ABSTRACT

Title of dissertation: A New Scheme for Monitoring Multivariate Process Dispersion

Xin Song, Doctor of Philosophy, 2009

Dissertation directed by: Professor Paul J. Smith
Department of Mathematics

Construction of control charts for multivariate process dispersion is not as straightforward as for the process mean. Because of the complexity of out of control scenarios, a general method is not available.

In this dissertation, we consider the problem of monitoring multivariate dispersion from two perspectives. First, we derive asymptotic approximations to the power of Nagao's test for the equality of a normal dispersion matrix to a given constant matrix under local and fixed alternatives. Second, we propose various unequally weighted sum of squares estimators for the dispersion matrix, particularly with exponential weights. The new estimators give more weights to more recent observations and are not exactly Wishart distributed. Satterthwaite's method is used to approximate the distribution of the new estimators.

By combining these two techniques based on exponentially weighted sums of squares and Nagao's test, we are able to propose a new control scheme MTNT, which is easy to implement. The control limits are easily calculated since they only depend on the dimension of the process and the desired in control average run length. Our

simulations show that compared with schemes based on the likelihood ratio test and the sample generalized variance, MTNT has the shortest out of control average run length for a variety of out of control scenarios, particularly when process variances increase.

A New Scheme for Monitoring Multvariate Process Dispersion

by

Xin Song

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Paul J. Smith, Chair/Advisor
Professor Frank B. Alt
Professor Benjemin Kedem
Professor Eric V. Slud
Professor Abram Kagan

To my parents, my brother and Jue Wang

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Paul Smith for his invaluable help and guidance on my thesis over the past four years. He has always made himself available for help and advice and there has never been an occasion when I've asked for help and he declined. It has been a pleasure to work with and learn from such an extraordinary individual.

I would also like to thank Professor Frank B. Alt. At the early stage of my research, he offered tremendous help.

Thanks are aslo due to Professor Eric V. Slud, Professor Benjemin Kedem and Professor Abram Kagan for agreeing to serve on my thesis committee and for sparing their time reviewing the manuscript.

My colleagues in the Mathematics Department have enriched my graduate life in many ways and deserve a special mention. Through the years, I have benifited greatly from many stimulating conversations with N. Ganesh, Denise Sam, Zhiwei Chen, Yabing Mai, Justin Brody, Greg Johnson.

I would also like to acknowledge help and support from some of the staff members, Haydee Hidalgo, Linette Berry, Tony Zhang, William Schildknecht, Sharon Welton, Celeste Regalado. Your effortless help has made my life in past years so enjoyable.

I owe my deepest thanks to my family - my mother, father and brother who

have always stood by me and guided me through my career, and have pulled me through against impossible odds at times. Words cannot express the gratitude I owe them.

Finally I would like to thank Jue Wang. Your love and encouragement have helped me through the hardest time.

It is impossible to remember all, and I apologize to those I've inadvertently left out.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SPC	Statistical Process Control.
EWMA	Exponentially Weighted Moving Average Control Chart.
CUSUM	Cumulative Sum Control Chart.
UCL	Upper Control Limit
LCL	Lower Control Limit
ARL	Average Run Length.
LR	Likelihood Ratio.
EWSS	Exponentially Weighted Sum of Squares.
PWSS	Polynomially Weighted Sum of Squares.
MTNT	Multivariate Nagao's Test Chart
MTLR	Multivariate Likelihood Ratio Chart
MTGV	Multivariate Generalized Variance Chart

Chapter 1

Introduction

Statistical Process Control (SPC) is a discipline that uses statistical techniques to measure and analyze the change in the mean and variance in processes. It is a primary tool for quality improvement. SPC began as a means for monitoring product quality in a manufacturing process and now its use has been extended to other areas like software design, health care, education, etc. Typically there are two sources that contribute to the variation in the process, the common causes that are inherent to the process and the assignable (special) causes that can be attributed to outside sources. Common causes happen by chance, are expected to happen and are not correctable. For example in metalworking, small measurement variation would happen because of the limited precision of a measurement gauge. Assignable causes, on the other hand, are unanticipated, not expected to happen and correctable. For example, an abnormal batch of product might be produced because of a sudden increase of temperature in a chemical reaction process. By collecting data from samples at various points within the production process, assignable causes that affect the quality of the end product or service can be detected and corrected.

A typical SPC implementation cycles through the following steps:

1. Depending on the goal and characteristics of the monitored process, pick an appropriate control scheme. Determine control limits.

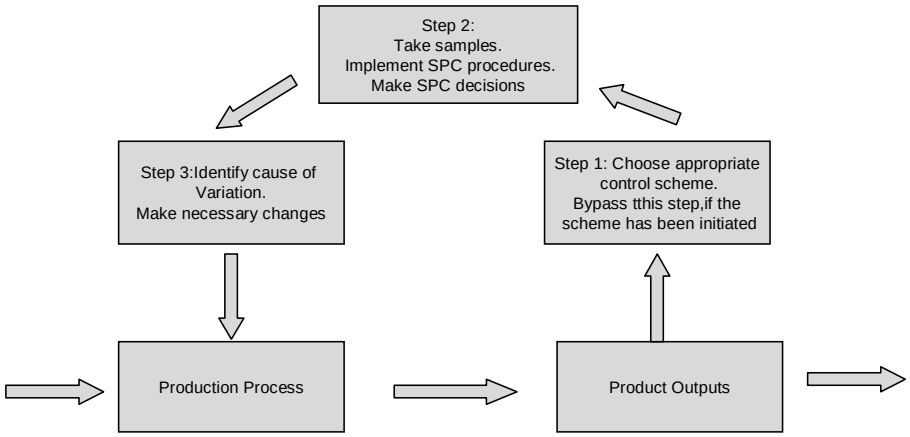

Figure 1.1: Graphic Representation of a SPC algorithm.

2. Collect samples from the process outputs. Implement the scheme with the collected samples.

3. If scheme indicates in-control, return to Step 2. If an out of control signal is given, identify the root cause of change. Make changes to the production process if it is deemed that some process attributes have changed.

A graphic representation for realizing the statistical process control is presented above in Figure 1.1.

There are four kinds of process shifting patterns, namely, the step shift, impulse shift, spike shift and trend shift, as in [53]. Figure 1.2 shows these four shifting patterns. The spike shift and the impulse shift represent situations when the process undergoes a shock and restores itself to the original condition. These typically are not concerns of SPC since if the process has the ability to correct itself then use of SPC is probably redundant. This is especially true for the spike shift case. The step shift represents the situation when the process has an instant and permanent

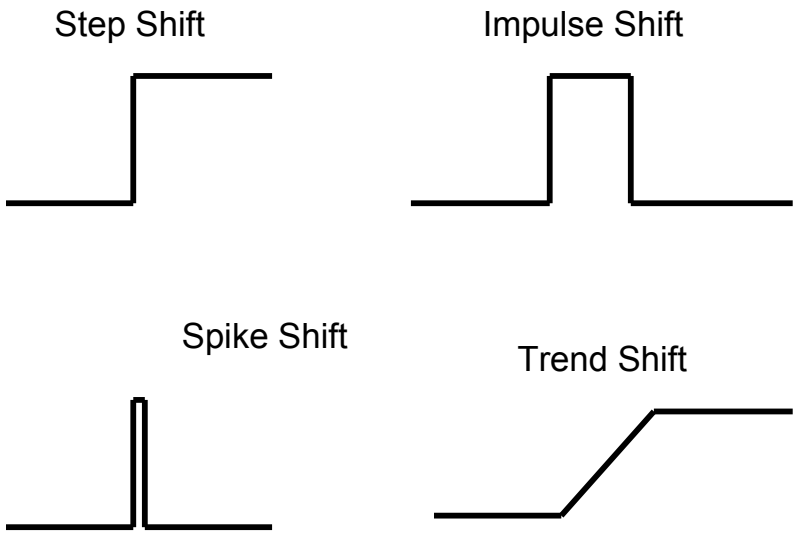

Figure 1.2: Four kinds of Shifting patterns.

shift and trend shifting means a gradual and permanent shift. In this paper we shall focus on the step shifting scenario, but the proposed new control scheme also applies to the trend shifting pattern.

The primary tool used for SPC is the control chart, which is a graphic representation of certain descriptive statistics for some specific quantitative measurements of the production process. Shewhart pioneered control charts in the early 1920s, and SPC has ever since received considerable attention. Many new methods have been proposed and studied, such as the exponentially weighted moving average (EWMA) chart and the Cumulative Sum (CUSUM) Chart. The Shewhart chart was initially intended for the monitoring of a univariate process mean, but now SPC applications have expanded from process mean to process variability, from univariate process to multivariate process, from $i.i.d.$ observations to correlated processes, and from continuous variables to attribute and count data.

In the univariate case, the well-known R, S, and S^2 charts are the standard

charts for the purpose of monitoring process variability. Various alternative techniques for monitoring process dispersion have been introduced and have gained wide acceptance in practice. Page [35] proposed a CUSUM chart based on the sample range for monitoring univariate normal process variance. Hawkins [14] discussed multivariate control schemes based on regression adjusted variables. Huang and Chen [15] proposed a synthetic control chart for monitoring process dispersion with sample standard deviation. There is also the method of utilizing the determinant of the sample covariance matrix $|S|$ as a multivariate dispersion measure. An example of this is Alt and Smith [3]. Yeh *et al.* [58] apply a likelihood ratio based EWMA control chart for monitoring variability of multivariate normal processes.

While extension of these techniques to the multivariate case is of great importance in practice, control chart procedures for monitoring the covariance of a multivariate process have received very little attention. Unlike the problem of monitoring multivariate process mean, it is not easy to define the changes in the covariance matrix that need to be detected. For a p dimensional multivariate process variance-covariance matrix, there are a total of $2^{p(p+1)/2} - 1$ ways that the matrix can change. Because of this, even though the designated multivariate control procedure for dispersion can issue a signal, it is difficult to identify the out-of-control process parameter(s). Also monitoring multivariate process dispersion tends to be computational intensive.

This dissertation is organized as follows. In Chapter 2, the three most widely used control schemes, the Shewhart chart, EWMA chart, CUSUM chart, and their properties will be reviewed. In Chapter 3, we examine the asymptotic distributions

of the traditional likelihood ratio test and another test proposed by Nagao for the equality of population variance covariance matrix to a given matrix under both local and fixed alternatives. We will then discuss the weighted moving average type of process variance covariance matrix estimator and its approximated distribution using Satterthwaite's method. In Chapter 4, a new control scheme to monitor multivariate process variability will be given, and finally some numerical studies to compare the proposed algorithm with existing ones. In Chapter 5, we summarize the results of this thesis and suggests topics for future research.

Chapter 2

Traditional Control Schemes and Average Run Length

2.1 Shewhart Control Chart

The simplest form of the control chart is the Shewhart chart ([43], 1931). In a Shewhart chart for monitoring the level of a univariate process, an observation at time t, x_t, is used to indicate whether or not the process has undergone some shift in its level. Assume that a sequence of observations x_1,x_2,... of a quality characteristic are independently distributed with mean μ and variance σ^2. The Shewhart control chart plots these individual observations against the control limits:

$$UCL = \mu + k\sigma$$

$$C = \mu$$

$$LCL = \mu - k\sigma,$$

LCL is the lower control limit, UCL is the upper control limit, [LCL, UCL] defines the in control region centered at the target value $C = \mu$, and the constant k is used to control the average run length(ARL). A value of $k = 3$ gives an in control average run length of 370. The Shewhart chart is easy to implement and sensitive to large shifts in the process mean. However it is slow to detect small or moderate changes.

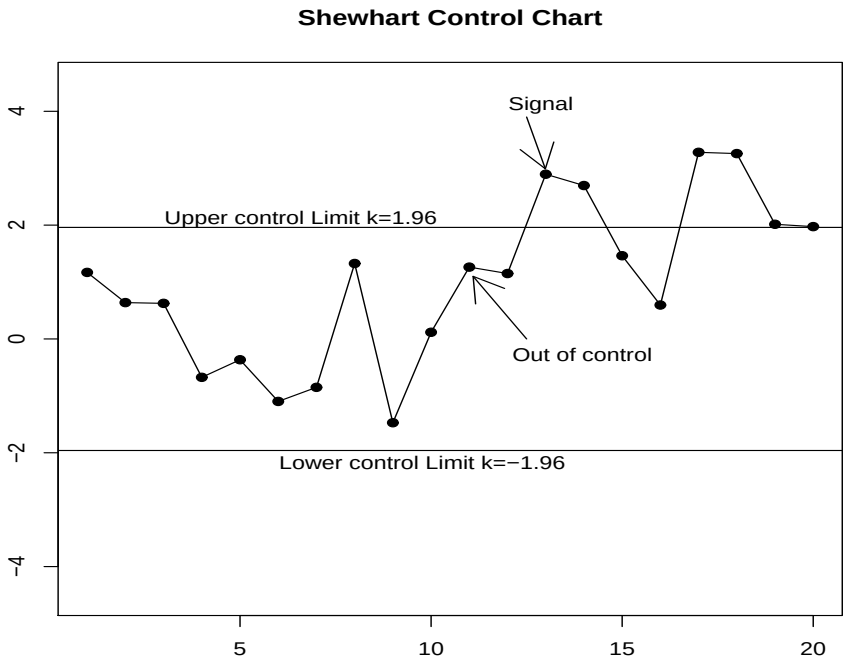

Figure 2.1: Target mean=0, variance=1. Mean shifts to 2 at the 11th observation, signal given at the 13th observation.

