alt (1985) cites an example from Johnson and Wichern using bivariate data resulting in "distinctly different correlations, r = 0.8, 0.9, and -0.8" although the sample covariances have the same generalized variance (p. 116). Similarly, Lowry and Montgomery (1997) note three covariance matrices:

$$S_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (2\text{-}4a)$$

$$S_2 = \begin{pmatrix} 2.32 & 0.40 \\ 0.40 & 0.50 \end{pmatrix}, \text{ and} \qquad (2\text{-}4b)$$

$$S_3 = \begin{pmatrix} 2.65 & -0.40 \\ -0.40 & 0.50 \end{pmatrix}, \qquad (2\text{-}4c)$$

each of which "conveys considerably different information about process variability and the correlation between the two variables" but result in $|\boldsymbol{S}_1| = |\boldsymbol{S}_2| = |\boldsymbol{S}_3| = 1$ (p. 804). Consequently, Lowry and Montgomery (1997), Alt (1985) and others suggest that the |***S***| chart not be used in isolation, but rather be used in conjunction with univariate dispersion control charts created from the multivariate data.

2.3.2 The W_i Chart

Some concerns related to use of the sample generalized variance, |***S***|, can be alleviated by employing a direct multivariate extension of the univariate S^2 control chart. The evolution of this control chart started with Alt in the mid-1970s and since 1985 it has been regularly mentioned in the literature as an alternative to the |***S***| chart.

Assuming that the true covariance matrix, **Σ**, is known (or estimated from a large, in-control sample), multiple comparisons are made between the sample covariance matrices from the process and the known covariance matrix. The multiple comparisons are a series of tests of significance of the form H_0: $\sigma^2 = \sigma_0^2$ vs. H_1: $\sigma^2 \neq \sigma_0^2$ (Alt, 1985, p. 116). In this case, the repeated hypothesis tests compare the known and the sample covariance matrices.

The test statistic computed and plotted for each sample, i, is

$$W_i = -pn + pn\ln(n) - n\ln\left(\frac{|A_i|}{|\Sigma|}\right) + tr(\Sigma^{-1}A_i) \qquad (2\text{-}5)$$

where

$$A_i = (n-1)S_i \qquad (2\text{-}6)$$

(Lowry and Montgomery, 1997, p. 804). In equations 2-5 and 2-6, p denotes the number of quality characteristics, n denotes the sample size, and $\boldsymbol{S}$ is the sample covariance matrix for the i^{th} sample.

The calculated point, W_i, is compared to an upper control limit (UCL) with a χ^2 distribution with $p(p+1)/2$ degrees of freedom at a significance level of α, as shown in equations 2-7,

$$\begin{cases} UCL = \chi^2_{\left(\frac{p(p+1)}{2},\alpha\right)} \\ LCL = 0 \end{cases} \qquad (2\text{-}7).$$

2.3.3 The *G* Chart

The *G* Chart is a recent application of the *G* statistic to multivariate statistical process control for dispersion introduced by Levinson *et al.* (2002). This work stems from attempts to improve the T^2 statistic (Holmes and Mergen, 1993) in view of variation and starts with the q^2 statistic reported by Hald (1952). Hald showed that q^2 is an unbiased estimator of σ^2 given by the equation:

$$q^2 = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2 \qquad (2\text{-}8),$$

where X_i and X_{i+1} are consecutive sample values, and n is the sample size. Equation 2-8 represents the Mean Squared Successive Differences (MSSD) estimation of the variances. Hald also proposed a ratio r (not to be confused with the correlation coefficient, r) that is equal to q^2/s^2. He suggested that this ratio could be used as an "hypothesis alternative to the hypothesis of statistical control" that would indicate a gradual change in the population mean, with small values of r being significant (Hald, p. 358). The distribution of r is approximately normal for $n > 20$; a table of fractiles is presented for values from $n = 4$ to $n = 20$.

Holmes and Mergen (1993) applied the concept of MSSD as an attempt to improve the sensitivity of the T^2 control chart (p. 261). Since q^2 is an unbiased estimator of σ^2, it can be used in place of s^2 to calculate the variance-covariance matrix $\boldsymbol{S}$. Consider two processes, X and Y, and the covariance calculated by the standard method and by the MSSD method. "If the X and Y processes are random and have no cross-correlation other than at zero lag, these two estimates of the universe covariance will be similar. If these conditions are not met, then there will be a difference in the estimators" (p. 622).

Holmes and Mergen (1998) extend the MSSD application as "a test for the equality of the regular covariance matrix and the MSSD covariance matrix to establish whether or not the multivariate process is stable" (p. 505). It employs a G statistic to compare the equality of two population covariance matrices calculated as

$$G = 2.3026(m)(M)\,, \tag{2-9}$$

where m is a constant defined as

$$m = 1 - \left[\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) - \left(\frac{1}{n_1 + n_2 - 2}\right)\right]\left(\frac{2p^2 + 3p - 1}{6(p+1)}\right) \tag{2-10}$$

and

$$M = (n_1 + n_2 - 2)\log|S| - (n_1 - 1)\log|S_1| - (n_2 - 1)\log|S_2| \tag{2-11}$$

where p represents the number of quality characteristics, n is the sample size, and $\boldsymbol{S}$ is the covariance matrix. Subscripts refer to the individual data for each of two covariance matrices, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ from samples of size n_1 and n_2, respectively, so that $\boldsymbol{S}$ represents a pooled covariance. In this equation the logarithm used is base 10.

The G statistic was proposed by Kramer and Jensen (1969b) and is based on the assumption that the determinant of the covariance matrix is the multivariate analog of the variance (Holmes and Mergen, 1998). It should be noted that this assumption comes from Wilks (1932), and is subject to the same arguments presented in Section 2.3.1.

Levinson *et al.* (2002) reason that a stable process would produce separate estimates of the covariance matrix that are approximately equal, and that the equivalence of these matrices could be tested using the G statistic. Two methods are presented for calculation of the covariance matrix, $\boldsymbol{S}$. The first is the standard calculation, referred to by Levinson *et al.* as the "full data set" method. The second method uses the MSSD (Holmes and Mergen, 1998) and is given as

$$S_{MSSD} = \frac{1}{2(n-1)} \sum_{j=2}^{n} (X_j - X_{j-1})(X_j - X_{j-1})^T \qquad \text{(2-12)},$$

where X_{j-1} and X_j are consecutive samples in a sample of size n, computed for k subgroups (number of samples), and T is the transpose operator. The value of the constant m remains the same as in equation 2-10 above, but the value of M changes to

$$M = (\nu_1 + \nu_2)\ln|S| - \nu_1 \ln|S_1| - \nu_2 \ln|S_2| \qquad \text{(2-13)}$$

where $|\boldsymbol{S}_2|$ is the sample generalized variance for the sample being compared to the in-control sample generalized variance, $|\boldsymbol{S}_1|$, and $\boldsymbol{S}$ is the pooled variance given by

$$S = \frac{(\nu_1 S_1 + \nu_2 S_2)}{\nu_1 + \nu_2} \qquad \text{(2-14)}$$

where

$$\nu_1 = k(n_2 - 1) \qquad \text{(2-15a)}$$

and

$$\nu_2 = (n_2 - 1) \qquad \text{(2-15b)}$$

where ν_1 and ν_2 are the degrees of freedom for the initial (control) sample, and subsequent samples, respectively, with n_2 representing the number of samples in the kth sample subset. The G statistic is then calculated as the product

$$G = (M)(m) \qquad \text{(2-16),}$$

where M and m are as defined above. Control limits are set as

$$\begin{cases} UCL = \chi^2_{\left(\frac{p(p+1)}{2},\frac{\alpha}{2}\right)} \\ LCL = \chi^2_{\left(\frac{p(p+1)}{2},\frac{1-\alpha}{2}\right)} \end{cases} \qquad \text{(2-17),}$$

where $\chi^2_{(p,q)}$ is the qth quantile (Levison *et al*., 2002, p. 541).

2.4 OTHER MEASURES OF DISPERSION

There are three other approaches found in the literature that attempt to capture the effect of dispersion on MSPC control charts. These are depicted on Figure 2 as the M (Hayter and Tsui, 1994) chart, the GS (Houshmand *et al*., 1997) chart, and the methods of partial and multiple correlation (Golnabi and Houshmand, 1997). While each of these approaches does address the dispersion issue, their directions diverge from that of this research since the objective here is to separate the dispersion from the location to characterize its behavior. Both the M chart and the GS chart combine dispersion and location measures into a single, chartable statistic, and are an attempt to address Alt's (1985) recommendation for the development of one control chart for the simultaneous monitoring of location and dispersion. The *M* chart and *GS* chart are reviewed in Appendix A and Appendix B, respectively.

The similar methods of partial correlation and multiple correlation do directly address changes in the correlation structure itself. However, the process makes iterative comparisons using a method that does not follow the template of the standard Shewhart chart which is the basis for MSPC charting. Thus, the practical difficulties and other implications of charting the generated parameters are not addressed. This algorithmic method is reviewed in Appendix C.

2.5 THE CORRELATION MATRIX AS A MEASURE OF DISPERSION

While three statistics ($|\boldsymbol{S}|$, W_i, and G) have been proposed for tracking changes in the dispersion, there has been no evaluation to compare the relative performances of these statistics. At best, these statistics have been "validated" using data from Jackson (1980) and others. The following section outlines the relationship between $\boldsymbol{R}$ and $\boldsymbol{S}$ as a mathematical justification to comparing the existing statistics ($|\boldsymbol{S}|$, W_i, and G) for monitoring changes in the correlation matrix.

Sample covariance provides a measure of linear association between two variables and is calculated as:

$$s_{ik} = \sum_{j=1}^{n} (X_{ij} - \overline{X}_i)(X_{kj} - \overline{X}_k) \qquad \text{(2-18)},$$

where X_i and X_k are values from two equal-sized (n) sets of variables with means $\overline{X}_i$ and $\overline{X}_k$, respectively, and j is an index. For the population, s_{ik} would be replaced by σ_{ik} and the average values $\overline{X}_i$ and $\overline{X}_k$ would be replaced by μ_i and μ_k, respectively. The sample correlation coefficient, r_{ik}, is the standardized sample covariance, where the product of the square roots of the sample variances provides the standardization. Thus, the sample correlation coefficient, r_{ik}, can also be viewed as a sample covariance (Johnson & Wichern, p.10).

The sample correlation matrix is comprised of p x p quality characteristics, r_{ik}, each of the form

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} \tag{2-19}$$

where $i = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, p$. In this equation, each entry r_{ik} is the Pearson's product moment correlation coefficient for the i^{th} and k^{th} entry in the p x p matrix. If the original values X_{ij} and X_{kj} are replaced by standardized values $\frac{(X_{ij} - \overline{X}_j)}{\sqrt{s_{ii}}}$ and $\frac{(X_{kj} - \overline{X}_k)}{\sqrt{s_{kk}}}$, respectively, then the standardized values are commensurable since both sets are centered at zero and expressed in standard deviation units. The sample covariance of the standardized observations, then, is the sample correlation coefficient, r_{ik} (Johnson & Wichern, p. 10).

Each sample correlation coefficient, r_{ik}, will possess several properties of interest. First, the value of r_{ik} must be between -1 and 1. Note that r_{ik} will equal r_{ki} for all values of i and k, and that any number correlated with itself will have a value of unity. Thus, the diagonals of the matrix $\boldsymbol{R}$ will be equal to one, and the matrix itself will be symmetric. This property means that, unlike the variance-covariance matrix $\boldsymbol{S}$, the correlation matrix $\boldsymbol{R}$ is bounded, allowing a comparison between calculated dispersion statistics.

Second, similar to covariance, the correlation coefficient measures the strength of linear association where $r_{ik} = 0$ indicates no association, and the sign of r_{ik} indicates direction. For correlation, $r_{ik} < 0$ indicates a tendency for one value in the pair to be larger than the mean when the other is smaller than its mean and $r_{ik} > 0$ indicates that values in each pair will tend to be either simultaneously large together or small together (Johnson & Wichern, p. 10).

Third, if the values of the i^{th} variable are changed to $y_{ij} = ax_{ij} + b$ and the those of the k^{th} variable are changed to $y_{kj} = cx_{kj} + d$, for $i = 1, 2, \ldots, n$ in both cases, then, provided that a and c have the same sign, there will be no change in r_{ik}.

Fourth, using either n or $n - 1$ (common to avoid bias) as the denominator for s_{ik} results in the same value for r_{ik}.

Despite the desirable qualities of the correlation matrix, $\boldsymbol{R}$, there are some items that remain of concern when using $\boldsymbol{R}$ in place of $\boldsymbol{S}$. Largely these relate to situations where $|\boldsymbol{S}| = 0$ (similarly, then, when $|\boldsymbol{R}| = 0$). When the determinant of a matrix is zero, the matrix is singular and cannot be inverted so that a dispersion statistic cannot be calculated. Note that this does not permit a calculation of the T^2 statistic for the mean vector either. This degenerate case can result from any row of the deviation matrix being expressible as a linear combination of the remaining rows, or in all cases where the number of quality characteristics, p, exceeds the sample size, n. The latter case can be avoided by proper design. In cases where the former situation occurs, Johnson and Wichern suggest:

> that the measurements on some variables be removed from the study as far as the mathematical computations are concerned. The corresponding reduced data matrix will then lead to a covariance matrix of full rank and a nonzero generalized variance. The question of which measurements to remove in degenerate cases is not easy to answer. When there is a choice, one should retain measurements on a (presumed) causal variable instead of those on a secondary characteristic (p. 105).

This was the case with the data analyzed by Holmes and Mergen (1993) using the G chart. The study looked at changes in the distribution of particle size in an industrial process, where size was categorized as small, medium, or large. The authors state, "The data on three different particle sizes are given . . . only the first two columns are used in the analysis since the total of the percentages is always 100 and the variance-covariance matrix will not invert under these conditions" (p. 622).

Large sample behavior tends to dominate the application of multivariate statistical process control, and the multivariate normal distribution is the distribution usually associated with MSPC processes. Johnson and Wichern note that there are two main reasons for this: (1) for certain natural phenomenon multivariate normal is the population model; and (2) for many statistics, the approximate sampling distribution is multivariate normal (p. 120). Therefore, provided the sample size, n, is large enough, the parent population does not need to be multivariate normal, but must have a mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$ (p. 144). If $\boldsymbol{\Sigma}$ is finite, $\boldsymbol{S}$ and $\boldsymbol{R}$ will be finite as well.

3.0 CORRELATION CONTROL CHART DEVELOPMENT

Since the three statistics, $|\boldsymbol{S}|$, W_i, and G, were developed at different times with different approaches they have not been appropriately compared to one another, nor have they been compared using the in-control *ARL*, the common metric of control charts in which shorter *ARL*s are generally desired (DeVore, 2002). To elaborate, Alt (1985) used a mathematical derivation to develop the W_i statistic and then used a small data set to compare it to the $|\boldsymbol{S}|$ statistic, the latter being a matrix determinant that was assumed to capture dispersion effects (Wilks, 1932). While Alt (1985) had mathematical justification for preferring the W_i statistic to the $|\boldsymbol{S}|$ statistic, closed-form solutions are generally not practical for the comparison of multiple statistics; rather simulations are used, and *ARL*s are determined as the measure of comparison (Montgomery, 1997; Holmes and Mergen, 1998; Levinson *et al.*, 2003). Accordingly, Levinson *et al.* (2003) derived the G statistic and used a simulation producing an *ARL* to show its effectiveness; however, they did not compare the G statistic to the $|\boldsymbol{S}|$ nor the W_i statistic. In the next chapter (Chapter 4), this research uses a terminating sequential simulation to compare the special case of correlation for two of these three statistics, using the in-control *ARL* as the measure of comparison. In this chapter, a mathematical evaluation develops modified versions of the $|\boldsymbol{S}|$ and W_i statistics and also shows that the G statistic cannot be modified for and applied to the correlation case.

3.1 RATIONALE FOR MODIFICATION OF THE COVARIANCE STATISTICS FOR THE APPLICATION TO CORRELATION MATRICES

The three statistics under consideration (|***S***|, W_i, and G) all utilize the generalized sample variance, |***S***|, in some fashion to capture the effects of covariance, a practice that can be traced to Wilks (1932). The concept presented here is that, as a special case of covariance (***S***), correlation (***R***), could be substituted into the equations for each of the statistics and control charts created therefrom. Because of the way the control limits are calculated, which will appear in subsequent sections, if a certain assumption is made these limits do not change when the generalized sample variance of the standardized variables, |***R***|, is substituted for the generalized sample variance of the non-standardized variables, |***S***|, and this is either beneficial or detrimental depending on the statistic considered.

An assumption is required since the generalized sample variance of the standardized variables, |***R***|, only captures a portion of the information contained by the generalized sample variance of the non-standardized variables, |***S***|. In addition to the correlation component, covariance also contains a scale factor component that provides for the standardization and is calculated as a product of the variances, s_{ii}, so that

$$\left|S\right| = (s_{11}s_{22}\cdots s_{pp})\left|R\right|, \tag{3-1}$$

for a matrix of i = 1 to p quality characteristics.

In the context of a control chart, then, the creation of charts based on |***R***| instead of |***S***| requires the assumption that the scale factor is the same in the samples as it is in the in-control condition and that out-of-control conditions are generated solely by changes in the correlation matrix, ***R***. While this assumption may not always be valid, it is reasonable in MSPC if it is the correlation component that is of interest and if it is the correlation component that is expected to provide the out-of-control condition. That is, the scale factor component is not being ignored; rather the concentration is on the correlation component's contribution to covariance. Such matrices occur in a number of processes ranging from chemical fractionating to anthropometric data to financial data and meet the requirements of an Exchange Structure which will be discussed in Chapter 5. Furthermore, the assumption of scale factor stability can be tested once an out-of-control condition is detected by the correlation control chart by also viewing the control chart for the covariance that utilizes $|\boldsymbol{S}|$, since $|\boldsymbol{S}|$ does not capture only the effects of correlation (Johnson & Wichern, 103). This approach is analogous to performing a linear regression and then examining the residuals to verify normality because that assumption is sometimes inconvenient to assess *a priori*. When the generalized sample variance is decomposed as shown in equation 3-1, the effects of the scale factor and the correlation are separated, as will be shown geometrically later, so that the correlation control charts become companions to the covariance control charts.

Because different correlation structures are not detected by the generalized sample variance of the non-standardized samples, and different correlation structures are detected by the generalized sample variance of the standardized samples, this suggests the following in regard to these control charts:

1. If the correlation chart is in-control and the covariance chart is in-control, then the process is in control.
2. If the correlation chart is out-of-control and the covariance chart is in-control, then the process *may be* out-of-control due to correlation.

3. If the correlation chart is in-control and the covariance chart is out-of-control, then the process *may be* out-of-control due to the scale factor.
4. If the correlation chart is out-of-control and the covariance chart is out-of-control, then the process is out-of-control due to both the correlation and the scale factor.

Thus, even though the correlation control chart could be seen mathematically as a special case of the covariance chart, the covariance chart would in practical applications be considered the companion chart to the correlation chart even if the primary objective is to detect changes in the correlation matrices generated from the process data.

3.2 RELATION OF $|R|$ TO $|S|$

Since the three covariance statistics being considered use $|\boldsymbol{S}|$ as part of their calculations, it is instructive to see the relationship between $|\boldsymbol{S}|$ and $|\boldsymbol{R}|$. Although equation 3-1 explains this relation mathematically, geometry of the bivariate or trivariate cases provides a visual explanation. It is extendable by induction for $p > 3$, where p represents the number of quality characteristics since, even if the p vectors are n-dimensional, p vectors can span no more than a p-dimensional space. If $p = 2$, a plane is defined. If $p = 3$, a volume is defined. Similarly, when extended for $p > 3$ the spaces become hypervolumes (specifically parallelotopes) that are difficult to visualize but nonetheless valid (Anderson, 1984).

Johnson & Wichern (1988), and similarly Anderson (1984), develop the notion of a deviation vector in (p =) 3-space that is useful in illustrating the relation of $|\boldsymbol{S}|$ to $|\boldsymbol{R}|$.

Define a vector

$$y_i'\left(\frac{1}{\sqrt{n}}A\right)\left(\frac{1}{\sqrt{n}}A\right)=\frac{x_{iA}+x_{iB}+\cdots+x_{in}}{n}A=\bar{x}_i A \tag{3-2}$$

where $\bar{x}_i$ is the sample mean and n is the dimension of an n x 1 vector $A'=[1,1,\ldots,1]$, so that $\frac{1}{n}y_i'A$ corresponds to the multiple of A that gives the projection of y_i onto the line determined by A. For each y_i, the geometric decomposition would be as shown in Figure 3.

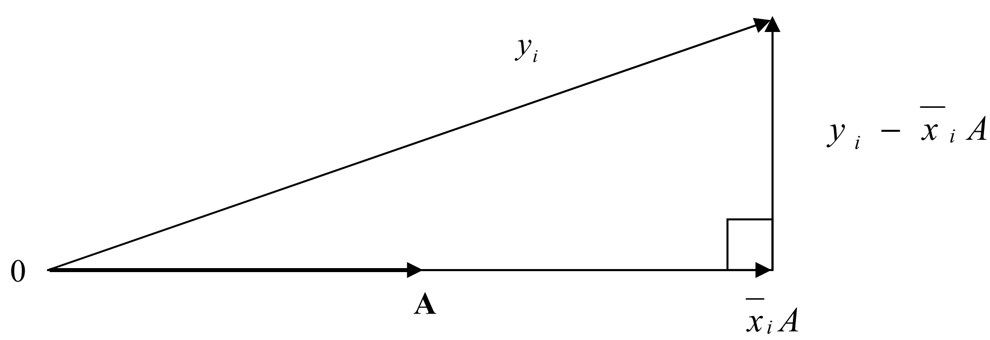


Figure 3 Decomposition of deviation vector

The deviations of the measurements on the i^{th} variable from their respective sample means, then, are given by the difference vector e_i

$$e_i = y_i - \bar{x}_i A = \begin{bmatrix} x_{iA}-\bar{x}_i \\ x_{iB}-\bar{x}_i \\ \vdots \\ x_{in}-\bar{x}_i \end{bmatrix}, \tag{3-3}$$

where the other variables are as defined previously.

The vector A has been defined purposefully to lie anywhere in n-space, as the axes of the deviation vectors, e_i, can now be translated to the origin without affecting their lengths nor their orientations. A typical view showing three deviation vectors is shown in Figure 4.

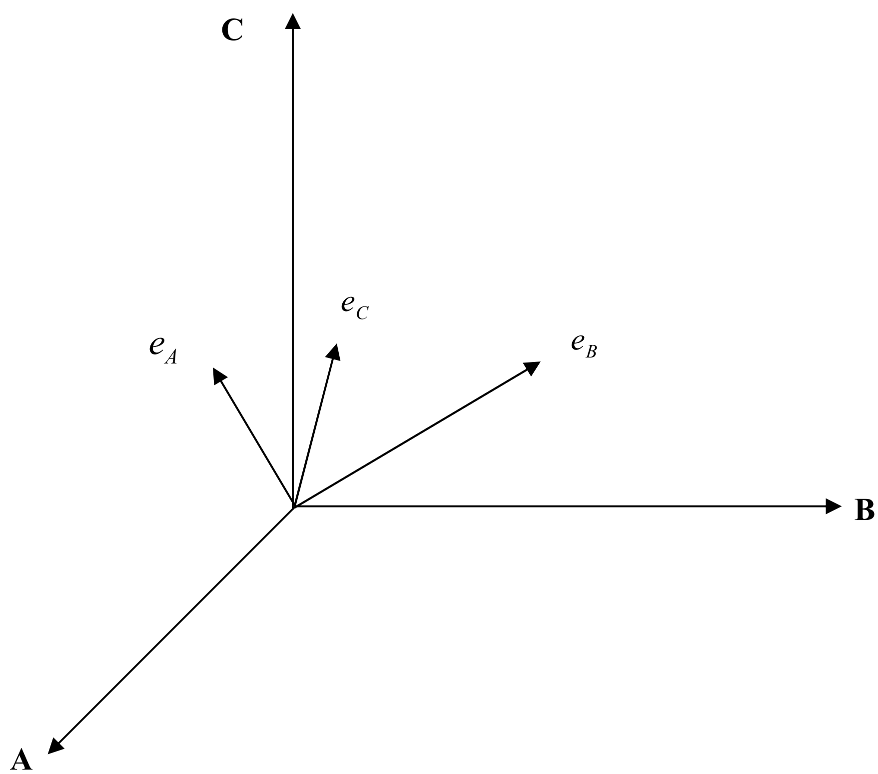


Figure 4 Deviation vectors on A-B-C axes

The squared lengths of the deviation vectors, which are proportional to the variance of the measurements on the i^{th} variable, are equivalent to the sum of the squared deviations, meaning that the lengths are proportional to the standard deviations. The cosine of the angle between pairs of deviation vectors is the correlation coefficient given by the equation

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \cos(\theta_{ik}) \qquad (3\text{-}4)$$

where r_{ik} denotes the correlation coefficient, s_{ik} is a variance, and the square roots of s_{ii} and s_{kk} are standard deviations. Figure 5 shows the geometric interpretation, in ($p = 2$) dimensions, between any two deviation vectors $e_1 = y_1 - \bar{x}_1 A$ and $e_2 = y_2 - \bar{x}_2 A$. Since the height of the trapezoidal region formed by the deviation vectors is $L_{e_1} \sin(\theta)$, the area bounded by the vectors and their projections is $L_{e_1} L_{e_2} \sqrt{1-\cos^2(\theta)}$. Johnson & Wichern show that, by using substitutions and extending to a multidimensional case by induction, that the volume generated in *n* space for *p*

deviation vectors is given by

$$|\boldsymbol{S}| = (n-1)^{-p} \, (\text{volume})^2. \tag{3-5}$$

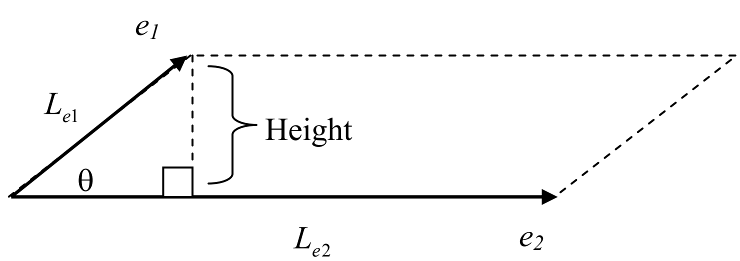


Figure 5 Trapezoidal region formed by deviation vectors

Looking at this in three dimensions, shown in Figure 6, it is evident that two components affect the volume—the length of the sides, and the angles between edges.

The correlation corresponds to the angle between deviation vectors and the lengths of the edges correspond to the scale factor (usually expressed in standard deviation units). Thus, in the covariance case, either the scale factor or the correlation may affect the volume and, therefore, the sample generalized variance, $|\boldsymbol{S}|$.

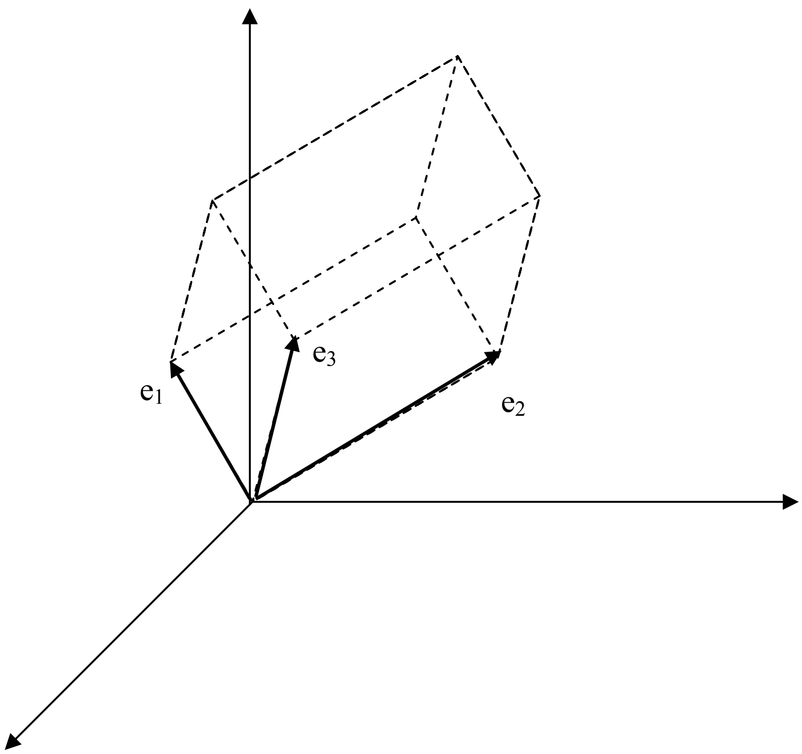


Figure 6 Parallelotope in Three Dimensions

When the covariance matrix is standardized, so that the standardized covariance, or correlation matrix, $\boldsymbol{R}$, is obtained, all of the vector lengths are equivalent and only the angle between vectors effects the volume. The geometric representation is the same as that of Figure 6, with all the e_i being of equal lengths. Assume, then, that the generalized sample variance of the standardized sample is used instead of the non-standardized version in calculation of the three covariance statistics ($|\boldsymbol{S}|$, W_i, and G), and that these statistics behave similarly as hypothesized. Then replacing $|\boldsymbol{S}|$ with $|\boldsymbol{R}|$ in the covariance statistics would lead to control charts that detect when correlation has gone out-of-control assuming the scale factor has remained constant between the in-control condition and subsequent samples. As stated, this assumption of constant scale factor should be investigated by comparing the correlation control chart with the covariance control chart when an out-of-control condition for correlation is detected.

3.3 COVARIANCE STATISTICS APPLIED TO CORRELATION[1]

Since correlation is a special case of covariance, the statistics used for MSPC monitoring of covariance should apply to the monitoring of correlation. To investigate the plausibility of correlation as a separate component, it is necessary that the scale factor does not change between the in-control condition and the samples—an assumption that must be checked by also using the covariance control chart once the correlation control chart detects an out-of-control condition. In doing so, the control limits for the correlation case are generally commensurate with those for the original covariance statistics and this is shown in subsequent sections. As Table 1 indicates, there is a slight modification of the control limits when considering $|\boldsymbol{R}|$ instead of $|\boldsymbol{S}|$, but no modification when considering W_R instead of W_i. These limits will be derived in the following sections. Issues with the G statistic will be addressed in Section 3.3.3.

[1] Thanks to Dr. Murat C. Testik, Ph.D. for reviewing the derivation of control limits and other calculations presented in Chapter 3. Dr. Testik received his doctorate from Arizona State University where he studied MSPC with Dr. Douglas Montgomery.

Table 1 The Statistics and Their Control Limits

Covariance Statistic	Correlation Statistic	Control Limits
$\lvert \boldsymbol{S} \rvert$	Not applicable	$\left[\lvert\Sigma\rvert \frac{\left(\chi^2_{2n-4,1-(\alpha/2)}\right)^2}{[4(n-1)^2]}, \lvert\Sigma\rvert \frac{\left(\chi^2_{2n-4,\alpha/2}\right)^2}{[4(n-1)^2]} \right]$
Not applicable	$\lvert \boldsymbol{R} \rvert$	$\left[\lvert\rho\rvert \frac{\left(\chi^2_{2n-4,1-(\alpha/2)}\right)^2}{[4(n-1)^2]}, \lvert\rho\rvert \frac{\left(\chi^2_{2n-4,\alpha/2}\right)^2}{[4(n-1)^2]} \right]$
W_i	W_R	$[0, \chi^2_{(p(p+1)/2, \alpha)}]$

3.3.1 Control Limits for |*R*|

When $p = 2$, $|S|$ is distributed exactly as chi-squared. However, for $p > 2$, Alt (1973) suggests using Anderson's approximation as

$$P\left[|\Sigma_0| \left(1 - z_{\alpha/2} \sqrt{\frac{2p}{n-1}} \right) \le |S| \le |\Sigma_0| \left(1 + z_{\alpha/2} \sqrt{\frac{2p}{n-1}} \right) \right] = 1 - \alpha \tag{3-5}$$

where Σ_0 is the in-control covariance matrix, S is the sample covariance matrix, p denotes the number of quality characteristics, n is the sample size, and $z_{\alpha/2}$ is from the standard normal distribution. Then the control limits for the $|S|$ chart are

$$UCL = |\Sigma_0| \left(1 + z_{\alpha/2} \sqrt{\frac{2p}{n-1}} \right) \tag{3-6}$$

and

$$LCL = |\Sigma_0|\left(1 - z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right) \tag{3-7}$$

where the variables are as noted above. Since $\Sigma_0 = V^{1/2}\rho V^{1/2}$, where V represents the diagonal matrix of variances so that $V^{½}$ is the diagonal matrix of standard deviations and ρ represents the correlation matrix for the in-control condition, substituting into the first equation gives

$$P\left[\left|V^{1/2}\rho V^{1/2}\right|\left(1 - z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right) \le \left|\hat{V}^{1/2}R\hat{V}^{1/2}\right| \le \left|V^{1/2}\rho V^{1/2}\right|\left(1 + z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right)\right] = 1 - \alpha \tag{3-8}$$

where R is the sample correlation matrix. Now, if A and B are k x k square matrices—which ***S*** and ***R*** have to be—and it has been shown that

$$|AB| = |A||B| \tag{3-9}$$

the equation now becomes

$$P\left[\left|V^{1/2}\right||\rho|\left|V^{1/2}\right|\left(1 - z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right) \le \left|\hat{V}^{1/2}\right||R|\left|\hat{V}^{1/2}\right| \le \left|V^{1/2}\right||\rho|\left|V^{1/2}\right|\left(1 + z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right)\right] = 1 - \alpha \tag{3-10}$$

Since the determinants are scalars and it is assumed that, under the in-control conditions, the standard deviation matrices for the sample and population are equivalent, and sample sizes are sufficiently large, dividing by the common factor leaves

$$P\left[|\rho|\left(1-z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right)\leq|R|\leq|\rho|\left(1+z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right)\right]=1-\alpha \tag{3-11}$$

meaning that the control limits for the $|R|$ chart would be

$$UCL=|\rho|\left(1+z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right) \tag{3-12}$$

and

$$LCL=|\rho|\left(1-z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right), \tag{3-13}$$

so that a control chart could be constructed for correlation based on the $|\boldsymbol{S}|$ covariance control chart.

However, a significant issue with the limits shown in Equations 3-12 and 3-13 is very evident. The determinant of a correlation matrix is bounded as [0,1] so that the upper and lower limits of Equations 3-12 and 3-13, respectively, will rarely be exceeded, especially when the magnitude of the correlation exceeds 0.4. This result is an indicator that the performance of the $|\boldsymbol{R}|$ statistic is questionable (or that correlation, thus measured, cannot be viewed as a major contributor to the covariance).

3.3.2 Control Limits for W_R

Use of the W_i statistic (Alt, 1973) is equivalent to making repeated hypothesis tests of the form

$$H_0 : \Sigma = \Sigma_0 \text{ versus } H_1 : \Sigma \neq \Sigma_0$$

where Σ and Σ_0 are the population covariance and matrix of in-control covariance values, respectively.

The statistic is a ratio of the maximum likelihood estimators for the multivariate normal distribution (μ, Σ) in the following sets:

$$\Omega = \{(\mu, \Sigma) : -\infty < \mu < \infty, \Sigma\_is\_positive\_definite\}$$

$$\omega = \{(\mu, \Sigma) : -\infty < \mu < \infty, \Sigma = \Sigma_0\}.$$

This implies that the mean is unspecified and the covariance matrix is specified, to be used in the likelihood ratio statistic to derive the control statistic. Note that this is important since the statistic now becomes invariant to changes in the mean vector.

It has been shown (Anderson, 1984) that the maximum likelihood estimators for the multivariate normal case are:

$$\hat{\mu}_\Omega = \bar{x} \qquad \hat{\Sigma}_\Omega = \frac{1}{n}A \qquad \text{(3-14a, b)}$$

$$\hat{\mu}_\omega = \bar{x} \qquad \hat{\Sigma}_\omega = \Sigma_0 \qquad \text{(3-15a, b)}$$

where $\bar{x}$ is the vector mean, n is the sample size, and $A = (n-1)S$ so that the likelihood functions based on the multivariate normal distribution are

$$L(\hat{\Omega}) = \left(\frac{2\pi}{n}\right)^{-pn/2} |A|^{-n/2} e^{-pn/2}, \qquad \text{(3-16)}$$

and

$$L(\omega) = (2\pi)^{-pn/2} |\Sigma_0|^{-n/2} e^{-\frac{1}{2}tr(\Sigma_0^{-1}A)}, \qquad \text{(3-17)}$$

(where *tr* is the trace operator) giving a maximum likelihood estimate

$$\Lambda(x) = \frac{1}{n} e^{pn/2} |\Sigma_0^{-1}A|^{n/2} e^{-\frac{1}{2}tr(\Sigma_0^{-1}A)}. \qquad \text{(3-18)}$$

Note here that the maximum likelihood of the covariance matrix is not an unbiased estimator. Taking the log likelihood, the statistic becomes

$$W_i = -2\ln(\Lambda) = -pn + pn\ln n - n\ln\left(\frac{|A|}{|\Sigma_0|}\right) + tr(\Sigma_0^{-1}A) \qquad \text{(3-19)}$$

which is rejected any time its value exceeds $\chi^2\left(\frac{p(p+1)}{2}, \alpha\right)$, an expected result for the log likelihood. This is the asymptotic general result, since the specification of Σ_0 requires $\frac{p(p+1)}{2}$ independent elements to be specified.

To approach this statistic from the standpoint of correlation in which the goal is repeated tests of the hypotheses

$$H_0 : \rho = \rho_0 \text{ versus } H_1 : \rho \neq \rho_0 ,$$

one could calculate the maximum likelihood ratio for a new Ω and ω defined in terms of the correlation. However, since the multivariate normal distribution requires introduction of the variance for full characterization, the result will be a test of the covariance.

If, instead, we look at the third and fourth terms of the W_i statistic, we see that they include the ratio $\frac{|A|}{|\Sigma_0|}$, which is equivalent to $\frac{|(n-1)S|}{|\Sigma_0|}$, or $(n-1)^p \frac{|S|}{|\Sigma_0|}$, where S is replacing Σ in equation 3-19, and depicts a scalar $(n\text{-}1)^p$ multiplied by the ratio of the determinants of the sample covariance matrix to the in-control covariance matrix. To convert this ratio for the in-control correlation matrix to the ratio of the determinants of the sample correlation matrix, $\frac{|R|}{|\rho_0|}$, requires making the assumption that the scale factors for both cases are equivalent, an assumption that has been noted must be checked once an out-of-control condition is indicated. The statistic, by substitution, then becomes

$$W_R = -2\ln(\Lambda) = -pn + pn\ln n - n\ln\left((n-1)^p \frac{|R|}{|\rho_0|} \right) + (n-1)tr(\rho_0^{-1}R) , \quad (3\text{-}20)$$

so that a control chart can be constructed for correlation, W_R, based on the W_i covariance control chart.

In addition to being intuitive, there is a justification for this substitution since the maximum likelihood of the correlation occurs at the same point as the maximum likelihood of the covariance.

This is shown by Anderson (1984, pp. 64-65) for the elements of the estimated correlation matrix, the correlation coefficients r_{ij}, and uses the following corollary:

> If on the basis of a given sample $\hat{\theta}_1,\ldots,\hat{\theta}_m$ are maximum likelihood estimates of the parameters $\theta_1,\ldots,\theta_m$ of a distribution, then $\phi_1(\hat{\theta}_1,\ldots,\hat{\theta}_m),\ldots,\phi_m(\hat{\theta}_1,\ldots,\hat{\theta}_m)$ are maximum likelihood estimators of $\phi_1(\theta_1,\ldots,\theta_m),\ldots,\phi_m(\theta_1,\ldots,\theta_m)$ if the transformation from $\theta_1,\ldots,\theta_m$ to $\phi_1,\ldots,\phi_m$ is one-to-one. If the estimators of $\theta_1,\ldots,\theta_m$ are unique, then the estimators of $\phi_1,\ldots,\phi_m$ are unique.

The maximum of the ratio of the correlation matrices, $\frac{|R|}{|\rho_0|}$, occurs at the same point as the maximum of the ratios of the covariance matrices, $\frac{|S|}{|\Sigma_0|}$, and W_R reaches its maximum at the same point as W_i and is thus commensurate. In other words, the control limits remain the same since they are based on the asymptotic approach of the maximum likelihood estimate.

Two numeric cases will be used to illustrate application of the modified W_i (named W_R) statistic. Assume that the following bivariate matrix, Σ_0, for a sample of size $n = 5$, and $\alpha = 0.05$ represents the in-control condition. The in-control correlation matrix, ρ_0, and the determinants of Σ_0 and ρ_0 are also shown below:

$$\Sigma_0 = \begin{bmatrix} 6.96 & 1.20 \\ 1.20 & 1.50 \end{bmatrix} \qquad |\Sigma_0| = 9$$

$$\rho_0 = \begin{bmatrix} 1 & 0.371 \\ 0.371 & 1 \end{bmatrix} \qquad |\rho_0| = 0.862 .$$

3.3.2.1 Case I: Covariance In-Control / Correlation Out-of-Control

Consider the following matrices

$$S_1 = \begin{bmatrix} 4.24 & 3 \\ 3 & 4.24 \end{bmatrix} \qquad |S_1| = 9$$

$$R_1 = \begin{bmatrix} 1 & 0.967 \\ 0.967 & 1 \end{bmatrix} \qquad |R_1| = 0.065$$

and then calculate the W_i statistic for covariance as

$$W_i = -pn + pn \ln n - n \ln\left((n-1)^p \frac{|S_1|}{|\Sigma_0|} \right) + (n-1) tr(\Sigma_0^{-1} S_1)$$

$$W_i = -2(5) + 2(5)\ln(5) - 5\ln\left((5-1)^2 \left(\frac{9}{9} \right) \right) + (5-1) tr \begin{bmatrix} 0.306 & 0.065 \\ 1.754 & 2.878 \end{bmatrix}$$

$$W_i = -10 + 16.09 - 13.86 + 12.736 = 4.966$$

which is compared to a chi-squared distribution

$$W_i = 4.966 < 7.81 = \chi^2_{3,0.05}$$

and found to be in-control.

Now, calculate the same statistic, modified for correlation

$$W_R = -pn + pn\ln n - n\ln\left((n-1)^p \frac{|R_1|}{|\rho_0|}\right) + (n-1)tr(\rho_0^{-1}R_1)$$

$$W_R = -2(5) + 2(5)\ln(5) - 5\ln\left((5-1)^2\left(\frac{0.065}{0.862}\right)\right) + (5-1)tr\begin{bmatrix} 0.743 & 0.691 \\ 0.691 & 0.743 \end{bmatrix}$$

$$W_R = -10 + 16.09 - 0.931 + 5.994 = 11.103$$

which is compared to the chi-squared distribution

$$W_R = 11.103 > 7.81 = \chi^2_{3,0.05}$$

and found to be out-of-control.

3.3.2.2 Case 2: Covariance Out-of-Control / Correlation In-Control

With the aforementioned in-control matrices established above, consider the following matrices

$$S_2 = \begin{bmatrix} 100 & 38.1 \\ 38.1 & 100 \end{bmatrix} \qquad |S_2| = 8548.39$$

$$R_2 = \begin{bmatrix} 1 & 0.381 \\ 0.381 & 1 \end{bmatrix} \qquad |R_2| = 0.854$$

and then calculate the W_i statistic for covariance as

$$W_i = -pn + pn\ln n - n\ln\left((n-1)^p \frac{|S_2|}{|\Sigma_0|} \right) + (n-1)tr(\Sigma_0^{-1} S_2)$$

$$W_i = -2(5) + 2(5)\ln(5) - 5\ln\left((5-1)^2 \left(\frac{8548.39}{9} \right) \right) + (5-1)tr\begin{bmatrix} 11.58 & -6.98 \\ 16.13 & 72.25 \end{bmatrix}$$

$$W_i = -10 + 16.09 - 48.14 + 335.32 = 293.27$$

which is compared to the chi-squared distribution

$$W_i = 293.27 > 7.81 = \chi^2_{3,0.05}$$

and found to be out-of-control.

Now, calculate the same statistic, modified for correlation

$$W_R = -pn + pn\ln n - n\ln\left((n-1)^p \frac{|R_2|}{|\rho_0|} \right) + (n-1)tr(\rho_0^{-1} R_2)$$

$$W_R = -2(5) + 2(5)\ln(5) - 5\ln\left((5-1)^2 \left(\frac{0.854}{0.862} \right) \right) + (5-1)tr\begin{bmatrix} 0.995 & 0.011 \\ 0.011 & 0.995 \end{bmatrix}$$

$$W_R = -10 + 16.09 - 13.81 + 7.96 = 0.23$$

which is compared to a chi-squared distribution

$$W_R = 0.23 < 7.81 = \chi^2_{3,0.05}$$

and found to be in-control. Similarly, examples can be developed whereby the covariance and correlation matrices are both either in-control or out-of-control at the same time. From a simulation study, the results for which appear in Chapter 4, it can be shown that the calculation of the aforementioned test statistics can be very dependent on the values of certain parameters. Thus, the cases shown above are simplistic examples designed only to show that correlation and covariance can be viewed separately.

3.3.3 The Problem with the *G* Statistic

Recall that the *G* statistic compares the equality of two covariance matrices for the purpose of detecting subtle shifts and is computed as

$$G = 2.3026(m)(M), \tag{3-21}$$

where *m* is a constant defined as

$$m = 1 - \left[\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) - \left(\frac{1}{n_1 + n_2 - 2}\right)\right]\left(\frac{2p^2 + 3p - 1}{6(p+1)}\right) \tag{3-22}$$

and

$$M = (n_1 + n_2 - 2)\log|S| - (n_1 - 1)\log|S_1| - (n_2 - 1)\log|S_2| \tag{3-23}$$

where p represents the number of quality characteristics from sample sizes n_1 and n_2, respectively. $\boldsymbol{S}_1$ was from the standard calculation for covariance, referred to by Levinson *et al.* as the "full data set" method and $\boldsymbol{S}_2$ used the mean squared successive differences, or MSSD, (Holmes and Mergen, 1998) formula and was given as

$$S_{MSSD} = \frac{1}{2(n-1)}\sum_{j=2}^{n}(X_j - X_{j-1})(X_j - X_{j-1})^T \tag{3-24}$$

so that $\boldsymbol{S}$ represents a pooled covariance. In this equation the logarithm used is base 10. In equation 3-24, X_{j-1} and X_j are consecutive samples in a sample of size *n*, and *T* is the transpose operator. While the calculation in 3-24 can be shown to be an unbiased estimator for the covariance, the challenge is to standardize this equation to form a correlation that is comparable to the correlation calculated as a standardized covariance using the regular approach. In other words, the development of a MSSD correlation is required and this type of calculation has not been demonstrated to be feasible.. As such, a modified version of the *G* statistic is discounted for use with the correlation control chart.

3.3.4 A Word About Bias

It should be noted that the sample correlation coefficient, r_{ik}, is not an unbiased estimator of the population correlation coefficient, ρ_{ik}, for any pair (i, k), so the correlation matrix, $\boldsymbol{R}$, is not an unbiased estimate of the correlation matrix $\boldsymbol{\rho}$. However, it has been noted by Johnson & Wichern that the bias given by

$$E(r_{ik}) - \rho_{ik} \qquad (3\text{-}25)$$

can be ignored if the sample size, n, is "moderately large" (p. 99). In the simulation studies of the next chapter (Chapter 4) sample sizes are increased from smaller values through the "moderately large" range up to a value of n = 30. As this range is necessary to use the simulation to characterize the behavior of the statistics, the bias that may be present for smaller sample sizes will be accepted herein since it cannot be accurately quantified.

4.0 INVESTIGATION OF CORRELATION STATISTIC BEHAVIOR

Based on the aforementioned analyses, the modified version of the G statistic is not suitable for development of correlation control charts and the proposed $|\boldsymbol{R}|$ statistic potentially exhibits many undesirable characteristics when $|\boldsymbol{R}| > 0.4$. In this chapter, the W_i Statistic, modified for correlation and denoted as W_R, and the control chart based on $|\boldsymbol{R}|$ are both investigated for a variety of conditions to characterize their behavior. A simulation study is used with the in-control *ARL* as the primary metric of performance evaluation.

The in-control *ARL* is defined as the average number of points that must be plotted before a point indicates an out-of-control condition even if the process is in-control (Montgomery, 1997). The average run length is given by the formula

$$ARL = \frac{1}{p} \tag{4-1}$$

where p is the probability that any point will cause an out-of-control signal. For example, the univariate X-bar chart has a probability $p = 0.0027$ that a plotted point falls outside the control limits when the process is in-control, given standard control limits of $\pm 3\sigma$. The *ARL* is then calculated as 1/0.0027 = 370. In other words, even if the process is in-control, an out-of-control signal will be generated, on average, every 370 points (Montgomery, 1997, p. 142). Use of the out-of-control *ARL* is not typical for comparisons of the type proposed herein for reasons presented with a description in Appendix D.

The in-control *ARL* is used to investigate the two covariance statistics that can be investigated for the special case of correlation matrices. Recall, from Chapter 3, that use of the W_i Statistic presumes the assumptions of consistent scale factor while the chart based on $|\boldsymbol{R}|$ does not consider the scale factor because it looks at $|\boldsymbol{R}|$ after the standardization has occurred. Although both methods could be used to monitor correlation alone if only the correlation matrix data is available, it is preferred to also employ a covariance monitoring chart, and/or other control charts, such as the T^2 (for the location vector) whenever feasible.

4.1 INTRODUCTION

For simulation purposes, it is expedient to directly create deviates of correlation matrices to test the behavior of the statistics. A method to calculate a data set starting from a desired correlation matrix, working backwards creating a covariance matrix and then an initial data set for simulation was not found in the literature—development of such an algorithm is left as future work. Note that this does not preclude the testing of the two methods with real-world data sets after the simulations have been performed. Specifically, multivariate normal deviates were generated at four levels of variance for two correlation matrices of common, special structures. Deviates were applied to one, two, and three entries in *p*-quality characteristic matrices (only one deviate could be applied when $p = 2$). A terminating sequential simulation was then used to apply the statistics of interest—modified $|\boldsymbol{S}|$ and W_i—to each of several combinations of the number of quality characteristics, p, and sample size, n, with the specified-tolerance *ARL* being the terminating event. The *ARL*s for each of these runs were tabulated and most were plotted against n and p to graphically compare the $|\boldsymbol{R}|$ and W_R statistics across parameters. An empirical equation was developed to assess the relative contributions of the parameters for the "moderate" correlation case—a case that has been shown to commonly exist in a variety of processes (Johnson & Wichern, 1988).

The two types of structures considered are reviewed in sections 4.2.1 and 4.2.2 along with the parameters tested. The terminating sequential simulation and its results are described in the remainder of the chapter.

4.2 CORRELATION MATRIX STRUCTURES

One challenge to making assessments of the various dispersion statistics is the adoption of in-control matrices from which to develop perturbations and then apply the statistics that are to be compared. Even though correlation matrices are bounded, there are still an infinite number of choices possible for the off-diagonal elements. Since correlated data often results in patterned matrices, this leads to the suggestion that the selection of common correlation structures allows an orderly method with which to conduct the analysis.

Six common structures for correlation matrices are Independence, Exchangeable, Unstructured, Fixed, Auto-regressive, and *M*-dependent. The time-dependency of the *M*-dependent and Auto-regressive cases lend themselves to other types of analyses whereas the Unstructured and Fixed structures tend to have elements of disorder that are not conducive to simulation studies. These four structures would not provide results that are comparable with each other and they are, therefore, not included. The Exchangeable structure, and Independence structure (which can be shown to be a subset of the Exchangeable structure), conversely, resemble structures that occur naturally in certain processes and are comparable so they are employed for the analyses.

4.2.1 Exchangable Structure

The *Exchangeable* correlation structure is defined as $R_{u,v} = 1$ if $u = v$, and $R_{u,v} = \alpha$ otherwise and is represented by

$$\begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{bmatrix} \qquad (4\text{-}2).$$

The Exchange Structure with α = 0.5 was selected for simulation because it expresses neither a weak nor a strong correlation and occurs naturally in a number of processes (Johnson & Wichern, 1988). This is consistent with the work of Hawkins (1991), whose data will be further considered in a later section in conjunction with this research. Deviates were generated at four levels of variance (0.25, 0.20, 0.15, and 0.10) from a multivariate normal distribution and applied to the Exchange (0.5) Structure for p = 2, 3, 5, and 8 quality characteristics for sample sizes, n = p + 1, . . . , 10, 12, 15, 20, 25, 30, 35, and 40. These ranges were chosen to capture a range commensurate with other MSPC studies. With the exception of the bivariate case where p = 2, three types of changes in the correlation matrix were considered. The first case, which is the only one that applies to the bivariate case, considered a change in the correlation coefficient between two variables, r_{12} (= r_{21}). The second case considered a change two correlation coefficients, meaning a change in correlation between three variables. The third case considered a change in three correlation coefficients, meaning a change in correlation between four variables. The following is an illustration of a perturbed Exchange (0.5) matrix for $p = 3$ with one change in correlation applied to $r_{12} = r_{21}$:

$$\begin{bmatrix} 1 & 0.43 & 0.5 \\ 0.43 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

This approach, used here to study the correlation matrix, is similar to the approach taken by Hawkins (1991) investigating test statistics that he claims are more powerful than the Hotelling T^2 for detecting shifts in the mean vector. Hawkins suggests that departures from control are more commonly restricted to changes in a minority of the variables, instead of to all the variables simultaneously, and applies several approaches based on maximum likelihood estimators. (p. 63).

The simulations were performed until the Average Run Length (*ARL*) was calculated for each combination to a standard error no greater than 0.05 of the *ARL* itself—this was the termination ratio used to stop the simulation—for both the $|R|$ and the W_R statistic and the results tabulated. A complete table of the results of each simulation run, including *ARL*, standard error, and number of repetitions until termination, appear in Appendix F.

4.2.2 Independence Structure

The *Independence* Structure is the Exchange Structure where alpha in equation 4-2 is equal to zero and it is an important structure to consider since the results of other types of analyses, such as principal components analysis (PCA), provide a relationship between variables where no correlation is expected to be present. As such, the detection of correlation is an indicator that something has changed in the expected relationship among the variables. The *Independence* correlation structure is defined as $R_{u,v} = 1$ if $u = v$, and $R_{u,v} = 0$ otherwise and is represented by

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad (4\text{-}3).$$

Deviates were generated from a multivariate normal distribution at up to four levels of variance (0.25, 0.20, 0.15, and 0.10) and applied to the Independence Structure for p = 2, 3, and 5 quality characteristics for sample sizes, $n = p + 1, \ldots, 10$, 12, 15, 20, 25, 30, 35, and 40. With the exception of the bivariate case where p = 2, three types of changes in the correlation matrix were considered, of the same nature as with the Exchange (0.5) Structure described in Section 4.2.1.

4.3 TERMINATING SEQUENTIAL SIMULATION

Since the *ARL* indicates the first time an out-of-control point is generated (Type I error), it meets the conditions of a "natural" event for which a terminating simulation is indicated (Law and Kelton, 2000). A point estimate of the *ARL*, for the various combinations of n and p described in sections 4.2.1 and 4.2.2, was obtained from each simulation. Table 2 summarizes these combinations. The trigger points that terminate the simulation are any values of the statistic that exceed the control limits as defined in Table 1, shown in Chapter 3.

To visualize how this simulation progresses, consider Figure 7. Note that, for clarity, in Figure 7 the value for n_0 is set at zero for initialization but that the actual value for the number of replications must, in practice, exceed two to employ the algorithm to be described. The various correlation matrices for the simulation are formed by the application of standard normal deviates to the selected number of off-diagonal entries of the in-control matrix, where $\rho \equiv R_0$ for each simulation. The test statistic (|***R***|, W_R) was then calculated and compared to the appropriate control limits as listed in Table 1. As long as the process remained "in-control" and the limits were not exceeded, the *ARL*(*n,p*) counter incremented and this process was repeated (the parenthetical notation indicates that the *ARL* is a function of *n* and *p*) until the control limits were exceeded. The value of *ARL*(*n,p*) was stored and the next replicate was performed. Replications continued until the simulation terminated by meeting or exceeding the established relative precision (an α = 0.05 was used for this research for reasons that will be explained in Section 4.4) or the value of *ARL*(*n,p*) exceeded 10,000 which is beyond a practical range for control charting.

Table 2 Combinations of *n* and *p* for Simulations

	|*R*| Statistic				W_R Statistic			
p->	**2**	**3**	**5**	**8***	**2**	**3**	**5**	**8***
n								
3	1	n/a	n/a	n/a	1	n/a	n/a	n/a
4	1, 2, 3	1, 2, 3	n/a	n/a	1, 2, 3	1, 2, 3	n/a	n/a
5	1, 2, 3	1, 2, 3	n/a	n/a	1, 2, 3	1, 2, 3	n/a	n/a
6	1, 2, 3	1, 2, 3	1, 2, 3	n/a	1, 2, 3	1, 2, 3	1, 2, 3	n/a
7	1, 2, 3	1, 2, 3	1, 2, 3	n/a	1, 2, 3	1, 2, 3	1, 2, 3	n/a
8	1, 2, 3	1, 2, 3	1, 2, 3	n/a	1, 2, 3	1, 2, 3	1, 2, 3	n/a
9	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
10	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
12	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
15	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
20	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
25	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
30	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
35	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
40	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3	1, 2, 3
*p = 8 not simulated for Independence Structure								
All combinations simulated for variances: 0.10, 0.15, 0.20, 0.25								

The values of the *ARL* as a function of n and p for the total number of simulations, n_0, was stored by the simulation. The average of these values was then taken as the final value for *ARL*(n,p). This is consistent with the description of the terminating sequential description described earlier.

A sequential procedure that added new replications one-at-a-time until the desired relative precision was attained was used to provide an estimate of the mean value of the run lengths, with a confidence interval of 100 $(1 - \alpha)$ percent and a relative error $0 < \gamma < 1$, given by γ' and was adapted from Law and Kelton (2000, pp. 513-514). The total number of required replications for each simulation was recorded, but this was found to give no good indication of the actual convergence of the MSPC test statistic for the fixed confidence interval. The speed of modern computers has made runtime a less important indicator for simulations of this nature.

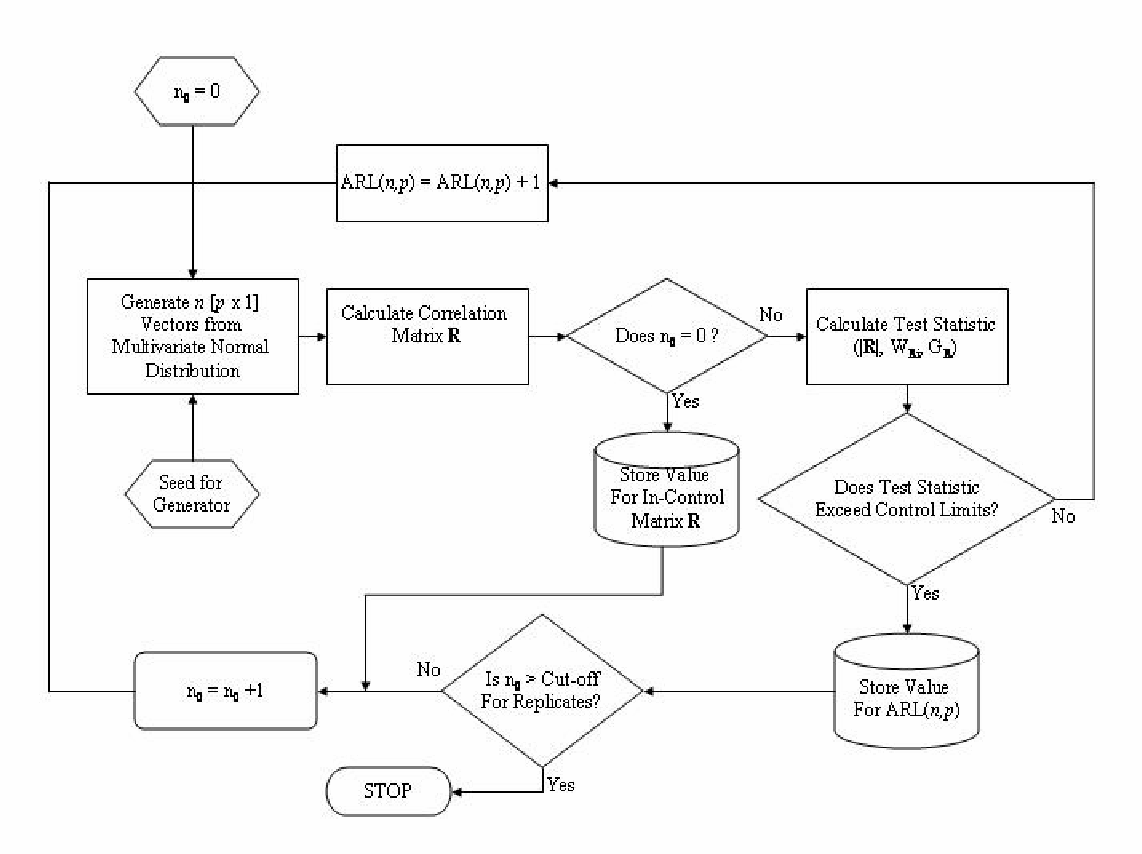


Figure 7 Simulation Flowchart

4.4 SIMULATION OF THE EXCHANGE STRUCTURE

The simulation was performed using MathCAD 2001 Professional software and an example of the code used is included as Appendix E. The tolerance level used was set at $\alpha = 0.05$ for two reasons. First, it provided relatively quick convergence of the algorithm while still providing a 95% confidence interval. Second, it was very robust to the ranges of matrix determinants used in the calculation. A line of code was also required to skip iterations (which were, therefore, not counted towards the *ARL*) in which singularities resulted. The *ARL*, standard error of the *ARL*, and number of iterations were recorded for each run. Summarized data appears in the tables of Chapter 4, and the entire data table appears as Appendix F.

The simulation results for the Exchange (0.5) Structure appear in sections 4.4.1. and 4.4.2 for the $|R|$ statistic and the W_R statistic, respectively. The two statistics are compared in Section 4.4.3. Results for the Independence Structure are considered in Section 4.5. Complete data for all of the Exchange Structure simulation runs appear in Appendix F, Table F-1. In the following sections, tabulated data will only be presented for the case where $p = 2$ as a means of introduction, after which only the charted data is used for discussion. Where discontinuities, or missing data points, occur on the charts, the simulation did not produce usable data for one of two reasons. The first reason for missing chart data would be that the simulation generated either a singularity or an “unknown error.” The second reason would be that the *ARL* generated by the simulation exceeded 10,000 and was, therefore, well beyond a practical upper limit for control charting.

4.4.1 The |*R*| Statistic

The results for the |R| Statistic are shown in Figures 8-10. In general, *ARL* decreases with increases in sample size (n), number of quality characteristics (p), and variance of the deviations. The *ARL* also decreases as the number of altered correlation coefficients in the matrix increases. In a number of cases, the combination of these varied parameters was significant enough to disallow the calculation of an *ARL*, or produced an *ARL* that exceeded 10,000 at which point the simulation was terminated automatically.

4.4.1.1 |*R*| Statistic for Number of Quality Characteristics *p* = 2 and One Change in the Correlation Matrix

Referring to Table 3 and Figure 8, it will be noticed that for a variance of 0.10, only sample sizes greater than 30 ($n > 30$) produced an *ARL* that was less than 10,000. Larger variances produced *ARL*s for larger range of sample sizes (*n*).

Table 3 *ARL*s for |*R*| Statistic for *p* = 2, with One Change to the Exchange (0.5) Matrix

	Variance			
Sample Size	0.25	0.20	0.15	0.10
3	>10,000	> 10,000	> 10,000	> 10,000
4	5044	> 10,000	> 10,000	> 10,000
5	1779	> 10,000	> 10,000	> 10,000
6	770	6985	> 10,000	> 10,000
7	411	4300	> 10,000	> 10,000
8	253	2329	> 10,000	> 10,000
9	180	1378	> 10,000	> 10,000
10	133	876	> 10,000	> 10,000
12	84	443	7185	> 10,000
15	53	206	3409	> 10,000
20	31	95	985	> 10,000
25	21	58	406	> 10,000
30	17	39	220	> 10,000
35	14	30	140	6384
40	12	24	99	3852

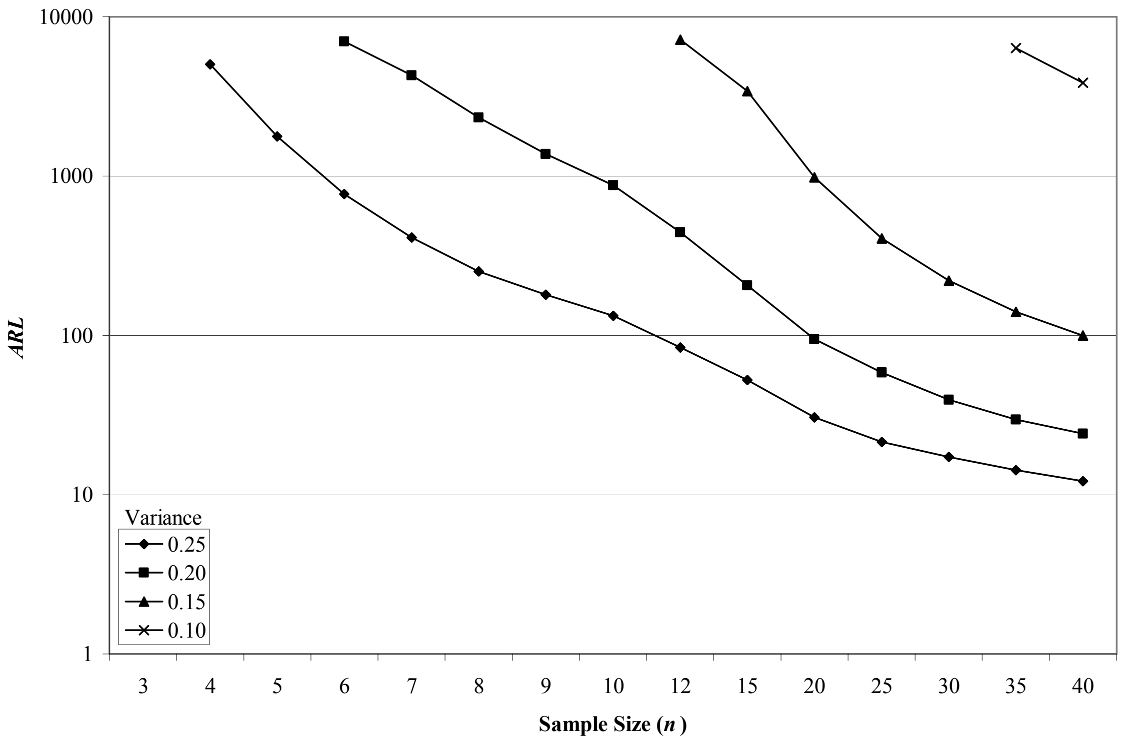


Figure 8 *ARL*s for |*R*| Statistic for *p* = 2, with One Change to the Exchange (0.5) Matrix

4.4.1.2 |*R*| Statistic for Number of Quality Characteristics *p* = 3 with 1-3 Changes in the Correlation Matrix

Referring to Figure 9, when the number of characteristics (p) was increased by one with a single change to the correlation matrix no simulations produced an *ARL* that was less than 10,000 for a variance of 0.10. For a variance of 0.15 only sample sizes greater than 15 ($n > 15$) produced an *ARL* that was less than 10,000. Comparing Figure 9(a) for a single change with $p = 3$ to Figure 8 for a single change with $p = 2$, the effect of increasing the number of quality characteristics (p) has increased the *ARL*s for the same range of sample sizes (n).