2.2 CUSUM Control Chart

The CUSUM control scheme was first proposed by Page ([35], 1954). Since then it has been widely researched and applied. It is a scheme that accumulates and monitors the deviations of the sample observations from a predetermined reference value r. Let the accumulated deviations be y_t, then

$$y_t = x_t - r + y_{t-1}$$

x_t could be a single observation at time t or the average of subgroup sampled at time t. Graphically, when the process is in control, the plot of y_t will be roughly a line with slope $\mu - r$. Hence if the mean of the process is larger than the reference value, the mean path is upward sloping. On the other hand, if the mean of the process is larger than the reference value, the sequence will show a downward sloping trend.

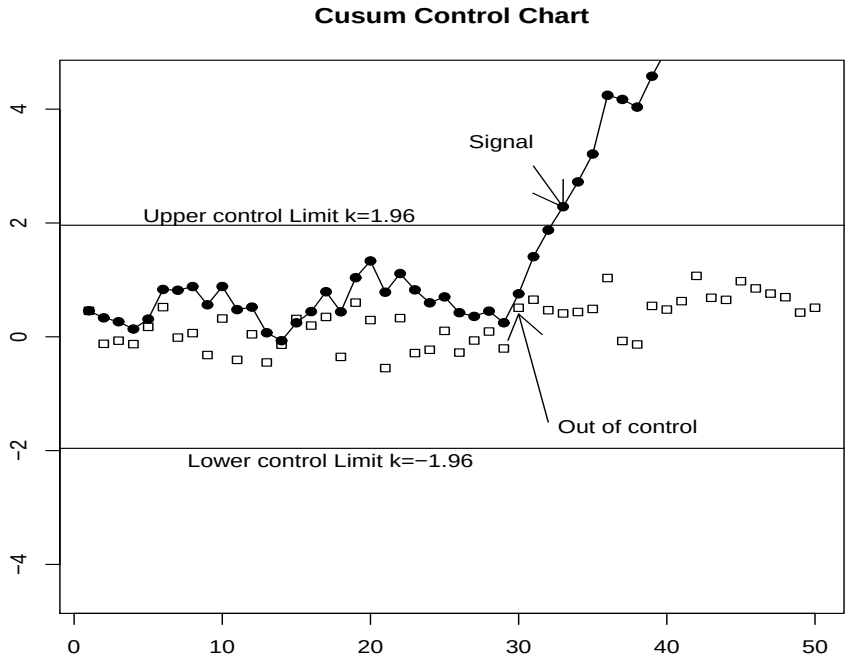

Figure 2.2: Target mean=0, variance=1, squares are the raw means of subgroups, and solid dots are the CUSUMs. Mean shifts to 0.5 at the 30th subgroup, and signal is given at the 33rd subgroup. Plot shows a clear change of slope after the mean shift.

A sudden change of slope is indication of change of process level. The CUSUM procedure can be derived from the perspective of the sequential probability ratio test. This derivation indicates that the value of h should be chosen as one-half of the shift in the level which should be detected quickly. CUSUM Control chart is sensitive to small shifts, and is not as effective as the Shewhart chart in detecting large shifts.

2.3 EWMA Control Chart

The exponentially weighted moving average (EWMA) control chart uses a moving average of observations of the process as the control statistic and is more sensitive than the Shewhart chart to small shifts in the level of the process. EWMA

was first described by Roberts ([39], 1959). The EWMA statistic is obtained recursively as

$$y_t = (1-\lambda)y_{t-1} + \lambda x_t, \qquad \text{for t = 1, 2, } \ldots. \tag{2.1}$$

The starting value for the statistic, y_0, can be set as the target parameter value or the average of the historical data. The constant λ determines the length of memory of the EWMA statistic. By expanding the right hand side of Equation 2.1.

$$\begin{aligned} y_t &= \lambda x_t + \lambda(1-\lambda)x_{t-1} + \cdots + \lambda(1-\lambda)^{t-1}x_1 + (1-\lambda)^t y_0 \\ &= \sum_{i=0}^{t-1} \lambda(1-\lambda)^i x_{t-i} + (1-\lambda)^t y_0 \\ &= \sum_{i=1}^{t} \lambda(1-\lambda)^{t-i} x_i + (1-\lambda)^t y_0 \end{aligned}$$

The sum of weights is

$$\sum_{i=0}^{t} w_i = 1 - (1-\lambda)^t + (1-\lambda)^t = 1$$

If $x_t,\ t = 1, 2, \ldots$ are $i.i.d.$ with mean μ and variance σ^2, then

$$\begin{aligned} &\mathrm{E}(y_t) = \mu \\ &\mathrm{Var}(y_t) = \frac{\lambda[1-(1-\lambda)^{2t}]}{2-\lambda}\sigma^2 \end{aligned}$$

If we let the process run long enough, that is as $t \to \infty$, y_t has asymptotic variance $\lambda\sigma^2/(2-\lambda)$. The EWMA control chart limits then are

$$\begin{aligned} &UCL = \mu + k\lambda\sqrt{\sigma/(2-\lambda)} \\ &Center = \mu \\ &LCL = \mu - k\lambda\sqrt{\sigma/(2-\lambda)} \end{aligned}$$

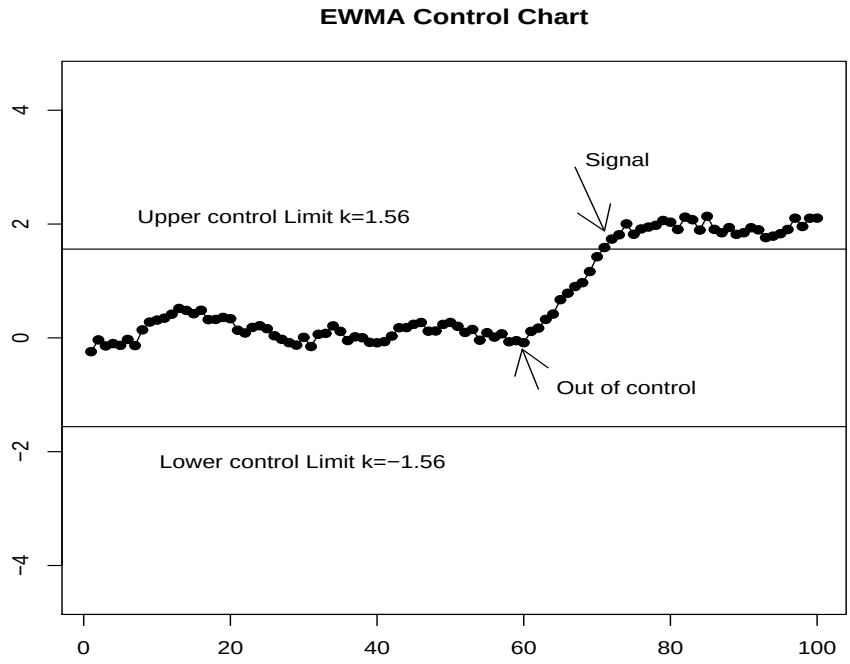

Figure 2.3: Target mean=0, variance=1, $\lambda = 0.1$, mean shifts to 2 at the 60th observation, and the signal is given at the 71th observation.

The factor k is chosen to give a desired in-control average run length. The parameter λ allows practitioners to tune the EWMA control scheme to best fit their needs. A small λ is desirable for small shifts and a large λ is preferred for large shifts, so it is a very flexible chart. When $\lambda = 1$, the scheme reduces to Shewhart control chart.

2.4 Average Run Length(ARL)

The Average Run Length (ARL) is the most frequently used performance measure for selecting and designing SPC procedures. There are two types of ARLs, the in-control ARL and the out-of-control ARL. If the process has no shift, the in-control ARL is the expected number of observations from the control statistic has stabilized until an out-of-control signal is given. If the process undergoes a shift, the out-of-control ARL is the expected number of observations from the time of shift

until an out-of-control signal is given. When one is setting up a process monitoring scheme, it is desirable that the scheme has a large in-control ARL when the process is stable around the target values and small out-of-control ARL when the process has a shift. The in-control and out-of-control ARLs are related to the Type I and Type II error probabilities in hypothesis testing.

Generally speaking, the run length of a control chart R is a non-negative discrete random variable. Then

$$\mathrm{ARL} = \sum_{i=1}^{\infty} \Pr(R \geq r) = 1 + \sum_{i=1}^{\infty} \Pr(R > r)$$

The Shewhart control chart gives a signal once an observation falls outside of the control region. Since all the observations are assumed to be $i.i.d.$, the in-control ARL follows a geometric distribution. Let q be the probability that a single observation falls outside the control limits, then the in-control ARL for the Shewhart scheme is

$$\mathrm{ARL}_{Shewhart} = 1/q$$

However, because of correlation among the chart statistics or high dimensionality, closed form expressions in terms of the design parameters for the ARL of a CUSUM chart, EWMA chart or other charts that monitor multivariate or correlated processes are hard to derive or totally intractable. Their ARLs are often computed as a numerical solution to an integral equation. There are mainly two methods for numerically approximation the ARL, the integral equations approach and the Markov chain approach. Robinson and Ho ([41], 1978) used numerical methods based on Edgeworth expansions also.

2.5 Numerical Treatment of ARL

Crowder ([10], 1987) derived the integral equation approach for approximating the EWMA chart run length conditional upon the outcome of the first observation. Suppose x_t has PDF $f(x)$. Let the average run length function with starting value $y_0 = b$ be $\mathrm{L}(b)$. With the first observation x_1, and if the process goes out of control, the run length is one, while if the process stays in control, then the additional run length will be $\mathrm{L}((1-\lambda)b + \lambda x_1)$. Hence

$$\begin{aligned}
\mathrm{L}(b) &= 1 \times \mathrm{P}(|(1-\lambda)b + \lambda x_1| > h) + \int_{|(1-\lambda)b+\lambda x|<h} (1 + \mathrm{L}((1-\lambda)b + \lambda x))\, f(x)\, dx \\
&= 1 + \int_{|(1-\lambda)b+\lambda x|<h} \mathrm{L}((1-\lambda)b + \lambda x) f(x)\, dx \\
&= 1 + 1/\lambda \int_{-h}^{h} \mathrm{L}(x) f(\frac{x-(1-\lambda)b}{\lambda})\, dx \qquad (2.2)
\end{aligned}$$

This is a Fredholm integral equation of the second kind. Crowder approximated $\mathrm{L}(b)$ by applying the Nystrom method [37] and Gaussian quadrature [5] method. He obtained a linear system by replacing the integral with an appropriate quadrature (Crowder used a Gaussian-Legendre one), considering the equation only at the related nodes. After solving the linear equation system we can evaluate $L(b)$. Further details about this method can be found in [10]. Rao *et al.* [36] were able to establish the uniqueness and convergence of the ARL function as the solution to the integral Equation 2.2.

Brook and Evans ([6], 1972) first used the Markov Chain method to approximate the average run length of the CUSUM charts. Lucas and Saccucci ([25], 1990) used it to analyze the performance of the EWMA control chart. Runger and Prabhu

([42], 1996) extended the method to discuss the properties of MEWMA(Multivariate EWMA) charts. The Markov chain method consists of discretizing the control region into small intervals and treating these intervals as states of a Markov Chain. The charting statistic would move among these states until it goes out of control (be absorbed). Champ and Rigdon ([8], 1991) showed that the Markov chain method is nothing other than a special case of the integral equation approach.

Chapter 3

Test Statistic

In this chapter, we will first briefly review the likelihood ratio test criterion and the test statistic proposed by Nagao ([32],[33],[34]). The distribution of Nagao's test under local and fixed alternatives will be derived in Section 3.1. In Section 3.2, we shall discuss improved estimation of the process variance covariance matrix, Satterthwaite's approximation technique will be used to approximate the distribution of proposed estimators.

In this chapter, we consider the following data structure. Let the vector-valued $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a sequence of independent identically distributed observations from a p-variate normal distribution, $\mathrm{N}_p(\mu, \mathbf{\Sigma})$. These observations represent measurements of p correlated process characteristics, where $\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$. Under the null hypothesis, $\mathbf{\Sigma} = \mathbf{\Sigma}_0$. Both μ and $\mathbf{\Sigma}_0$ are assumed to be known. If we don't know μ and $\mathbf{\Sigma}_0$, then a long burn-in process run is needed to estimate these values. Since we are interested in monitoring process dispersion, without loss of generality, we may assume that $\mu = 0$. The sample variance covariance matrix estimator is assumed to be Wishart distributed and has k degrees of freedom.

3.1 Test criteria

3.1.1 Likelihood Ratio Test and Nagao's Test

Given the observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, we wish to test

$$\mathrm{H}_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0 \qquad \text{vs.} \qquad \mathrm{H}_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0 \tag{3.1}$$

for a given positive definite in control process dispersion matrix $\mathbf{\Sigma}_0$. The likelihood ratio test criterion, by the arguments of Anderson [1], rejects the null hypothesis for small values of

$$\mathrm{LR} = \left(\frac{e}{k}\right)^{kp/2} \left|k\mathbf{S}\mathbf{\Sigma}_0^{-1}\right|^{k/2} \mathrm{etr}\left\{-\left(\frac{k}{2}\right)\Sigma_0^{-1}\mathbf{S}\right\} \tag{3.2}$$

where etr means the exponential of a matrix trace, $k = n - 1$,

$$\mathbf{S} = \sum_{i=1}^{n} \frac{(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T}{k},$$
$$\bar{\mathbf{X}} = \frac{\sum_{i=1}^{n} \mathbf{X}_i}{n}.$$

According to the Wilks' theorem [54], under the null hypothesis, the asymptotic distribution of the log likelihood ratio test statistic T_{LR}

$$\mathrm{T}_{LR} = -2\ln(LR) \tag{3.3}$$

is χ^2 with $p(p+1)/2$ degrees of freedom.

Sugiura ([50], [47]) discussed the asymptotic expansion of the non-null distribution of $\mathbf{T}_{LR}$ under both local and fixed alternatives. He showed that, under local alternatives that converge to the null hypothesis at the rate of $k^{-1/2}$

$$\mathrm{H}_1^{\star} : \mathbf{\Sigma}_0^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1/2} = \mathbf{I} + k^{-1/2}\mathbf{\Theta}, \tag{3.4}$$

$\boldsymbol{\Theta}$ is a symmetric matrix that indicates the size and direction of the shifts, the asymptotic distribution of T_{LR} is χ^2 with $p(p+1)/2$ degrees of freedom and non-centrality parameter $\mathrm{tr}\boldsymbol{\Theta}^2/2$. While under fixed alternatives H_1, he showed that the characteristic function of $k^{-1/2}\mathrm{T}_{LR}$ has the form:

$$\mathrm{C}_{H_1}(t) = \mathrm{E}\Big[\exp(itk^{-1/2}\mathrm{T}_{LR})\Big] = \left(\frac{k}{2e}\right)^{itpk^{1/2}} \frac{\Gamma_p(k/2 - itk^{1/2})}{\Gamma_p(k/2)} \times \frac{|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|^{-itk^{1/2}}}{|\mathbf{I} - 2itk^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|^{k/2-itk^{1/2}}} \tag{3.5}$$

where $\Gamma_p(a)$ is the multivariate Gamma function with the property that

$$\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(a + (1-j)/2)$$

Sugiura did further expansion of (3.5),

$$\begin{aligned}\ln(\mathrm{C}_{H_1}(t)) =& it\sqrt{k}\Big(\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}) - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big) \\ &- k^2\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2 + \mathrm{O}(k^{-1/2}),\end{aligned}$$

which implies that

$$k^{-1/2}\left[\mathrm{T}_{LR} - k\Big(\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}) - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big)\right] \xrightarrow[k\to\infty]{d} \mathrm{N}(0, 2\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2).$$

As noted by Sugiura, the variance quantity

$$\tau^2 = 2\mathrm{tr}\left[(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2\right] \tag{3.6}$$

can be regarded as a measure of the departure from the null hypothesis H_0. This motivated Nagao [34] to propose the following test criterion, which in this paper, we call Nagao's test and abbreviate it as T_{NT}. By replacing $\boldsymbol{\Sigma}$ in τ^2 by its unbiased

estimator $\mathbf{S}$ and multiplying by $k/2$, Nagao's test criterion for the equality of process variance covariance matrix to a given matrix $\mathbf{\Sigma}_0$ is:

$$\mathrm{T}_{NT} = (k/2)\mathrm{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2. \tag{3.7}$$

Nagao's test rejects the null hypothesis when the observed value of T_{NT} is larger than a predetermined threshold.

Nagao ([33],[34]) derived the asymptotic expansion of T_{NT} under the null hypothesis H_0 and concluded that the asymptotic null distribution of T_{NT} when $k = n - 1$ is large is χ^2 with $p(p+1)/2$ degrees of freedom.

3.1.2 Some optimal properties of Nagao's test

In this section, we shall indicate some properties of T_{NT} and sketch briefly the proofs of these results.

Theorem 1 (Invariance) *Nagao's test T_{NT} is invariant with respect to the transformations*

$$\mathbf{Y} = \mathbf{Q}\mathbf{X}$$

where $\mathbf{Q}$ *is a square invertible* $p \times p$ *matrix.*

Suppose the Nagao's test statistic associated with Y is $\mathrm{T}^{\star}_{NT}$. Let the variance covariance matrix of Y and the corresponding estimator be $\mathbf{\Sigma}^{\star}$ and $\mathbf{S}^{\star}$ respectively. Then

$$\mathbf{\Sigma}^{\star} = \mathbf{Q}\mathbf{\Sigma}\mathbf{Q}^T, \qquad \mathbf{S}^{\star} = \mathbf{Q}\mathbf{S}\mathbf{Q}^T.$$

Hence

$$\begin{aligned}
\mathrm{T}^{\star}_{NT} &= (k/2)\mathrm{tr}\Big(\mathbf{S}^{\star}(\mathbf{\Sigma}^{\star})^{-1} - \mathbf{I}\Big)^2 \\
&= (k/2)\mathrm{tr}\Big(\mathbf{Q}\mathbf{S}\mathbf{Q}^T(\mathbf{Q}\mathbf{\Sigma}_0\mathbf{Q}^T)^{-1} - \mathbf{I}\Big)^2 \\
&= (k/2)\mathrm{tr}\Big(\mathbf{Q}\mathbf{S}\mathbf{Q}^T(\mathbf{Q}^T)^{-1}\mathbf{\Sigma}_0^{-1}\mathbf{Q}^{-1} - \mathbf{I}\Big)^2 \\
&= (k/2)\mathrm{tr}\Big(\mathbf{Q}\mathbf{S}\mathbf{\Sigma}_0^{-1}\mathbf{Q}^{-1} - \mathbf{I}\Big)^2 \\
&= (k/2)\mathrm{tr}(\mathbf{S}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2 = \mathrm{T}_{NT}.
\end{aligned}$$

If we let $\mathbf{Q} = \mathbf{\Sigma}_0^{-1/2}$, or in other words, if the observations are normalized, T_{NT} is invariant to such standardization, so that when the in control process dispersion matrix is not $\mathbf{I}$, we may choose to normalize the observations first before carrying out the test. We will replace $\mathbf{S}\mathbf{\Sigma}_0^{-1}$ by the standardized sample variance covariance matrix $\mathbf{S}^{\star}$ and call it the standardized Nagao's test

$$\mathrm{T}_{NT} = (k/2)\mathrm{tr}(\mathbf{S}^{\star} - \mathbf{I})^2$$

where

$$\begin{aligned}
\mathbf{Y}_i &= \mathbf{\Sigma}_0^{-1/2}\mathbf{X}_i \\
\mathbf{S}^{\star} &= \frac{\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T}{n-1} \\
\bar{\mathbf{Y}} &= \frac{\sum_{i=1}^{n}\mathbf{Y}_i}{n}.
\end{aligned}$$

It is to be noted that the log likelihood criterion is also invariant to this kind of transformation. This can be verified by making the substitution in Equation 3.2.

Proposition 1 (Frobenius Norm) *Let the Frobenius norm (distance) between two matrices,* $\mathbf{A}$ *and* $\mathbf{B}$*, be* $\left\|\mathbf{A}-\mathbf{B}\right\|_F$*. Then*

$$T_{NT} = (k/2)\left\|\mathbf{S}-\mathbf{I}\right\|_F^2.$$

This is true since

$$\begin{aligned}\left\|\mathbf{S}-\mathbf{I}\right\|_F &= \sqrt{\mathrm{tr}(\mathbf{S}-\mathbf{I})(\mathbf{S}-\mathbf{I})^T} \\ &= \sqrt{\mathrm{tr}(\mathbf{S}-\mathbf{I})^2}, \quad \text{(Symmetricity of } \mathbf{S},\ \mathbf{I})\end{aligned}$$

hence the result.

The Frobenius norm, or the Frobenius distance between two matrices, is a common and in many ways the simplest way of measuring the discrepancy between two matrices. In matrix factorization analysis, Frobenius distance is often used as a criterion of studying the closeness of approximations of a matrix by the product of two matrices ([44], 2004). Nagao's test T_{NT} is the squared Frobenius distance between the estimated process variance covariance matrix $\mathbf{S}$ and the target identity matrix $\mathbf{I}$, scaled by one half the degrees of freedom of $\mathbf{S}$.

Notice also that

$$\mathrm{T}_{NT} = (k/2)\sum_i\sum_j(\mathbf{S}_{ij}-\mathbf{I}_{ij})^2 = (k/2)\|\mathrm{Vec}(\mathbf{S}-\mathbf{I})\|_E^2,$$

$\|\quad\|_E$ means the Euclidean norm of a vector. Hence Nagao's test T_{NT} is also measure of the Euclidean distance between $\mathrm{Vec}(\mathbf{S})$ and $\mathrm{Vec}(\mathbf{I})$.

Nagao's test is also a function of the eigenvalues of sample variance covariance matrix $\mathbf{S}$. Let the sequence of eigenvalues for the standardized sample variance-

covariance matrix $\mathbf{S}$ be $l_1 \leq l_2 \leq \cdots \leq l_p$. Then

$$\begin{aligned} \mathrm{T}_{NT} &= (k/2)\mathrm{tr}(\mathbf{S}-\mathbf{I})^2 \\ &= (k/2)(\mathrm{tr}\mathbf{S}^2 - 2\mathrm{tr}\mathbf{S} + p) \\ &= (k/2)\Big(\sum_{i=1}^{p} l_i^2 - 2\sum_{i=1}^{p} l_i + p\Big) \\ &= (k/2)\sum_{i=1}^{p}(l_i - 1)^2. \end{aligned}$$

That is to say, T_{NT} is also a measure of the discrepancy between the eigenvalues of the normalized sample variance covariance matrix and the eigenvalues of the identity matrix.

3.2 Asymptotic expansions of likelihood ratio test under fixed alternatives

In Section 3.1.1, we briefly went over the asymptotic distribution of $\mathrm{T}_{LR}/\sqrt{k}$ derived by Sugiura [50] under the fixed alternative $H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0$. In this section, we will try to refine Sugiura's results.

Let $\tau^2 = 2\mathrm{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2$ and define

$$\mathrm{Z} = \left[\mathrm{T}_{LR} - k\left\{\mathrm{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}) - p - \ln|\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}|\right\}\right]/\big(\sqrt{k}\tau\big).$$

Let $\mathrm{tr}_j = \mathrm{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1})^j$ and let $\Phi^{(r)}(z)$ be the rth derivative of the standard normal cumulative distribution function $\Phi(z)$. The first seven derivatives of the standard

normal distribution $\Phi(z)$ can be calculated iteratively as

$$\Phi^{(1)}(z) = \phi(z),$$

$$\Phi^{(2)}(z) = -z\phi(z),$$

$$\Phi^{(3)}(z) = \phi(z)(z^2 - 1),$$

$$\Phi^{(4)}(z) = \phi(z)(3z - z^3),$$

$$\Phi^{(5)}(z) = \phi(z)(z^4 - 6z^2 + 3),$$

$$\Phi^{(6)}(z) = \phi(z)(-z^5 + 10z^3 - 15z),$$

$$\Phi^{(7)}(z) = \phi(z)(z^6 - 15z^4 + 45z^2 - 15).$$

Then Sugiura ([47], 1969) proved that the cumulative distribution function (CDF) G(z) of Z has the form

$$\begin{aligned}
\mathrm{G}(z) =& \Phi(z) - \frac{\phi(z)}{6\sqrt{k}}\left[\frac{3p(p+1)}{\tau} + \frac{4(p + 2\mathrm{tr}_3 - 3\mathrm{tr}_2)(z^2-1)}{\tau^3}\right] \\
&+ \frac{\phi(z)}{72k}\left[\frac{-18p(p+1)(p^2+p+4)}{\tau^2}\right. \\
&\qquad + \frac{24(3z - z^3)}{\tau^4} \times \{p(p^2+p+2) - 3p(p+1)\mathrm{tr}_2 \\
&\qquad\qquad + 2(p^2+p-4)\mathrm{tr}_3 + 6\mathrm{tr}_4\} \\
&\qquad \left. + \frac{16(-z^5 + 10z^3 - 15z)}{\tau^6}(p + 2\mathrm{tr}_3 - 3\mathrm{tr}_2)^2\right] \\
&+ \mathrm{O}(k^{-3/2}).
\end{aligned}$$

Then the density $g(z)$ satisfies,

$$\begin{aligned} g(z) =& \phi(z) + \frac{z\phi(z)}{6\sqrt{k}}\left[\frac{3p(p+1)}{\tau} + \frac{4(p+2\mathrm{tr}_3-3\mathrm{tr}_2)(z^2-1)}{\tau^3}\right. \\ &\left. - \frac{8(p+2\mathrm{tr}_3-3\mathrm{tr}_2)}{\tau^3}\right] \\ &+ \frac{\phi(z)}{72k}\left[\frac{9(z^2-1)p(p+1)(p^2+p+4)}{\tau^2}\right. \\ &+ \frac{24(z^4-6z^2+3)}{\tau^4}\{p(p^2+p+2) - 3p(p+1)\mathrm{tr}_2 \\ &+ 2(p^2+p-1)\mathrm{tr}_3 + 6\mathrm{tr}_4\} \\ &\left. + \frac{16(z^6-15z^4+45z^2-15)}{\tau^6} \times (p+2\mathrm{tr}_3-3\mathrm{tr}_2)^2\right] \\ &+ \mathrm{O}(k^{-3/2}). \end{aligned}$$

Since all the odd moments of standard normal are 0,

$$\begin{aligned} \mathrm{E}(z) &= \int_{-\infty}^{\infty} \frac{z^2\phi(z)}{6\sqrt{k}}\left[\frac{3p(p+1)}{\tau} + \frac{4(p+2\mathrm{tr}_3-3\mathrm{tr}_2)(z^2-1)}{\tau^3}\right. \\ &\qquad \left. - \frac{8(p+2\mathrm{tr}_3-3\mathrm{tr}_2)}{\tau^3}\right] dz + \mathrm{O}(k^{-3/2}) \\ &= \frac{1}{6\sqrt{k}}\left[\frac{p(p+1)}{\tau} + \frac{8(p+2\mathrm{tr}_3-3\mathrm{tr}_2)}{\tau^3} - \frac{8(p+2\mathrm{tr}_3-3\mathrm{tr}_2)}{\tau^3}\right] \\ &\quad + \mathrm{O}(k^{-3/2}) \\ &= \frac{p(p+1)}{2\tau\sqrt{k}} + \mathrm{O}(k^{-3/2}). \end{aligned}$$

and also

$$\int_{-\infty}^{\infty} z^2 g(z)\,dz = 1 + \int_{-\infty}^{\infty} \frac{z^2\phi(z)}{72k}\Bigg[\frac{9(z^2-1)p(p+1)(p^2+p+4)}{\tau^2} + \frac{24(z^4-6z^2+3)}{\tau^4}\Big\{p(p^2+p+2) - 3p(p+1)\mathrm{tr}_2 + 2(p^2+p-1)\mathrm{tr}_3 + 6\mathrm{tr}_4\Big\} + \frac{16(z^6-15z^4+45z^2-15)}{\tau^6}(p+2\mathrm{tr}_3-3\mathrm{tr}_2)^2\Bigg]\,dz$$

$$+ \mathrm{O}(k^{-2})$$

$$= 1 + \frac{p(p+1)(p^2+p+4)}{4k\tau^2} + \mathrm{O}(k^{-2}).$$

Hence

$$\begin{aligned}\mathrm{Var}(z) &= 1 + \frac{p(p+1)(p^2+p+4)}{4k\tau^2} - \frac{p^2(p+1)^2}{4k\tau^2} + \mathrm{O}(k^{-3/2}) \\ &= \frac{k\tau^2 + p(p+1)}{k\tau^2} + \mathrm{O}(k^{-3/2}).\end{aligned}$$

Combining the above results, we have the following theorem.