In Figure 9(c), it will be noticed that for a variance of 0.10, only the simulation with a sample size, $n = 35$, produced an *ARL* that was less than 10,000 since the sample size, $n = 40$, produced an error. For a variance of 0.15 the *ARL* was greater than 10,000 or errors were generated until sample sizes were greater than 20 ($n > 20$). Thus, Figure 9 shows that the *ARL* decreases as the number of changes in the correlation matrix increases, holding the value of the variance and the number of quality characteristics (p) constant.

a. One change

b. Two changes

c. Three changes

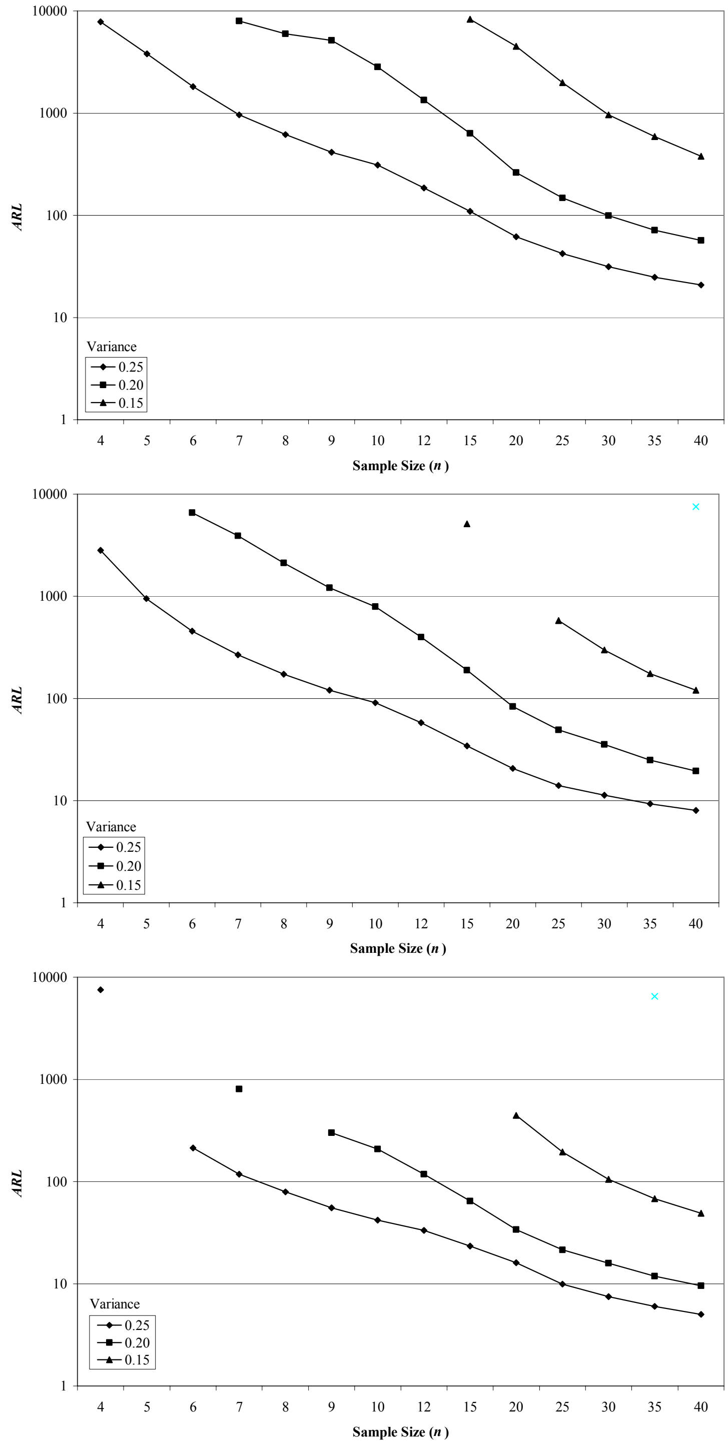


Figure 9 ***ARL*s for |*R*| Statistic for *p* = 3, with 1- 3 Changes to the Exchange (0.5) Matrix**

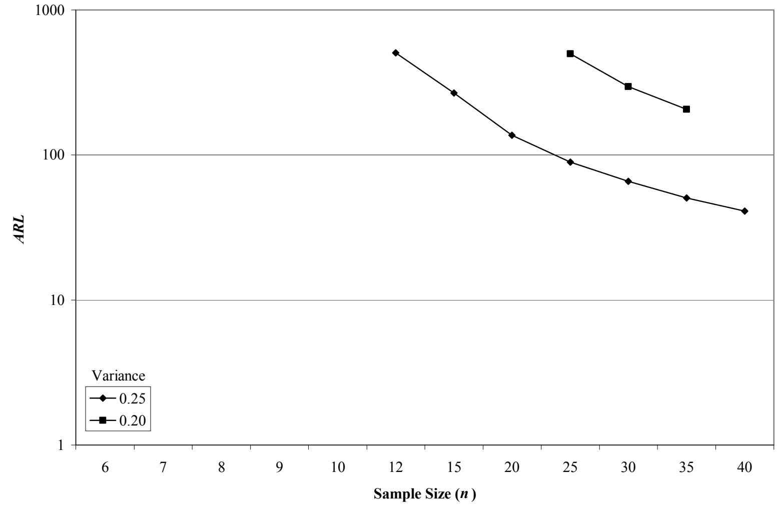
4.4.1.3 |*R*| Statistic for Number of Quality Characteristics *p* = 5 and 1-3 Changes in the Correlation Matrix

In Figure 10(a), for variances 0.10 and 0.15, all of the simulations returned an error or generated an *ARL* that was over 10,000. Compared to $p = 2$ quality characteristics for one change in the correlation matrix, Figure 8, and $p = 3$ quality characteristics for one change in the correlation matrix, Figure 9(a), as the number of quality characteristics, *p*, increases, with all other parameters held constant, the corresponding *ARL*s increase, creating the changes shown by consecutive comparison. It is also worth noting that errors were generated for the smaller sample sizes (*n*).

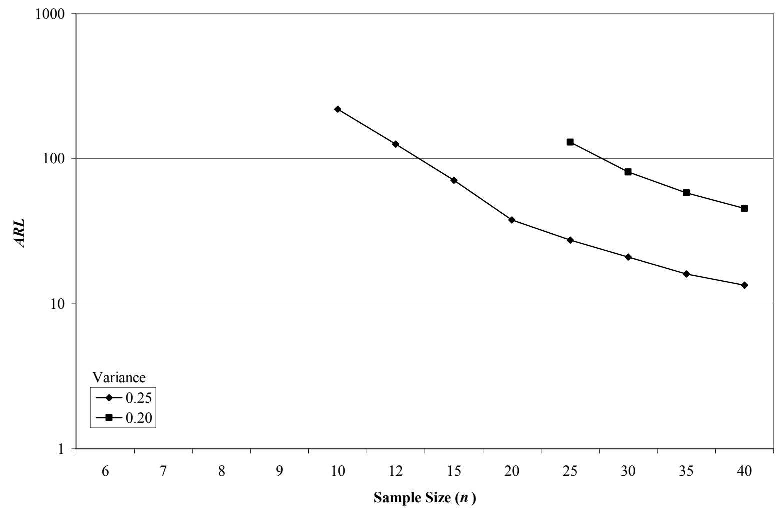
Referring to Figure 10(b), it will be noticed that a larger sample size (*n*) was required to get meaningful *ARL*s for the range of variances considered. A flattening of the curve is illustrated for the larger sample sizes (*n*) in Figure 10(b).

In Figure 10(c) the number of meaningful *ARL*s increased for the range of variances considered by changing an additional correlation coefficient in the matrix versus the case of only two changes to the correlation matrix. Although *ARL*s for a variance of 0.10 were all greater than 10,000, some meaningful *ARL*s were generated by simulations using a variance of 0.15, suggesting that a certain amount of "total variation" needs to be present in the correlation matrix for a change to be detected. Increasing the amount of "total variation," however, appears also to decrease the range for the *ARL*s generated, as illustrated by consecutively comparing the curves in Figure 10 for the variance 0.25.

a. One change



b. Two changes



c. Three changes

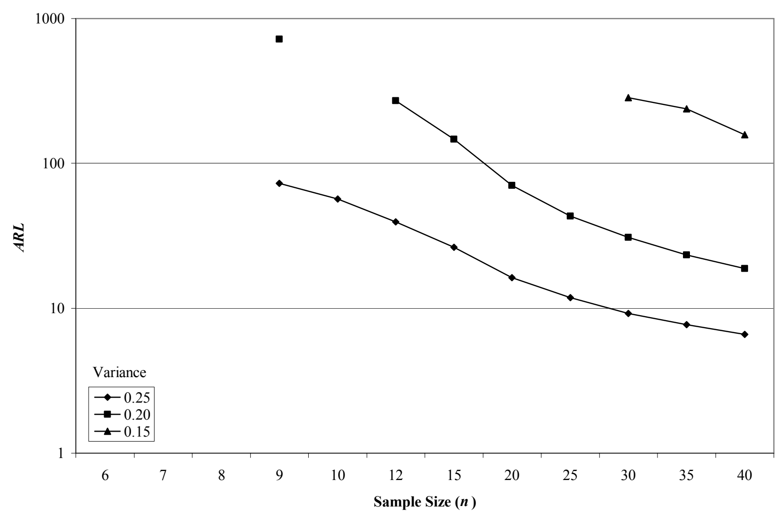


Figure 10 *ARL*s for |*R*| Statistic for *p* = 5, with 1-3 Changes to the Exchange (0.5) Matrix

4.4.1.4 |*R*| Statistic for Number of Quality Characteristics $p = 8$ and 1-3 Changes in the Correlation Matrix

The amount of "total variation" generated with one change in the correlation matrix and $p = 8$ quality characteristics was insufficient to produce meaningful *ARL*s for the range of sample sizes (n) and variances considered. The exception occurred with a variation of 0.25 and sample sizes of 25 or more ($n > 20$), but the range was contracted as has been illustrated previously. This data was not charted nor were simulations run for variances less than 0.15 due to the trends exhibited by the larger variances.

4.4.2 The W_R Statistic

The results for the W_R Statistic are shown in and Figures 11-14. In general, the W_R Statistic behaves similarly to the $|R|$ Statistic where *ARL* decreases with increases in sample size (n), number of quality characteristics (p), and variance applied for the perturbations. The *ARL* also decreases as the number of altered correlation coefficients in the matrix increases. However, the range of values for both statistics are not equivalent for the Exchange (0.5) Structure, as will be explored in Section 4.4.3.

In a number of cases where the |R| Statistic would not calculate for the combination of varied parameters presented, the W_R Statistic did. Again, complete data is presented in Table F-1 of the Appendix.

4.4.2.1 W_R Statistic for Number of Quality Characteristics p = 2 and One Change in the Correlation Matrix

In Table 4 and Figure 11 for the range of variances considered most of the simulation runs returned *ARL*s less than 10,000 although combinations of small sample size (n) and small variance tended to be insufficient to cause *ARL*s within a practical range.

Table 4 *ARL*s for W_R Statistic for p = 2, with One Change to the Exchange (0.5) Matrix

	Variance			
Sample Size	0.25	0.20	0.15	0.10
3	197	471	3369	> 10,000
4	109	255	1717	> 10,000
5	70	161	1008	> 10,000
6	53	106	646	9603
7	41	85	451	9430
8	32	65	328	9106
9	29	56	248	8651
10	24	50	205	7577
12	19	35	131	5868
15	14	27	90	3022
20	10	18	52	1071
25	10	12	35	492
30	6	10	26	269
35	5	8	20	172
40	4	7	17	121

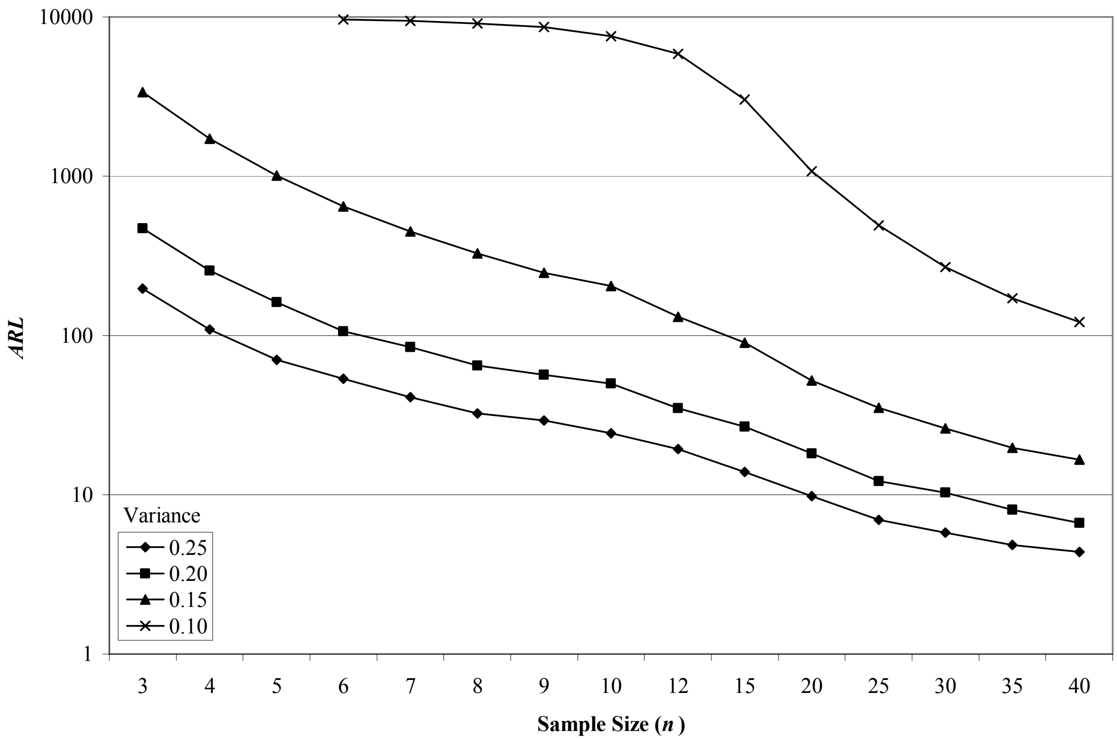


Figure 11 ***ARLs*** **for** W_R **Statistic for** $p = 2$**, with One Change to the Exchange (0.5) Matrix**

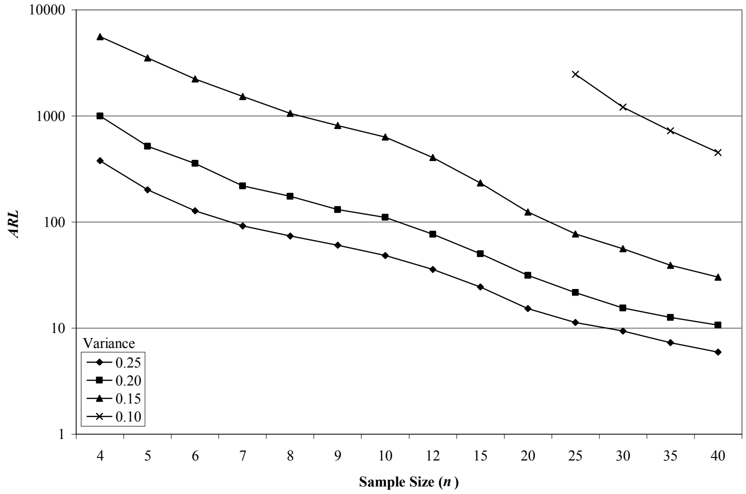
4.4.2.2 W_R Statistic for Number of Quality Characteristics $p = 3$ and 1-3 Changes in the Correlation Matrix

In Figure 12(a), it will be noticed that for a variance of 0.10, only sample sizes greater than 20 ($n > 20$) produced an *ARL* that was less than 10,000. Comparing Figure 12(a) to Figure 11, the effect of increasing the number of quality characteristics (p) has increased the *ARL*s for the same range of sample sizes (n) and is similar to the phenomenon seen with the $|R|$ Statistic in Figures 8 and 9(a).

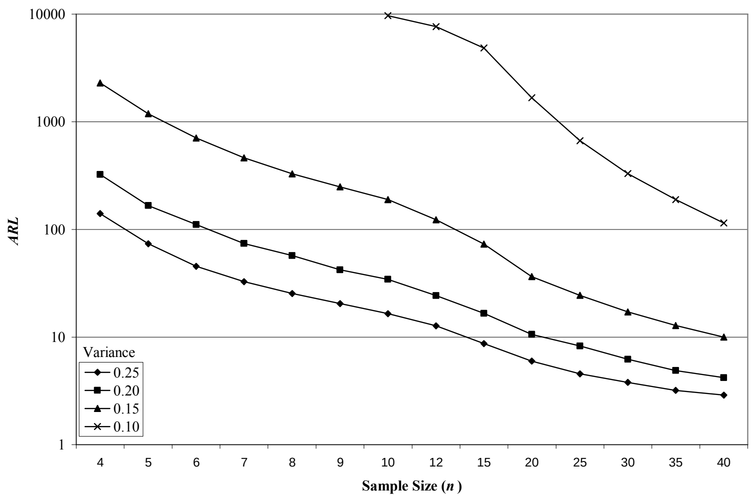
With two changes in the correlation matrix, Figure 12(b) shows, for a variance of 0.10 only sample sizes greater than 9 ($n > 9$) produced an *ARL* that was less than 10,000.

As the number of changes increases to three, Figure 12(c), for the smallest tested variance of 0.10, only the simulation with a sample size, n = 4, produced an *ARL* that was greater than 10,000. Comparing Figures 12(a) through (c), it will be noticed that the *ARL* curve shifts downward as the number of changes in the correlation matrix increases, holding the value of the variance and the number of quality characteristics (p) constant.

a. One change



b. Two changes



c. Three changes

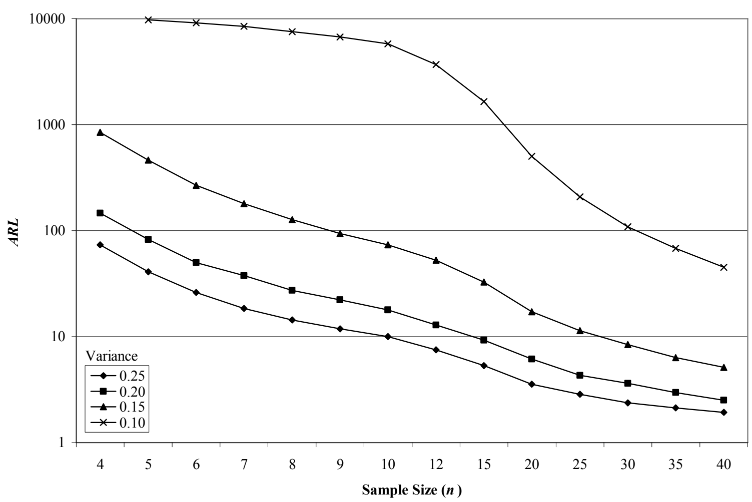
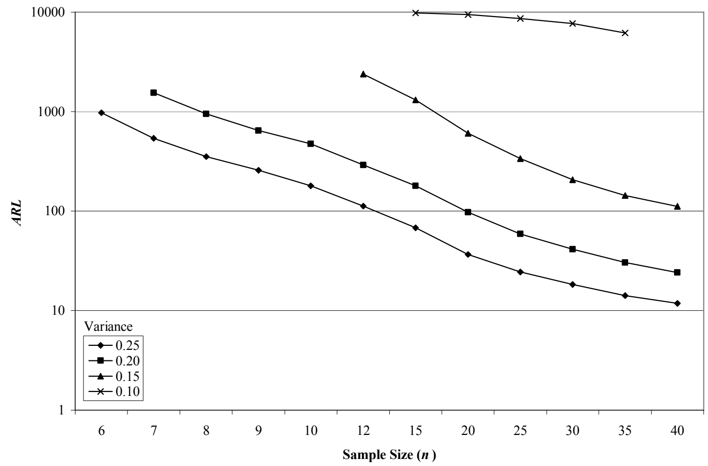


Figure 12 *ARL*s for W_R Statistic for p = 3, with 1-3 Changes to the Exchange (0.5) Matrix

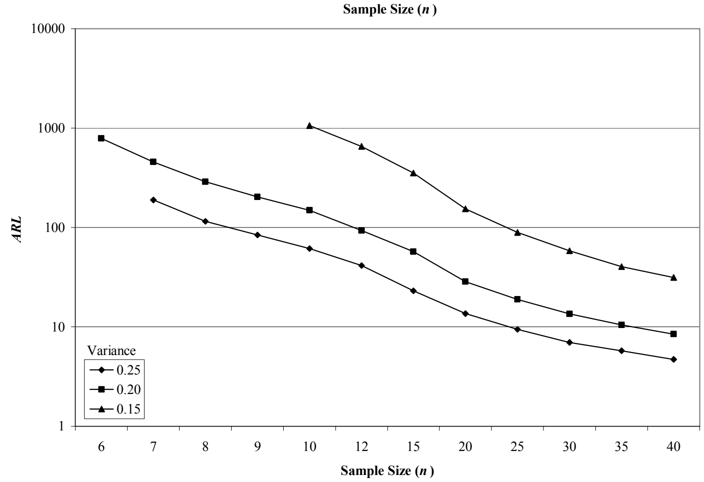
4.4.2.3 W_R Statistic for Number of Quality Characteristics $p = 5$ and 1-3 Changes in the Correlation Matrix

Figure 13 shows the effect of increasing the number of quality characteristics to $p = 5$. For variance 0.10 most of the simulations returned an *ARL* that was either over 10,000 or close to it, and for variance 0.15 errors were returned for the smaller sample sizes ($n < 12$). Compared to $p = 2$ quality characteristics for one change in the correlation matrix, Figure 11, and $p = 3$ quality characteristics for one change in the correlation matrix, Figure 12, it will be noticed that, as the number of quality characteristics, p, increases, with all other parameters held constant, the corresponding *ARL*s increase, creating a steeper curve as shown by comparison. It is also worth noting that errors were generated for the smaller sample sizes (n). Figure 13(b) indicates that a larger sample size (n) was required to get meaningful *ARL*s for the range of variances considered. The number of meaningful *ARL*s increased for the range of variances considered by changing an additional correlation coefficient in the matrix versus the case of Section 4.4.2.2 as shown in Figure 13(c). A flattening of the curve is illustrated for the larger sample sizes (n) in Figure 13. Similarly to the $|R|$ Statistic, increasing the amount of "total variation," however, appears also to contract the range for the *ARL*s generated, as illustrated by consecutively comparing the curves in Figure 13 for any of the variances.

a. One change



b. Two changes



c. Three changes

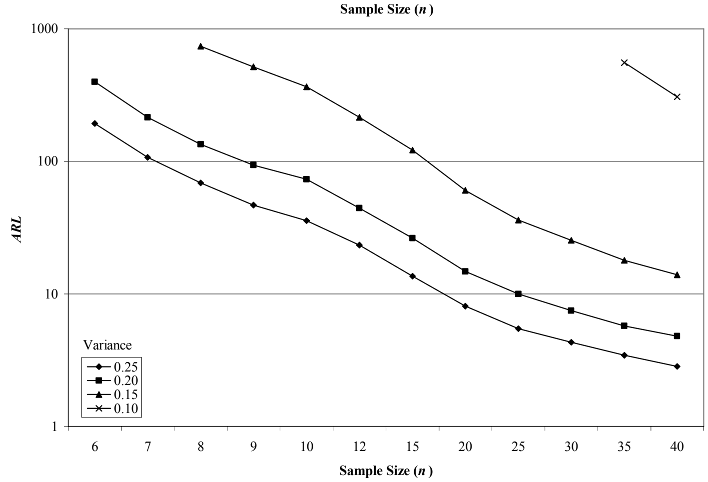
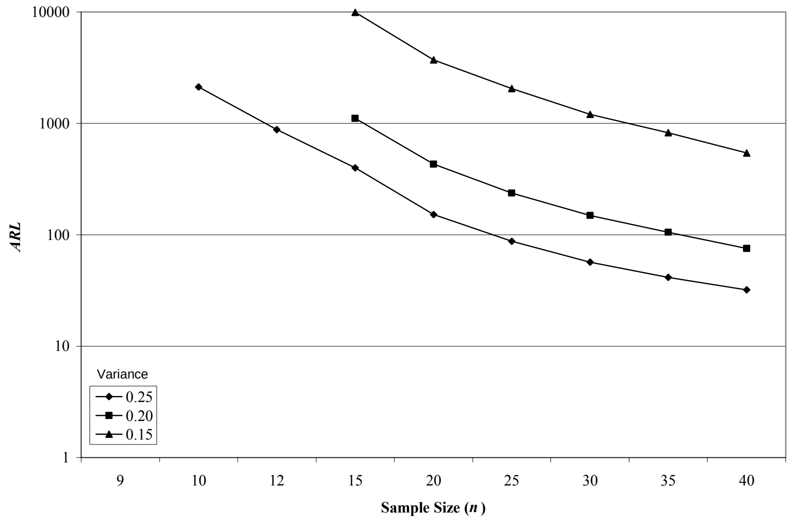


Figure 13 *ARL*s for W_R Statistic for *p* = 8, with One Change to the Exchange (0.5) Matrix

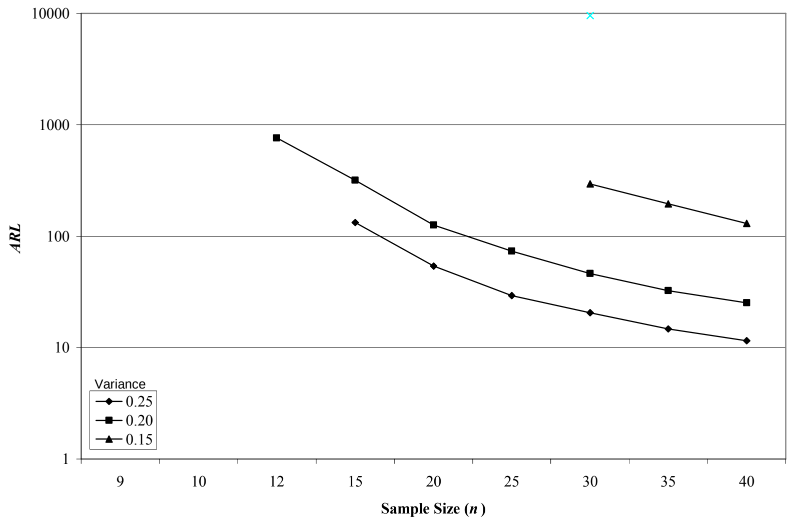
4.4.2.4 W_R Statistic for Number of Quality Characteristics p = 8 and 1-3 Changes in the Correlation Matrix

In Figure 14(a) the amount of "total variation" generated with one change in the correlation matrix and p = 8 quality characteristics produced meaningful *ARL*s for the higher ends of the ranges of sample sizes (n) and variances considered. As sample sizes (n) and variances simultaneously decreased, the simulations returned either errors or *ARL*s exceeding 10,000. No simulations were run for variances less than 0.15 due to the trends exhibited by the larger variances.

a. One change



b. Two changes



c. Three changes

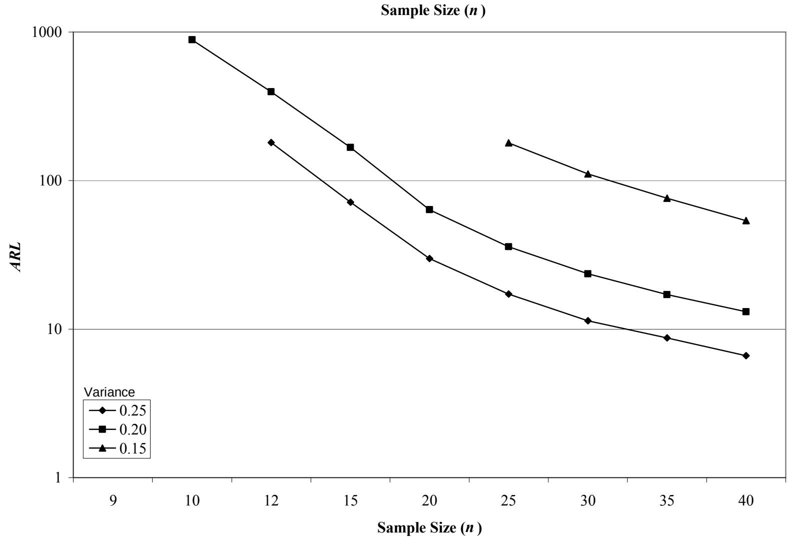


Figure 14 *ARL*s for W_R Statistic for *p* = 8, with 1-3 Changes to the Exchange (0.5) Matrix

In Figure 14(b), recognize that the performance of the *ARL*s was similar to that for $p = 8$ quality characteristics and one change in the correlation. Again, it will be noticed that the amount of "total variation" generated with two changes in the correlation matrix and $p = 8$ quality characteristics produced meaningful *ARL*s for the higher ends of the ranges of sample sizes (n) and variances considered. As sample sizes (n) and variances simultaneously decreased, the simulations returned either errors or *ARL*s exceeding 10,000.

With three changes to the correlation matrix and $p = 8$, the majority of the simulations runs produced meaningful *ARL*s except for cases of small sample size (n) and/or small variance. For a variance of 0.10 all simulation runs produced either errors or an *ARL* that exceeded 10,000 with the larger sample sizes producing the former and the smaller sample sizes the latter.

4.4.3 Comparing the $|R|$ Statistic to the W_R Statistic

The qualitative behavior of the $|R|$ Statistic and the W_R Statistic was similar for the Exchange Structure with alpha = 0.5. In general, for both statistics:

- the *ARL* decreases with increases in sample size (n),
- the *ARL* decreases with increases in the number of quality characteristics (p)
- the *ARL* decreases with increases in the variance of the deviations
- the *ARL* decreases as the number of altered correlation coefficients in the matrix increases.

One of the important differences between the performance of the $|R|$ Statistic and the W_R Statistic concerns the range of the *ARL*s. As the previously listed characteristics increase, the *ARL* decreases more rapidly for the $|R|$ Statistic than for the W_R Statistic. Additionally, for the smaller end of the range of variances, the *ARL* for the $|R|$ Statistic exceeds 10,000 in many of the cases. The practical significance of this is that an out-of-control condition may never be recognized by the $|R|$ Statistic. Additional insight into this characteristic of the $|R|$ Statistic will be discussed in Section 4.5, where the Independence Structure is considered, and a deeper examination of the W_R Statistic is provided in Section 4.6.

4.5 SIMULATIONS FOR THE INDEPENDENCE STRUCTURE

The following sections review the results for the $|R|$ Statistic and the W_R statistic for the Independence Structure. Complete data for all of the Independence Structure simulation runs appear in Appendix F, Table F-2. In the following sections, tabulated data will only be presented for the case where $p = 2$ as a means of introduction, after which the charted data is used for discussion. Where discontinuities, or missing data points, occur on the charts, the simulation did not produce usable data for one of two reasons. The first reason for missing chart data would be that the simulation generated either a singularity or an "unknown error." The second reason would be that the *ARL* generated by the simulation exceeded 10,000 and was, therefore, well beyond a reasonable number for charting.

4.5.1 The |*R*| Statistic

The $|R|$ Statistic did not perform well for the Independence Structure. In most cases, the simulation returned either an error or a value exceeding 10,000. The implication of these results is that use of the $|R|$ Statistic will rarely show an out-of-control condition. This is not unexpected when one considers that only very gross changes in the individual correlation coefficients, r_{ij}, will ever result in a determinant, $|R|$, that varies significantly from unity. As the control limits are calculated by

$$UCL = |\rho|\left(1 + z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right) \tag{4-3}$$

and

$$LCL = |\rho|\left(1 - z_{\alpha/2}\sqrt{\frac{2p}{n-1}}\right), \tag{4-4}$$

where ρ is the in-control correlation matrix, p denotes the number of quality characteristics, n is the sample size, and $z_{\alpha/2}$ is a value from the standard normal distribution, the upper and lower limits of Equations 4-3 and 4-4, respectively will never be exceeded in most practical applications (see description in Section 3.3.1).

4.5.2 The W_R Statistic

The results for the W_R Statistic are shown in Figures 15-17. In general, the W_R Statistic behaves similarly when applied to the Independence Structure as when applied to the Exchange (0.5) Structure described in Section 4.1.2. The *ARL* decreases with increases in sample size (n), number of quality characteristics (p), and variance of the perturbations. The *ARL* also decreases as the number of altered correlation coefficients in the matrix increases.

4.5.2.1 W_R Statistic for Number of Quality Characteristics p = 2 and One Change in the Correlation Matrix

Referring to Table 5 and Figure 15 for the range of variances considered most of the simulation runs returned *ARL*s less than 10,000 although combinations of small sample size (n) and small variance tended to be insufficient to cause *ARL*s within a practical range.

Table 5 *ARL*s for W_R Statistic for p = 2, with One Change to the Independence Matrix

	Variance			
Sample Size	0.25	0.20	0.15	0.10
3	7373	> 10,000	> 10,000	> 10,000
4	4900	> 10,000	> 10,000	> 10,000
5	2807	9414	> 10,000	> 10,000
6	1622	8995	> 10,000	> 10,000
7	990	8254	> 10,000	> 10,000
8	625	7005	> 10,000	> 10,000
9	435	5316	> 10,000	> 10,000
10	308	3774	> 10,000	> 10,000
12	175	1791	> 10,000	> 10,000
15	89	680	8927	> 10,000
20	44	221	5143	> 10,000
25	26	98	1799	> 10,000
30	17	58	706	> 10,000
35	14	39	353	9455
40	11	28	200	8942

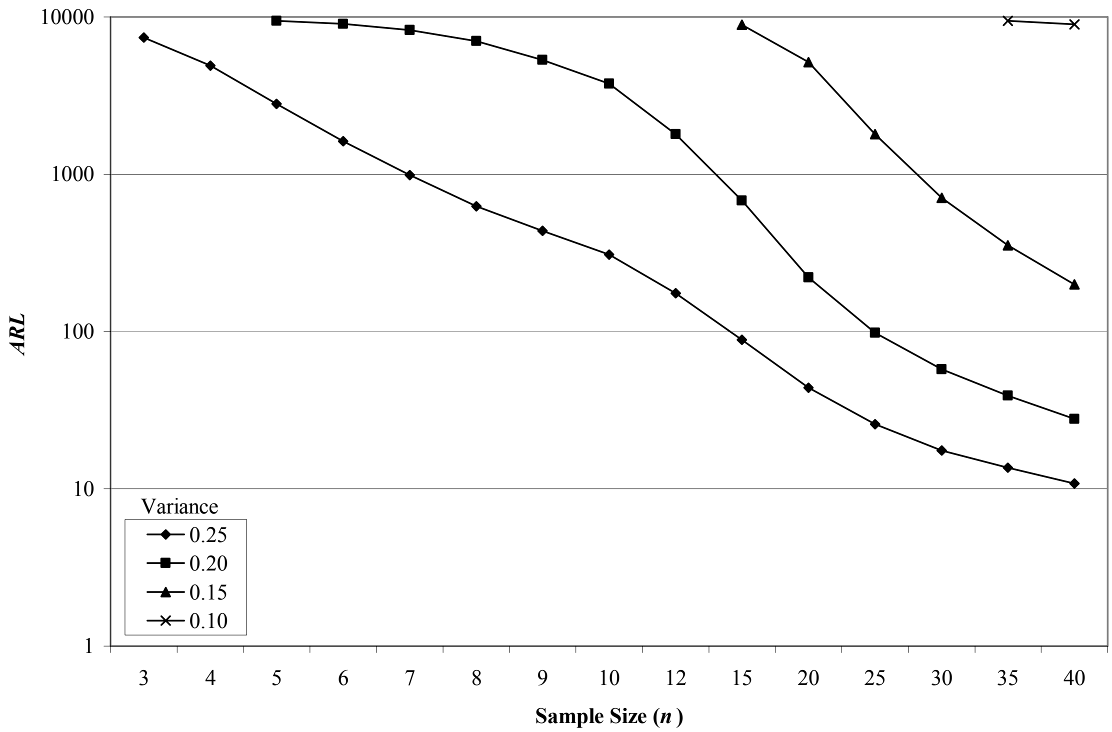


Figure 15 *ARL*s for W_R Statistic for *p* = 2, with One Change to the Independence Matrix

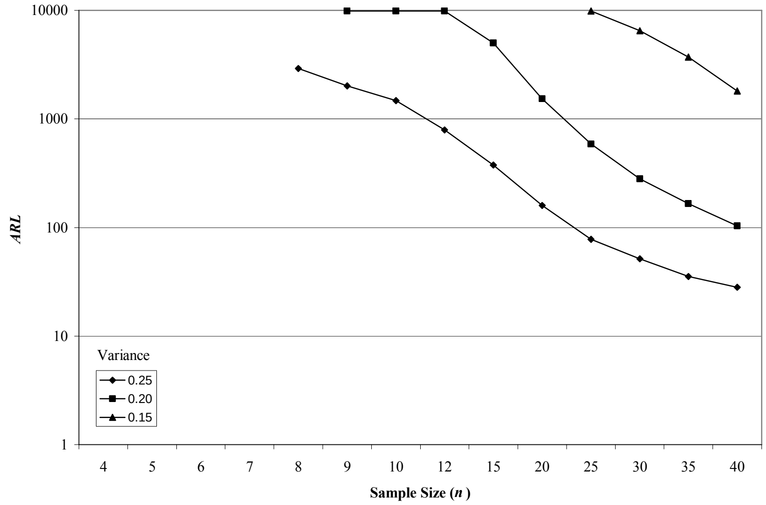
4.5.2.2 W_R Statistic for Number of Quality Characteristics *p* = 3 and 1-3 Changes in the Correlation Matrix

In Figure 16(a) for variances of 0.10 and 0.15, the *ARL*s were large or exceeded 10,000 for most of the sample sizes. Comparing Figure 15 to Figure 16(a), the effect of increasing the number of quality characteristics (*p*) has increased the *ARL*s for the same range of sample sizes (*n*), the same as with the Exchange (0.5) Structure shown Figure 8. This is also illustrated by comparing Figure 16(a) and Figure 16(b) and is similar to the phenomenon seen with the Exchange (0.5) Structure in Figures 12(a) and 12(b).

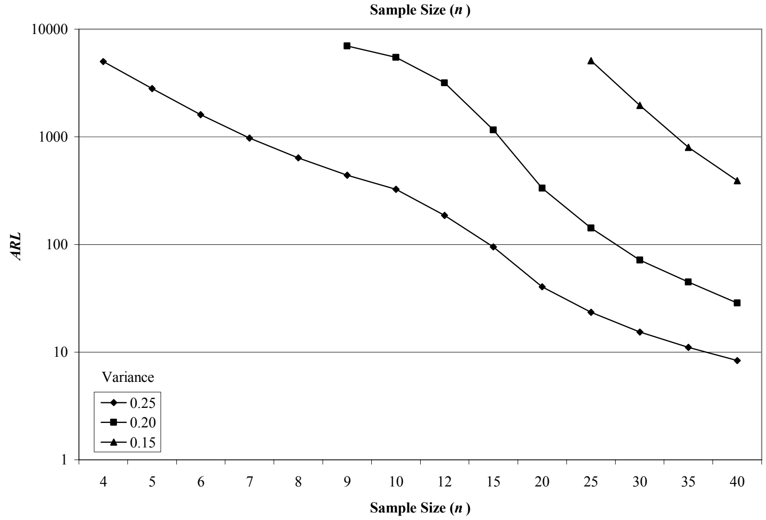
Figure 16 (b) shows that, for the smaller variances of 0.10 and 0.15, the *ARL*s generated were impractically large, with most exceeding 10,000.

In Figure 16(c) variances of 0.10 and 0.15 in most of the simulations produced an *ARL* exceeding 10,000 for the range of sample sizes (*n*) considered. Comparing Figure 16(a) though 16(c) the *ARL* curve for the Independence Structure shifts as the number of changes in the correlation matrix increases, holding the value of the variance and the number of quality characteristics (*p*) constant, similar to the case for the Exchange (0.5) Structure shown in Figure 14(a) through 14(c).

a. One change



b. Two changes



c. Three changes

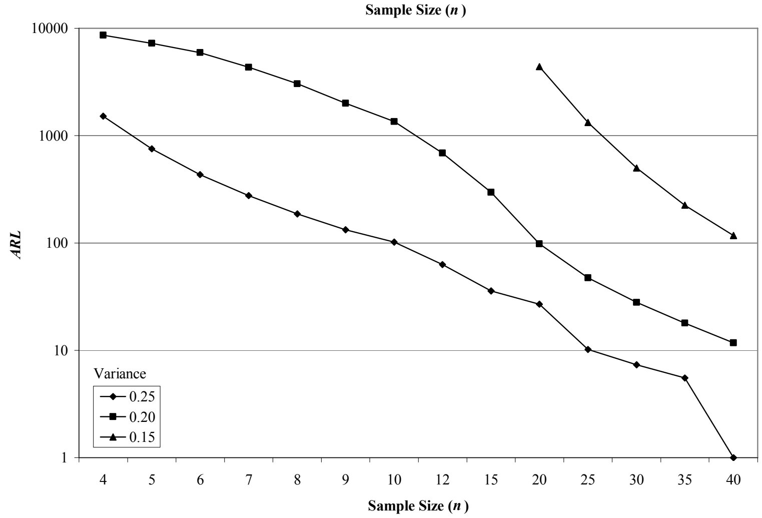


Figure 16 ***ARL*s for W_R Statistic for p = 3, with 1-3 Changes to the Independence Matrix**

4.5.2.3 W_R Statistic for Number of Quality Characteristics $p = 5$ and 1-3 Changes in the Correlation Matrix

When the number of quality characteristics was increased to $p = 5$, only the simulations using a variance of 0.25 returned an *ARL* that was in a reasonable range, Figure 17(a), although sample sizes greater than 15 ($n > 15$) returned *ARL*s less than 10,000 for a variance of 0.20. Compared to $p = 2$ quality characteristics for one change in the correlation matrix, Figure 15, and $p = 3$ quality characteristics for one change in the correlation matrix, Figure 16(a), as the number of quality characteristics, p, increases, with all other parameters held constant, the corresponding *ARL*s increase, shifting the curves as shown by consecutively comparing Figures 17(a) through 17(c).

With two changes, a larger sample size (n) was required to get meaningful *ARL*s for a variance of 0.25. Smaller variances (0.20, 0.15, and 0.10) returned either errors or an *ARL* exceeding 10,000.

Figure 17 shows how the number of meaningful *ARL*s increased for the range of variances considered by changing an additional correlation coefficient in the matrix versus the case of only two changes to the correlation matrix. A flattening of the curve is illustrated for the larger sample sizes (n) in Figure 17(c). Thus, increasing the amount of "total variation," however, appears also to contract the range for the *ARL*s generated, as illustrated by consecutively comparing the curves in Figure 17 for any of the variances. This is also true in the Exchange (0.5) Structure, as shown in Figure 14.

a. One change

b. Two changes

c. Three changes

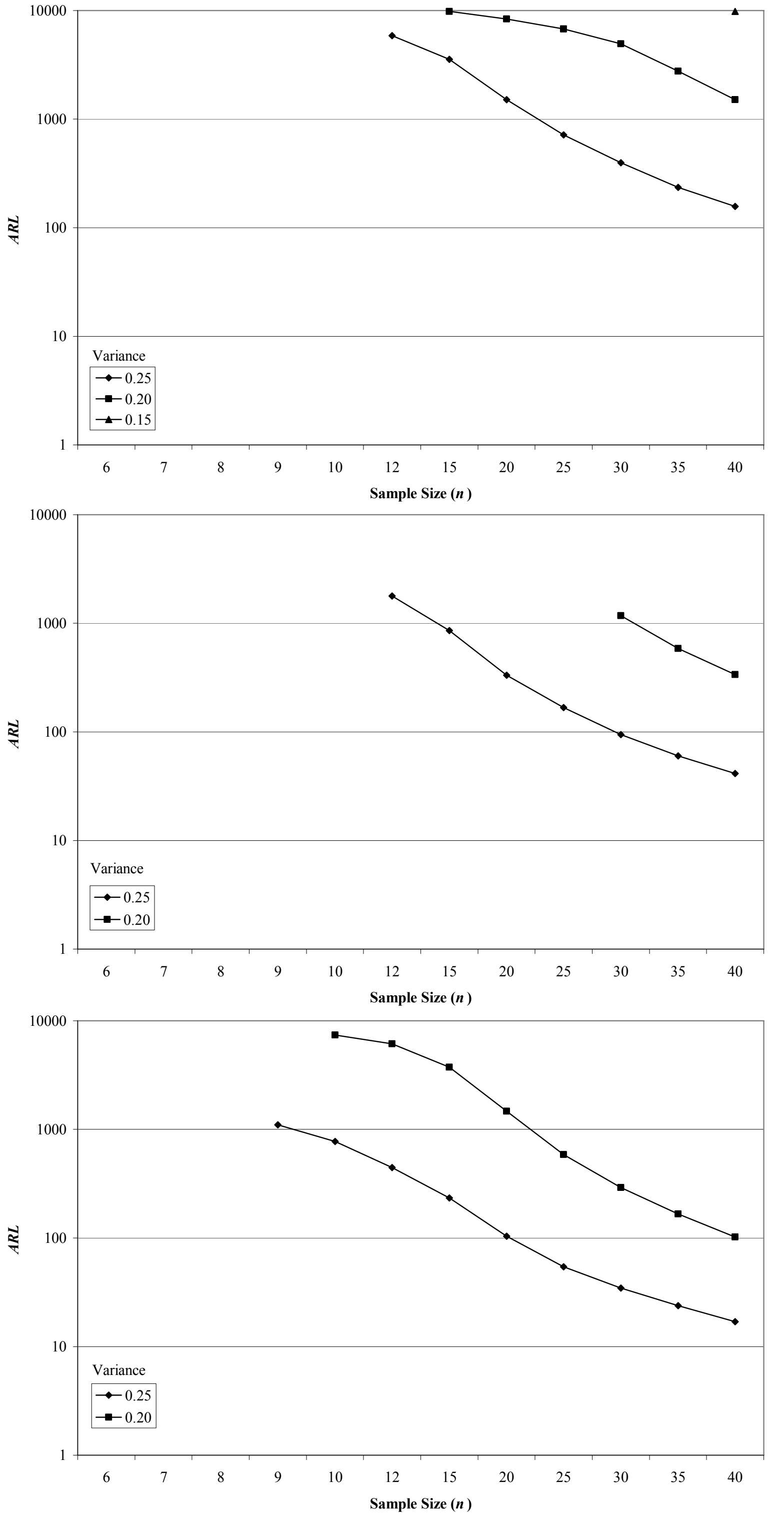


Figure 17 *ARL*s for W_R Statistic for p = 5, with 1-3 Changes to the Independence Matrix

4.5.2.4 W_R Statistic for Number of Quality Characteristics $p = 8$ and One to Three Changes in the Correlation Matrix

The amount of "total variation" generated with one to three changes in the correlation matrix and $p = 8$ quality characteristics did not produce meaningful *ARL*s, even for the higher ends of the ranges of sample sizes (n) and variances considered. This is in contrast to the Exchange (0.5) Structure (Figures 11-14) where the *ARL*s did not exceed 10,000 for all of the combinations as in the Independence Structure case. Higher values for the number of quality characteristics (*e.g.* $p = 8$) were, therefore, not simulated.

4.6 EMPIRICAL EQUATION

Based on the results of the simulations, it appears that the W_R Statistic is superior to the $|R|$ statistic since it can be calculated for both the Independence Structure and the Exchange (0.5) Structure, whereas the $|R|$ Statistic is not capable of detecting an out-of-control condition as the Exchange (0.5) Structure approaches the Independence Structure (as alpha 0.5 → 0.0). In addition, the ranges of values was also broader. In order to make recommendations for the application of the W_R Statistic and to further understand the effect of the various parameters, the *ARL*s for all cases where the *ARL* obtained was less than 10,000 for the simulation results for the Exchange (0.5) Structure were regressed on the number of changes to the correlation matrix, the variances, the number of quality characteristics (p), and the sample sizes (n).

An examination of the residuals revealed that regressing on the natural logarithm of the *ARL* was more appropriate, and two equations were generated. When a regression equation is developed for prediction, either a natural or standardized version can be used, but to interpret the contribution of the individual regressor variables, the standardized regression equation should be employed (Myers, 1990, pp. 384-385).

4.6.1 Natural Equation

For the sake of parsimony, the natural form of the regression equation is considered for determining the sample size based on the desired *ARL*. The natural regression equation is

$$\ln(ARL) = 12.011 + 0.467p - 0.810 * changes - 0.113n - 30.388 * \mathrm{var}iance \qquad (4\text{-}6),$$

where the variables are as listed above. Considering that the simulations were run with a specified standard error of 0.05 for the termination ratio, the coefficient of determination for the regression being 0.868 is reasonable. It will be shown that this value can be improved when the standardized equation is calculated in the next section for purposes of characterization. A Kolmogorov-Smirnov test was performed to verify normality.

Since the number of quality characteristics (*p*) is known from the size of the matrix, and the number of changes and variance can be assumed based on process conditions, a suitable sample size can be chosen based on the *ARL* desired, using this equation as an approximation. For example, assume that it is desired to have an *ARL* equivalent to that of the univariate Shewhart Chart, (*ARL* = 370), and that the expected amount of variation between two variables is 0.20 when five quality characteristics (*p*) are being considered.

Since *ln* (370) = 5.913, the required sample size can be estimated from 4-6 as

$$5.913 = 12.011 + 0.467 * (5) - 0.810 * (1) - 0.113 * n - 30.388 * (0.20) \quad (4\text{-}7),$$

or, in other words, choose a sample size of *n* = 13.67, or approximately *n* = 14. Due to process constraints or economic considerations, it may not be possible to specify a desired *ARL* and then calculate a corresponding sample size. In these cases, Equation 4-7 can give an estimate of the *ARL* that can be expected (and, hence, the probability of a Type I error), noting that the Exchange (0.5) Structure on which it is based assumes neither a strong nor a weak correlation but, rather, a moderate correlation between all variables concerned.

4.6.2 Standardized Equation

For purposes of characterization, the standardized regression equation was explored. The standardized version of Equation 4-6 is

$$\ln(ARL) = 0.441p - 0.335 * changes - 0.648n - 0.750 * \mathrm{var}iance \quad (4\text{-}8)$$

which, necessarily, also has a coefficient of determination of 0.868. Residuals were assessed for normality. In an attempt to better understand the contributions of each parameter, interaction terms were investigated. The only term that showed no significance (p = 0.596) for the full model was the interaction between the number of changes and the sample size (*n*).

As shown in Figures 11-14, the *ARL* increases with the number of quality characteristics (*p*) but decreases with increases in the number of changes to the correlation matrix, the sample size (*n*) and the degree of the change as represented by the variance. Equation 4-6 shows that the largest effect on the *ARL* comes from the amount of change introduced—this would be the preferred situation inasmuch as this is the change that the W_R Statistic is expected to detect.

Using forward, backward, and stepwise regression with probability of F to enter ≤ 0.050 and to leave ≥ 0.10, the same results were obtained and are as shown in Table 6A. From this table it is also evident that the variance is the largest contributor to the calculation for ln(*ARL*) shown in equation 4-8. The marginal impact to the model resulting from the inclusion of interaction terms may be foregone for the sake of parsimony when calculating a sample size estimate based on desired *ARL* and for such estimations, Equation 4-6 is recommended.

Table 6 Improvements to Regression Equation by Addition of Interaction Terms

Coefficient of Determination	Parameters
0.383	variance
0.645	variance, n
0.761	variance, n, p
0.868	variance, n, p, changes
0.931	variance, n, p, changes, $(\text{variance})^2$
0.958	variance, n, p, changes, $(\text{variance})^2$, $(\text{n})^2$
0.963	variance, n, p, changes, $(\text{variance})^2$, $(\text{n})^2$, $p \times n$
0.968	variance, n, p, changes, $(\text{variance})^2$, $(\text{n})^2$, $p \times n$, $n \times$ variance
0.975	variance, n, p, changes, $(\text{variance})^2$, $(\text{n})^2$, $p \times n$, $n \times$ variance,
0.976	variance, n, p, changes, $(\text{variance})^2$, $(\text{n})^2$, $p \times n$, $n \times$ variance,

5.0 IMPLEMENTING THE CORRELATION CONTROL CHART

The interpretation and evaluation of multivariate control charts is more complicated than for univariate control charts. The increased number of variables to consider, their associated parameters and correlation among the variables can possibly trigger or mask the presence of a shift, and the non-scalar nature of the shift itself can have the same effect. As a result, it is usually recommended that several multivariate charts be employed together, or in unison with univariate charts on the variables of interest (Alt, 1985; Johnson & Wichern, 1988; Flury & Riedwyl, 1988). The theoretical implementation of a correlation matrix control chart is presented in this Chapter.

In Section 5.1, some general recommendations for the application of correlation matrix control charts are discussed. Guidelines from the interpretation of location MSPC charts are then extended to the (dispersion) correlation control chart. As the *ARL* performance of dispersion control charts is a continuing area of research, in Section 5.2, the *ARL* performance of the correlation control chart using W_R, the modified version of the W_i statistic, is compared to the *ARL* performance of MSPC charts for location. Section 5.3 presents an example of shift detection with the correlation control chart and Section 5.4 comments on *ARL* behavior. The practical applicability of the correlation control chart is explored later, in Chapter 7.

5.1 DATA REQUIREMENTS FOR USE OF THE CORRELATION CONTROL CHART

The correlation control chart developed herein could more precisely be called a correlation matrix control chart as it does not seek to mitigate the effects of correlation between variables but, rather, attempts to detect shifts in the correlation between them via the standardized covariance (*a.k.a.* correlation) matrix. Data that is suited to this technique is common in chemical engineering where multivariate process control is used to monitor reactors and columns. This type of data is also naturally occurring such as in the movement of stock returns and in the biological sciences where measurements taken on numerous specimens of living things reveal fairly consistent correlation matrices (Johnson & Wichern, p. 366). The form of the correlation matrix is the Exchange Structure special case where the correlation between all variables is the same. In practice, the values are not all exactly the same, but the matrix entries are close enough to one another that techniques of principal components analysis are easily applied. While no formal definition is provided for “close enough to,” Johnson and Wichern present one illustrative example for the body weights of 150 female mice after the births of their first four litters. The correlation matrix provided is shown to be

$$R = \begin{bmatrix} 1.000 & 0.7501 & 0.6329 & 0.6363 \\ 0.7501 & 1.000 & 0.6925 & 0.7386 \\ 0.6329 & 0.6925 & 1.000 & 0.6625 \\ 0.6363 & 0.7386 & 0.6625 & 1.000 \end{bmatrix}$$

and the principal components analysis is performed. Although the entries are not all equal, they are all within 95% of the average value of 0.690 and the structure is considered of the Exchange type.

Instead of, or in addition to, the principal components approach, a correlation control chart could theoretically be used to check the first four litters of groups of five (since $n > p$) mice over time to monitor for shifts changes in weight outside the expected relationship that may then indicate illness or other factors.

The correlation control chart appears to detect subtle changes in the correlation matrix, $\boldsymbol{R}$, that differ from the in-control condition, $\boldsymbol{\rho}$. Thus, as the number of quality characteristics, p, increases, the correlation control chart decreases in efficacy because the effects of the variation become masked. Similarly, increasing the amount of total variation in the system has the effect of decreasing the efficacy of the W_R statistic because the *ARL*s of the resultant chart are decreased (quantifying this total system variation in a single parameter is considered future work). The opposite is also true; too little variation cannot be detected by the W_R statistic. The aforementioned phenomena are not unique and are seen with various other MSPC control charts as will be discussed in Chapter 6. In short, there are, for most types of control charts, specific ranges of application that are more suited to the chart of interest than to other charts, and *vice versa*.

When the whole data set is available, and it is possible to calculate the covariance matrix, $\boldsymbol{S}$, the appropriateness of the correlation control chart can be explored. The correlation control chart is designed to detect changes in the correlation matrix, assuming the scale factor remains nearly constant, and changes to the correlation matrix, therefore, should have the dominant effect on changes to the covariance matrix. This happens when the entries in the correlation matrix are relatively large compared to the entries in the covariance matrix and the scale factor is correspondingly large. For example, assuming an Exchange (0.5) structure would give correlation entries, by definition, of approximately 0.5. Thus, a typical covariance entry would be in the 10^{-4} range, say 0.0003 and the scale factor large, in this case around 1600.

5.2 CHOICE OF SAMPLE SIZE

Regarding the application of the W_R statistic (the modified W_i statistic) for the development of correlation control charts, the desired *ARL* is selected and the required sample size, *n*, is then calculated from the empirical equation for the moderate Exchange structure shown in equation 4-6, based on expectations of the other parameters. Like the similar approach of Lowry *et al.* (1992) for the MEWMA chart, the empirical equation is only considered as a general guideline, which may require some modification depending on the data under consideration. This method of setting up the control chart based on the desired *ARL* is also used by Crowder (1989) and Lucas and Saccucci (1990), (see also a summary in Lowry *et al.*, page 50), although tabulated data from simulations instead of an empirical regression equation is used as a basis in those examples.

Despite the desire to base sample size on desired *ARL*, the preferred multivariate sample size is sometimes stated to be three to four times the number of quality characteristics (Flury and Riedwyl, 1988). Duncan (1986) makes an argument for choosing control chart sample sizes of four or five whenever possible, as the chance of the change occurring during the sample taking procedure is minimized when the sample size itself is small. In other words, the effects of taking averages on the samples should not be allowed to mask changes to the cause system (p. 497). As the simulations for both the Exchange (0.5) correlation structure and the Independence Structure show, Figures 11 to 13, the *ARL*s decrease asymptotically with the sample size and the "better" values for *ARL* are often associated with smaller sample sizes, specifically when $n = p + 1$. While modern computers and data collection methods can often reduce the costs associated with sample gathering, it is recommended that the sample sizes remain smaller to take advantage of larger *ARL*s (recognizing that these will be specified for the process under consideration), and the automation be used to collect more frequent, instead of larger, samples. This recommendation is captured using $n = p + 1$ and also lends itself to application of tests for special causes which will be mentioned in Section 5.3.

5.3 DETECTING SHIFTS WITH THE CORRELATION CONTROL CHART

As MSPC charts for detecting shifts in the dispersion vector are an emerging area of research, it is insightful to examine the recommendations for various MSPC charts that are employed to detect shifts in the location vector. While the Hotelling T^2 chart is the basic multivariate location chart, additional statistics and charts have been developed to improve upon the detection capabilities that the Hotelling T^2 chart can provide. Some of these are modified versions of the Hotelling T^2 chart, while others are multivariate extensions of univariate techniques. In both cases, it is recognized that the presence of correlation, captured by the covariance matrix, Σ, may be present (see Chapter 2).

In striving to detect gradual trends away from in-control conditions using specific types of charts, some of the assumptions underlying the Hotelling T^2 chart are relaxed, such as assuming that all variables change simultaneously following a normal distribution (Hawkins, 1991). This approach allows the separation of the location and dispersion components which can then be evaluated independently. Most of the research to date has concentrated on the location vector once the dispersion effects are identified or quantified. From the decomposition, it follows that many of the same techniques applicable to the location vector can also be considered for the dispersion scenarios.

First, a shift may be present in multivariate data, due to inertia, even if an out-of-control point has not yet been detected. This can occur if two consecutive points are calculated from vectors in opposing directions. The application of some Shewhart tests for special causes to the chart can be used to find these shifts (see Nelson, 1984). Lowry *et al.* (1992) discuss this approach with regard to their MEWMA charts as well as noting that Crossier (1988) suggested a similar approach for the multivariate CUSUM chart (p. 51-52). This approach will be advocated for the correlation control charts since these tests for special causes do not rely on specification of the underlying distribution.

Second, it is good practice to use additional control charts when attempting to detect shifts. One of the more common examples is the use of univariate charts to notice a shift due to correlation that may not be detected by the single statistic of a multivariate chart. In the case of the correlation control chart, it is the shift in the correlation control chart that is of interest, but it is nonetheless prudent to employ a second chart, perhaps a MEWMA chart or Hotelling T^2 chart. Recall, however, that the correlation control chart assumes that the correlation component of the covariance matrix is changing while the scale factor remains constant and this assumption should be verified whenever the data allows. If the scale factor is known to be constant and/or only the correlation data is available, the correlation control chart would, of necessity, have to be used in isolation.

In summary, it is desired to have data where the correlation matrix approximates the moderate Exchange structure, the scale factor is large in comparison to the covariance matrix entries, the number of quality characteristics does not exceed $p = 5$, and the sample size is chosen to be $n = p + 1$. Detecting changes using the correlation statistics should also employ tests for special causes, and the correlation control charts should be used in conjunction with other types of control charts such as the W_i chart for covariance or the Hotelling T^2 chart for location, similar to the companion univariate Shewhart X-bar and R charts.

5.4 THE ISSUE OF *ARL*

The logarithmic nature of the W_R statistic naturally presents some concerns because the range on the *ARL* for a given situation can become quite large. This issue is not unique to the correlation control charts discussed herein but is also present with similar location control charts that are discussed in Chapter 6. To some extent, this effect is mitigated by choosing a sample size $n = p + 1$ but, as Table 7 shows, the *ARL*s for various combinations of the Exchange (0.5) Structure, based on empirical regression equation 4-6, still have a wide range even when this advice is applied.

If it is assumed that an *ARL* of 370 is a reasonable benchmark, based on its use with univariate Shewhart charts, then Table 7 illustrates how the choice of sample size (n) may be help to mitigate the effects of the logarithmic nature of the *ARL* for the W_R statistic. Also, in view of this, the chart should be interpreted with the expectation of more Type I errors when the total system variation is large. Note that sample sizes of $p = 8$ were not used for computation of the empirical regression equation.

Table 7 Selected *ARL*s based on the Empirical Regression Equation

p	number of changes	variance	*ARL* for $n = p + 1$	*ARL* for $n = 3p$	*ARL* for $n = 4p$
2	1	0.25	59	42	33
2	1	0.20	275	196	157
2	1	0.15	1291	919	733
2	1	0.10	6046	4308	3436
3	1	0.25	84	48	34
3	1	0.20	392	223	159
3	1	0.15	1839	1045	745
3	1	0.10	8614	4896	3488
3	2	0.25	37	21	15
3	2	0.20	175	99	71
3	2	0.15	818	465	331
3	2	0.10	3832	2178	1552
3	3	0.25	17	9	7
3	3	0.20	78	44	31
3	3	0.15	364	207	147
3	3	0.10	1705	969	690
5	1	0.25	170	61	35
5	1	0.20	797	288	164
5	1	0.15	3732	1350	767
5	1	0.10	17487	6325	3595
5	2	0.25	76	27	16
5	2	0.20	354	128	73
5	2	0.15	1660	601	341
5	2	0.10	7779	2814	1599
5	3	0.25	34	12	7
5	3	0.20	158	57	32
5	3	0.15	739	267	152
5	3	0.10	3461	1252	711

5.5 SUMMARY

It is recommended to design the correlation control chart by choosing the sample size, n, based on the known and expected values of the other parameters of the empirical regression equation,

Equation 4-6, for the desired *ARL*, with $n = p + 1$ often giving the most optimal condition in the presence of large total expected system variation. The chart is then evaluated by looking for not only out-of-control signals but also for those tests for special causes that would be applicable to a chart without a centerline value and then determining the reasonableness of shifts knowing the expected *ARL*. Additional control charts for the process should be used in conjunction with the W_R correlation control chart, especially the covariance control chart based on the unmodified W_i statistic, whenever the data required to construct them is available. Chapter 7 will illustrate these issues by applying the W_R statistic to three data sets. First, Chapter 6 will compare the W_R statistic to other MSPC approaches.

6.0 COMPARISON WITH OTHER MSQC TECHNIQUES

The correlation control charts developed herein are based on W_R, a modified version of Alt's (1985) $\boldsymbol{W_i}$ statistic for covariance control charts. Since the $|\boldsymbol{S}|$ statistic has been shown to have drawbacks for control charting (see discussion in Section 2.3.1) and the G statistic is designed to detect gradual shifts instead of instantaneous ones, comparisons of either to the W_i statistic in terms of *ARL* performance have not been undertaken (see Section 2.3.3). As a result and due to the differences between correlation and covariance, a one-to-one benchmark for the performance of the developed correlation control charts does not exist as such. Significantly more attention has been devoted to the evaluation of multivariate measures of location than to those for dispersion and, compared with these, the behavior of the W_R statistic can be characterized as typical, as can the $\boldsymbol{W_i}$ statistic, by induction.

Consider, for example, the multivariate exponentially weighted moving average (MEWMA) chart developed by Lowry *et al.* (1992). The *ARL* of their MEWMA chart[2] is compared to that for Crossier's (1988) MCUSUM chart, Pignatiello and Runger's (1990) multivariate CUSUM (MC1) chart, and the Hotelling χ^2 chart. Several approaches are considered to make the comparison, with either the *ARL* and or the number of quality characteristics, *p*, being held constant. One of the key differences between the location charts

---

[2] Two versions of the MEWMA chart are considered—one using the actual covariance matrix and the other an asymptotic approximation.

reviewed by Lowry *et al*. and the dispersion charts considered herein (or elsewhere, for that matter), is the necessary consideration of the noncentrality parameter, λ, given by

$$\lambda = \left(\mu' \Sigma^{-1} \mu\right)^{1/2} \qquad (6\text{-}1),$$

where **μ** is the mean vector and Σ is the covariance matrix. With the *ARL* performance of the location charts considered by Lowry *et al*. shown to depend only on this parameter, a comparison amongst the location charts is thus possible (pp. 47-48, 51). In the case of dispersion charts, including the correlation control charts developed with the modified W_i statistic, the shift one is interested in detecting is a part of the noncentrality parameter, meaning that comparison of the dispersion chart *ARL*s with those of the location charts is not wholly commensurate. Nonetheless, a review of the *ARL*s for the location charts, noting the effect of the noncentrality parameter, indicates that the behavior of the dispersion charts can be considered typical for multivariate control charts in general.

Since the covariance matrix is composed of a scale factor and a correlation matrix, related by the equation

$$|\Sigma| = \left(\sigma_{11}\sigma_{22}\cdots\sigma_{pp}\right)|\rho| \qquad (6\text{-}2),$$

where ρ is the correlation matrix and the product of the σ_{ii} individual covariances on the matrix diagonal form the scale factor, one can see that changes in either component will affect the value of the noncentrality parameter, λ. If it is assumed that the scale factor remains constant then changes to the correlation matrix invoke changes in the covariance matrix, Σ, and therefore the noncentrality parameter.

For the location charts evaluated by Lowry *et al*., with the number of quality characteristics, *p*, held constant, the *ARL*s decreased when the noncentrality parameter, λ, was increased. Thus, as the amount of dispersion increased, the *ARL*s of the location charts decreased. Comparatively, this same phenomenon was exhibited by the W_i statistic modified to the W_R statistic to address shifts in the correlation matrix. With the number of quality characteristics, *p*, held constant, the *ARL*s for the correlation control charts decreased when the variance of the deviates applied to the in-control correlation matrix were increased. The same effect occurred when the variance was held constant but the number of entries in the correlation matrix to which the deviates was applied was increased (refer to Sections 4.4 and 4.5).

Another comparison can be made regarding the ranges of the *ARL*s produced by the location control charts described in Lowry *et al*. and those of the W_R correlation control charts based on the modified W_i statistic. The same comparison can be made with data from Hawkins (1991). Both methods were proposed to be more sensitive to shifts in the location vector than the Hotelling T^2 approach whereas the correlation control chart is highly sensitive to small shifts, due partially to the small range for the determinant of the correlation matrix that is employed in the calculations. Thus, a comparison with the Hotelling T^2 chart would not be appropriate. Again, while is not possible to make a one-to-one comparison between the *ARL* performance of the location charts and the dispersion charts, the behaviors of both show them to be typical of multivariate control charts in general. In particular, the logarithm relationship of the MEWMA chart described by Lowry *et al*. is of interest. The MEWMA extends the univariate EWMA case by defining vectors of the form

$$Z_i = RX_i + (I - R)Z_{i-1} \quad (6\text{-}3),$$

where *i* is an index,

$$Z_0 = 0 \quad (6\text{-}4),$$

and

$$R = diag(r_1, r_2, \ldots, r_p) \quad (6\text{-}5)$$

for $0 < r_j \leq 1$, j = 1, 2, . . . , *p*.