Theorem 2 *Under the alternative* H_1: $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$, *the log likelihood ratio criterion* T_{LR} *is distributed as*

$$P\Big(\frac{T_{LR}-\mu}{\sqrt{\tau^\star}} \leq z\Big) = \Phi(z) + O(k^{-1/2}) \tag{3.8}$$

as $k = n-1 \to \infty$, *where* $\mu = k\Big\{tr\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big\} + p(p+1)/2$ *and* $\tau^\star = 2ktr(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2 + p(p+1)$.

We modified Sugiura's results by adding an extra $p(p+1)/2$ to the mean and $p(p+1)$ to the variance. This modification is trivial for distant alternatives when $k\Big\{\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big\}$ and $2k\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2$ are significantly bigger

than $p(p+1)/2$ and $p(p+1)$ respectively, but significant for local alternatives when $k\Big\{\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big\}$ and $2k\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2$ are comparable with $p(p+1)/2$ and $p(p+1)$.

As an example, we made 10,000 simulations of T_{LR} and compare with 10000 simulations of normal data. In the simulation, $\boldsymbol{\Sigma}_0 = \mathbf{I}_3$ and $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+a \end{smallmatrix}\right)$, $k = 199$. Table 3.1 compares the sample quantiles with theoretical normal quantiles.

Table 3.1: Sample quantiles of T_{LR} vs. normal approximation. $a = 1$

Quantile	1%	5%	10%	25%	50%	75%	90%	95%	99%
T_{LR}	27.35	36.87	42.65	52.58	65.47	79.76	93.97	102.87	120.68
Normal	21.12	34.41	41.46	53.34	66.85	80.27	93	100.3	114.29

Figure 3.1 is the QQ plot. Figure 3.2 shows the density plots.

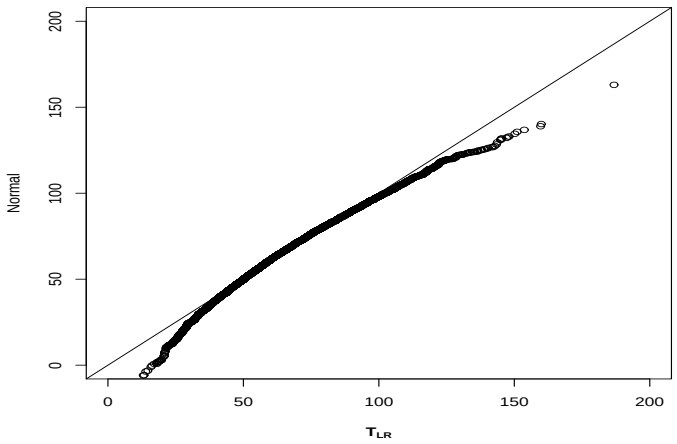

Figure 3.1: QQ plot of T_{LR} vs. normal with mean $k\Big\{\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - p - \ln|\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1}|\Big\} + p(p+1)/2$ and variance $2k\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2 + p(p+1)$. $\boldsymbol{\Sigma}_0 = \mathbf{I}$, a=1, k=199

Overall we see good fit except for low-end and high-end tails where the true quantiles are greater than normal theory quantiles, which suggests inflated Type I

error rates for hypothesis testing.

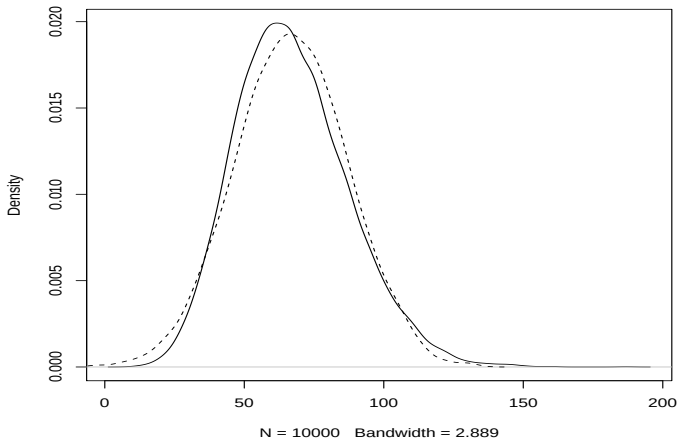

Figure 3.2: Density plot. Solid line: T_{LR}. Dotted line: Normal

Table 3.2 compares the power of the log likelihood ratio test T_{LR} calculated from simulation with theoretical power for various alternatives. The first five columns are considered as small shifts (variance of the third dimension shifts to 1.07, 1.21, 1.35, 1.5, 1.71 respectively), while the last five are considered as medium or large shifts.

Table 3.2: Power of T_{LR} and normal approximation for various shifts. Significance level=0.05

a	$1/\sqrt{199}$	$3/\sqrt{199}$	$5/\sqrt{199}$	$7/\sqrt{199}$	$9/\sqrt{199}$	1	1.25	1.5	1.75	2
T_{LR}	0.072	0.274	0.63	0.886	0.979	1	1	1	1	1
Normal	0.051	0.321	0.678	0.872	0.953	0.996	0.999	1	1	1
Difference	0.021	-0.047	- 0.048	0.014	0.026	0.004	0.001	0	0	0

In the first column, the shift is really small, from 1 to 1.07, T_{LR} has a higher rate of rejection than theoretically suggested. As shift size increases, the likelihood

ratio test doesn't perform as well as the normal approximation suggests. For medium to large shifts, the true and approximate powers converge.

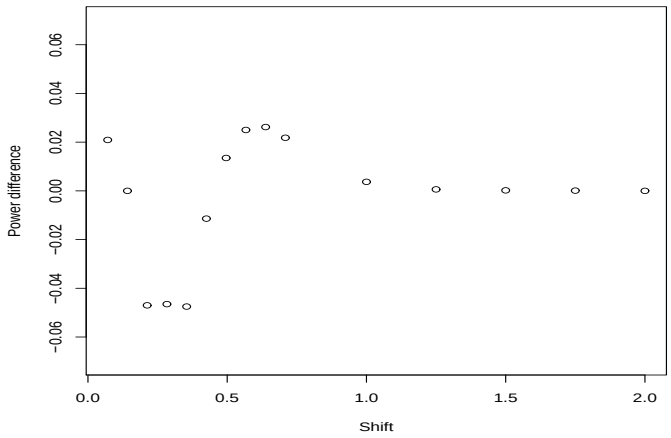

Figure 3.3: Plot of the difference between the two power functions of $\mathbf{T}_{LR}$ and normal theoretical for a variety of shifts. Significance level 0.05.

Finally let us examine T_{LR} from the perspective of SPC. When the process is under control, the mean of T_{LR} is $p(p+1)/2$. If the variance covariance matrix shifts to $\mathbf{\Sigma} \neq \mathbf{\Sigma}_0$, then the mean of T_{LR} becomes

$$\mathrm{E}(\mathrm{T}_{LR}) = k\Big\{\mathrm{tr}\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - p - \ln|\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}|\Big\} + p(p+1)/2$$

In terms of the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$, $l_1, l_2, \ldots, l_p$

$$\mathrm{E}(\mathrm{T}_{LR}) = k\sum_{i=1}^{p}(l_i - 1 - \ln(l_i)) + p(p+1)/2. \tag{3.9}$$

The term $l_i - 1 - \ln(l_i)$ is a convex function of l_i and has its minimum at $l_i = 1$. In other words, T_{LR} increases as l_i departs from its in-control value 1. Both increase and decrease of l_i contribute positively to $\mathrm{E}(\mathrm{T}_{LR})$. Equation 3.9 suggests that these contributions are additive too.

3.3 Asymptotic expansions of Nagao's test under fixed alternatives

For Nagao's testing criteria, consider the standardized T_{NT}, assuming the process mean is $\mathbf{0}$. let the normalized variance covariance matrix estimator be

$$\begin{aligned}\mathbf{S} &= \sum_{i=1}^{k} \mathbf{Y}_i \mathbf{Y}_i^T / k \\ &= \boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^{-1/2} + \mathbf{U}\end{aligned}$$

where

$$\boldsymbol{\Sigma}_0^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^{-1/2} = [\sigma_{ij}]_{p\times p}, \qquad \mathbf{U} = [u_{ij}]_{p\times p} = \left[\sum_{l=1}^{k} y_{li} y_{lj}/k - \sigma_{ij}\right]_{p\times p}$$

$$\mathrm{E}(\mathbf{U}) = 0, \qquad \mathrm{Var}(\mathbf{U}) = \mathrm{O}(1/k).$$

The matrix $\mathbf{U}$ captures the randomness of T_{NT}. Then we can rewrite T_{NT} as

$$\begin{aligned}(1/\sqrt{k})\mathrm{T}_{NT} &= \big(\sqrt{k}/2\big)\mathrm{tr}(\mathbf{S}-\mathbf{I})^2 \\ &= \big(\sqrt{k}/2\big)\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1/2} - \mathbf{I} + \mathbf{U})^2 \\ &= \big(\sqrt{k}/2\big)\Big\{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2 + 2\mathrm{tr}\Big((\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})\mathbf{U}\Big) + \mathrm{tr}\mathbf{U}^2\Big\},\end{aligned}$$

hence

$$\frac{1}{\sqrt{k}}\big\{\mathrm{T}_{NT} - k\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2/2\big\} = \mathrm{tr}\Big((\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})\sqrt{k}\mathbf{U}\Big) + \mathrm{O}(k^{-1/2})$$

Vectorize $\mathbf{U} = \mathbf{S} - \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1/2}$ by stacking the columns of the matrix $\mathbf{U}$ on top of one another

$$\mathrm{Vec}(\mathbf{U}) = \begin{bmatrix} u_{11} & u_{21} & \dots & u_{p1} & u_{12} & \dots & u_{pp}\end{bmatrix}^T_{p^2\times p^2}.$$

The covariance between any two entries u_{st} and u_{kl} of Vec($\mathbf{U}$) is

$$\begin{aligned}
Cov(u_{ml}, u_{st}) &= \mathrm{E}\Big\{ \Big(\sum_{i=1}^{k} y_{im}y_{il}/k - \sigma_{ml} \Big) \Big(\sum_{j=1}^{k} y_{js}y_{jt}/k - \sigma_{st} \Big) \Big\} \\
&= \mathrm{E}\Big\{ \Big(\sum_{i=1}^{k} y_{im}y_{il} \sum_{j=1}^{k} y_{js}y_{jt} \Big)/k^2 - \sigma_{ml}\sigma_{st} \Big\} \\
&= \mathrm{E}\Big\{ \Big(\sum_{i=1}^{k} \sum_{j=1}^{k} y_{im}y_{il}y_{js}y_{jt} \Big)/k^2 - \sigma_{ml}\sigma_{st} \Big\} \\
&= \mathrm{E}\Big\{ \Big(\sum_{i=1}^{k} y_{im}y_{il}y_{is}y_{it} + \sum_{i\neq j} \sum_{j=1}^{k} y_{im}y_{il}y_{js}y_{jt} \Big)/k^2 - \sigma_{ml}\sigma_{st} \Big\} \\
&= \Big\{ k\mathrm{E}(y_m y_l y_s y_t) + k(k-1)\mathrm{E}(y_m y_l)\mathrm{E}(y_s y_t) \Big\}/k^2 - \sigma_{ml}\sigma_{st} \\
&= \Big\{ \sigma_{ml}\sigma_{st} + \sigma_{ms}\sigma_{lt} + \sigma_{mt}\sigma_{ls} + (k-1)\sigma_{ml}\sigma_{st} \Big\}/k - \sigma_{ml}\sigma_{st} \\
&= (\sigma_{ms}\sigma_{lt} + \sigma_{mt}\sigma_{ls})/k.
\end{aligned}$$

Under the regularity conditions, the Central Limit Theorem gives

$$\sqrt{k}\mathrm{Vec}(\mathbf{U}) \xrightarrow[k\to\infty]{d} \mathrm{N}(0, \mathbf{\Sigma}_u)$$

where

$$\mathbf{\Sigma}_u = \Big[\sigma_{ms}\sigma_{lt} + \sigma_{mt}\sigma_{ls} \Big]_{p^2\times p^2}$$

also

$$\mathrm{tr}\Big((\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})\sqrt{k}\mathbf{U} \Big) = \sum_{i=1}^{k}\sum_{j=1}^{k} \sqrt{k}(\sigma_{ij} - \mathbf{I}_{ij})u_{ji} = \Big(\mathrm{Vec}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^T \Big)^T \sqrt{k}\mathrm{Vec}(\mathbf{U}).$$

A simple application of the delta method here gives us the following result.