The MEWMA chart signals an out-of-control condition when

$$T_i^2 = Z_i^{'}\Sigma_{Z_i}^{-1}Z_i > h \qquad (6\text{-}6),$$

where *h* is chosen to achieve a specified *ARL*. The covariance matrix, $\Sigma_{Z_i}^{-1}$, is derived from $\boldsymbol{Z}_i$ (p. 48). The simulations showed that the logarithm of the in-control *ARL* is nearly a linear function of *h*, and this is used to approximate appropriate control limits for other desired *ARL*s using the MEWMA approach. Similarly, the empirical regression equation (discussed in Section 4.6),

$$\ln(ARL) = 12.011 + 0.467p - 0.810 * changes - 0.113n - 30.388 * \text{var}iance \qquad (6\text{-}7),$$

developed from the simulations for the Exchange (0.5) correlation structure also exhibits a logarithmic form that can be used to approximate appropriate parameters, such as sample size, *n*, for a desired *ARL*, given a known number of quality characteristics, *p*, expected number of *changes*, and expected amount of change to the correlation matrix, *variance*.

Furthermore, the numeric range for the expected *ARL*s of the MEWMA, CUSUM and Hotelling χ^2 charts, holding various parameters constant and changing others, is similar to that found for the W_R statistic used to detect shifts in the correlation matrix. For example, with h = 14.78, r = 0.10, and p = 3 quality characteristics, the MEWMA *ARL* ranges from 1007 for a centrality parameter λ = 0.0 down to 3.75 for a for a centrality parameter λ = 3.0 (Lowry *et al.*, Table 3, p. 49). For the correlation control chart, with *p* = 3 quality characteristics, Charts 14-16 show similar ranges and patterns. In both cases, more variation translates to a lower *ARL* in logarithmic fashion. Note that Lowry *et al*. (1992) suggest an *ARL* of 200, an *ARL* reasonably achieved in most correlation control chart applications when the sample size, *n*, is twice to three times the number of quality characteristics, *p.* While this may hold for the location vector, the *ARL* for the correlation control chart is more reasonable when the sample size is closer to the number of quality characteristics (i.e. when $n = p + 1$), as illustrated by viewing the entries in Table 7 for various combinations of the parameters.

Hawkins (1991) compares five location charts using a regression-adjusted variables approach. These include Crosier's (1988) CUSUM of *T*, Crossier's Multivariate CUSUM, Woodall and Ncube's (1985) MCUSUM applied to *X*, Woodall and Ncube's (1985) MCUSUM applied to *Z* as a transform of *X*, and a Euclidean norm of the *Z* CUSUMs for $p = 5$ quality characteristics (note that two of the scenarios utilize variants of the same Crossier CUSUM chart considered by Lowry *et al*.). With multiple changes applied to the in-control condition of a $p = 5$ quality characteristics scenario, *ARL*s are in a range that rarely exceeds 100 and, for many of the approaches, are less than ten. While Hawkins's point is to illustrate the superiority of his regression-adjusted method that provides a more concise look than that which would be obtained from global signals by essentially decoupling the variables so that the shift in each can be considered individually, the same patterns and similar ranges emerge as those seen with the correlation control charts as evidenced by the data for the moderate Exchange Structure case.

In summary, while there are no one-to-one comparisons for the performance of the correlation control chart performance, other control chart statistics have been similarly derived. Those with logarithmic calculations produce *ARL*s in ranges that are commensurate with that for the correlation statistic, W_R.

7.0 DEMONSTRATION WITH DATA SETS

Three data sets are considered to illustrate the application of correlation control charts and the disparity between theoretical validity and practical applications. The first involves a manipulation of the Flury-Riedwyl (1988) data set used by Hawkins (1991) and considers relative dimensions on switch drums. The second example uses a data set from Johnson & Wichern (1988) for stock return data based on the work of King (1966). The third example uses data from a distillation column simulator developed at the University of Delaware (Doyle, Gatzke and Parker, 1999). Only the W_R statistic (the modified W_i statistic) is considered for these cases, as the $|\boldsymbol{R}|$ statistic was shown in Chapter 4 to perform poorly in the simulation study. These data sets demonstrate that the W_R statistic must be applied with caution, as it does not offer reasonable statistical process control monitoring in all situations. Recall that it is desired to have data where the correlation matrix approximates the moderate Exchange structure, the scale factor is large in comparison to the covariance matrix entries, the number of quality characteristics does not exceed $p = 5$, and the sample size is chosen to be $n = p + 1$. Three data sets are considered in the following sections, with the basic characteristics listed in Table 8.

Table 8 Data Sets for Testing WR Correlation Statistic

Data Set	Characteristics	Sample Size	Variation
Flury-Riedwyl	$p = 5$	n = 6	Moderate
King	$p = 2$	$n = 3$	Low
	$p = 3$	$n = 4$	Large
	$p = 5$	$n = 6$	Large
Doyle, Gatzke & Parker	$p = 2$	$n = 3$	Low

7.1 FLURY-RIEDWYL DATA SET

A simulated data set, based on data from Flury and Riedwyl (1988), was generated by Hawkins (1991) to compare several statistics for MSQC of the location vector. The original data represents the dimensions of a switch drum where X_1 is the inside diameter of the drum and $X_2, \ldots, X_5$ are the head-to-edge distances of four sectors cut into the drum. Hawkins uses the sample mean and covariance to be the true mean and covariance and then generates fifty data points based on a multinormal distribution about the center. After point 35, a quarter of a standard

deviation shift was introduced into X_5, chosen specifically by Hawkins because it had neither weak nor strong correlation with the other variables. Also, after point 35, the marginal standard deviation of X_1 was increased by 50%. While Hawkins's purpose was evaluation of mean vector statistics, it is evident that his approach to the Flury-Riedwyl data provides a data set for $p = 5$ quality characteristics that includes a shift in the correlation between X_1 and X_5 after point 35.

The W_R Statistic behaves predictably for the Flury-Riedwyl / Hawkins data set based on the simulation results of Section 4.4. Since this data set contained 50 points, it was divided into samples of size $n = 6$ with the last sample under the value of n dropped from the analysis. The eight samples that were generated were used to calculate the values of the covariance statistic, W_{i}, and the correlation statistic, W_{R}, and these are shown in Table 8. Hawkins (1991) objective was to develop control charts that were more sensitive to shifts in the mean vector than the traditional T^2 chart and, as a result, the traditional T^2 chart is not capable of detecting the shifts that were introduced after point 35. However, since the shifts were created by manipulating the covariance, the W_i Statistic and the proposed W_R correlation statistic would be expected to indicate a shift on their respective control charts. Referring to Table 8 and also to Figure 18, the sample containing point 35 and its associated shift do indicate the out-of-control condition with the upper control limit of 34.39 exceeded for both. Note that while the covariance statistic slightly exceeds the upper control limit for samples two and three, the correlation statistic does not, meaning that the scale factor is the likely cause for a shift that barely exceeds the control limit.

Table 9 Flury-Riedwyl Data Set Results

Sample Number	W_i Statistic	W_R Statistic
1	23.72	9.48
2	36.64	33.08
3	34.56	26.61
4	22.00	22.80
5	19.23	13.69
6	78.88	100.19
7	11.87	9.11
8	32.18	22.09

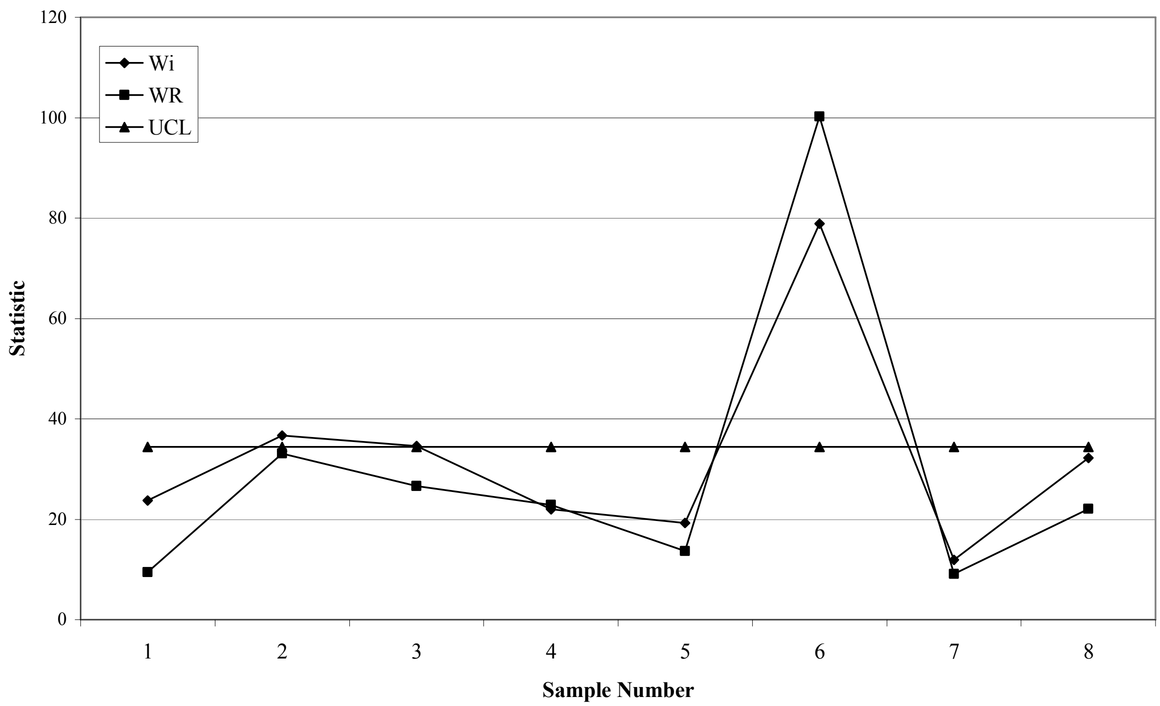


Figure 18 Correlation Control Chart for Flury-Riedwyl / Hawkins (1991) Data Set

7.2 FINANCIAL DATA FROM KING (1966) AND JOHNSON & WICHERN (1988)

Another example of the application of the correlation control chart is provided by financial data that was originally approached using principal components analysis by King (1966) and considers the observed weekly rates of return for five stocks (a simplified example is given in Johnson and Wichern, p. 376-377). Three of the stocks are from chemical companies and two from petroleum companies. King's analysis revealed two major components to the changes in the returns: the *market* component affected all stocks generally whereas the *industry* component noted a contrast between the chemical and petroleum company stocks. The correlation control chart is applied to the correlation matrix of the observed weekly returns either to detect either a market or an industry shift. In practice, it is more likely to have one or two correlations changing more severely than to have all the correlation relationships shifting simultaneously (Hawkins, 1991) and these are the type of shifts that are assumed to be indicators of out-of-control conditions.

Three separate approaches were used to analyze the weekly stock return data, illustrating the effect of the data set itself on the practicality of applying the W_R statistic. The first approach considered the correlation matrix for all five stocks: Allied Chemical, DuPont, Union Carbide, Exxon, and Texaco. The second approach considered the correlation matrix for only the three chemical companies. The third approach considered only the correlation matrix for the two petroleum companies. In all cases, the sample size, n, was chosen to be the number of characteristics (the stocks) plus one, or $n = p + 1$, to keep the *ARL* as high as possible in view of the presence of much expected variation. For each case, both the W_R statistic for correlation and the unmodified W_i statistic, for covariance, were calculated and plotted. The data set consisted of 100 data points, or $100/n$ samples of size n. Thus, each sample group roughly represents one "month" of return data and the control charts generated are plotted as Figures 19 through 21.

King's (1966) analysis showed, using principal components analysis, that the stocks of various industries moved by primarily two forces. Primarily, stock returns moved with the stock market as a whole, and these were called market effects. Secondarily, stock returns moved within industries such as chemical and petroleum, or moved relative to other industries, such as contrasting chemical with petroleum. These were called industry effects (see Johnson and

Wichern, p. 376-377 for additional interpretation). Both types of changes are germane to the correlation control chart, as the monitoring of these effects lend insight into the movement of the stock prices.

7.2.1 All Stocks

The first analysis used the in-control correlation and covariance matrices for all five stocks, which has a moderately correlated Exchange Structure with the covariance matrices having small but non-zero entries. Referring back to Table 7, based on the empirical regression equation, for five characteristics, a sample size of six, three expected changes to the matrix entries, and changes of variance 0.25 or greater, the *ARL* will be, at best, approximately 34. Examining Figure 19, the first half of the groups exhibit a lack of control, with the covariance coming into control after point seven. The correlation, while not in control for much of the chart, does track with the covariance from point eight until the end. Although not it is not meant to be a rigorous statement, the scale factor remains relatively constant from point nine forward. Note that the chart shows the correlation and covariance statistics, not the correlation and covariance themselves, so that the gap between the two curves does not have a one-to-one representation of the scale factor. Considering the small expected *ARL* for this chart, the two points of largest concern are points two and seven, both of which show a correlation statistic that exceeds the covariance statistic, meaning that the scale factor assumption has been violated. In consideration of the value of $p = 5$ and the relatively large amount of expected variation, construction of this chart is not of much value, and it is included to illustrate this.

Table 10 Statistics Calculated for All Stocks

Sample Number	W_i Statistic	W_R Statistic
1	174.60	115.21
2	47.03	110.78
3	28.32	-16.05
4	164.94	173.83
5	99.70	107.87
6	-78.35	39.25
7	84.65	24.19
8	3.95	131.15
9	-15.59	55.56
10	-14.37	50.63
11	-18.10	46.37
12	-117.07	39.79
13	-69.39	76.64
14	-122.06	56.02
15	-42.10	30.95
16	-73.27	6.19

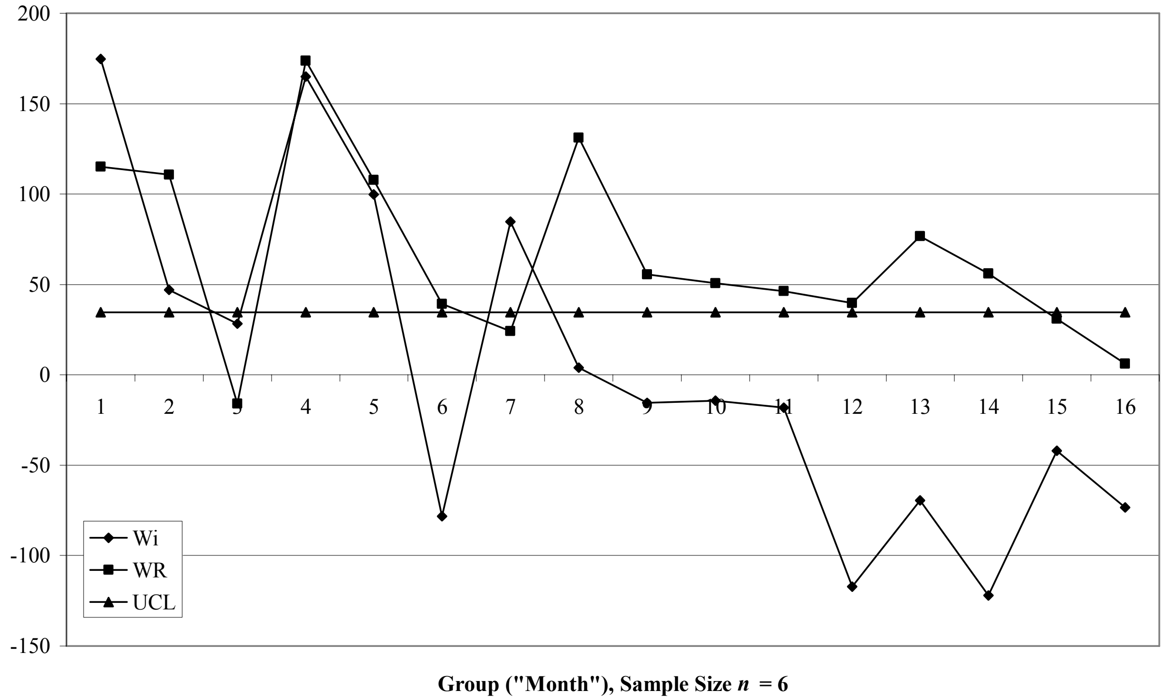


Figure 19 Correlation Control Chart for Allied Chemical, DuPont, Union Carbide, Exxon, and Texaco Weekly Stock Returns (Johnson & Wichern, 2002, Table 8.4)

7.2.2 Chemical Stocks

The data for only the chemical company stocks (Allied Chemical, DuPont, and Union Carbide) is a better representation of the limited utility of the correlation control chart and is shown in Table 11 and Figure 20. Since the matrices for the chemical stocks case are submatrices of the overall case, the same characteristics of Exchange Structure and small, non-zero covariances apply. In some respects, this chart represents a "worst-case" scenario since the diverse product mix of the three companies coupled with general volatility of stock price within a market segment leads to the expectation of out-of-control conditions on a more frequent basis than may be expected with companies producing a more homogenous product mix.

Referring back to Table 7, based on the empirical regression equation, for three characteristics, a sample size of four, three expected changes to the matrix entries, and changes of variance 0.25 or greater, the *ARL* will be, at best, approximately 17, so at least one Type I error can be expected from this data set. With the exception of points two and ten, the covariance remains in control for all sample groups so that the effect of scale factor, in those cases where the correlation statistic is out-of-control, is to bring the covariance back into control.

If the covariance statistic was not available, and since the correlation matrix is the dominant feature, those sample groups where a shift is indicated would be reviewed to determine possible causes based on the correlation statistic itself. However, since both the covariance and correlation statistics are available and plotted, only point two would qualify as signaling an out-of-control condition for both covariance and correlation. If the correlation matrix underlying this sample group is viewed, it can be seen that the returns for Allied Chemical have increased compared to DuPont and Union Carbide for that sample group. Another point on the chart may be of interest, between samples nine and ten where the covariance is in-control and the correlation is out-of-control at point nine and then the situation reverses at point ten. In all likelihood, the correlation, as a component of covariance, will not be out-of-control if the covariance is not. Thus, point ten shows an out-of-control condition where the scale factor is a more likely cause than the correlation matrix.

Table 11 Statistics Calculated for Chemical Stocks

Sample Number	W_i Statistic	W_R Statistic
1	8.10	-0.17
2	34.64	21.08
3	-57.03	-16.88
4	-24.86	-13.02
5	-29.23	-0.01
6	-10.86	10.27
7	-28.41	-12.31
8	-30.82	-14.35
9	5.31	34.30
10	40.19	7.08
11	-94.67	-37.82
12	-13.99	26.47
13	-18.15	-18.91
14	9.55	20.38
15	-37.50	-21.93
16	-29.77	33.78
17	-1.66	5.40
18	-57.64	19.23
19	-34.60	-43.26
20	-34.66	39.62
21	-55.38	-20.20
22	-31.91	14.24
23	8.89	13.56
24	-1.52	28.40
25	6.05	42.45

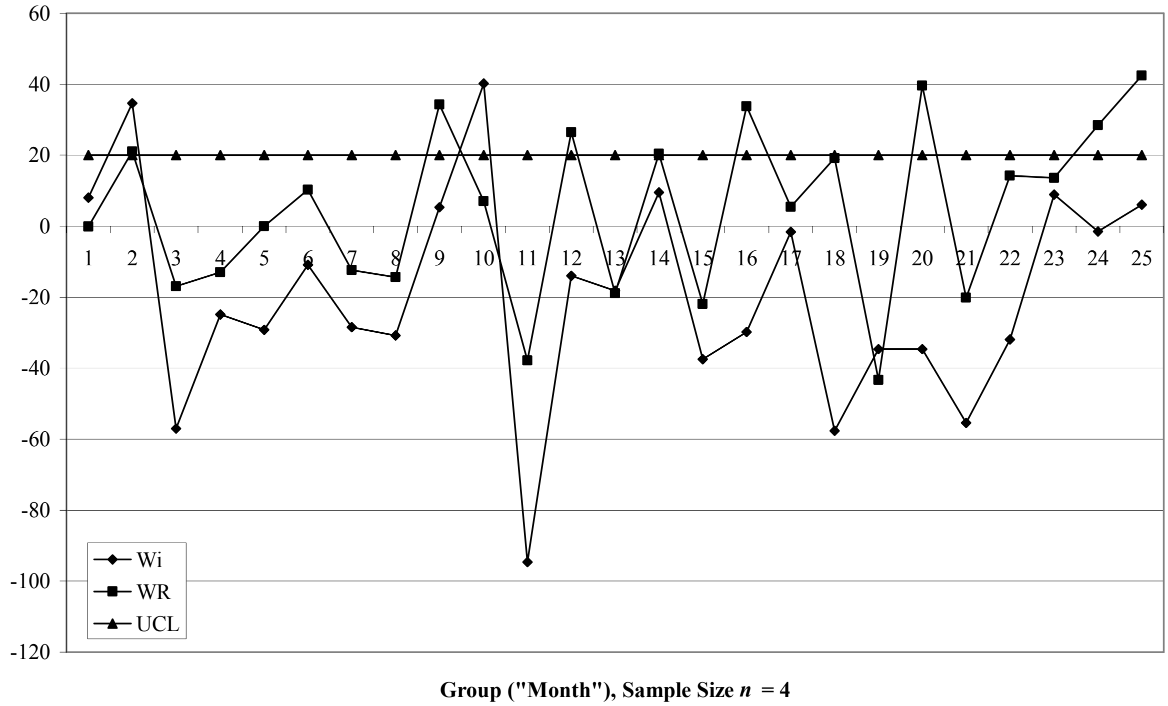


Figure 20 Correlation Control Chart for Allied Chemical, DuPont, and Union Carbide Weekly Stock Returns (Johnson & Wichern, 2002, Table 8.4)

7.2.3 Petroleum Stocks

Figure 21 shows the control charts for only the petroleum stocks (Exxon and Texaco), with data in Table 12. Using Table 7 as a reference, based on the empirical regression equation, for two characteristics, a sample size of three, only one possible change to the matrix entries, and changes of variance 0.25 or greater, the *ARL* will be, at best, approximately 59. As the product mix of the oil companies is more homogenous than that for the chemical companies, less volatility is expected in the correlation between weekly stock returns.

Other than the first point, for which the data reveals a zero entry that explains why the covariance statistic is out-of-control, both the covariance and correlation statistics remain in-control for all samples. Furthermore, tests for specific causes (Nelson, 1984) applied to both charts do not reveal any suspicious patterns.

Table 12 Statistics Calculated for Petroleum Stocks

Sample Number	W_i Statistic	W_R Statistic
1	16.20	7.49
2	-2.50	-8.34
3	-19.94	7.09
4	-7.25	-0.01
5	-13.51	5.34
6	13.74	7.44
7	-16.62	-33.65
8	8.61	8.99
9	-13.16	-22.73
10	-29.51	5.04
11	-31.42	-12.04
12	-26.01	-21.78
13	-12.86	-16.19
14	-69.56	-56.14
15	-6.69	9.82
16	-6.32	9.84
17	-25.21	-13.37
18	-32.83	-2.39
19	-14.94	-13.61
20	-15.29	9.14
21	-11.39	9.42
22	5.45	1.60
23	-7.54	7.54
24	-13.73	10.00
25	-37.39	9.30
26	-18.29	9.60
27	-26.79	-5.07
28	-30.93	-16.49
29	-50.69	-13.93
30	-10.92	10.04
31	-19.38	3.05
32	-35.76	-4.23
33	-22.57	8.04

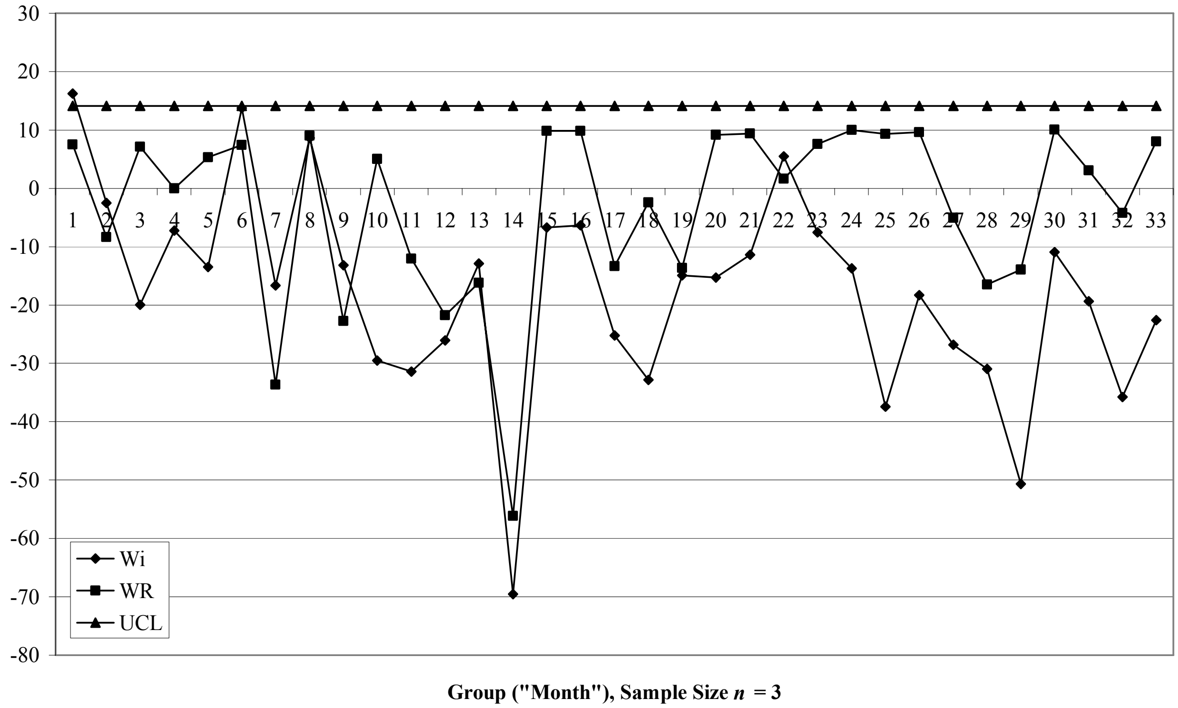


Figure 21 Correlation Control Chart for Exxon, and Texaco Weekly Stock Returns (Johnson & Wichern, 2002, Table 8.4)

7.3 DISTILLATION COLUMN (DOYLE, GATZKE & PARKER, 1999)

While the smaller data sets considered in Sections 7.1-7.2 provided some encouraging results, it was desired to apply the W_R statistic to a larger data set developed from a real or (validated) simulated process to assess performance under conditions that could be expected in industrial settings. As previously noted, chemical processes often produce data for which the correlation matrix remains nearly constant. As an example consider a distillation column where the stoichiometric relationship between outputs is expected to remain in control unless a process disruption occurs at the column input.

The Department of Chemical Engineering at the University of Delaware developed, under direction of F. J. Doyle III, a software-based simulator for both an annealing furnace and a distillation column (1999). The modules of this simulator, and the accompanying text, are used for the instruction of process control, primarily at the undergraduate level. The distillation column was originally developed by K. Weischedel and T. J. McAvoy and has been validated and used by several researchers since its introduction in 1980. The simulator models a 27-tray column into which a 50%-50% ethanol-methanol mixture is fed at the 14th tray with the objective of producing 85% methanol and ethanol output streams at the top and bottom, respectively. There are four column inputs and four column outputs, as shown in Table 13. For purposes of research into the behavior of the W_R statistic, the steady-state analysis section of the distillation column text applies. The correlation between Overhead and Bottom MeOH Composition was considered for testing plausibility of the W_R statistic.

As the objective of using the column model was to evaluate the efficacy of the W_R statistic for MSPC monitoring of a chemical process, not an analysis of the process itself, only a brief summary of the employment of the model will be provided before considering the results obtained.

Table 13 Inputs and Outputs for Distillation Column Model

Inputs	Outputs
Feed Flow Rate	Overhead MeOH Composition
Feed MeOH Composition	Overhead MeOH Flow Rate
Vapor Flow Rate	Bottom MeOH Composition
Reflux Ratio	Bottom MeOH Flow Rate

7.3.1 Creation of Test Data Set from Column Simulator

For steady state operation, the column is started with initial input values of 0.025 for the Feed Flow Rate, 0.5 for the Feed MeOH Composition, 0.033 for the Vapor Flow Rate, and 1.75 for the Reflux Ratio. These values can be changed as the simulator runs, and the responses in the output are visible on four continuously updating screens. With minor programming, the data underlying the graphic responses are available in tabular form along with the value of the clock which runs throughout the simulation. For purposes of testing the W_R statistic, the values of the Overhead and Bottom MeOH Composition were captured along with the clock.[3] The actual output data appears as Appendix G, and is graphed in Figure 22. Essentially, the input value of the Feed MeOH Composition was changed from 0.50 to 0.33 somewhere prior to clock number 50 shown on the x-axis. The curves of Figure 37 show the resultant change to the Overhead and Bottom MeOH compositions at this transition point. Thus, up to time zero (clock number zero), the in-control conditions are represented. A large sample of data leading up to time zero was used to establish the in-control correlation matrix, in-control covariance matrix, and in-control mean vector for the process. Statistical process control was then attempted with the data shown in Figure 22 and Appendix G. It was hypothesized that the W_R statistic would detect the transition and then continue to show an out-of-control condition beyond time clock 145 where the process had stabilized at a new point.

[3] Additional studies were performed on other combinations of inputs and outputs with similar results. The information presented here is that which was deemed most illustrative of the phenomena concerned.

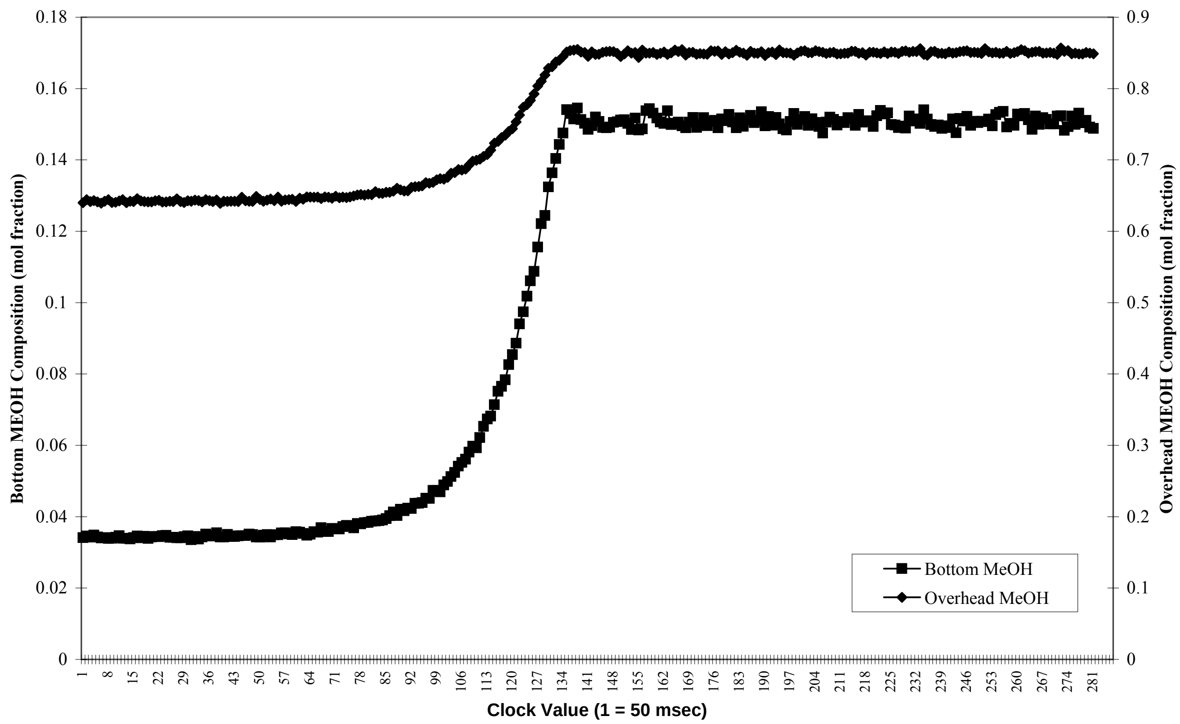


Figure 22 Relation of Overhead and Bottom MeOH Compositions at Transition Point

7.3.2 Analysis of the Statistical Process Control

Using the in-control conditions, a tolerance of $\alpha = 0.0027$, and a sample size of $n = p + 1 = 3$, based on the recommendations of Section 5.2, the Hotelling T^2 statistic for the mean vector, the W_i statistic for covariance, and the W_R statistic for correlation were calculated for each sample in order to create a control chart for the distillation column process. An abbreviated table of the calculated statistics is included as Table 14 (again, the raw data from the simulator is presented in Appendix G).

Table 14 Abbreviated Data from Distillation Column

Clock Start	Clock End	T^2 Statistic	W_i Statistic	W_R Statistic
1	3	19.40581	-24.69999	-20.56958
4	6	101.7249	-15.2121	-17.76114
7	9	16.2472	-7.061518	8.230153
…	…	…	…	…
46	48	123.3706	-21.83342	6.304022
49	51	21.62375	14.75854	6.157492
52	54	364.0311	-9.543662	-1.723563
55	57	452.3275	6.246241	0.0133
58	60	3.18E+04	-27.41963	-25.97202
61	63	157.2806	8.547046	6.384029
64	66	7.31E+03	-21.68386	6.233781
67	69	7.91E+03	-3.855103	-10.01732
70	72	1.31E+04	-22.63011	7.363401
73	75	9.95E+03	-27.49375	4.846569
76	78	605.0046	-2.12742	-8.349874
79	81	2.34E+03	-15.67689	7.289592
82	84	2.61E+04	-18.72286	3.073323
85	87	2.15E+03	12.38127	-2.610835
88	90	3.33E+03	33.82547	2.744224
91	93	487.2216	32.69232	4.543911
94	96	2.44E+04	-2.961295	-25.28779
97	99	563.8493	50.90036	3.049007
…	…	…	…	…
256	258	1.82E+05	139.7199	7.300016
259	261	4.28E+04	76.7424	3.192868
262	264	7.14E+04	128.1519	5.572952
265	267	2.90E+05	48.37465	7.551111
268	270	1.04E+06	7.642254	8.350875
271	273	1.88E+04	165.1768	7.94836
274	276	2.79E+05	82.3648	3.863147
277	279	2.13E+05	62.11641	8.14695

Two MSPC charts were created from the statistical data. The first chart, included here as Figure 23, is the classical Hotelling T^2 chart used for monitoring the location via the mean vector. It is evident that the Hotelling T^2 chart falls short of being able to properly detect shifts in the mean vector for this process. The upper control limit has a value of 12.40, which is exceeded in almost every case.

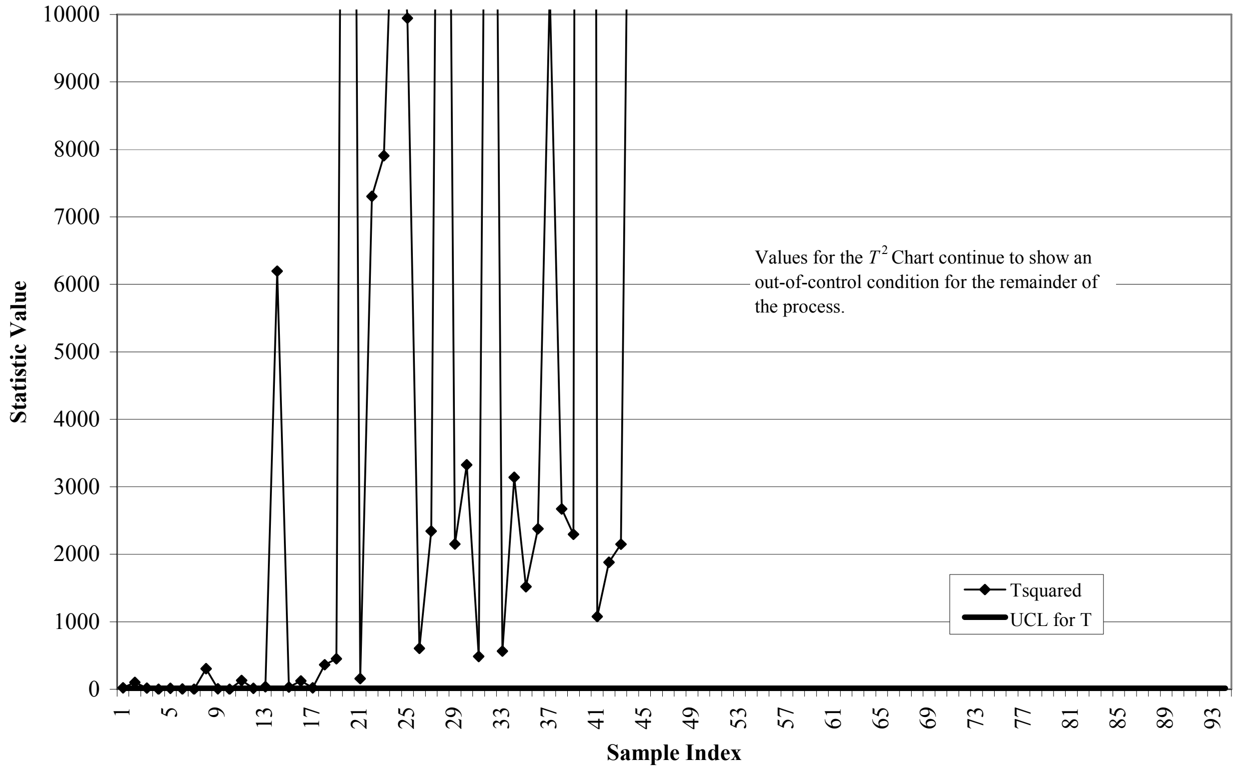


Figure 23 Hotelling T^2 Chart for Distillation Column

Halfway through the data set, when the transition occurs, the T^2 statistic exceeds 10,000 or 10^3 times the control limit, and is no longer plotted on the chart. While the correlation dispersion statistics are of primary consideration, the location statistic is included to illustrate that it is incapable of detecting the shift.

Looking next at the dispersion control chart for the same process and data, Figure 24, it is obvious that there are issues with the W_i statistic and W_R statistic as well. Of the two, the W_i (covariance) statistic appears to perform better than the W_R (correlation) statistic.

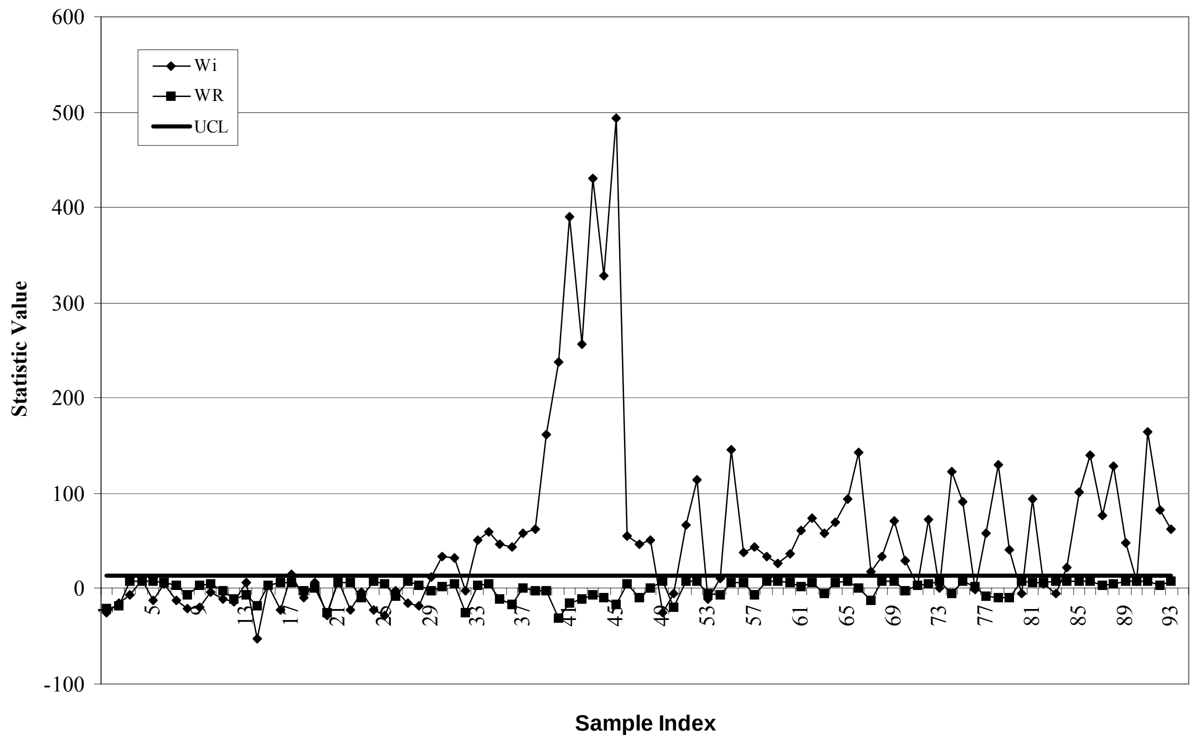


Figure 24 Dispersion Control Chart for Distillation Column Simulator

In the dispersion control chart, the W_i statistic shows, with the process having an out-of-control condition near the transition point (approximately sample index 33). The system remains out-of-control beyond the transition point but, although the statistic is out-of-control beyond sample index 52, the process itself has stabilized with regard to the new output concentrations. The W_R statistic, however, never indicates an out-of-control condition, meaning that the correlation has—according to the W_R statistic—remained in control throughout this period of the process. Unfortunately, the data could not be perturbed to show otherwise. In fact, various changes were made to the output data of Figure 22 in the range between clock number ten and 25. The W_R statistic *never* was able to detect an out-of-control condition.

7.4 EVALUATION OF THE W_R STATISTIC

What does application of these various MSPC statistics to the various data sets indicate? First, it confirms the conceptual idea that, if the correlation (or other dispersion measure) is of primary interest then the mean vector should not be the statistic that is monitored. Specifically, if the dispersion is the variable of interest, then a location MSPC chart such as the Hotelling T^2 may not be appropriate. Second, it confirms the efficacy of Alt's (1973) W_i statistic, at least in some cases, for the monitoring of covariance as the combined effect of both the correlation matrix and a scale factor. Third, the W_R statistic has limited applicability. The bounded nature of the correlation matrix which, unlike the covariance matrix, allows the behavior of its associated statistic to be characterized, severely limits its ability to indicate process changes even if correlation is the dominant component of covariance. Fourth, these insights indicate that the scale factor and correlation components of covariance should not always be separated when using MSPC statistics and charts to monitor for changes in dispersion.

The statistic appeared to perform, to some degree, for the Flury-Riedwyl and King data sets but did not perform at all for the Doyle, Gatzke & Parker data set. This is disturbing since both the first and last data sets shared a commonality inasmuch as the number of characteristics, sample size, and expected variation were similar. For the chemical distillation column, it is important to note that, by one definition, the column is always in control because of the stoiciometric relationship between the inputs and outputs. The amount of variation may have been masked by the randomness of the simulator in this case.

8.0 CONCLUSION

The research presented herein provides several contributions to the existing body of knowledge surrounding Multivariate Statistical Process Control Charts. Specifically, it adds to the neglected area of charting dispersion parameters that have evolved from research associated with the Hotelling T^2 location chart. While one contribution of this research is the assessment of current dispersion statistics ($|\boldsymbol{S}|$, W_i, and G) for the correlation matrix, another is the consideration of charting changes in the correlation matrix itself. The former aspect has implications directly to dispersion charts as well as for adding further understanding to Hotelling T^2 charts. The latter has additional applications beyond engineering in areas of business and biological science where correlated data is not only common, but expected. This chapter summarizes the major findings of this research.

8.1 EVALUATION OF THE COVARIANCE STATISTICS FOR THE SPECIAL CASE OF CORRELATION

Three statistics that were originally developed for the control charting of covariance were considered for the special case of the covariance matrix. The first, the $|\boldsymbol{S}|$ statistic developed by Wilks (1932), was modified to become the $|\boldsymbol{R}|$ statistic by isolating the scale factor. Through a mathematical analysis, it was demonstrated that the control limits for the $|\boldsymbol{R}|$ statistic were, in most cases, so wide as to be incapable of showing an out-of-control condition. A simulation model confirmed this finding. The second, the W_i statistic developed by Alt (1973), was modified to become the W_R statistic by isolating the scale factor. It was demonstrated both mathematically and geometrically that the correlation and scale factor components were independent and could be separated from the covariance. The preference of the W_R statistic for MSPC monitoring of the correlation matrix was confirmed through the simulation study for various levels of the number of quality characteristics and sample sizes at four levels of variance for two commonly occurring forms of correlation matrix structure. The third, the G statistic developed by Holmes and Mergen (1993) was shown mathematically to be infeasible for application to correlation matrices. The G statistic relies on the Mean Squared Successive Differences calculation of Hald (1952) to evaluate changes in covariance and an analog for correlation matrices could not be developed.

8.2 THE W_R STATISTIC

The W_R statistic is a modified form of the W_i statistic developed by Alt (1973) and was evaluated both mathematically and via a simulation study that employed three permutations to two standard correlation structure matrices for a variety of parameters, with the permutation scheme suggested by previous work in other areas of MSPC (Hawkins, 1991). The parameters were varied over common ranges of values that were also suggested by previous work in various areas of MSPC. While the W_R statistic appeared to be superior to the $|\boldsymbol{R}|$ statistic for the control charting of correlation matrices, it was also shown that certain assumptions and requirements are necessary for its application. The primary assumption is that the scale factor component of covariance is expected to remain constant while the changes to the process are attributed to the changes in the correlation matrix. In practice, this assumption should be recognized or verified by using the W_R chart in conjunction with the W_i chart whenever feasible. This approach is similar to verifying the normality of residuals after running a linear regression. The condition of “constant” scale factor and changing correlation matrix occurs naturally in some processes, to which the W_R chart can be applied. While quantification of “large” remains as future work, the W_R statistic is most suitable when the scale factor is “large” compared to the correlation component of the covariance matrix.

8.2.1 W_R Statistic for the Exchangeable Structure

The W_R statistic was evaluated for several permutations of the Exchangeable Correlation Structure for the case of moderate correlation to be consistent with previous work in other areas of MSPC. This case represents the situation most similar to the types of matrices that would be considered for monitoring in various process control settings. Performance of the statistic was judged via the Average Run Length (*ARL*) for changes over a range of parameter changes and the relative effect of each of these parameters to the expected *ARL* was evaluated through the development of an empirical regression equation. This equation demonstrated that the largest amount of influence on *ARL*, at a 30:1 ratio, was due to changes in the variance of the permutated entries in the matrix, which is the result that would be expected and preferred. As expected, increases in variance, the number of entries permutated, and the sample size all resulted in decreases to the *ARL*. The *ARL* was increased, albeit slightly, with increases in the number of quality characteristics. This is consistent with the behavior of multivariate location control charts.

One of the major issues found with the W_R statistic is that, unlike the W_i statistic from which it was developed, is that the size of the matrix has a major effect on the utility of the W_R statistic itself. Due to the bounded nature of the correlation matrix and the logarithmic nature of the calculation for the W_R statistic, the W_R chart works best to detect small changes; this is consistent with previous work in other areas of MSPC where logarithmic calculations are involved. Recommendations for the determination of sample size and chart interpretation were generated from these results, recognizing that the W_R statistic is more sensitive to small changes than the W_i statistic from which it was derived.

8.2.2 W_R Statistic for the Independence Structure

The W_R statistic was also evaluated for several permutations of the Independence Correlation Structure in which there is no correlation between any of the variables in the correlation matrix. This scenario would occur when a Principal Components Analysis, or similar procedure, had rotated the axes so as to render the variables independent of one another. The correlation control chart using the W_R statistic would be useful to detect any changes that may violate this assumed independence. Like the moderate Exchange Structure, increases in variance, the number of entries permutated, and the sample size all resulted in decreases to the *ARL* while the *ARL* was increased with increases in the number of quality characteristics. The range of utility was more severely limited in the Independence case versus the Exchangeable case, which was expected due to the logarithmic nature of the calculation and is consistent with previous work in other areas of MSPC.

8.3 APPLICATION RECOMMENDATIONS

Based on the results presented herein, the W_R statistic is recommended for the MSPC control charting of correlation matrices that follow a basic Exchangeable Structure with the entries in an approximate range (0.4 – 0.6) for $p = 5$ quality characteristics or less. The sample size is recommended to be $n = p + 1$ unless the amount of variation to the correlation matrix is known with certainty, in which case the empirical regression equation (4.6) can be used to estimate the sample size for the desired *ARL*. This approach is consistent with previous work in similar areas of MSPC and ensures the largest *ARL* considering the logarithmic nature of the calculation for the W_R statistic.

Even when these recommendations are followed, results from larger data sets are mixed—a phenomenon common to other types of MSPC techniques where the *ARL* is logarithmic with respect to the parameters. Partly, this can be attributed to the bounded nature of the correlation matrix. Also, all these statistics use a matrix determinant calculation which can be traced to Wilks (1932) but has been subsequently shown to have drawbacks as a single measure that cannot capture all the information contained by the matrix despite its claim to do so (Johnson & Wichern, 1988).

8.4 TESTS WITH THREE DATA SETS

Two of the three data sets tested with the W_R correlation statistic showed encouraging results, but the third—the distillation column of Section 7.3—yielded discouraging results. This indicates that application of the correlation control chart should be approached with caution, as the amount of correlation, and how much it is expected to change, are both factors that influence the capabilities of the chart.

8.5 CONTRIBUTION

Several new concepts have been introduced through this research. Starting with the separation of the covariance matrix into two components—the correlation matrix and a scale factor—to determine an out-of-control dispersion condition via matrix analysis. The potentially unbounded nature of the scale factor component encourages study of the correlation matrix which is bounded and resulted in the investigation of a method to perform MSPC on correlation matrices.

The W_R statistic, developed herein as a modified form of the W_i statistic, provides a method to analyze some correlation matrices "as-is." While it does not require additional control charts it could be used to supplement other control charts, similar to the use of R charts along with X-bar charts for the univariate case. The development of an empirical equation from a simulation study for the W_R correlation statistic supported the concept that the amount of variation introduced into the correlation matrix is the parameter that has the largest effect on the *ARL* for the process and that this factor outweighs the effects of choice of sample size and number of characteristics considered. Recommendations for limited use of the W_R statistic for the monitoring of correlation matrices via control charts were presented. Applications to data sets showed that the application range is narrow, commensurate with similarly derived MSPC statistics.

The mathematical and simulation demonstration that the proposed $|R|$ statistic is not suitable for control charting under many conditions is a non-trivial result since several dispersion measures in MSPC have been built around the determinant of the correlation matrix. A discussion and mathematical analysis of why the G statistic cannot be modified for application to correlation whereas the $|R|$ and W_i statistic can is explained by the origin of the G statistic being based on a mean squares successive differences calculation that has no analog with a correlation matrix. These analyses lend insight into the nature of correlation matrices with respect to an MSPC orientation.

An overall contribution is a demonstration that the correlation component of covariance may not be as influential as previously believed or, conversely, that the scale factor component of covariance is of primary influence on dispersion in many situations.

8.6 RECOMMENDATIONS FOR FUTURE WORK

During the research undertaken for this dissertation, several areas for future work have been identified. First would be the development of an algorithm to generate correlation matrix test data. That is, starting with an in-control correlation matrix, generating deviates of it, and then generating covariance matrices, mean vectors and, finally, raw data that can be used for testing when the correlation matrix is the item of interest. Second, building upon the development of the empirical regression equation for the Exchange Structure, it was determined that there is a need for a single measure to quantify the amount of variation in a multivariate process that incorporates the number of characteristics (p), the number of changes to matrix entries, and the variation of the expected changes. Third, although the W_R statistic can be implemented for small matrices, methods to monitor for changes in the correlation matrix when the number of characteristics (p) becomes large (> 5) are needed. Fourth, a chartable statistic developed from the methods of partial and multiple correlation (Appendix C) could be adapted for use in MSPC. Fifth, as with other statistics from the MSPC arena, the limits of applicability for the W_R statistic should be extended via future studies.

APPENDIX A

THE *M* CHART

The *M* chart (Hayter and Tsui, 1994) starts with the defining of a critical point noted by the authors as $C_{\mathbf{R},\alpha}$ that captures the simultaneous confidence intervals with respect to the correlation matrix, $\boldsymbol{R}$. The control limits are given by

$$\begin{cases} UCL = C_{R,\alpha} \\ LCL = 0 \end{cases} \qquad \text{(A-1).}$$

An out-of-control point is signaled when the calculated value of *M* exceeds the critical point, where *M* is given by the equation

$$M = \underset{1 \le i \le k}{Max}\left(\frac{x_i - \mu_i^{\ 0}}{\sigma_i}\right) \qquad \text{(A-2).}$$

.

In equation A-2, x_i and σ_i are the mean and standard deviation for the *i*th sample, and μ_i^0 is the symbol for the in-control mean at sample *i*. When $\mu = \mu^0$, the probability that each of the confidence intervals contains the value μ_i^0 is $1 - \alpha$. Although their critical regions differ, in comparing the *M* chart to the χ^2 chart (the χ^2 chart is used in place of the T^2 chart in this article because the covariance is assumed to be known), the authors note that both the χ^2 and M statistics control the overall error rate at exactly α, and usually reach the same conclusion. However, they note it is also possible for one statistic to trigger and the other not to trigger. The reason for this is shown geometrically and the overall conclusion is that neither statistic is superior in terms of relative power or relative sensitivity.

Historically, the χ^2 (or T^2) statistic is preferred because the required critical point is independent of correlation structure and is readily available. The developers note that the calculation of the critical point, $C_{\mathbf{R},\alpha}$, is now possible with modern computers. However, numerical integration techniques are not feasible, due to the dimensionality of the integration region, for quality characteristics greater than or equal to five, so extensive simulations are required.

APPENDIX B

THE *GS* CHART

To understand the *GS* chart, it helps to envision the simplest multivariate case where there are only two variables of interest and the variances are equal. Plotted on a Cartesian plane, the confidence intervals for the two variables would intersect to form a square. The critical region for the same variables on a Hotelling T^2 chart would be a circle. Houshmand *et al*. (1997) relate these regions using calculus, noting the instances where the square region can indicate an out-of-control point when the circle does not, and *vice versa*.

When correlation is introduced to this bivariate case, the square becomes a parallelogram, and the circle becomes an ellipse. Applying a Fisher-*Z* transform to relate the two cases, the procedure is then expanded to encompass the multivariate situation. Houshmand *et al.* (1997) use the resulting statistic to develop the *GS* chart, or "Generalized Shewhart" chart. They further demonstrate that the familiar *X*-bar chart can be a subset when the number of characteristics is given by $p = 1$. The procedure starts by assigning the desired overall Type I error of the process, α, and the individual Type I errors for each of the *i* characteristics, α_i. Then, for each quality characteristic $i = 1 \ldots p$ calculate the value

$$b_i = z_{\alpha i} \qquad \text{(B-1)}$$

where $z_{\alpha i}$ is obtained from the standard normal distribution.

The mean is calculated as

$$\overline{X} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p) \tag{B-2}$$

for $i = 1 \ldots p$ quality characteristics and the statistic

$$Z = (Z_1, Z_2, \ldots, Z_p)^T = \sqrt{n} R^{-1/2} (\overline{X}) \tag{B-3}$$

where n is the sample size, $R^{-1/2}$ is the inverse of the square root of the correlation matrix, T is the transpose operator, and

$$\overline{X}^T = \left[\frac{\overline{X}_1 - \mu_1}{\sigma_{11}}, \frac{\overline{X}_2 - \mu_2}{\sigma_{22}}, \ldots, \frac{\overline{X}_p - \mu_p}{\sigma_{pp}} \right] \tag{B-4}$$

In equation B-4, $\overline{X}_i$ is the mean value of the ith characteristic for a sample of size n, μ_i is the in-control value for that mean value, and σ_{ii} is the covariance of the ith characteristic.

Continuing, for each of the characteristics, $i = 1 \ldots p$, δ_i is calculated as

$$\delta_i = \frac{Z_i}{b_i} \tag{B-5}$$

for each of the Z_i and b_i as defined in equations B-1 and B-3 above.

The single statistic that is then calculated for the control chart is given by

$$\delta = \frac{1}{2} \left(p - \sum_{k=1}^{p} \frac{|1 - |\delta_k||}{1 - \delta_k} \right) \tag{B-6}$$

where k is an index for the $i = 1 \ldots p$ quality characteristics. If any of the $|\delta_k| < 1 \forall k = 1,2, \ldots p$, then $\delta = 0$. Otherwise, $\delta = r$ if r of the variables are out-of-control. Thus, the control limits cannot be expressed in the traditional manner with an upper and lower limit. Instead, if $\delta = 0$, the process is in-control; if $\delta > 0$, the process is out-of-control for $k = 1, 2, \ldots, p$.

APPENDIX C

METHODS OF PARTIAL AND MULTIPLE CORRELATION

The methods of partial correlation and multiple correlation attempt to directly detect changes in the correlation matrix but are not designed for control charts. For an example based on a trivariate normal population with a sample of size n = 10 and p = 3 quality characteristics, Golnabi and Houshmand (1997) present a table with nine parameters to be compared to two sets of control limits.

Both methods are algorithmic in nature. The method of partial correlations starts by calculating the covariance, ***S***, from a sample of size n. The standardized covariance matrix is then calculated as

$$S_S = R^{-1/2} S R^{-1/2} \qquad \text{(C-1)}$$

where $R^{-1/2}$ is the inverse of the square root of the correlation matrix. From S_S, all of the sample partial correlations are calculated and compared to a Beta distribution with parameters α = (n – k – 1) / 2 and β = ½, where n is the sample size, and k is an index value for the partial correlations

$$r_{12}, r_{13}, \ldots, r_{1p}, r_{23.1}, \cdots, r_{(p-1)p.12,\cdots(p-2)} \qquad \text{(C-2)}$$

of the p quality characteristics. Each sample variance, s_{ii} (also calculated from S_S , and where i is an index) is compared to a χ^2 distribution with (n-1) degrees of freedom. Comparing the calculated parameters to the control limits, it is possible to detect whether there has been a shift in a particular partial correlation (shown in equation C-2) or sample variance, s_{ii}.

The method of multiple correlations begins similarly to the method of partial correlations, differing after the standardized covariance matrix, S_S, has been calculated in equation 2-27. From the standardized covariance matrix, the sample variances, standardized sample correlation matrix, R_S, and the statistics $y_1, y_2, \ldots, y_{p-1}$ are extracted. The latter values are calculated by the formula

$$y_k = \sqrt{1 - \frac{|R_S|}{|R_k|}} \tag{C-3}$$

where R_k is the k by k upper left triangular matrix of the standardized sample correlation matrix R_S. Each of the y_k are calculated and compared to a Beta distribution with parameters $\alpha = (n - k - 1) / 2$ and $\beta = k/2$, where n is the sample size, and k is an index value. Each sample variance, s_{ii} is calculated in the same manner as the method of partial correlations and is compared to a χ^2 distribution with (n-1) degrees of freedom. Comparing the calculated parameters to the Beta distribution or χ^2 distribution, it is possible to detect whether there has been a shift in a particular multiple correlation or variance, respectively. For both methods, the calculated statistics are tabulated for comparison to tolerance limits; charts are not used. The number of calculations required for either method increases rapidly as the number of quality characteristics increases.

APPENDIX D

OUT-OF-CONTROL *ARL*

The use of in-control Average Run Length (*ARL*), as opposed to out-of-control *ARL* is predominant in the literature as a means to evaluate the performance of statistics proposed for statistical process control, whether univariate or multivariate.* The in-control *ARL* is the inverse probability of detecting a shift in the mean when the process is actually in-control and essentially denotes a "false alarm." On the other hand, the out-of-control *ARL* measures the inverse probability of not detecting a shift in the mean when the process is actually out-of-control (or a shift in the mean has occurred). To make an assessment of out-of-control *ARL*, it is necessary to define what is considered to be the out-of-control condition.

In the univariate case, for example, shifts in the mean vector can be defined by assigning a value that is known to be out-of-control. This information is then compared to an operating characteristic curve to get a probability that is then used to calculate the out-of-control *ARL*. This type of calculation is subsequently used to determine the appropriate sampling frequency.

There are a myriad of factors that could contribute to an out-of-control condition in the multivariate case and choice of the appropriate out-of-control condition as a start point for analysis is more complicated. When making comparisons amongst statistics, factors that shift the distribution of the data set are often a function of the parameters, such as sample size (*n*) and number of characteristics (*p*), and may not be uniformly applied to all the statistics under consideration for the comparison. Specifically, the value of *p* is not used in determination of the control limits for the |***S***| statistic, and the value of *n* is not used when the control limits for the W_i and *G* statistics are determined. As a result, investigation into the performance of the dispersion statistics using out-of-control *ARL* as a measure of performance is left as future work.

* When in-control ARL and out-of-control ARL are both used, the customary notation is ARL_0 for the in-control ARL and ARL_1 for the out-of-control ARL (Montgomery, 1997).

APPENDIX E

MATHCAD SIMULATION PROGRAM

The following is an example of the MathCAD 2001 code that was used for the simulation study described in Section 4, the data for which appears in Tables 15 and 16 of Appendix F. The example depicted here is for the Independence Structure with (p = 2) quality characteristics.

Declared Variables that are Global for Simulation (varied for the study):

These variables are entered prior to running the simulation.

$$\rho \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$n := 4$

$var := 0.1($

$\alpha := 0.0\text{\textquotedbl}$

$\gamma := 0.02$

Normal Deviate Generator:

Uses the built-in MathCAD random normal deviate generator to generate individual deviates, called by Randomize Code.

$$\mathrm{NORRN}(\mathrm{var}) := \mathrm{rnorm}(1, 0, \mathrm{var})$$

Randomize Code

Applies normal deviates of the specified variance (from the generator) to a form a deviation matrix that, when added to the in-control correlation matrix using the Sampler Code, produces a correlation matrix that is a normal deviate of the in-control condition.

```
Randomize(B, var) := | for  i ∈ 1.. rows(B)
                     |   for  j ∈ 1.. cols(B)
                     |     | R_(i,j) ← B_(i,j) + NORRN(var)_1
                     |     | R_(j,i) ← R_(i,j)
                     |     | R_(i,i) ← 1
                     | return  R
```

Sampler Routine:

Applies the normal deviates of the specified variance from the Randomize Code to the in-control correlation matrix to produce a correlation matrix that is a normal deviate of the in-control condition.

$$\text{Sampler}(n, B) := \left|\begin{array}{l} U \leftarrow B \\ \text{for } k \in (1..n) \\ \quad U \leftarrow U + \text{Randomize}(B, var) \\ \dfrac{U}{n+1} \end{array}\right.$$

SubRoutineR Code:

Compares a deviate of the correlation matrix to the control limits (UCLR, LCLR) until an out-of-control condition is detected, then returns the number of trials before this condition is reached (ARLR) and sets a flag to stop the iterations. A flag is also set if the value of ARLR exceeds 10000.

$$\text{SubRoutineR}(n, \rho, \alpha) := \left|\begin{array}{l} flag \leftarrow 0 \\ ARLR \leftarrow 0 \\ \text{while } flag \neq 1 \\ \quad \left|\begin{array}{l} R \leftarrow \text{Sampler}(n, \rho) \\ p \leftarrow \text{rows}(R) \\ UCLR \leftarrow |\rho| \cdot \left(1 + \left|\text{qnorm}\left(\dfrac{\alpha}{2}, 0, 1\right)\right| \cdot \sqrt{\dfrac{2 \cdot p}{n-1}}\right) \\ LCLR \leftarrow |\rho| \cdot \left(1 - \left|\text{qnorm}\left(\dfrac{\alpha}{2}, 0, 1\right)\right| \cdot \sqrt{\dfrac{2 \cdot p}{n-1}}\right) \\ flag \leftarrow 1 \text{ if } |R| \geq UCLR \vee |R| \leq LCLR \\ flag \leftarrow 1 \text{ if } ARLR \geq 10000 \\ ARLR \leftarrow ARLR + 1 \end{array}\right. \\ U_{1,1} \leftarrow ARLR \\ U_{2,1} \leftarrow |R| \\ U_{3,1} \leftarrow flag \\ U \end{array}\right.$$

SubRoutineW Code:

Compares a deviate of the correlation matrix to the control limits (UCLR, LCLR) until an out-of-control condition is detected, then returns the number of trials before this condition is reached (ARLW) and sets a flag to stop the iterations. A flag is also set if ARLW exceeds 10000.

$$
\text{SubRoutineW}(n,\rho,\alpha) := \left|
\begin{array}{l}
\text{flag} \leftarrow 0 \\
\text{ARLW} \leftarrow 0 \\
\text{while flag} \neq 1 \\
\quad \left|
\begin{array}{l}
R \leftarrow \text{Sampler}(n,\rho) \\
p \leftarrow \text{rows}(R) \\
A \leftarrow (n-1)\cdot R \\
W \leftarrow -p\cdot n + p\cdot n\cdot \ln(n) - n\cdot \ln\left(\dfrac{|A|}{|\rho|}\right) + \text{tr}\left(\rho^{-1}\cdot A\right) \\
\text{UCLW} \leftarrow \text{qchisq}\left[\alpha, \dfrac{[p\cdot(p+1)]}{2}\right] \\
\text{LCLW} \leftarrow 0 \\
\text{flag} \leftarrow 1 \ \text{ if } \ W \geq \text{UCLW} \vee W \leq \text{LCLW} \\
\text{flag} \leftarrow 1 \ \text{ if } \ \text{ARLW} \geq 10000 \\
\text{ARLW} \leftarrow \text{ARLW} + 1
\end{array}
\right. \\
\text{return ARLW}
\end{array}
\right.
$$

Delta Routine:

Calculates the standard error of the estimate. This is used by the TestR and TestW routines to determine the relative precision that is used to terminate the simulation for each.

$$
\delta(\alpha, X) := \left|
\begin{array}{l}
n \leftarrow \text{rows}(X) \\
df \leftarrow n-1 \\
t_0 \leftarrow \text{qt}\left(1-\dfrac{\alpha}{2}, df\right) \\
s \leftarrow \sqrt{\dfrac{n}{n-1}\cdot \text{Var}(X)} \\
\delta \leftarrow t_0\cdot \dfrac{s}{\sqrt{n}} \\
\text{return } \delta
\end{array}
\right.
$$

SimTest Routine (adapted to |R|)

If the relative precision is outside of the tolerance, it increments and returns the value of the replicate.

Xbar = ARL for the combination
delta = standard error of estimate
ratio = γ (termination ratio)
rep = number of repetitions to get to specified relative precision

$$
\mathrm{TestR}(\alpha,\gamma,n,\rho) := \left|
\begin{array}{l}
ratio \leftarrow 1 \\
rep \leftarrow 0 \\
U \leftarrow \mathrm{SubRoutineR}(n,\rho,\alpha) \\
\text{while } ratio \ge \gamma \\
\quad \left|
\begin{array}{l}
V \leftarrow \mathrm{SubRoutineR}(n,\rho,\alpha) \\
U \leftarrow \mathrm{augment}(U,V) \\
ARL \leftarrow \left(U^{T}\right)^{\langle 1\rangle} \\
Xbar \leftarrow \mathrm{mean}(ARL) \\
delta \leftarrow \delta(\alpha, ARL) \\
ratio \leftarrow \dfrac{delta}{\left|\left|Xbar\right|\right|} \\
rep \leftarrow rep + 1
\end{array}
\right. \\
Q_{1,1} \leftarrow Xbar \\
Q_{2,1} \leftarrow delta \\
Q_{3,1} \leftarrow ratio \\
Q_{4,1} \leftarrow rep \\
Q
\end{array}
\right.
$$

SimTest Routine (adapted to W_R)

If the relative precision is outside of the tolerance, it increments and returns the value of the replicate.