Theorem 3 *Under the alternative H_1: $\mathbf{\Sigma} \neq \mathbf{\Sigma}_0$, T_{NT} is asymptotically distributed as*

$$(1/\sqrt{k})\Big\{ T_{NT} - ktr(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2/2 \Big\} \xrightarrow[k\to\infty]{d} N\Bigg(0, \Big(Vec(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^T \Big)^T \mathbf{\Sigma}_u \, Vec(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^T \Bigg)$$

where k is the degrees of freedom of the variance covariance estimator in T_{NT}.

Using the same simulation structure as that of T_{LR}, we generated 10000 replications of T_{NT} and compared the sample quantiles with the normal theoretical quantiles when $a = 2$ in Table 3.3.

Table 3.3: Sample quantiles of $T_{NT}/\sqrt{k}$ Vs. normal approximation. $a = 2$.

Quantile	1%	5%	10%	25%	50%	75%	90%	95%	99%
$T_{NT}/\sqrt{k}$	13.17	16.94	19.04	23.31	28.69	34.82	41.08	45.35	53.42
Normal	8.94	14.33	17.39	22.56	28.27	33.93	39.11	42.22	47.89

Figure 3.5 shows the corresponding quantile-quantile plot and Figure 3.6 compares the density plots. The worst discrepencies occur in the tails.

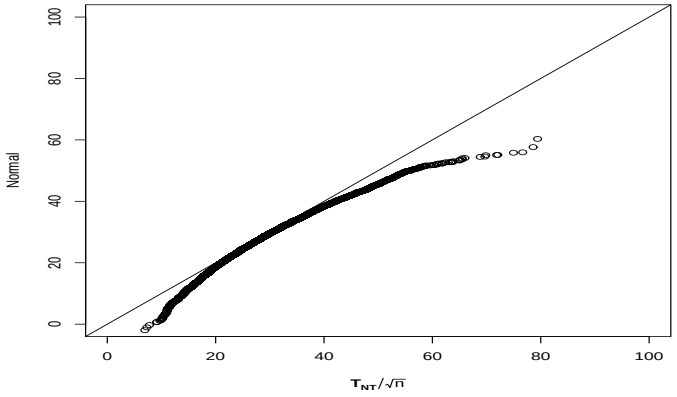

Figure 3.4: QQ plot of $T_{NT}/\sqrt{k}$ Vs. normal with mean $\sqrt{k}\text{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0 - \mathbf{I})^2/2$. $\mathbf{\Sigma}_0 = \mathbf{I}$, k=199.

Table 3.4 compares the power of Nagao's test T_{NT} calculated from simulation with theoretical power for various alternatives. Both Table 3.4 and Figure 3.6 suggest that the normal approximation is very crude for local alternatives. In fact, the limiting non-null distribution is degenerate at the null hypothesis so that the asym-

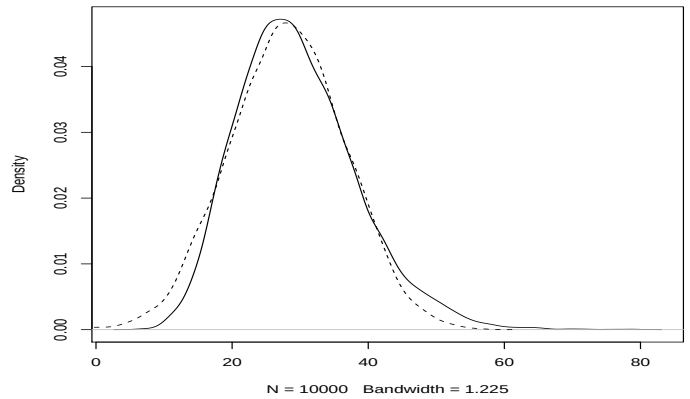

Figure 3.5: Density plot. Solid line: $T_{NT}/\sqrt{k}$. Dotted line: Normal.

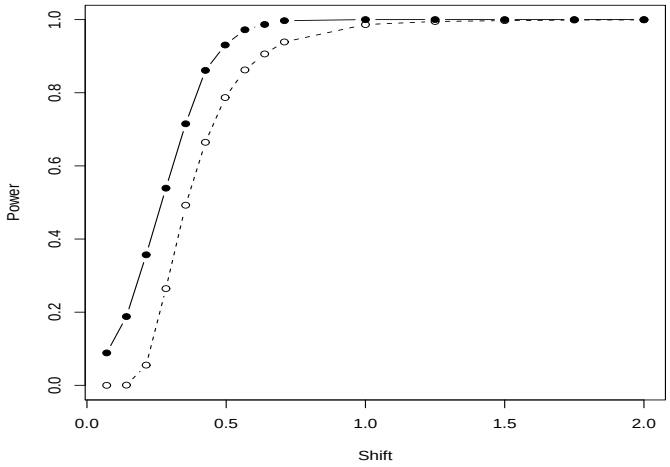

Figure 3.6: Power plot. Solid line: T_{NT}. Dashed line: Normal approximation. Significance level=0.05.

Table 3.4: Power of T_{NT} vs. asymptotic normal approximation for various shifts. Significance level=0.05.

Shift	$1/\sqrt{199}$	$3/\sqrt{199}$	$5/\sqrt{199}$	$7/\sqrt{199}$	$9/\sqrt{199}$	1	1.25	1.5	1.75	2
T_{NT}	0.089	0.36	0.715	0.931	0.987	1	1	1	1	1
Normal	0	0.055	0.493	0.787	0.906	0.986	0.995	0.997	0.999	1
Difference	0.089	0.302	0.223	0.144	0.081	0.014	0.005	0.003	0.001	0

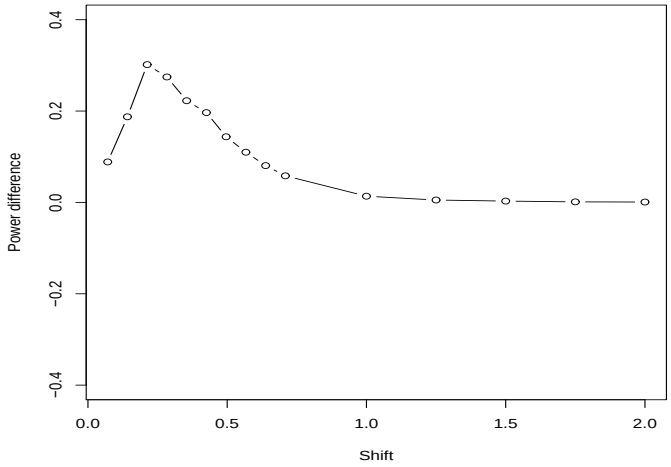

Figure 3.7: Plot of the difference between the two power functions of T_{NT} and normal theoretical for a variety of shifts. Significance level 0.05.

potic formulas do not give good approximations when the alternative is close to the null hypothesis. Some further discussion about the limiting non-null distributions under sequences of local alternatives is needed. We will extend this result in the next section.

3.4 Asymptotic distribution of Nagao's test criteria under local alternatives

In this section, we will continue our discussion of the distributional properties of T_{NT} in last section and show that under local alternatives, the statistic $T_{NT} = k\text{tr}(\mathbf{S} - \mathbf{I})^2/2$ follows a χ^2 distribution with $p(p+1)/2$ degrees of freedom and nocentrality paramenter $\text{tr}(\Theta)^2/2$. Because of the invariance, we consider the test using the standardized observations $\mathbf{Y}_t = \mathbf{\Sigma}_0^{-1/2}\mathbf{X}_t$, which has variance covariance matrix $\mathbf{\Sigma}^\star$.

In the following discussion, we will use the matrix exponential and matrix

logarithm, which are defined as

$$\exp(\mathbf{A}) = \sum_{i=0}^{\infty} \frac{1}{i!}\mathbf{A}^i,$$

$$\mathbf{B} = \ln(\mathbf{A}),$$

if $\mathbf{A}$ is a real symmetric positive definite matrix, then $\ln(\mathbf{A}) = \mathbf{B}$ is uniquely defined as the inverse of the matrix exponential function (restricted to the domain of real symmetric matrices), that is $\exp(\ln(\mathbf{A})) = \mathbf{A}$ if $\mathbf{A}$ is real symmetric and positive definite.

Let us consider testing

$$\mathrm{H}_0 : \mathbf{\Sigma}^\star = \mathbf{\Sigma}_0^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1/2} = \mathbf{I}_{p\times p}$$

vs. the sequence of alternatives:

$$\mathrm{H}_1 : \mathbf{\Sigma}^\star = \mathbf{\Sigma}_0^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1/2} = \mathbf{I}_{p\times p} + k^{-1/2}\mathbf{\Theta}$$

where k is the degrees of freedom of the sample variance covariance matrix and $\mathbf{\Theta}$ is a symmetric matrix.

Under the alternative hypothesis, the distribution of $\mathbf{S}$ is Wishart$\Big(k, k^{-1}(\mathbf{I} + k^{-1/2}\mathbf{\Theta})\Big)$. The characteristic function of T_{NT} is

$$\begin{aligned}\mathrm{C}(t) = c_{p,k} \int \exp\{it\mathrm{T}_{NT}\}|k^{-1}(I + k^{-1/2}\mathbf{\Theta})|^{-\frac{k}{2}}|\mathbf{S}|^{\frac{1}{2}(k-p-1)} \\ \times \mathrm{etr}\Big\{(-\frac{1}{2}(k^{-1}(\mathbf{I} + k^{-1/2}\mathbf{\Theta}))^{-1}\mathbf{S})\Big\}\, d\mathbf{S} \qquad (3.10)\end{aligned}$$

where the constant

$$c_{p,k} = \left\{ 2^{kp/2}\pi^{p(p-1)/4} \prod_{\alpha=1}^{p} \mathbf{\Gamma}[\frac{1}{2}(k+1-\alpha) \right\}^{-1}.$$

By expressing T_{NT} in terms of

$$\mathbf{Y} = (k/2)^{1/2} \ln \mathbf{S} \qquad \text{or} \qquad \mathbf{S} = \exp(\sqrt{2/k}\mathbf{Y}),$$

we have

$$\begin{aligned}
\mathrm{T}_{NT} &= k\mathrm{tr}(\mathbf{S}-\mathbf{I})^2/2 \\
&= k\mathrm{tr}\Big(\exp(\sqrt{2/k}\mathbf{Y}) - \mathbf{I}\Big)^2/2 \\
&= \mathrm{tr}\mathbf{Y}^2 + (2/k)^{1/2}\mathrm{tr}\mathbf{Y}^3 + (7/6k)\mathrm{tr}\mathbf{Y}^4 + \mathrm{O}(k^{-3/2}).
\end{aligned}$$

According to Nagao [34], the Jacobian of such a transformation is

$$\left|\frac{\partial \mathbf{S}}{\partial \mathbf{Y}}\right| = \left(\frac{2}{k}\right)^{p(p+1)/4} \mathrm{etr}\{(2/k)^{1/2}\mathbf{Y}\} \prod_{i>j}^{p} \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}$$

Nagao [34] showed that the asymptotic expansion of the distribution of $\mathbf{Y}$ is

$$\begin{aligned}
f(\mathbf{Y}) = c^{\star}_{p,k} &\times \mathrm{etr}\left\{(1/2)(k-p+1)(2/k)^{1/2}\mathbf{Y} - (k/2)e^{(2/k)^{1/2}\mathbf{Y}}\right\} \\
&\times \left\{1 + \frac{1}{2}(p-1)(\frac{k}{2})^{1/2}\mathrm{tr}\mathbf{Y} + \frac{1}{12k}((3p^2-6p+1)(\mathrm{tr}\mathbf{Y})^2 + p\mathrm{tr}\mathbf{Y}^2)\right\} \qquad (3.11)
\end{aligned}$$

where

$$c^{\star}_{p,k} = \left\{\prod_{\alpha=1}^{p} \mathbf{\Gamma}[\frac{1}{2}(k+1-\alpha)]\right\}^{-1} \left(\frac{k}{2}\right)^{p(2k-p+1)/4} \pi^{-p(p-1)/4}$$

Using (3.10) and (3.11), we can rewrite the characteristic function $\mathrm{C}(t)$ as

$$\begin{aligned}
\mathrm{C}(t) = c^{\star}_{p,k}&|\mathbf{I} + k^{-1/2}\mathbf{\Theta}|^{-k/2} \\
&\times \int \exp\left\{(it)\mathrm{tr}\mathbf{Y}^2 + \sqrt{\frac{2}{k}}(it)\mathrm{tr}\mathbf{Y}^3 + \frac{7}{6k}(it)\mathrm{tr}\mathbf{Y}^4\right. \\
&\qquad \left. + \frac{1}{2}(k-p+1)\sqrt{\frac{2}{k}}\mathrm{tr}\mathbf{Y} - \frac{k}{2}\mathrm{tr}\Big((\mathbf{I}+k^{-1/2}\mathbf{\Theta})^{-1}\exp(\sqrt{2/k}\mathbf{Y}\Big)\right\} \\
&\times \left\{1 + \frac{1}{2}(p-1)\sqrt{\frac{2}{k}}\mathrm{tr}\mathbf{Y}\right. \\
&\qquad \left. + \frac{1}{12k}\Big((3p^2-6p+2)(\mathrm{tr}\mathbf{Y})^2 + p\ tr\mathbf{Y}^2\Big) + \mathrm{O}(k^{-3/2})\right\} d\mathbf{Y},
\end{aligned}$$

where the integration region is the set of all symmetric $p \times p$ matrices.

Using the Neumann Series expansion:

$$(\mathbf{I} + k^{-1/2}\mathbf{\Theta})^{-1} = \mathbf{I} - \frac{\mathbf{\Theta}}{\sqrt{k}} + \frac{\mathbf{\Theta}^2}{k} - \frac{\mathbf{\Theta}^3}{k^{3/2}} + \frac{\mathbf{\Theta}^4}{k^2} + \mathrm{O}(k^{-5/2})$$

and

$$\exp(\sqrt{2/k}\mathbf{Y}) = \mathbf{I} + \sqrt{\frac{2}{k}}\mathbf{Y} + \frac{\mathbf{Y}^2}{k} + \frac{\sqrt{2}\mathbf{Y}^3}{3k^{3/2}} + \frac{\mathbf{Y}^4}{6k^2} + \mathrm{O}(k^{-5/2})$$

We have

$$\begin{aligned}
(\mathbf{I} + k^{-1/2}\mathbf{\Theta})^{-1}\exp(\sqrt{2/k}\mathbf{Y}) = &\left(\mathbf{I} - \frac{\mathbf{\Theta}}{\sqrt{k}} + \frac{\mathbf{\Theta}^2}{k} - \frac{\mathbf{\Theta}^3}{k^{3/2}} + \frac{\mathbf{\Theta}^4}{k^2} + \mathrm{O}(k^{-\frac{5}{2}})\right) \\
&\times \left(\mathbf{I} + \sqrt{\frac{2}{k}}\mathbf{Y} + \frac{\mathbf{Y}^2}{k} + \frac{\sqrt{2}\mathbf{Y}^3}{3k^{3/2}} + \frac{\mathbf{Y}^4}{6k^2} + \mathrm{O}(k^{-\frac{5}{2}})\right) \\
= &\,\mathbf{I} + \frac{1}{\sqrt{k}}\big(\sqrt{2}\mathbf{Y} - \mathbf{\Theta}\big) + \frac{1}{k}\big(\mathbf{Y}^2 - \sqrt{2}\mathbf{\Theta}\mathbf{Y} + \mathbf{\Theta}^2\big) \\
&+ \frac{1}{k^{3/2}}\left(\frac{\sqrt{2}}{3}\mathbf{Y}^3 - \mathbf{\Theta}\mathbf{Y}^2 + \sqrt{2}\mathbf{\Theta}^2\mathbf{Y} - \mathbf{\Theta}^3\right) \\
&+ \frac{1}{k^2}\left(\frac{\mathbf{Y}^4}{6} - \frac{\sqrt{2}}{3}\mathbf{\Theta}\mathbf{Y}^3 + \mathbf{\Theta}^2\mathbf{Y}^2 - \sqrt{2}\mathbf{\Theta}^3\mathbf{Y} + \mathbf{\Theta}^4\right) \\
&+ \mathrm{O}(k^{-5/2}).
\end{aligned}$$

Also

$$\begin{aligned}
|\mathbf{I} + k^{-1/2}\mathbf{\Theta}|^{-k/2} &= \exp\Big(-\frac{k}{2}\ln|\mathbf{I} + k^{-1/2}\mathbf{\Theta}|\Big) \\
&= \exp\left\{\frac{k}{2}\Big(\frac{\mathrm{tr}\mathbf{\Theta}}{k^{1/2}} + \frac{\mathrm{tr}\mathbf{\Theta}^2}{2k} - \frac{\mathrm{tr}\mathbf{\Theta}^3}{3k^{3/2}} + \frac{\mathrm{tr}\mathbf{\Theta}^4}{4k^2} + \mathrm{O}(k^{-5/2})\Big)\right\} \\
&= \exp\left\{-\frac{\sqrt{k}}{2}\mathrm{tr}\mathbf{\Theta} + \frac{\mathrm{tr}\mathbf{\Theta}^2}{4} - \frac{\mathrm{tr}\mathbf{\Theta}^3}{6\sqrt{k}} + \frac{\mathrm{tr}\mathbf{\Theta}^4}{8k} + \mathrm{O}(k^{-3/2})\right\}
\end{aligned}$$

After some calculation, we have

$$
\begin{aligned}
\mathrm{C}(t) &= c^{\star}_{p,k} \exp\Big(-\frac{kp}{2}\Big) \exp\Big(-\frac{\mathrm{tr}\boldsymbol{\Theta}^2}{4}\Big) \\
&\quad \times \int \exp\left\{(it)\mathrm{tr}\mathbf{Y}^2 - \frac{1}{2}\mathrm{tr}\mathbf{Y}^2 + \frac{\sqrt{2}}{2}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}\right\} \\
&\qquad \times \exp\left\{\sqrt{\frac{2}{k}}\Big((it)\mathrm{tr}\mathbf{Y}^3 - \frac{\mathrm{tr}\mathbf{Y}^3}{6} + \frac{\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2}{2\sqrt{2}} - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}}{2} + \frac{1}{2}(1-p)\mathrm{tr}\mathbf{Y} + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}}\Big)\right. \\
&\qquad\quad + \frac{1}{k}\Big(\frac{7}{6}(it)\mathrm{tr}\mathbf{Y}^4 - \frac{\mathrm{tr}\mathbf{Y}^4}{12} + \frac{\sqrt{2}}{6}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^3 - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}^2}{2} + \frac{\sqrt{2}}{2}\mathrm{tr}\boldsymbol{\Theta}^3\mathbf{Y} - \frac{3\mathrm{tr}\boldsymbol{\Theta}^4}{8}\Big) \\
&\qquad\quad \left. + \mathrm{O}(k^{-3/2})\right\} \\
&\qquad \times \left\{1 + \frac{1}{2}(p-1)\sqrt{\frac{2}{k}}\mathrm{tr}\mathbf{Y} + \frac{1}{12k}\big((3p^2-6p+2)(\mathrm{tr}\mathbf{Y})^2 + p\ tr\mathbf{Y}^2\big)\right. \\
&\qquad\quad \left. + \mathrm{O}(k^{-3/2})\right\} d\mathbf{Y} \\
&= c^{\star}_{p,k} \exp\Big(-\frac{kp}{2}\Big) \exp\Big(-\frac{\mathrm{tr}\boldsymbol{\Theta}^2}{4}\Big) \int \mathrm{E}_1 \times \mathrm{E}_2 \times \mathrm{E}_3 \, d\mathbf{Y}.
\end{aligned}
$$

By completing the square for E_1

$$
\begin{aligned}
\mathrm{E}_1 &= \exp\left\{(it)\mathrm{tr}\mathbf{Y}^2 - \frac{1}{2}\mathrm{tr}\mathbf{Y}^2 + \frac{\sqrt{2}}{2}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}\right\} \\
&= \exp\left\{-\frac{1}{2}(1-2it)\mathrm{tr}\Big(\mathbf{Y} - \frac{\boldsymbol{\Theta}}{\sqrt{2}(1-2it)}\Big)^2\right\} \exp\Big(\frac{\mathrm{tr}\boldsymbol{\Theta}^2}{4(1-2it)}\Big). \qquad (3.12)
\end{aligned}
$$

By carrying out further expansion for E_2, we have

$$\begin{aligned}
\mathrm{E}_2 = 1 + \sqrt{\frac{2}{k}}\Bigg\{ & \Big(it - \frac{1}{6}\Big)\mathrm{tr}\mathbf{Y}^3 + \frac{\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2}{2\sqrt{2}} - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}}{2} + \frac{1}{2}(1-p)\mathrm{tr}\mathbf{Y} + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}}\Bigg\} \\
& + \frac{1}{k}\Bigg\{ \Big(it - \frac{1}{6}\Big)^2(\mathrm{tr}\mathbf{Y}^3)^2 + \frac{it - \frac{1}{6}}{\sqrt{2}}\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2 + \Big(it - \frac{1}{6}\Big)(1-p)\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\mathbf{Y} \\
& \quad + (\frac{7}{6}it - \frac{1}{12})\mathrm{tr}\mathbf{Y}^4 - (it - \frac{1}{6})\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y} + \frac{(\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2)^2}{8} + \frac{\sqrt{2}}{6}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^3 \\
& \quad + \frac{2(it - \frac{1}{6})}{3\sqrt{2}}\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}^3 + \frac{1-p}{2\sqrt{2}}\mathrm{tr}\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2 - \frac{1}{2\sqrt{2}}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2 \\
& \quad - \frac{1-p}{2}\mathrm{tr}\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y} - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}^2}{2} + \frac{1}{6}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2\mathrm{tr}\boldsymbol{\Theta}^3 + \frac{1}{4}(p-1)^2(\mathrm{tr}\mathbf{Y})^2 \\
& \quad + \frac{(\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y})^2}{4} + \frac{\sqrt{2}}{2}\mathrm{tr}\boldsymbol{\Theta}^3\mathbf{Y} - \frac{1}{3\sqrt{2}}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}^3 + \frac{1-p}{3\sqrt{2}}\mathrm{tr}\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}^3 \\
& \quad - \frac{3}{8}\mathrm{tr}\boldsymbol{\Theta}^4 + \frac{(\mathrm{tr}\boldsymbol{\Theta}^3)^2}{18}\Bigg\} \\
& + \mathrm{O}(k^{-3/2}),
\end{aligned}$$

hence

$$\begin{aligned}
\mathrm{E}_2 \times \mathrm{E}_3 = 1 + \sqrt{\frac{2}{k}}\Bigg\{ & \Big(it - \frac{1}{6}\Big)\mathrm{tr}\mathbf{Y}^3 + \frac{\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2}{2\sqrt{2}} - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}}{2} + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}}\Bigg\} \\
& + \frac{1}{k}\Bigg\{ \Big(it - \frac{1}{6}\Big)^2(\mathrm{tr}\mathbf{Y}^3)^2 + \Big(\frac{it - \frac{1}{6}}{\sqrt{2}}\Big)\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2 - \Big(it - \frac{1}{6}\Big)\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y} \\
& \quad + \frac{(\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2)^2}{8} + \Big(\frac{7}{6}it - \frac{1}{12}\Big)\mathrm{tr}\mathbf{Y}^4 - \frac{1}{2\sqrt{2}}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2 \\
& \quad + \frac{2(it - \frac{1}{6})}{3\sqrt{2}}\mathrm{tr}\mathbf{Y}^3\mathrm{tr}\boldsymbol{\Theta}^3 + \frac{\sqrt{2}}{6}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^3 - \frac{1}{3\sqrt{2}}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}\mathrm{tr}\boldsymbol{\Theta}^3 + \frac{1}{6}\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2\mathrm{tr}\boldsymbol{\Theta}^3 \\
& \quad + \frac{(\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y})^2}{4} - \frac{1}{2}\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}^2 + \frac{p}{12}\mathrm{tr}\mathbf{Y}^2 - \frac{1}{12}(\mathrm{tr}\mathbf{Y})^2 + \frac{\sqrt{2}}{2}\mathrm{tr}\boldsymbol{\Theta}^3\mathbf{Y} \\
& \quad + \frac{(\mathrm{tr}\boldsymbol{\Theta}^3)^2}{18} - \frac{3}{8}\mathrm{tr}\boldsymbol{\Theta}^4\Bigg\} \\
& + \mathrm{O}(k^{-3/2}) \\
= 1 + \sqrt{\frac{2}{k}}\mathrm{F}_1 & + \frac{1}{k}\mathrm{F}_2 + \mathrm{O}(k^{-3/2}).
\end{aligned}$$

Let $g(\mathbf{\Theta}) = \mathbf{\Theta}/(\sqrt{2}(1-2it)) = (g_{ij})$. Then the first term of E_1 can be rewritten as

$$\exp\left\{ -\frac{\sum_{i=1}^{p}(\mathbf{Y}_{ii} - g_{ii})^2 + 2\sum_{i>j}(\mathbf{Y}_{ij} - g_{ij})^2}{2(1-2it)^{-1}} \right\} \tag{3.13}$$

As in Nagao [34], the special structure of E_1 in (3.13) prompts us to consider the vector of the unique elements in $\mathbf{Z} = \mathbf{Y} - \mathbf{\Theta}/(\sqrt{2}(1-2it))$, Hvec($\mathbf{Z}$), as having a $p(p+1)/2$ dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $[\sigma_{ij,kl}]$ with

$$\text{Hvec}(\mathbf{Z}) = \Big[z_{11}, z_{22}, \ldots, z_{pp}, z_{12}, z_{13}, \ldots, z_{p-1,p} \Big]^{T}_{p(p+1)/2 \times 1}$$

and

$$\sigma_{ij,kl} = (1-2it)^{-1}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$$

where δ_{ij} is the Kronecker Delta function with

$$\delta_{ij} = \begin{cases} 1 & i = j \\ \\ 0 & \text{otherwise} \end{cases}$$

Here the variance quantity $\sigma_{ij,kl}$ is complex valued. The " Covariance Matrix " of Hvec($\mathbf{Z}$) is block diagonal

$$\Big[\sigma_{ij,kl}\Big] = \begin{bmatrix} (1-2it)^{-1}\mathbf{I}_{p\times p} & \mathbf{0} \\ \\ \mathbf{0} & \frac{1}{2}(1-2it)^{-1}\mathbf{I}_{\frac{p(p-1)}{2}\times\frac{p(p-1)}{2}} \end{bmatrix}_{\frac{p(p+1)}{2}\times\frac{p(p+1)}{2}}.$$