Xbar = ARL for the combination
delta = standard error of estimate
ratio = γ (termination ratio)
rep = number of repetitions to get to specified relative precision

```
TestW(α, γ, n, ρ) :=  | ratio ← 1
                      | rep ← 0
                      | U ← SubRoutineW(n, ρ, α)
                      | while ratio ≥ γ
                      |   | V ← SubRoutineW(n, ρ, α)
                      |   | U ← augment(U, V)
                      |   | ARL ← (U^T)^<1>
                      |   | Xbar ← mean(ARL)
                      |   | delta ← δ(α, ARL)
                      |   | ratio ← delta / ||Xbar||
                      |   | rep ← rep + 1
                      | Q[1,1] ← Xbar
                      | Q[2,1] ← delta
                      | Q[3,1] ← ratio
                      | Q[4,1] ← rep
                      | Q
```

Summary Table for Simulation:

The following table provides the summary information of the data that was input prior to beginning the simulation.

In-Control Correlation Matrix:

ρ = ▮

Variance of Deviates:

var = ▮

Sample Size:

n = ▮

Number of Quality Characteristics:

p = ▮

Tolerance:

α = ▮

Relative Precision:

γ = ▮

Simulation Start Code:

The following two lines start the simulations for the |R| and WR statistic using the parameters declared at the beginning and provide columnar output with the labels defined below.

Xbar = ARL for the combination
delta = standard error of estimate
ratio = γ (termination ratio)
rep = number of repetitions to get to specified relative precision

$\text{TestR}(\alpha, \gamma, n, \rho)$ = ▮

$\text{TestW}(\alpha, \gamma, n, \rho)$ = ▮

APPENDIX F

MATHCAD SIMULATION RESULTS

This appendix provides complete results of the simulation study described in Section 4. In the first four columns of both tables, *p* is the number of quality characteristics, "perturbs" is the number of non-diagonal matrix entries to which deviates were applied, *n* is the sample size, and "var" is the variance used to generate the deviates. For the $|R|$ and W_R statistics, *ARL* gives the average run length calculated by the simulation, at the tolerance provided by the simulation and "std err" is the standard error of the simulation, at the tolerance provided by the simulation (and specified through the design of the simulation algorithm). The column marked "reps" refers to the number of replications required before the sequential simulation automatically terminated at the specified tolerance ratio.

For the *ARL*, non-numeric entries in the table indicate the following: "over" means that the *ARL* calculated exceeded 10,000 and the simulation was automatically terminated; "sing" indicates that the simulation returned a singularity after several attempted runs; "unk" represents an "unknown error" returned by the simulation after several attempted runs; and "n/s" means that the combination of parameters listed was not simulated since the trend that developed from simulations with previous parameters indicated that meaningful results would not be obtained.

Table 15 Simulation Results for the Exchange (0.5) Structure

				|R|			W_R		
p	perturbs	n	var	ARL	std err	reps	ARL	std err	reps
2	1	3	0.25	over	----	----	196	10	1527
2	1	4	0.25	5044	252	718	109	5	1620
2	1	5	0.25	1779	89	1466	70	4	1435
2	1	6	0.25	771	39	1539	53	3	1529
2	1	7	0.25	411	21	1528	41	2	1497
2	1	8	0.25	253	13	1577	32	2	1548
2	1	9	0.25	180	9	1557	29	1	1297
2	1	10	0.25	133	7	1541	24	1	1473
2	1	12	0.25	84	4	1600	19	1	1438
2	1	15	0.25	53	3	1515	14	1	1379
2	1	20	0.25	31	2	1507	10	0	1351
2	1	25	0.25	21	1	1385	7	0	1398
2	1	30	0.25	17	1	1377	6	0	1334
2	1	35	0.25	14	1	1377	5	0	1125
2	1	40	0.25	12	1	1369	4	0	1190
2	1	3	0.20	over	----	----	471	24	1490
2	1	4	0.20	over	----	----	255	13	1592
2	1	5	0.20	over	----	----	161	8	1554
2	1	6	0.20	6985	349	394	106	5	1457
2	1	7	0.20	4300	215	884	85	4	1599
2	1	8	0.20	2329	116	1426	65	3	1564
2	1	9	0.20	1378	69	1550	56	3	1357
2	1	10	0.20	876	44	1542	50	2	1433
2	1	12	0.20	443	20	1934	35	2	1709
2	1	15	0.20	206	10	1541	27	1	1374
2	1	20	0.20	95	5	1556	18	1	1413
2	1	25	0.20	58	3	1600	12	1	1335
2	1	30	0.20	39	2	1409	10	1	1370
2	1	35	0.20	30	1	1499	8	0	1325
2	1	40	0.20	24	1	1479	7	0	1380
2	1	3	0.15	over	----	----	3369	168	1100
2	1	4	0.15	over	----	----	1717	86	1584
2	1	5	0.15	over	----	----	1008	50	1639
2	1	6	0.15	over	----	----	646	32	1594
2	1	7	0.15	over	----	----	451	23	1592
2	1	8	0.15	over	----	----	328	16	1645
2	1	9	0.15	over	----	----	248	12	1616
2	1	10	0.15	over	----	----	205	10	1573
2	1	12	0.15	7185	359	366	131	7	1518
2	1	15	0.15	3409	170	1139	90	5	1558
2	1	20	0.15	985	49	1556	52	3	1638
2	1	25	0.15	406	20	1494	35	2	1420
2	1	30	0.15	220	11	1506	26	1	1405
2	1	35	0.15	140	7	1551	20	1	1504
2	1	40	0.15	99	5	1518	17	1	1481
2	1	3	0.10	over	----	----	over	----	----

Table 15 (continued)

2	1	4	0.10	over	----	----	over	----	----
2	1	5	0.10	over	----	----	over	----	----
2	1	6	0.10	over	----	----	9603	479	43
2	1	7	0.10	over	----	----	9430	467	71
2	1	8	0.10	over	----	----	9106	452	85
2	1	9	0.10	over	----	----	8651	430	140
2	1	10	0.10	over	----	----	7577	378	296
2	1	12	0.10	over	----	----	5868	293	580
2	1	15	0.10	over	----	----	3022	151	1219
2	1	20	0.10	over	----	----	1071	54	1532
2	1	25	0.10	over	----	----	492	25	1484
2	1	30	0.10	over	----	----	269	13	1547
2	1	35	0.10	6384	319	487	172	9	1557
2	1	40	0.10	3852	193	1001	121	6	1617
3	1	4	0.25	7807	389	260	379	19	1508
3	1	5	0.25	3808	190	976	202	10	1524
3	1	6	0.25	1816	91	1336	128	6	1370
3	1	7	0.25	960	48	1425	92	5	1469
3	1	8	0.25	616	31	1525	74	4	1653
3	1	9	0.25	415	21	1582	61	3	1611
3	1	10	0.25	309	15	1465	48	2	1357
3	1	12	0.25	185	9	1579	36	2	1536
3	1	15	0.25	109	5	1675	24	1	1431
3	1	20	0.25	62	3	1568	15	1	1394
3	1	25	0.25	42	2	1647	11	1	1426
3	1	30	0.25	31	2	1573	9	0	1291
3	1	35	0.25	25	1	1581	7	0	1377
3	1	40	0.25	21	1	1484	6	0	1352
3	1	4	0.20	over	----	----	999	50	1523
3	1	5	0.20	over	----	----	520	26	1517
3	1	6	0.20	over	----	----	357	17	1604
3	1	7	0.20	8000	399	223	219	11	1588
3	1	8	0.20	5988	299	541	175	9	1630
3	1	9	0.20	5146	207	905	131	7	1529
3	1	10	0.20	2834	142	1231	111	6	1682
3	1	12	0.20	1347	67	1383	77	4	1521
3	1	15	0.20	632	32	1489	50	3	1502
3	1	20	0.20	263	13	1525	31	2	1499
3	1	25	0.20	148	7	1611	22	1	1539
3	1	30	0.20	99	5	1639	15	1	1291
3	1	35	0.20	72	4	1575	13	1	1469
3	1	40	0.20	57	3	1503	11	1	1458
3	1	4	0.15	over	----	----	5608	280	635
3	1	5	0.15	over	----	----	3532	176	1135
3	1	6	0.15	over	----	----	2235	112	1365
3	1	7	0.15	over	----	----	1532	77	1494
3	1	8	0.15	over	----	----	1062	53	1500
3	1	9	0.15	over	----	----	813	41	1500

Table 15 (continued)

3	1	10	0.15	over	----	----	634	32	1528
3	1	12	0.15	over	----	----	405	20	1298
3	1	15	0.15	8310	414	201	234	12	1502
3	1	20	0.15	4515	226	829	124	6	1442
3	1	25	0.15	1985	99	1313	77	4	1421
3	1	30	0.15	965	48	1423	56	3	1605
3	1	35	0.15	590	30	1538	39	2	1544
3	1	40	0.15	378	19	1539	30	2	1415
3	1	4	0.10	over	----	----	over	----	----
3	1	5	0.10	over	----	----	over	----	----
3	1	6	0.10	over	----	----	over	----	----
3	1	7	0.10	over	----	----	over	----	----
3	1	8	0.10	over	----	----	over	----	----
3	1	9	0.10	over	----	----	over	----	----
3	1	10	0.10	over	----	----	over	----	----
3	1	12	0.10	over	----	----	over	----	----
3	1	15	0.10	over	----	----	over	----	----
3	1	20	0.10	over	----	----	over	----	----
3	1	25	0.10	over	----	----	2481	124	1254
3	1	30	0.10	over	----	----	1214	61	1423
3	1	35	0.10	over	----	----	724	36	1478
3	1	40	0.10	over	----	----	453	23	1586
3	2	4	0.25	2821	141	1287	140	7	1505
3	2	5	0.25	952	48	1548	74	4	1475
3	2	6	0.25	456	23	1613	45	2	1511
3	2	7	0.25	268	13	1533	33	2	1629
3	2	8	0.25	173	9	1626	25	1	1314
3	2	9	0.25	120	6	1505	20	1	1644
3	2	10	0.25	91	5	1457	17	1	1440
3	2	12	0.25	58	3	1550	13	1	1415
3	2	15	0.25	34	2	1475	9	0	1209
3	2	20	0.25	21	1	1522	6	0	1318
3	2	25	0.25	14	1	1501	5	0	1296
3	2	30	0.25	11	1	1393	4	0	1036
3	2	35	0.25	9	0	1353	3	0	1001
3	2	40	0.25	8	0	1695	3	0	1244
3	2	4	0.20	over	----	----	323	16	1540
3	2	5	0.20	over	----	----	166	8	1533
3	2	6	0.20	6582	329	455	111	6	1492
3	2	7	0.20	3911	195	1009	74	4	1494
3	2	8	0.20	2122	106	1430	57	3	1427
3	2	9	0.20	1209	60	1583	42	2	1482
3	2	10	0.20	794	40	1505	34	2	1583
3	2	12	0.20	399	20	1541	24	1	1715
3	2	15	0.20	190	9	1558	17	1	1503
3	2	20	0.20	83	4	1406	11	1	1528
3	2	25	0.20	49	2	1537	8	0	1419
3	2	30	0.20	36	2	1420	6	0	1341

Table 15 (continued)

3	2	35	0.20	25	1	1442	5	0	1239
3	2	40	0.20	20	1	1467	4	0	1201
3	2	4	0.15	over	----	----	2296	115	1416
3	2	5	0.15	over	----	----	1185	59	1418
3	2	6	0.15	over	----	----	709	35	1439
3	2	7	0.15	over	----	----	461	23	1587
3	2	8	0.15	over	----	----	329	16	1521
3	2	9	0.15	over	----	----	248	12	1574
3	2	10	0.15	over	----	----	190	9	1449
3	2	12	0.15	over	----	----	123	6	1336
3	2	15	0.15	5128	256	751	73	4	1428
3	2	20	0.15	unk	----	----	36	2	1554
3	2	25	0.15	580	29	1527	24	1	1503
3	2	30	0.15	299	13	1947	17	1	1830
3	2	35	0.15	175	9	1481	13	1	1423
3	2	40	0.15	120	6	1412	10	1	1278
3	2	4	0.10	over	----	----	over	----	----
3	2	5	0.10	over	----	----	over	----	----
3	2	6	0.10	over	----	----	over	----	----
3	2	7	0.10	over	----	----	over	----	----
3	2	8	0.10	over	----	----	over	----	----
3	2	9	0.10	over	----	----	over	----	----
3	2	10	0.10	over	----	----	9666	451	13
3	2	12	0.10	over	----	----	7643	381	268
3	2	15	0.10	over	----	----	4841	242	797
3	2	20	0.10	over	----	----	1670	83	1461
3	2	25	0.10	over	----	----	667	33	1538
3	2	30	0.10	over	----	----	330	17	1587
3	2	35	0.10	over	----	----	189	9	1600
3	2	40	0.10	7546	377	302	114	6	1651
3	3	4	0.25	unk	----	----	73	4	1597
3	3	5	0.25	213	11	1510	41	2	1427
3	3	6	0.25	118	6	1479	26	1	1593
3	3	7	0.25	79	4	1580	18	1	1418
3	3	8	0.25	55	3	1621	14	1	1426
3	3	9	0.25	42	2	1588	12	1	1332
3	3	10	0.25	33	2	1539	10	1	1292
3	3	12	0.25	23	1	1493	7	0	1410
3	3	15	0.25	16	1	1461	5	0	1315
3	3	20	0.25	10	0	1428	4	0	1093
3	3	25	0.25	7	0	1685	3	0	1296
3	3	30	0.25	6	0	1217	2	0	821
3	3	35	0.25	5	0	1118	2	0	809
3	3	40	0.25	4	0	1136	2	0	724
3	3	4	0.20	over	----	----	147	7	1550
3	3	5	0.20	sing	----	----	82	4	1425
3	3	6	0.20	unk	----	----	50	2	1605
3	3	7	0.20	802	40	1525	38	2	1508

Table 15 (continued)

3	3	8	0.20	unk	----	----	27	1	1565
3	3	9	0.20	300	15	1457	22	1	1639
3	3	10	0.20	208	10	1518	18	1	1381
3	3	12	0.20	118	6	1562	13	1	1343
3	3	15	0.20	65	3	1523	9	0	1508
3	3	20	0.20	34	2	1922	6	0	1553
3	3	25	0.20	22	1	1516	4	0	1119
3	3	30	0.20	16	1	1517	4	0	1112
3	3	35	0.20	12	1	1399	3	0	966
3	3	40	0.20	10	0	1699	3	0	1000
3	3	4	0.15	over	----	----	847	42	1491
3	3	5	0.15	over	----	----	461	23	1532
3	3	6	0.15	over	----	----	268	13	1478
3	3	7	0.15	over	----	----	180	9	1559
3	3	8	0.15	over	----	----	127	6	1515
3	3	9	0.15	over	----	----	94	5	1598
3	3	10	0.15	over	----	----	73	4	1540
3	3	12	0.15	sing	----	----	52	3	1603
3	3	15	0.15	unk	----	----	33	2	1548
3	3	20	0.15	445	22	1547	17	1	1546
3	3	25	0.15	195	10	1509	11	1	1360
3	3	30	0.15	105	5	1646	8	0	1423
3	3	35	0.15	68	3	1580	6	0	1288
3	3	40	0.15	49	2	1544	5	0	1358
3	3	4	0.10	over	----	----	over	----	----
3	3	5	0.10	over	----	----	9763	477	19
3	3	6	0.10	over	----	----	9115	452	105
3	3	7	0.10	over	----	----	8477	423	178
3	3	8	0.10	over	----	----	7514	375	340
3	3	9	0.10	over	----	----	6730	336	451
3	3	10	0.10	over	----	----	5778	260	752
3	3	12	0.10	over	----	----	3700	166	1275
3	3	15	0.10	over	----	----	1656	83	1617
3	3	20	0.10	over	----	----	502	25	1529
3	3	25	0.10	over	----	----	208	10	1563
3	3	30	0.10	over	----	----	109	5	1582
3	3	35	0.10	6517	326	478	68	3	1445
3	3	40	0.10	sing	----	----	45	2	1702
5	1	6	0.25	sing	----	----	977	49	1464
5	1	7	0.25	sing	----	----	539	27	1658
5	1	8	0.25	unk	----	----	353	18	1585
5	1	9	0.25	unk	----	----	257	13	1603
5	1	10	0.25	unk	----	----	180	9	1592
5	1	12	0.25	506	25	1635	112	6	1577
5	1	15	0.25	268	13	1474	68	3	1555
5	1	20	0.25	137	7	1509	37	2	1427
5	1	25	0.25	89	4	1632	24	1	1557
5	1	30	0.25	66	3	1680	18	1	1594

Table 15 (continued)

5	1	35	0.25	51	3	1556	14	1	1418
5	1	40	0.25	41	2	1602	12	1	1370
5	1	6	0.20	over	----	----	sing	----	----
5	1	7	0.20	over	----	----	1544	77	1624
5	1	8	0.20	over	----	----	951	48	1537
5	1	9	0.20	over	----	----	645	32	1619
5	1	10	0.20	over	----	----	475	24	1571
5	1	12	0.20	sing	----	----	291	16	1471
5	1	15	0.20	unk	----	----	180	9	1522
5	1	20	0.20	unk	----	----	98	5	1583
5	1	25	0.20	499	25	1660	59	3	1457
5	1	30	0.20	297	15	1483	41	2	1502
5	1	35	0.20	207	10	1432	30	2	1502
5	1	40	0.20	unk	----	----	24	1	1484
5	1	6	0.15	over	----	----	over	----	----
5	1	7	0.15	over	----	----	over	----	----
5	1	8	0.15	over	----	----	over	----	----
5	1	9	0.15	over	----	----	sing	----	----
5	1	10	0.15	over	----	----	sing	----	----
5	1	12	0.15	over	----	----	2387	119	1451
5	1	15	0.15	over	----	----	1313	66	1551
5	1	20	0.15	over	----	----	606	30	1561
5	1	25	0.15	over	----	----	337	17	1517
5	1	30	0.15	sing	----	----	207	10	1568
5	1	35	0.15	unk	----	----	144	7	1624
5	1	40	0.15	unk	----	----	111	6	1513
5	1	6	0.10	over	----	----	over	----	----
5	1	7	0.10	over	----	----	over	----	----
5	1	8	0.10	over	----	----	over	----	----
5	1	9	0.10	over	----	----	over	----	----
5	1	10	0.10	over	----	----	over	----	----
5	1	12	0.10	over	----	----	over	----	----
5	1	15	0.10	over	----	----	9783	465	20
5	1	20	0.10	over	----	----	9408	464	60
5	1	25	0.10	over	----	----	8622	430	165
5	1	30	0.10	over	----	----	7692	384	283
5	1	35	0.10	over	----	----	6172	308	522
5	1	40	0.10	over	----	----	sing	----	----
5	2	6	0.25	unk	----	----	sing	----	----
5	2	7	0.25	unk	----	----	189	9	1557
5	2	8	0.25	unk	----	----	116	6	1403
5	2	9	0.25	unk	----	----	84	4	1584
5	2	10	0.25	219	11	1589	61	3	1444
5	2	12	0.25	126	6	1528	41	2	1528
5	2	15	0.25	71	4	1400	23	1	1430
5	2	20	0.25	38	2	1858	14	1	1847
5	2	25	0.25	27	1	1548	9	0	1447
5	2	30	0.25	21	1	1520	7	0	1266

Table 15 (continued)

5	2	35	0.25	16	1	1455	6	0	1356
5	2	40	0.25	13	1	1547	5	0	1248
5	2	6	0.20	over	----	----	788	39	1417
5	2	7	0.20	sing	----	----	456	23	1551
5	2	8	0.20	sing	----	----	289	14	1461
5	2	9	0.20	sing	----	----	203	10	1609
5	2	10	0.20	sing	----	----	149	7	1419
5	2	12	0.20	unk	----	----	93	5	1487
5	2	15	0.20	unk	----	----	57	3	1612
5	2	20	0.20	unk	----	----	28	1	1620
5	2	25	0.20	130	6	1483	19	1	1486
5	2	30	0.20	81	4	1523	14	1	1720
5	2	35	0.20	58	3	1518	10	1	1336
5	2	40	0.20	46	2	1496	8	0	1346
5	2	6	0.15	over	----	----	sing	----	----
5	2	7	0.15	over	----	----	sing	----	----
5	2	8	0.15	over	----	----	sing	----	----
5	2	9	0.15	over	----	----	sing	----	----
5	2	10	0.15	over	----	----	1063	53	1526
5	2	12	0.15	over	----	----	656	33	1521
5	2	15	0.15	over	----	----	352	18	1497
5	2	20	0.15	sing	----	----	154	8	1539
5	2	25	0.15	sing	----	----	89	4	1518
5	2	30	0.15	sing	----	----	58	3	1451
5	2	35	0.15	unk	----	----	40	2	1658
5	2	40	0.15	unk	----	----	31	2	1448
5	2	6	0.10	over	----	----	over	----	----
5	2	7	0.10	over	----	----	over	----	----
5	2	8	0.10	over	----	----	over	----	----
5	2	9	0.10	over	----	----	over	----	----
5	2	10	0.10	over	----	----	over	----	----
5	2	12	0.10	over	----	----	over	----	----
5	2	15	0.10	over	----	----	over	----	----
5	2	20	0.10	over	----	----	over	----	----
5	2	25	0.10	over	----	----	over	----	----
5	2	30	0.10	over	----	----	sing	----	----
5	2	35	0.10	over	----	----	sing	----	----
5	2	40	0.10	over	----	----	unk	----	----
5	3	6	0.25	unk	----	----	193	10	1388
5	3	7	0.25	unk	----	----	107	5	1428
5	3	8	0.25	unk	----	----	69	3	1565
5	3	9	0.25	73	4	1487	47	2	1602
5	3	10	0.25	57	3	1349	36	2	1454
5	3	12	0.25	39	2	1446	23	1	1388
5	3	15	0.25	26	1	1405	14	1	1439
5	3	20	0.25	16	1	1396	8	0	1248
5	3	25	0.25	12	1	1479	5	0	1181
5	3	30	0.25	9	0	1411	4	0	1056

Table 15 (continued)

5	3	35	0.25	8	0	1356	3	0	1013
5	3	40	0.25	7	0	1337	3	0	963
5	3	6	0.20	sing	----	----	398	20	1651
5	3	7	0.20	sing	----	----	215	11	1619
5	3	8	0.20	unk	----	----	134	6	1856
5	3	9	0.20	721	36	1528	94	4	1815
5	3	10	0.20	unk	----	----	73	4	1352
5	3	12	0.20	270	14	1455	44	2	1532
5	3	15	0.20	146	7	1624	26	1	1472
5	3	20	0.20	70	4	1449	15	1	1453
5	3	25	0.20	43	2	1362	10	1	1337
5	3	30	0.20	31	2	1389	7	0	1326
5	3	35	0.20	23	1	1443	6	0	1325
5	3	40	0.20	19	1	1401	5	0	1127
5	3	6	0.15	over	----	----	sing	----	----
5	3	7	0.15	over	----	----	sing	----	----
5	3	8	0.15	over	----	----	738	37	1543
5	3	9	0.15	over	----	----	518	26	1626
5	3	10	0.15	sing	----	----	365	18	1705
5	3	12	0.15	sing	----	----	215	11	1579
5	3	15	0.15	sing	----	----	121	6	1675
5	3	20	0.15	sing	----	----	60	3	1559
5	3	25	0.15	unk	----	----	36	2	1409
5	3	30	0.15	284	19	1489	25	1	1437
5	3	35	0.15	237	12	1476	18	1	1446
5	3	40	0.15	158	8	1538	14	1	1476
5	3	6	0.10	over	----	----	over	----	----
5	3	7	0.10	over	----	----	over	----	----
5	3	8	0.10	over	----	----	over	----	----
5	3	9	0.10	over	----	----	over	----	----
5	3	10	0.10	over	----	----	over	----	----
5	3	12	0.10	over	----	----	over	----	----
5	3	15	0.10	over	----	----	over	----	----
5	3	20	0.10	over	----	----	sing	----	----
5	3	25	0.10	over	----	----	sing	----	----
5	3	30	0.10	over	----	----	sing	----	----
5	3	35	0.10	over	----	----	555	28	1540
5	3	40	0.10	over	----	----	307	15	1468
8	1	9	0.25	sing	----	----	unk	----	----
8	1	10	0.25	sing	----	----	2127	106	1519
8	1	12	0.25	sing	----	----	877	44	1532
8	1	15	0.25	unk	----	----	397	20	1433
8	1	20	0.25	unk	----	----	152	8	1404
8	1	25	0.25	179	9	1651	87	4	1458
8	1	30	0.25	125	6	1538	57	3	1538
8	1	35	0.25	97	5	1514	41	2	1590
8	1	40	0.25	76	4	1478	32	2	1383
8	1	9	0.20	9948	278	2	sing	----	----

Table 15 (continued)

8	1	10	0.20	9948	278	2	sing	----	----
8	1	12	0.20	9948	278	2	sing	----	----
8	1	15	0.20	sing	----	----	1107	55	1556
8	1	20	0.20	sing	----	----	430	21	1485
8	1	25	0.20	sing	----	----	237	12	1518
8	1	30	0.20	sing	----	----	149	7	1904
8	1	35	0.20	unk	----	----	105	5	1477
8	1	40	0.20	406	20	1589	76	4	1501
8	1	9	0.15	sing	----	----	sing	----	----
8	1	10	0.15	over	----	----	over	----	----
8	1	12	0.15	over	----	----	over	----	----
8	1	15	0.15	over	----	----	9966	186	2
8	1	20	0.15	over	----	----	3707	167	1324
8	1	25	0.15	9948	278	2	2058	93	1863
8	1	30	0.15	8989	403	157	1202	54	1920
8	1	35	0.15	9948	278	2	824	41	1420
8	1	40	0.15	sing	----	----	541	27	1564
8	2	9	0.25	unk	----	----	sing	----	----
8	2	10	0.25	unk	----	----	sing	----	----
8	2	12	0.25	unk	----	----	sing	----	----
8	2	15	0.25	unk	----	----	132	7	1488
8	2	20	0.25	77	4	1449	54	3	1660
8	2	25	0.25	47	2	1738	29	1	1650
8	2	30	0.25	35	2	1539	21	1	1414
8	2	35	0.25	27	1	1530	15	1	1434
8	2	40	0.25	unk	----	----	11	1	1371
8	2	9	0.20	over	----	----	sing	----	----
8	2	10	0.20	over	----	----	sing	----	----
8	2	12	0.20	sing	----	----	764	38	1532
8	2	15	0.20	sing	----	----	318	16	1545
8	2	20	0.20	sing	----	----	126	6	1663
8	2	25	0.20	unk	----	----	73	4	1538
8	2	30	0.20	unk	----	----	46	2	1436
8	2	35	0.20	145	7	1447	33	2	1449
8	2	40	0.20	102	5	1453	25	1	1439
8	2	9	0.15	over	----	----	over	----	----
8	2	10	0.15	over	----	----	over	----	----
8	2	12	0.15	over	----	----	sing	----	----
8	2	15	0.15	over	----	----	sing	----	----
8	2	20	0.15	over	----	----	sing	----	----
8	2	25	0.15	over	----	----	sing	----	----
8	2	30	0.15	sing	----	----	294	15	1588
8	2	35	0.15	sing	----	----	195	10	1500
8	2	40	0.15	sing	----	----	130	7	1401
8	2	9	0.10	over	----	----	over	----	----
8	2	10	0.10	over	----	----	over	----	----
8	2	12	0.10	over	----	----	over	----	----
8	2	15	0.10	over	----	----	over	----	----

Table 15 (continued)

8	2	20	0.10	over	----	----	over	----	----
8	2	25	0.10	over	----	----	over	----	----
8	2	30	0.10	over	----	----	9553	466	34
8	2	35	0.10	over	----	----	sing	----	----
8	2	40	0.10	over	----	----	sing	----	----
8	3	9	0.25	unk	----	----	sing	----	----
8	3	10	0.25	unk	----	----	sing	----	----
8	3	12	0.25	unk	----	----	181	9	1542
8	3	15	0.25	unk	----	----	71	4	1434
8	3	20	0.25	27	1	1360	30	1	1402
8	3	25	0.25	19	1	1387	17	1	1462
8	3	30	0.25	5	0	1	11	1	1414
8	3	35	0.25	5	0	1	9	0	1307
8	3	40	0.25	5	0	1	7	0	1308
8	3	9	0.20	sing	----	----	sing	----	----
8	3	10	0.20	sing	----	----	888	44	1521
8	3	12	0.20	sing	----	----	396	20	1579
8	3	15	0.20	unk	----	----	167	8	1585
8	3	20	0.20	unk	----	----	64	3	1512
8	3	25	0.20	unk	----	----	36	2	1493
8	3	30	0.20	unk	----	----	24	1	1425
8	3	35	0.20	unk	----	----	17	1	1514
8	3	40	0.20	36	2	1554	13	1	1411
8	3	9	0.15	over	----	----	sing	----	----
8	3	10	0.15	over	----	----	sing	----	----
8	3	12	0.15	9679	468	18	sing	----	----
8	3	15	0.15	sing	----	----	unk	----	----
8	3	20	0.15	sing	----	----	sing	----	----
8	3	25	0.15	sing	----	----	180	9	1506
8	3	30	0.15	sing	----	----	111	6	1485
8	3	35	0.15	unk	----	----	76	4	1673
8	3	40	0.15	unk	----	----	54	3	1597
8	3	9	0.10	over	----	----	over	----	----
8	3	10	0.10	over	----	----	over	----	----
8	3	12	0.10	over	----	----	over	----	----
8	3	15	0.10	over	----	----	over	----	----
8	3	20	0.10	over	----	----	over	----	----
8	3	25	0.10	over	----	----	over	----	----
8	3	30	0.10	over	----	----	sing	----	----
8	3	35	0.10	over	----	----	sing	----	----
8	3	40	0.10	over	----	----	sing	----	----

The following table, Table 16, provides the results for simulations of the Independence correlation matrix structure and uses the same nomenclature as that for Table 15.