Hence

$$\left|\Big[\sigma_{ij,kl}\Big]\right| = (1-2it)^{-p(p+1)/2}\ 2^{-p(p-1)/2}.$$

Thus we have the asymptotic expansion for the characteristic function

$$\begin{aligned}
\mathrm{C}(t) &= c^{\star}_{p,k}(2\pi)^{\frac{p(p+1)}{4}}2^{-\frac{p(p-1)}{4}}(1-2it)^{-\frac{p(p+1)}{4}}\exp\Big(-\frac{kp}{2}\Big)\exp\Big(-\frac{\mathrm{tr}\boldsymbol{\Theta}^2}{4}+\frac{\mathrm{tr}\boldsymbol{\Theta}^2}{4(1-2it)}\Big)\\
&\quad\times \mathrm{E}_Y\Big(1+\sqrt{\frac{2}{k}}\mathrm{F}_1+\frac{1}{k}\mathrm{F}_2+\mathrm{O}(k^{-3/2})\Big)\\
&= c^{\star\star}_{p,k}(1-2it)^{-\frac{p(p+1)}{4}}\exp\Big(\frac{1}{2}it\mathrm{tr}\boldsymbol{\Theta}^2(1-2it)^{-1}\Big)\\
&\quad\times \mathrm{E}_Y\Big(1+\sqrt{\frac{2}{k}}\mathrm{F}_1+\frac{1}{k}\mathrm{F}_2+\mathrm{O}(k^{-3/2})\Big) \qquad (3.14)
\end{aligned}$$

where

$$\begin{aligned}
c^{\star\star}_{p,k} &= \Big\{\prod_{\alpha=1}^{p}\boldsymbol{\Gamma}[\frac{1}{2}(k+1-\alpha)]\Big\}^{-1}\Big(\frac{k}{2}\Big)^{p(2k-p-1)/4}\pi^{-p(p-1)/4}\\
&\quad\times(2\pi)^{\frac{p(p+1)}{4}}2^{-\frac{p(p-1)}{4}}\exp\Big(-\frac{kp}{2}\Big) \qquad (3.15)
\end{aligned}$$

Using the Stirling's formula in (3.15), we get

$$c^{\star\star}_{p,k} = 1-\frac{p(2p^2+3p-1)}{24k}+\mathrm{O}(k^{-2})$$

The structure of (3.14) resembles the characteristic function of a non-central χ^2 distribution with some correction terms. The degrees of freedom and non-centrality parameter are $p(p+1)/2$ and $\frac{1}{2}\mathrm{tr}\boldsymbol{\Theta}^2$ respectively. To evaluate the correction terms, we will substitute $\mathbf{Z}+g(\boldsymbol{\Theta})$ for $\mathbf{Y}$. Note that matrix trace is a linear operator, then all the odd moments of $\mathbf{Z}$ are zero. We will also use the following results:

$$\begin{aligned}
\mathrm{E}_Z(\mathbf{Z}^2) &= \frac{p+1}{2(1-2it)}\mathbf{I}\\
\mathrm{E}_Z(\mathrm{tr}\mathbf{Z}^2) &= \frac{p(p+1)}{2(1-2it)}\\
\mathrm{E}_Z\left(\mathrm{tr}(\mathbf{Z}^2\boldsymbol{\Theta})\right) &= \frac{(p+1)\mathrm{tr}\boldsymbol{\Theta}}{2(1-2it)},
\end{aligned}$$

so that for the $k^{-1/2}$ order term F_1

$$
\begin{aligned}
\mathrm{E}_Y(\mathrm{F}_1) &= \mathrm{E}_Y\Big\{ \Big(it-\frac{1}{6}\Big)\,\mathrm{tr}\mathbf{Y}^3 + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}} + \frac{\mathrm{tr}\boldsymbol{\Theta}\mathbf{Y}^2}{2\sqrt{2}} - \frac{\mathrm{tr}\boldsymbol{\Theta}^2\mathbf{Y}}{2} \Big\} \\
&= \mathrm{E}_Z\Big\{ \Big(it-\frac{1}{6}\Big)\,\mathrm{tr}\Big(3\mathbf{Z}^2 g(\boldsymbol{\Theta}) + g(\boldsymbol{\Theta})^3\Big) - \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{6\sqrt{2}} \\
&\qquad + \frac{\mathrm{tr}\big(\boldsymbol{\Theta}(\mathbf{Z}^2 + g(\boldsymbol{\Theta})^2)\big)}{2\sqrt{2}} - \frac{\mathrm{tr}\big(\boldsymbol{\Theta}^2 g(\boldsymbol{\Theta})\big)}{2} \Big\} \\
&= \mathrm{E}_Z\Big\{ \frac{3(it-\frac{1}{6})\mathrm{tr}\mathbf{Z}^2\boldsymbol{\Theta}}{\sqrt{2}(1-2it)} + \frac{(it-\frac{1}{6})\mathrm{tr}\boldsymbol{\Theta}^3}{2\sqrt{2}(1-2it)^3} + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}} \\
&\qquad + \frac{\mathrm{tr}\boldsymbol{\Theta}\mathbf{Z}^2}{2\sqrt{2}} + \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{4\sqrt{2}(1-2it)^2} - \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{2\sqrt{2}(1-2it)} \Big\} \\
&= \mathrm{E}_Z\Big\{ \Big(-\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}(1-2it)}\Big)\mathrm{tr}\mathbf{Z}^2\boldsymbol{\Theta} \\
&\qquad + \Big(\frac{1}{3\sqrt{2}} - \frac{1}{2\sqrt{2}(1-2it)} + \frac{1}{6\sqrt{2}(1-2it)^3}\Big)\mathrm{tr}\boldsymbol{\Theta}^3 \Big\} \\
&= \Big(-\frac{1}{\sqrt{2}(1-2it)} + \frac{1}{\sqrt{2}(1-2it)^2}\Big)\frac{p+1}{2}\mathrm{tr}\boldsymbol{\Theta} \\
&\qquad + \Big(\frac{1}{3\sqrt{2}} - \frac{1}{2\sqrt{2}(1-2it)} + \frac{1}{6\sqrt{2}(1-2it)^3}\Big)\mathrm{tr}\boldsymbol{\Theta}^3 \\
&= b_0 + b_1(1-2it)^{-1} + b_2(1-2it)^{-2} + b_3(1-2it)^{-3} \qquad (3.16)
\end{aligned}
$$

where

$$
\begin{aligned}
b_0 &= \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{3\sqrt{2}}, \\
b_1 &= -\frac{(p+1)\mathrm{tr}\boldsymbol{\Theta}}{2\sqrt{2}} - \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{2\sqrt{2}}, \\
b_2 &= \frac{(p+1)\mathrm{tr}\boldsymbol{\Theta}}{2\sqrt{2}}, \\
b_3 &= \frac{\mathrm{tr}\boldsymbol{\Theta}^3}{6\sqrt{2}}.
\end{aligned}
$$

Let $P(d,\delta^2)$ be the lower tail probability of the noncentral χ^2-distribution with d degrees of freedom and non-centrality parameter δ^2. By combining the results from

Equation 3.14, 3.16 and inverting the characteristic function, we obtain the limiting distribution of Nagao's test T_{NT} under the sequence of local alternatives H_1.

Theorem 4 *Under the sequence of local alternatives H_1, as $k \to \infty$, the limiting distribution of Nagao's test statistic $T_{NT} = ktr(\mathbf{S}-\mathbf{I})^2/2$ is χ^2 with $d = p(p+1)/2$ degrees of freedom and non-centrality parameter $\delta^2 = tr\mathbf{\Theta}^2/2$. More accurately,*

$$P(\,T_{NT} < \chi^2) = P(d,\delta^2) + \sqrt{\frac{2}{k}}\Big\{b_0 P(d,\delta^2) + b_1 P(d+2,\delta^2) + b_2 P(d+4,\delta^2) + b_3 P(d+6,\delta^2)\Big\} + O(k^{-1}) \tag{3.17}$$

where $b_i, i = 0,1,2,3$, are given above.

In terms of the original observations $\mathbf{X}_t$, the asymptotic distribution should be χ^2 with $d = p(p+1)/2$ degrees of freedom and noncentrality parameter $\delta^2 = k\mathrm{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2/2$. The $k^{-1/2}$ order expansion terms involve χ^2-distributions with the same non-centrality parameter δ^2 and degrees of freedom $d, d+2, d+4, d+6$. Note also that

$$\sum_{i=0}^{3} b_i = 0,$$

implying

$$\mathrm{E}(\mathrm{T}_{NT}) = \tau^2 + d + \sum_{i=0}^{3} 2\sqrt{2} i b_i/\sqrt{k} + \mathrm{O}(k^{-1}).$$

In Section 3.1, we pointed out the monotonicity of the log likelihood ratio test with respect to the departure of the eigenvalues $l_1, l_2, \ldots, l_p$ of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$ from 1. The result also holds asymptotically for Nagao's test since the non-centrality parameter

$$\delta^2 = k\mathrm{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2/2 = k\sum_{i=1}^{p}(l_i - 1)^2/2.$$

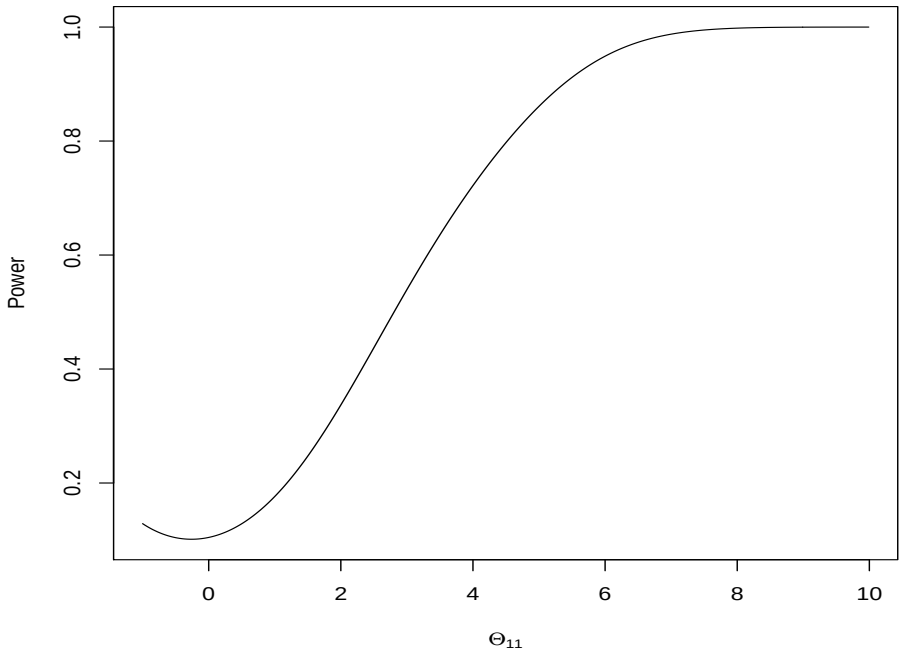

Figure 3.8: Power function of Nagao's test T_{NT} against the variance change of the first variable of a two dimensional observation vector. The x-axis $\mathbf{\Theta}_{11}$ is the (1,1) entry of $\mathbf{\Theta}$ matrix, which indicates the size and direction of change. Significance level = 0.05.

Hence the further the eigenvalue $l_i, (i = 1, 2, \ldots, p)$ departs from 1, the greater the value of δ^2. Also similar to the log likelihood ratio test, both increase and decrease of l_i contribute to δ^2 since δ^2 is a sum of squares of deviations. These individual contributions are additive as well.

Figure 3.8 plots the power function of T_{NT} against various changes of the variability of the first dimension in a two dimensional vector. The graph clearly shows that the Nagao's test T_{NT} has power against both variability increase and decrease.

3.5 Improved Estimation of Covariance Matrix

In the implementation of a SPC scheme, batch (subgroup) samples are taken in a temporal order. Let the batch size be m. To monitor the process variability, one needs an estimator of $\mathbf{\Sigma}$ which can best reflect the current state of the process, in control or out of control.

Let $\mathbf{X}_{ij}, j = 1, 2, \ldots m$ be the observations of the ith subgroup. The traditional and widely used estimator at time i, assuming the mean is known and zero, is

$$\mathbf{S}_i^{\star} = \sum_{j=1}^{m} \frac{X_{ij} X_{ij}^T}{m}.$$

We usually want the batch sample size m to be always larger than the dimensionality p so that the sample variance-covariance matrix has full rank. However sometimes due to economical or practical restrictions, the batch size is small, and in some cases, $m = 1$. If we only use samples from one subgroup to estimate process variance covariance matrix, the estimator $\mathbf{S}_i$ is often singular ($m < p$) and unstable.

A new estimator is needed. It is natural to "borrow strength" from the past and use a combination of past and current observations in the estimation of variance-covariance matrix $\mathbf{\Sigma}$. Here we suggest using a weighted sum of squares. For the ith subgroup with observations $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{im}$, let

$$\mathbf{V}_i = \sum_{j=1}^{m} \mathbf{X}_{ij} \mathbf{X}_{ij}^T / m.$$

Then the overall weighted average estimator of $\mathbf{\Sigma}$ at time t is

$$\begin{aligned}\mathbf{S}_t &= \sum_{i=t-n+1}^{t} w_i \mathbf{V}_i \\ &= \sum_{i=t-n+1}^{t} \sum_{j=1}^{m} w_i X_{ij} X_{ij}^T / m, \qquad w_i \leq w_j. \end{aligned} \tag{3.18}$$

The window size n measures how far we look back. We use $n-1$ past subgroups and current subgroup, and w_i is the weight of the ith subgroup. Since the more recent subgroups carry the more up-to-date information about the process, it therefore makes sense to give higher weights to more recent subgroups and lower weights to older ones, hence the requirements $w_i \leq w_j$. All the observations within each subgroup share the same weight w_i/m. The estimator $\mathbf{S}_t$ no longer has an exact Wishart distribution.

Here we consider two weighted sums, exponentially weighted sum of squares (EWSS) and polynomially weighted sum of squares (PWSS). The forms and properties of these two kinds of weighting schemes are introduced in the next two sections.