Table 16 Simulation Results for the Independence Structure

				|R|			W_R		
p	perturbs	n	var	ARL	std err	reps	ARL	std err	reps
2	1	3	0.25	over	----	----	7373	369	339
2	1	4	0.25	over	----	----	4900	245	800
2	1	5	0.25	over	----	----	2807	140	1306
2	1	6	0.25	over	----	----	1622	81	1550
2	1	7	0.25	over	----	----	990	49	1489
2	1	8	0.25	over	----	----	625	31	1421
2	1	9	0.25	over	----	----	435	22	1338
2	1	10	0.25	over	----	----	308	15	1398
2	1	12	0.25	over	----	----	175	9	1415
2	1	15	0.25	over	----	----	89	4	1455
2	1	20	0.25	5322	266	712	44	2	1427
2	1	25	0.25	2848	142	1275	26	1	1473
2	1	30	0.25	1579	79	1476	17	1	1360
2	1	35	0.25	981	49	1499	14	1	1516
2	1	40	0.25	663	33	1418	11	1	1188
2	1	3	0.20	over	----	----	over	----	----
2	1	4	0.20	over	----	----	over	----	----
2	1	5	0.20	over	----	----	9414	468	69
2	1	6	0.20	over	----	----	8995	446	128
2	1	7	0.20	over	----	----	8254	411	223
2	1	8	0.20	over	----	----	7005	350	375
2	1	9	0.20	over	----	----	5316	266	711
2	1	10	0.20	over	----	----	3774	189	1060
2	1	12	0.20	over	----	----	1791	90	1497
2	1	15	0.20	over	----	----	680	34	1451
2	1	20	0.20	over	----	----	221	11	1501
2	1	25	0.20	over	----	----	98	5	1455
2	1	30	0.20	over	----	----	58	3	1530
2	1	35	0.20	over	----	----	39	2	1675
2	1	40	0.20	over	----	----	28	1	1625
2	1	3	0.15	over	----	----	over	----	----
2	1	4	0.15	over	----	----	over	----	----
2	1	5	0.15	over	----	----	over	----	----
2	1	6	0.15	over	----	----	over	----	----
2	1	7	0.15	over	----	----	over	----	----
2	1	8	0.15	over	----	----	over	----	----
2	1	9	0.15	over	----	----	over	----	----
2	1	10	0.15	over	----	----	over	----	----
2	1	12	0.15	over	----	----	over	----	----
2	1	15	0.15	over	----	----	8927	443	135
2	1	20	0.15	over	----	----	5143	257	746
2	1	25	0.15	over	----	----	1799	90	1489
2	1	30	0.15	over	----	----	706	35	1510
2	1	35	0.15	over	----	----	353	18	1412

Table 16 (continued)

2	1	40	0.15	over	----	----	200	10	1536
2	1	3	0.10	over	----	----	over	----	----
2	1	4	0.10	over	----	----	over	----	----
2	1	5	0.10	over	----	----	over	----	----
2	1	6	0.10	over	----	----	over	----	----
2	1	7	0.10	over	----	----	over	----	----
2	1	8	0.10	over	----	----	over	----	----
2	1	9	0.10	over	----	----	over	----	----
2	1	10	0.10	over	----	----	over	----	----
2	1	12	0.10	over	----	----	over	----	----
2	1	15	0.10	over	----	----	over	----	----
2	1	20	0.10	over	----	----	over	----	----
2	1	25	0.10	over	----	----	over	----	----
2	1	30	0.10	over	----	----	over	----	----
2	1	35	0.10	over	----	----	9455	472	56
2	1	40	0.10	over	----	----	8942	447	131
3	1	4	0.25	over	----	----	over	----	----
3	1	5	0.25	over	----	----	over	----	----
3	1	6	0.25	over	----	----	over	----	----
3	1	7	0.25	over	----	----	over	----	----
3	1	8	0.25	over	----	----	2909	145	1202
3	1	9	0.25	over	----	----	2018	101	1389
3	1	10	0.25	over	----	----	1480	74	1441
3	1	12	0.25	over	----	----	793	40	1559
3	1	15	0.25	over	----	----	377	19	1516
3	1	20	0.25	over	----	----	159	8	1495
3	1	25	0.25	7430	371	300	78	4	1556
3	1	30	0.25	5234	262	716	51	3	1499
3	1	35	0.25	3353	167	1145	35	2	1571
3	1	40	0.25	2128	106	1346	28	1	1437
3	1	4	0.20	over	----	----	over	----	----
3	1	5	0.20	over	----	----	over	----	----
3	1	6	0.20	over	----	----	over	----	----
3	1	7	0.20	over	----	----	over	----	----
3	1	8	0.20	over	----	----	over	----	----
3	1	9	0.20	over	----	----	9845	440	5
3	1	10	0.20	over	----	----	9845	440	5
3	1	12	0.20	over	----	----	9845	440	5
3	1	15	0.20	over	----	----	4988	249	778
3	1	20	0.20	over	----	----	1530	76	1422
3	1	25	0.20	over	----	----	587	29	1487
3	1	30	0.20	over	----	----	281	14	1533
3	1	35	0.20	over	----	----	166	8	1474
3	1	40	0.20	over	----	----	103	5	1583
3	1	4	0.15	over	----	----	over	----	----
3	1	5	0.15	over	----	----	over	----	----
3	1	6	0.15	over	----	----	over	----	----
3	1	7	0.15	over	----	----	over	----	----

Table 16 (continued)

3	1	8	0.15	over	----	----	over	----	----
3	1	9	0.15	over	----	----	over	----	----
3	1	10	0.15	over	----	----	over	----	----
3	1	12	0.15	over	----	----	over	----	----
3	1	15	0.15	over	----	----	over	----	----
3	1	20	0.15	over	----	----	over	----	----
3	1	25	0.15	over	----	----	9845	440	5
3	1	30	0.15	over	----	----	6494	324	475
3	1	35	0.15	over	----	----	3709	185	1074
3	1	40	0.15	over	----	----	1814	91	1415
3	1	4	0.10	over	----	----	over	----	----
3	1	5	0.10	over	----	----	over	----	----
3	1	6	0.10	over	----	----	over	----	----
3	1	7	0.10	over	----	----	over	----	----
3	1	8	0.10	over	----	----	over	----	----
3	1	9	0.10	over	----	----	over	----	----
3	1	10	0.10	over	----	----	over	----	----
3	1	12	0.10	over	----	----	over	----	----
3	1	15	0.10	over	----	----	over	----	----
3	1	20	0.10	over	----	----	over	----	----
3	1	25	0.10	over	----	----	over	----	----
3	1	30	0.10	over	----	----	over	----	----
3	1	35	0.10	over	----	----	over	----	----
3	1	40	0.10	over	----	----	over	----	----
3	2	4	0.25	over	----	----	5007	250	740
3	2	5	0.25	over	----	----	2804	140	1286
3	2	6	0.25	over	----	----	1610	81	1526
3	2	7	0.25	over	----	----	976	49	1572
3	2	8	0.25	over	----	----	637	32	1630
3	2	9	0.25	over	----	----	439	22	1525
3	2	10	0.25	over	----	----	326	16	1588
3	2	12	0.25	over	----	----	186	9	1512
3	2	15	0.25	8445	421	189	95	5	1516
3	2	20	0.25	unk	----	----	40	2	1457
3	2	25	0.25	unk	----	----	23	1	1513
3	2	30	0.25	unk	----	----	15	1	1405
3	2	35	0.25	unk	----	----	11	1	1465
3	2	40	0.25	460	23	1536	8	0	1444
3	2	4	0.20	over	----	----	over	----	----
3	2	5	0.20	over	----	----	over	----	----
3	2	6	0.20	over	----	----	over	----	----
3	2	7	0.20	over	----	----	over	----	----
3	2	8	0.20	over	----	----	over	----	----
3	2	9	0.20	over	----	----	6975	348	371
3	2	10	0.20	over	----	----	5467	273	643
3	2	12	0.20	over	----	----	3163	158	1217
3	2	15	0.20	over	----	----	1162	58	1452
3	2	20	0.20	over	----	----	334	17	1440

Table 16 (continued)

3	2	25	0.20	over	----	----	142	7	1414
3	2	30	0.20	over	----	----	72	4	1410
3	2	35	0.20	8480	423	186	45	2	1662
3	2	40	0.20	7582	379	286	29	1	1679
3	2	4	0.15	over	----	----	over	----	----
3	2	5	0.15	over	----	----	over	----	----
3	2	6	0.15	over	----	----	over	----	----
3	2	7	0.15	over	----	----	over	----	----
3	2	8	0.15	over	----	----	over	----	----
3	2	9	0.15	over	----	----	over	----	----
3	2	10	0.15	over	----	----	over	----	----
3	2	12	0.15	over	----	----	over	----	----
3	2	15	0.15	over	----	----	over	----	----
3	2	20	0.15	over	----	----	over	----	----
3	2	25	0.15	over	----	----	5098	255	722
3	2	30	0.15	over	----	----	1961	98	1459
3	2	35	0.15	over	----	----	798	40	1537
3	2	40	0.15	over	----	----	391	20	1481
3	2	4	0.10	n/s	----	----	n/s	----	----
3	2	5	0.10	n/s	----	----	n/s	----	----
3	2	6	0.10	n/s	----	----	n/s	----	----
3	2	7	0.10	n/s	----	----	n/s	----	----
3	2	8	0.10	n/s	----	----	n/s	----	----
3	2	9	0.10	n/s	----	----	n/s	----	----
3	2	10	0.10	n/s	----	----	n/s	----	----
3	2	12	0.10	n/s	----	----	n/s	----	----
3	2	15	0.10	n/s	----	----	n/s	----	----
3	2	20	0.10	n/s	----	----	n/s	----	----
3	2	25	0.10	n/s	----	----	n/s	----	----
3	2	30	0.10	n/s	----	----	n/s	----	----
3	2	35	0.10	n/s	----	----	n/s	----	----
3	2	40	0.10	n/s	----	----	n/s	----	----
3	3	4	0.25	over	----	----	1522	76	1629
3	3	5	0.25	over	----	----	752	36	1626
3	3	6	0.25	over	----	----	432	32	1532
3	3	7	0.25	over	----	----	277	14	1449
3	3	8	0.25	over	----	----	187	9	1443
3	3	9	0.25	over	----	----	133	7	1434
3	3	10	0.25	over	----	----	102	5	1516
3	3	12	0.25	sing	----	----	63	3	1559
3	3	15	0.25	unk	----	----	36	2	1473
3	3	20	0.25	unk	----	----	27	0	1
3	3	25	0.25	unk	----	----	10	1	1305
3	3	30	0.25	273	14	1406	7	0	1353
3	3	35	0.25	176	9	1446	6	0	1287
3	3	40	0.25	122	6	1508	1	0	1
3	3	4	0.20	over	----	----	8608	430	149
3	3	5	0.20	over	----	----	7261	363	347

Table 16 (continued)

3	3	6	0.20	over	----	----	5949	297	582
3	3	7	0.20	over	----	----	4322	216	888
3	3	8	0.20	over	----	----	3044	152	1193
3	3	9	0.20	over	----	----	2000	100	1412
3	3	10	0.20	over	----	----	1354	68	1486
3	3	12	0.20	over	----	----	687	34	1645
3	3	15	0.20	over	----	----	296	15	1366
3	3	20	0.20	over	----	----	98	5	1540
3	3	25	0.20	7155	357	375	48	2	1369
3	3	30	0.20	sing	----	----	28	1	1490
3	3	35	0.20	unk	----	----	18	1	1346
3	3	40	0.20	unk	----	----	12	1	1413
3	3	4	0.15	over	----	----	over	----	----
3	3	5	0.15	over	----	----	over	----	----
3	3	6	0.15	over	----	----	over	----	----
3	3	7	0.15	over	----	----	over	----	----
3	3	8	0.15	over	----	----	over	----	----
3	3	9	0.15	over	----	----	over	----	----
3	3	10	0.15	over	----	----	over	----	----
3	3	12	0.15	over	----	----	over	----	----
3	3	15	0.15	over	----	----	over	----	----
3	3	20	0.15	over	----	----	4385	218	881
3	3	25	0.15	over	----	----	1323	66	1408
3	3	30	0.15	over	----	----	499	25	1548
3	3	35	0.15	over	----	----	225	11	1304
3	3	40	0.15	over	----	----	117	6	1465
3	3	4	0.10	n/s	----	----	n/s	----	----
3	3	5	0.10	n/s	----	----	n/s	----	----
3	3	6	0.10	n/s	----	----	n/s	----	----
3	3	7	0.10	n/s	----	----	n/s	----	----
3	3	8	0.10	n/s	----	----	n/s	----	----
3	3	9	0.10	n/s	----	----	n/s	----	----
3	3	10	0.10	n/s	----	----	n/s	----	----
3	3	12	0.10	n/s	----	----	n/s	----	----
3	3	15	0.10	n/s	----	----	n/s	----	----
3	3	20	0.10	n/s	----	----	n/s	----	----
3	3	25	0.10	n/s	----	----	n/s	----	----
3	3	30	0.10	n/s	----	----	n/s	----	----
3	3	35	0.10	n/s	----	----	n/s	----	----
3	3	40	0.10	n/s	----	----	n/s	----	----
5	1	6	0.25	over	----	----	over	----	----
5	1	7	0.25	over	----	----	over	----	----
5	1	8	0.25	over	----	----	over	----	----
5	1	9	0.25	over	----	----	over	----	----
5	1	10	0.25	over	----	----	over	----	----
5	1	12	0.25	over	----	----	5847	292	577
5	1	15	0.25	over	----	----	3558	178	1106
5	1	20	0.25	over	----	----	1506	75	1585

Table 16 (continued)

5	1	25	0.25	over	----	----	715	36	1600
5	1	30	0.25	over	----	----	395	20	1662
5	1	35	0.25	over	----	----	235	12	1460
5	1	40	0.25	over	----	----	156	8	1462
5	1	6	0.20	over	----	----	over	----	----
5	1	7	0.20	over	----	----	over	----	----
5	1	8	0.20	over	----	----	over	----	----
5	1	9	0.20	over	----	----	over	----	----
5	1	10	0.20	over	----	----	over	----	----
5	1	12	0.20	over	----	----	over	----	----
5	1	15	0.20	over	----	----	9783	564	20
5	1	20	0.20	over	----	----	8366	417	189
5	1	25	0.20	over	----	----	6765	338	422
5	1	30	0.20	over	----	----	4923	246	770
5	1	35	0.20	over	----	----	2759	138	1260
5	1	40	0.20	over	----	----	1508	75	1558
5	1	6	0.15	over	----	----	over	----	----
5	1	7	0.15	over	----	----	over	----	----
5	1	8	0.15	over	----	----	over	----	----
5	1	9	0.15	over	----	----	over	----	----
5	1	10	0.15	over	----	----	over	----	----
5	1	12	0.15	over	----	----	over	----	----
5	1	15	0.15	over	----	----	over	----	----
5	1	20	0.15	over	----	----	over	----	----
5	1	25	0.15	over	----	----	over	----	----
5	1	30	0.15	over	----	----	over	----	----
5	1	35	0.15	over	----	----	over	----	----
5	1	40	0.15	over	----	----	9783	465	20
5	1	6	0.10	n/s	----	----	n/s	----	----
5	1	7	0.10	n/s	----	----	n/s	----	----
5	1	8	0.10	n/s	----	----	n/s	----	----
5	1	9	0.10	n/s	----	----	n/s	----	----
5	1	10	0.10	n/s	----	----	n/s	----	----
5	1	12	0.10	n/s	----	----	n/s	----	----
5	1	15	0.10	n/s	----	----	n/s	----	----
5	1	20	0.10	n/s	----	----	n/s	----	----
5	1	25	0.10	n/s	----	----	n/s	----	----
5	1	30	0.10	n/s	----	----	n/s	----	----
5	1	35	0.10	n/s	----	----	n/s	----	----
5	1	40	0.10	n/s	----	----	n/s	----	----
5	2	6	0.25	over	----	----	over	----	----
5	2	7	0.25	over	----	----	sing	----	----
5	2	8	0.25	over	----	----	sing	----	----
5	2	9	0.25	over	----	----	sing	----	----
5	2	10	0.25	over	----	----	sing	----	----
5	2	12	0.25	over	----	----	1780	89	1485
5	2	15	0.25	over	----	----	859	43	1433
5	2	20	0.25	over	----	----	333	17	1482

Table 16 (continued)

5	2	25	0.25	over	----	----	168	8	1508
5	2	30	0.25	sing	----	----	94	5	1439
5	2	35	0.25	unk	----	----	60	3	1483
5	2	40	0.25	sing	----	----	41	2	1524
5	2	6	0.20	over	----	----	over	----	----
5	2	7	0.20	over	----	----	over	----	----
5	2	8	0.20	over	----	----	over	----	----
5	2	9	0.20	over	----	----	over	----	----
5	2	10	0.20	over	----	----	over	----	----
5	2	12	0.20	over	----	----	over	----	----
5	2	15	0.20	over	----	----	over	----	----
5	2	20	0.20	over	----	----	sing	----	----
5	2	25	0.20	over	----	----	sing	----	----
5	2	30	0.20	over	----	----	1173	59	1365
5	2	35	0.20	over	----	----	588	29	1431
5	2	40	0.20	over	----	----	337	17	1534
5	2	6	0.15	over	----	----	over	----	----
5	2	7	0.15	over	----	----	over	----	----
5	2	8	0.15	over	----	----	over	----	----
5	2	9	0.15	over	----	----	over	----	----
5	2	10	0.15	over	----	----	over	----	----
5	2	12	0.15	over	----	----	over	----	----
5	2	15	0.15	over	----	----	over	----	----
5	2	20	0.15	over	----	----	over	----	----
5	2	25	0.15	over	----	----	over	----	----
5	2	30	0.15	over	----	----	over	----	----
5	2	35	0.15	over	----	----	over	----	----
5	2	40	0.15	over	----	----	over	----	----
5	2	6	0.10	over	----	----	over	----	----
5	2	7	0.10	over	----	----	over	----	----
5	2	8	0.10	over	----	----	over	----	----
5	2	9	0.10	over	----	----	over	----	----
5	2	10	0.10	over	----	----	over	----	----
5	2	12	0.10	over	----	----	over	----	----
5	2	15	0.10	over	----	----	over	----	----
5	2	20	0.10	over	----	----	over	----	----
5	2	25	0.10	over	----	----	over	----	----
5	2	30	0.10	over	----	----	over	----	----
5	2	35	0.10	over	----	----	over	----	----
5	2	40	0.10	over	----	----	over	----	----
5	3	6	0.25	n/s	----	----	sing	----	----
5	3	7	0.25	n/s	----	----	sing	----	----
5	3	8	0.25	n/s	----	----	sing	----	----
5	3	9	0.25	n/s	----	----	1102	55	1633
5	3	10	0.25	n/s	----	----	773	39	1508
5	3	12	0.25	n/s	----	----	446	20	1910
5	3	15	0.25	n/s	----	----	234	12	1684
5	3	20	0.25	n/s	----	----	104	5	1461

Table 16 (continued)

5	3	25	0.25	n/s	----	----	54	3	1307
5	3	30	0.25	n/s	----	----	35	2	1440
5	3	35	0.25	n/s	----	----	24	1	1499
5	3	40	0.25	n/s	----	----	17	1	1452
5	3	6	0.20	n/s	----	----	over	----	----
5	3	7	0.20	n/s	----	----	over	----	----
5	3	8	0.20	n/s	----	----	over	----	----
5	3	9	0.20	n/s	----	----	over	----	----
5	3	10	0.20	n/s	----	----	7386	369	316
5	3	12	0.20	n/s	----	----	6125	306	513
5	3	15	0.20	n/s	----	----	3733	187	1064
5	3	20	0.20	n/s	----	----	1470	73	1413
5	3	25	0.20	n/s	----	----	588	29	1616
5	3	30	0.20	n/s	----	----	291	15	1556
5	3	35	0.20	n/s	----	----	166	8	1594
5	3	40	0.20	n/s	----	----	102	5	1490
5	3	6	0.15	n/s	----	----	n/s	----	----
5	3	7	0.15	n/s	----	----	n/s	----	----
5	3	8	0.15	n/s	----	----	n/s	----	----
5	3	9	0.15	n/s	----	----	n/s	----	----
5	3	10	0.15	n/s	----	----	n/s	----	----
5	3	12	0.15	n/s	----	----	n/s	----	----
5	3	15	0.15	n/s	----	----	n/s	----	----
5	3	20	0.15	n/s	----	----	n/s	----	----
5	3	25	0.15	n/s	----	----	n/s	----	----
5	3	30	0.15	n/s	----	----	n/s	----	----
5	3	35	0.15	n/s	----	----	n/s	----	----
5	3	40	0.15	n/s	----	----	n/s	----	----
5	3	6	0.10	n/s	----	----	n/s	----	----
5	3	7	0.10	n/s	----	----	n/s	----	----
5	3	8	0.10	n/s	----	----	n/s	----	----
5	3	9	0.10	n/s	----	----	n/s	----	----
5	3	10	0.10	n/s	----	----	n/s	----	----
5	3	12	0.10	n/s	----	----	n/s	----	----
5	3	15	0.10	n/s	----	----	n/s	----	----
5	3	20	0.10	n/s	----	----	n/s	----	----
5	3	25	0.10	n/s	----	----	n/s	----	----
5	3	30	0.10	n/s	----	----	n/s	----	----
5	3	35	0.10	n/s	----	----	n/s	----	----
5	3	40	0.10	n/s	----	----	n/s	----	----

APPENDIX G

DISTILLATION COLUMN SIMULATOR DATA

Table 17 presents the numeric data obtained from the distillation column simulator described in Section 7.3. Two clock numbers are shown. The first gives the original time stamp from the simulation run and the second is simply the renumbering applied to the extracted portion of the simulation run that was used for the analysis of Section 7.3. The upper control limit for the T^2 statistic is 12.407. The upper control limit for the W_i and W_R statistics is 14.156.

Table 17 MeOH Concentrations from Distillation Column Simulator

Original Clock (msec)	Clock	Bottom MeOH	Overhead MeOH
172000	1	0.034145	0.64001
171950	2	0.034579	0.6443
171900	3	0.034249	0.64146
171850	4	0.034878	0.64302
171800	5	0.034324	0.64112
171750	6	0.034023	0.63943
171700	7	0.034162	0.64214
171650	8	0.03394	0.64441
171600	9	0.033991	0.64041
171550	10	0.03421	0.64062
171500	11	0.03471	0.64236
171450	12	0.033946	0.64418
171400	13	0.033966	0.64014
171350	14	0.033805	0.6421

Table 17 (continued)

171300	15	0.034147	0.6416
171250	16	0.03461	0.64577
171200	17	0.034078	0.64256
171150	18	0.03452	0.64147
171100	19	0.033926	0.64126
171050	20	0.034316	0.64124
171000	21	0.034445	0.64265
170950	22	0.034417	0.64344
170900	23	0.0347	0.64102
170850	24	0.034784	0.64113
170800	25	0.034246	0.64208
170750	26	0.034189	0.64148
170700	27	0.034125	0.64562
170650	28	0.034073	0.64149
170600	29	0.034376	0.64039
170550	30	0.034646	0.64291
170500	31	0.033575	0.64229
170450	32	0.03445	0.6436
170400	33	0.033728	0.64312
170350	34	0.034362	0.64137
170300	35	0.035116	0.64422
170250	36	0.034628	0.64297
170200	37	0.034585	0.64154
170150	38	0.035505	0.64388
170100	39	0.034354	0.63909
170050	40	0.034349	0.64209
170000	41	0.03508	0.64186
169950	42	0.034452	0.64208
169900	43	0.034433	0.64204
169850	44	0.034638	0.64144
169800	45	0.034724	0.64768
169750	46	0.034707	0.64264
169700	47	0.035154	0.64223
169650	48	0.035002	0.6418
169600	49	0.03435	0.64886
169550	50	0.034268	0.64445
169500	51	0.034812	0.64231
169450	52	0.035008	0.6437
169400	53	0.034356	0.64517
169350	54	0.034971	0.64296
169300	55	0.034941	0.64792
169250	56	0.035411	0.64203
169200	57	0.035488	0.64393

Table 17 (continued)

169150	58	0.035211	0.64441
169100	59	0.035027	0.64502
169050	60	0.035766	0.642
169000	61	0.035652	0.64633
168950	62	0.035308	0.64475
168900	63	0.034794	0.64818
168850	64	0.035074	0.64792
168800	65	0.0358	0.64755
168750	66	0.035664	0.64792
168700	67	0.036962	0.64595
168650	68	0.035843	0.64818
168600	69	0.035872	0.64769
168550	70	0.036698	0.64638
168500	71	0.036614	0.64915
168450	72	0.036597	0.64689
168400	73	0.03726	0.64807
168350	74	0.037573	0.64732
168300	75	0.037453	0.64814
168250	76	0.036891	0.64932
168200	77	0.038092	0.65098
168150	78	0.037972	0.65144
168100	79	0.038335	0.65053
168050	80	0.038429	0.65178
168000	81	0.038734	0.65119
167950	82	0.038809	0.65541
167900	83	0.038884	0.65352
167850	84	0.03907	0.65314
167800	85	0.039488	0.65462
167750	86	0.040306	0.65458
167700	87	0.04131	0.65608
167650	88	0.040354	0.6605
167600	89	0.042099	0.65787
167550	90	0.041696	0.65638
167500	91	0.04238	0.65676
167450	92	0.042278	0.66156
167400	93	0.04375	0.66257
167350	94	0.043813	0.66275
167300	95	0.043982	0.66381
167250	96	0.045205	0.66822
167200	97	0.04516	0.66722
167150	98	0.047408	0.66857
167100	99	0.047325	0.67221
167050	100	0.046933	0.67376

Table 17 (continued)

167000	101	0.048911	0.67304
166950	102	0.049868	0.67547
166900	103	0.051264	0.68143
166850	104	0.05242	0.68225
166800	105	0.054163	0.68643
166750	106	0.055236	0.68571
166700	107	0.056139	0.68696
166650	108	0.058094	0.69269
166600	109	0.059772	0.69815
166550	110	0.059305	0.69955
166500	111	0.062157	0.70094
166450	112	0.065354	0.70506
166400	113	0.067415	0.70766
166350	114	0.068203	0.71365
166300	115	0.071422	0.7234
166250	116	0.075171	0.72584
166200	117	0.076576	0.73129
166150	118	0.078391	0.73478
166100	119	0.082642	0.73992
166050	120	0.085402	0.74381
166000	121	0.088628	0.75315
165950	122	0.094039	0.7628
165900	123	0.097461	0.77378
165850	124	0.10181	0.77758
165800	125	0.1061	0.78352
165750	126	0.10879	0.79257
165700	127	0.11558	0.80347
165650	128	0.12213	0.81009
165600	129	0.12443	0.81913
165550	130	0.13244	0.8285
165500	131	0.13636	0.83065
165450	132	0.14039	0.83736
165400	133	0.14435	0.83965
165350	134	0.14757	0.84489
165300	135	0.1541	0.85102
165250	136	0.15301	0.85322
165200	137	0.15151	0.854
165150	138	0.15451	0.85492
165100	139	0.15133	0.85126
165050	140	0.15026	0.85048
165000	141	0.14858	0.84514
164950	142	0.14959	0.85148
164900	143	0.15203	0.84785

Table 17 (continued)

164850	144	0.15076	0.848
164800	145	0.1491	0.85074
164750	146	0.14899	0.85167
164700	147	0.14916	0.85211
164650	148	0.1505	0.85188
164600	149	0.15069	0.84931
164550	150	0.15116	0.84485
164500	151	0.15126	0.8484
164450	152	0.15021	0.85289
164400	153	0.14854	0.84972
164350	154	0.1517	0.85057
164300	155	0.14838	0.84369
164250	156	0.14871	0.85385
164200	157	0.15382	0.84969
164150	158	0.15434	0.84951
164100	159	0.15315	0.85057
164050	160	0.15171	0.8478
164000	161	0.15065	0.84945
163950	162	0.15024	0.85107
163900	163	0.15381	0.84797
163850	164	0.14989	0.84998
163800	165	0.1503	0.85403
163750	166	0.15058	0.85164
163700	167	0.14957	0.85452
163650	168	0.14902	0.84702
163600	169	0.1506	0.85049
163550	170	0.15194	0.8509
163500	171	0.14909	0.84819
163450	172	0.14994	0.84891
163400	173	0.15187	0.84835
163350	174	0.14961	0.84798
163300	175	0.1507	0.85312
163250	176	0.15104	0.85208
163200	177	0.14901	0.85274
163150	178	0.15149	0.84791
163100	179	0.151	0.85183
163050	180	0.15276	0.84831
163000	181	0.15061	0.8504
162950	182	0.14901	0.85364
162900	183	0.1518	0.85074
162850	184	0.14943	0.84966
162800	185	0.15121	0.84656
162750	186	0.15245	0.85193

Table 17 (continued)

162700	187	0.15044	0.84978
162650	188	0.15106	0.8492
162600	189	0.1535	0.85174
162550	190	0.14954	0.84616
162500	191	0.15218	0.8503
162450	192	0.14975	0.85129
162400	193	0.15193	0.84754
162350	194	0.15009	0.85416
162300	195	0.14878	0.85024
162250	196	0.14841	0.84987
162200	197	0.15074	0.84962
162150	198	0.15303	0.84659
162100	199	0.15	0.84995
162050	200	0.15126	0.85226
162000	201	0.15213	0.85287
161950	202	0.14966	0.85038
161900	203	0.15153	0.85028
161850	204	0.1505	0.85302
161800	205	0.14936	0.85185
161750	206	0.14753	0.85084
161700	207	0.15059	0.84896
161650	208	0.15198	0.85011
161600	209	0.15136	0.85074
161550	210	0.14976	0.84831
161500	211	0.15075	0.8488
161450	212	0.15047	0.84944
161400	213	0.15185	0.84976
161350	214	0.15149	0.85249
161300	215	0.14969	0.85204
161250	216	0.15274	0.84946
161200	217	0.15058	0.85014
161150	218	0.15093	0.84703
161100	219	0.15096	0.85087
161050	220	0.14934	0.85074
161000	221	0.15242	0.84996
160950	222	0.15383	0.8489
160900	223	0.15287	0.85137
160850	224	0.15318	0.84901
160800	225	0.14989	0.85103
160750	226	0.14977	0.85092
160700	227	0.14905	0.84904
160650	228	0.1495	0.852
160600	229	0.14894	0.85265

Table 17 (continued)

160550	230	0.15226	0.85101
160500	231	0.15091	0.852
160450	232	0.1515	0.85173
160400	233	0.15021	0.85609
160350	234	0.15404	0.84768
160300	235	0.1514	0.84656
160250	236	0.1499	0.85166
160200	237	0.14937	0.85205
160150	238	0.14981	0.84932
160100	239	0.14881	0.8492
160050	240	0.14953	0.8486
160000	241	0.14907	0.85116
159950	242	0.15146	0.84959
159900	243	0.14763	0.85038
159850	244	0.15163	0.85167
159800	245	0.15088	0.85229
159750	246	0.15219	0.85287
159700	247	0.14965	0.85064
159650	248	0.15082	0.85095
159600	249	0.15048	0.85032
159550	250	0.15082	0.84912
159500	251	0.15071	0.85596
159450	252	0.1514	0.85157
159400	253	0.14953	0.84988
159350	254	0.15267	0.85089
159300	255	0.15332	0.84946
159250	256	0.1536	0.84974
159200	257	0.14925	0.85202
159150	258	0.15021	0.84919
159100	259	0.14967	0.85026
159050	260	0.15281	0.85182
159000	261	0.15231	0.85453
158950	262	0.15303	0.85298
158900	263	0.15127	0.84966
158850	264	0.14855	0.85099
158800	265	0.15227	0.85185
158750	266	0.14996	0.85221
158700	267	0.15188	0.84979
158650	268	0.15137	0.84991
158600	269	0.15012	0.85058
158550	270	0.15009	0.8496
158500	271	0.15229	0.84796
158450	272	0.15242	0.85702

Table 17 (continued)

158400	273	0.14827	0.85193
158350	274	0.14929	0.85327
158300	275	0.15227	0.84876
158250	276	0.14998	0.84955
158200	277	0.15315	0.84833
158150	278	0.15037	0.8485
158100	279	0.15109	0.85061
158050	280	0.14918	0.84962
158000	281	0.14879	0.84885

BIBLIOGRAPHY

Alt, F., "Aspects of Multivariate Control Charts," M. S. Thesis, Georgia Institute of Technology, Atlanta, Georgia, 1973.

Alt, F., "Multivariate Statistical Quality Control," in The Encyclopedia of Statistical Sciences, edited by Kotz, S., Johnson, N., and Read, C., (New York: John Wiley, 1985): 110-122.

Anderson, T. An Introduction to Multivariate Statistical Analysis, (New York: John Wiley & Sons, 1984.

Crossier, R., "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," Technometrics, Vol. 30, No. 3, (August 1988): 291-303.

Crowder, S., "Design of Exponentially Weighted Moving Average Schemes," Journal of Quality Technology, Vol. 21, (1989): 135-162.

DeVore, Jay L., Probability and Statistics for Engineering and the Sciences, 5th Edition, (Pacific Grove, CA: Duxbury, 2000).

Doganoksoy, N., Faltin, F., and Tucker, W., "Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment," Communications in Statistics—Theory and Methods, Vol. 20, No. 9, (1991): 2775-2790.+

Doyle, F., Gatzke, E., and Parker, R., Process Control Modules, (Newark, DE: University of Delaware, Department of Chemical Engineering, 1999).

Duncan, A., Quality Control and Industrial Statistics, 5th ed., (Chicago: Irwin, 1986).

Flury, B. and Riedwyl, H., Multivariate Statistics: A Practical Approach, (London: Chapman & Hall, 1988).

Golnabi, S., and Houshmand, A., "Monitoring Mulivariate Process Variability," IIE 6th Industrial Engineering Research Conference Proceedings, 66-71.

Hald, A., Statistical Theory with Engineering Applications, (New York: John Wiley & Sons, 1952).

Hayter, A. and Tsui, K., "Identification and Quantification in Multivariate Quality Control Problems," Journal of Quality Technology, Vol. 26, No. 3, (July 1994): 197-208.

Hawkins, D., "Multivariate Quality Control Based on Regression-Adjusted Variables," Technometrics, Vol. 33, No. 1, (February 1991): 61-75.

Hawkins, D., "Regression Adjustment for Variables in Multivariate Quality Control," Journal of Quality Technology, Vol. 25, No. 3, (July 1993): 170-182.

Holmes, D. and Mergen, A., "Improving the Performance of the T^2 Control Chart," Quality Engineering, Vol. 5, No. 4, (1993): 619-625.

Holmes, D. and Mergen, A., "A Multivariate Test for Randomness," Quality Engineering, Vol. 10, No. 3, (1998): 505-508.

Hotelling, H., "Multivariate Quality Control, Illustrated by the Air Testing of Sample Bombsights," in Techniques of Statistical Analysis, edited by C. Eisenhart, M. Hastay, and W. Wallis, (New York: McGraw-Hill, 1947): 113-184.

Houshmand, A., Golnabi, S., and Frosh, M., "Generalization of Shewhart X-bar Chart," IEEE Proceedings of Annual Reliability and Maintainability Symposium, (1997): 314-319.

Jackson, J., "Quality Control Methods for Several Related Variables," Technometrics, Vol. 1, No. 4, (November, 1959): 359-377.

Jackson, J., "Principal Components and Factor Analysis: Part I—Principal Components," Journal of Quality Technology, Vol. 12, No. 4, (October, 1980): 201-213.

Jackson, J., "Multivariate Quality Control," Communications in Statistics—Theory and Methods, Vol. 14, No. 11, (1985): 2657-2688.

Johnson, R., and Wichern, D., Applied Multivariate Statistical Analysis, 2nd Ed., (Englewood Cliffs, NJ: Prentice-Hall, 1988).

King, B., "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, Vol. 39, No. 1, Part 2: Supplement on Security Prices, (January 1966): 139-190.

Kramer, C., and Jensen, D., "Fundamentals of Multivariate Analysis Part I. Inference About Means," Journal of Quality Technology, Vol. 1, No. 2, (April 1969): 120-133.

Kramer, C., and Jensen, D., "Fundamentals of Multivariate Analysis Part II. Inference About Two Treatments," Journal of Quality Technology, Vol. 1, No. 3, (1969): 189-204.+

Law, A., and Kelton, W., Simulation Modeling and Analysis, 3rd Ed., (Boston: McGraw-Hill, 2000).

Levison, W., Holmes, D., and Mergen, A., "Variation Charts for Mulitvariate Processess," Quality Engineering, Vol. 14, No. 4, (2002): 539-545.

Lowry, C., and Montgomery, D., "A Review of Multivariate Control Charts," IIE Transactions, Vol. 27, (1995): 800-810.

Lowry C., Woodall, W., Champ, C., & Rigdon, S., "A Multivariate Exponentially Weighted Moving Average Control Chart," Technometrics, Vol. 34, No. 1, (February 1992): 46-53.

Lucas, J. and Saccuci, M., "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," Technometrics, Vol. 32, (1990): 1-12.

Mason, R., Tracy, N., and Young, J., "Decomposition of T^2 for Multivariate Control Chart Interpretation," Journal of Quality Technology, Vol. 27, No. 2, (April 1995): 99-108.

Montgomery, D., Introduction to Statistical Quality Control, 3rd ed., (New York: John Wiley & Sons, 1997).

Montgomery, D., and Wadsworth, H., "Some Techniques for Multivariate Quality Control Applications," ASQC Technical Conference Transactions, Washington, DC, (1972).

Myers, R., Classical and Modern Regression with Applications, 2nd ed., (Belmont, CA: Duxbury Press, 1990).

Nelson L., "The Shewhart Control Chart—Tests for Special Causes," Journal of Quality Technology, Vol. 16, No. 4, (October 1984): 237-239.

Pignatello, J. and Runger, G., "Comparisons of Multivariate CUSUM Charts," Journal of Quality Technology, Vol. 22, No. 3, (July 1990): 173-186.

Sullivan, J. and Woodall, W., "A Comparison of Multivariate Control Charts for Individual Observations," Journal of Quality Technology, Vol. 28, No. 4, (October 1996): 398-408.

Wade, M., and Woodall, W., "A Review and Analysis of Cause-Selecting Control Charts," Journal of Quality Technology, Vol. 25, No. 3, (July 1993): 161-169.

Wilks, S., "Certain Generalizations in the Analysis of Variance," Biometrika, Vol. 24, (1932): 471-494.

Woodall, W., and Ncube, M., "Multivariate CUSUM Quality-Control Procedures," Technometrics, Vol. 27, (August 1985): 285-292.