3.5.1 Exponentially Weighted Sum of Squares (EWSS)

The exponentially weighted sum of squares, $\mathbf{S}_t^{EWSS}$, like the EWMA control scheme, is defined recursively as

$$\begin{aligned}\mathbf{S}_t^{EWSS} &= \lambda \sum_{j=1}^{m} X_{tj} X_{tj}^T/m + (1-\lambda)\mathbf{S}_{t-1} \\ &= \lambda \sum_{j=1}^{m} X_{tj} X_{tj}^T/m + \lambda(1-\lambda)/m \sum_{j=1}^{m} X_{(t-1)j} X_{(t-1)j}^T \\ &\quad + \cdots + (1-\lambda)^t/m \sum_{j=1}^{m} X_{0j} X_{0j}^T \\ &= \sum_{i=1}^{t} \sum_{j=1}^{m} \lambda(1-\lambda)^{t-i}/m X_{ij} X_{ij}^T + (1-\lambda)^t/m \sum_{j=1}^{m} X_{0j} X_{0j}^T\end{aligned}$$

where $\mathbf{S}_0^{EWSS} = \sum_{j=1}^{m} X_{0j} X_{0j}^T/m$ is not from observed subgroup and usually taken as the in control process dispersion matrix $\mathbf{\Sigma}_0$.

The sum of weights is

$$\sum_{i=0}^{t} \sum_{j=1}^{m} w_i/m = 1 - (1-\lambda)^t + (1-\lambda)^t = 1$$

The ratio of the weights between two adjacent subgroups is a constant:

$$\frac{w_i}{w_{i-1}} = \frac{1}{1-\lambda} > 1.$$

3.5.2 Polynomially Weighted Sum of Squares (PWSS)

In polynomially weighted sum of squares (PWSS), we use a moving window of size n to estimate $\mathbf{\Sigma}$:

$$\mathbf{S}_t^{PWSS} = \sum_{i=t-n+1}^{t} \sum_{j=1}^{m} \frac{(i-t+n)^c X_{ij} X_{ij}^T}{(1+2^c+\cdots+n^c)m}.$$

The weight of the ith subgroup is $w_i = (i - t + n)^c/(1 + 2^c + \cdots + t^c)$, and the weight of the jth observation in the ith subgroup is w_i/m. Here the parameter c determines the rate at which "older" data enter into the calculation of the SWSS statistic. A large value of c gives more weight to recent data and less weight to older data; a small value of c gives more weight to older data. The parameter n is the window size, a large value of n means more past subgroups are considered in the weighting scheme.

Unlike in EWSS, the ratio of weights is not a constant:

$$\frac{w_i}{w_{i-1}} = \Big(\frac{i}{i-1}\Big)^c > 1$$

Both EWSS and PWSS are unbiased estimators of $\mathbf{\Sigma}$. EWSS considers all the past available observations while PWSS only considers those observations that fall into the fixed size moving window. One natural advantage EWSS has against PWSS is that its calculation only needs the most recent EWSS value and current subgroup, while PWSS needs all the past and current n (window size) subgroups. Also EWSS only has one tuning parameter λ while PWSS has two, n and c.

3.5.3 Satterthwaite's Approximation

In this section, we shall study the distributional behavior of $\mathbf{S}_t^{EWSS}$, $\mathbf{S}_t^{PWSS}$, or more generally any weighted sum of squares $\mathbf{S}_t = \sum_{i=t-n+1}^{t} \sum_{j=1}^{m} w_i X_{ij} X_{ij}^T/m$. Recall that when the process is under control, the observations $\mathbf{X}_{ij}$ are assumed to be independent normally distributed with mean $\mathbf{0}$ and variance covariance matrix

$\mathbf{\Sigma}$. Under these assumptions,

$$\mathbf{X}_{ij}\mathbf{X}_{ij}^T \backsim \mathbf{W}_p(1, \mathbf{\Sigma}).$$

Let

$$\mathbf{X} = \Big[X_{(t-n+1)1},\ X_{(t-n+1)2},\ \ldots,\ X_{(t-n+1)m},\ \ldots,\ X_{tm}\Big]_{p\times nm}.$$

Then we say that the random matrix $\mathbf{X}$ is normally distributed with mean zero and covariance matrix $\mathbf{\Sigma}_X = \text{Var-cov}(\text{Vec}(\mathbf{X}))$. Then $\mathbf{S}_t$ can be rewritten as

$$\mathbf{S}_t = \mathbf{X}\mathbf{W}\mathbf{X}^T \tag{3.19}$$

where

$$\mathbf{W} = 1/m \times \begin{bmatrix} w_{t-n+1}\mathbf{I}_m & \mathbf{0} & \ldots & \\ \mathbf{0} & w_{t-n+2}\mathbf{I}_m & \ldots & \\ \vdots & & \vdots & \\ \mathbf{0} & \mathbf{0} & \ldots & w_t\mathbf{I}_m \end{bmatrix}_{nm\times nm}$$

w_i each has multiplicity of m in $\mathbf{W}$. Hence $\mathbf{S}_t$ is a matrix quadratic form with diagonal weight matrix $\mathbf{W}$.

Many authors have studied the distributional behavior of the quadratic forms like above. $\mathbf{S}_t$ is a nonnegative linear combination of Wishart distributed random matrices with degrees of freedom 1 and scale matrix $\mathbf{\Sigma}$, we are natually motivated to think that the linear combination itself might also be exactly distributed as Wishart. This is true for some special cases; a simple example is when $\mathbf{W} = Diag(1/n)$. When w_i's are not identical, for the univariate case, the celebrated Cochran's therem gives the conditions under which a quadratic form is χ^2 distributed. Many others

have since extended the scope to more general setups. Anderson [1] discussed the multivariate analogue of Cochran's theorem. Wong *et al.* [55] [56] derived more general conditions under which a linear combination of Wishart distributed random matrices is also exactly distributed as Wishart.

Mathew *et al.* [28] showed that a general matrix quadratic form $\mathbf{XBX}^T$, where $\mathbf{B} = \mathbf{AA}^T$ is a nonnegative definite $n \times n$ matrix of rank r and $\mathbf{A}$ is an $n \times r$ matrix of rank r, follows Wishart$(b, \mathbf{\Sigma}/b)$ if and only if the following condition holds:

$$(\mathbf{A^T} \otimes \mathbf{I_p})\mathbf{\Sigma_X}(\mathbf{A} \otimes \mathbf{I_p}) = \mathbf{C} \otimes \mathbf{\Sigma}$$

with $\mathbf{C}$ being a symmetric and idempotent $r \times r$ matrix and $b = rank(\mathbf{C})$. However some simple verifications showed that neither EWSS nor PWSS satisfies the required conditions.

In the univariate case, Satterthwaite [40] proposed a technique to approximate the distribution of a linear combination of χ^2 variates by that of a single scaled χ^2 with appropriate degrees of freedom. Univariate Satterthwaite's Approximation matches the first two moments of the linear combination and another χ^2 variate, hence it's a method of moments estimation.

When $p = 1$, that is when we have univariate observations, the approximation of S_t is given by

$$S_t \backsim a\chi^2_k,$$

where

$$\mathrm{E}(S_t) = \sigma^2 \qquad\qquad \mathrm{E}(a\chi^2_k) = ak$$

$$\mathrm{Var}(S_t) = 2\sum_{i=t-n+1}^{t} w_i^2\sigma^4/m \qquad\qquad \mathrm{Var}(a\chi^2_k) = 2a^2k.$$

By equating the first two moments, we have

$$\sigma^2 = ak$$

$$\sum_{i=t-n+1}^{t} w_i^2\sigma^4/m = a^2k$$

The solution is

$$a = \sum_{i=t-n+1}^{t} w_i^2\sigma^2/m$$

$$k = m/\sum_{i=t-n+1}^{t} w_i^2.$$

When $p > 1$, it is natural to extend the idea of the univariate Satterthwaite Approximation to multivariate data, that is to approximate the distribution of $\mathbf{S}_t$ by

$$\mathbf{S}_t \backsim \mathrm{Wishart}(k, \mathbf{\Sigma}')$$

However this approximation is complicated by the fact that there are more expectations, variances and covariances (p+$p(p$+$1)/2$ of them) than the number ($p(p$+$1)/2$) of unknowns in $\mathbf{\Sigma}$. Tan and Gupta [52] suggested solving the problem by equating the expectations and generalized variances.

Mortarino [31] summarized the multivariate Satterthwaite estimation results as:

$$\mathbf{U} = \sum_{i}^{n} \frac{w_i}{m_i}\mathbf{W}_p(m_i, \mathbf{\Delta}_i)$$

That is, $\mathbf{U}$ is a linear combination of Wishart distributed random matrices $\mathbf{W}_p$ with weights, degrees of freedom and scale matrices shown above. Then the distribution of $\mathbf{U}$ can be approximated as

$$\mathbf{U} \backsim \mathbf{W}_p(k, \mathbf{\Delta})$$

with

$$k = \left\{ \frac{(det(\sum_{i=1}^{n} w_i \mathbf{\Delta}_i)^{p+1}}{det(\Delta^\star)} \right\}^{\frac{2}{p(p+1)}},$$
$$\mathbf{\Delta} = \frac{\sum_{i=1}^{n} \mathbf{\Delta}_i}{k}$$

where $\mathbf{\Delta}^\star = \mathbf{K}_p^- \left\{ \sum_{i=1}^{n} w_i^2 (\mathbf{\Delta}_i \otimes \mathbf{\Delta}_i)/m_i \right\} \mathbf{K}_p$, the matrix $\mathbf{K}_p$ and its Moore-Penrose inverse $\mathbf{K}_p^-$ have the property that

$$\text{Det}\left\{ \mathbf{K}_p^- (\mathbf{D} \otimes \mathbf{D}) \mathbf{K}_p \right\} = \{\text{Det}(\mathbf{D})\}^{p+1}$$

for any symmetric matrix $\mathbf{D}$ of rank p.

Applying the above technique to $\mathbf{S}_t$ in (3.19), then

$$\mathbf{S}_t \backsim \mathbf{W}_p(k, \mathbf{\Sigma}'), \tag{3.20}$$

and the degrees of freedom k and the corresponding scale matrix can be additionally simplified as

$$k = \left\{ (1/(\sum_{i=t-n+1}^{t} w_i^2/m)^{\frac{p(p+1)}{2}} \right\}^{\frac{2}{p(p+1)}} = m/ \sum_{i=t-n+1}^{t} w_i^2 \tag{3.21}$$
$$\mathbf{\Sigma}' = \mathbf{\Sigma}/k.$$

The approximate degrees of freedom does not depend on p.

Applying (3.18) to EWSS and PWSS, we get:

$$k^{EWSS} = \frac{m}{(\sum_{i=0}^{t}(\lambda(1-\lambda)^{t-i})^2)} \rightarrow \frac{m(2-\lambda)}{\lambda}, \qquad \text{as } t \rightarrow \infty$$

and

$$k^{PWSS} = \frac{m(\sum_{i=t-n+1}^{t} i^c)^2}{\sum_{i=1}^{t} i^{2c}}$$

. Since

$$\frac{1}{c+1} = \frac{x^{c+1}}{c+1}\bigg|_0^1 = \int_0^1 x^c\,dx \approx \sum_{i=1}^{n} \frac{1}{n}\Big(\frac{i}{n}\Big)^c, \qquad \text{when } n \text{ is large,}$$

$$\sum_{i=t-n+1}^{t} i^c \approx \frac{n^{c+1}}{c+1}.$$

We have

$$k^{SWSS} \approx \frac{m(\frac{n^{c+1}}{c+1})^2}{\frac{n^{2c+1}}{2c+1}} = \frac{mn(2c+1)}{(c+1)^2}.$$

We will refer to k^{EWSS} and k^{PWSS} as the respective "effective degrees of freedom" in EWSS and PWSS. The corresponding scale matrix can be derived accordingly.

Table 3.5 shows the "effective degrees of freedom" for different choices of λ for the EWSS scheme with subgroup size 1. Figure 3.9 displays the weights for various

Table 3.5: Effective degrees of freedom for various λ

λ	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.3	0.5
k	199	99	66	49	39	33	28	24	22	19	9	7	3

choice of λ and different lags. Table 3.6 shows the "effective degrees of freedom" for

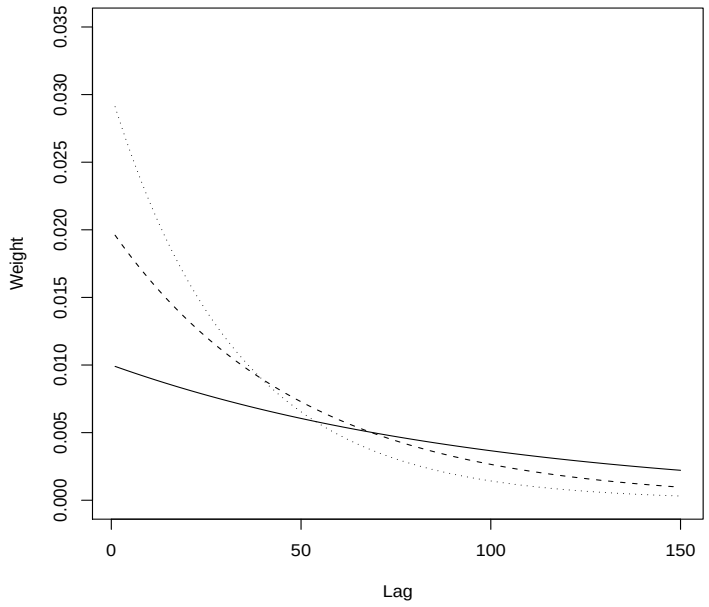

Figure 3.9: Weight vs. lag for various λ. Solid line: $\lambda = 0.01$. Dashed line: $\lambda = 0.02$. Dotted line: $\lambda = 0.03$

different choice of window size n and weighting parameter c for the PWSS scheme with subgroup size 1. Figure 3.10 compares the weight curves for various weighting factors c when window size is 100.

Table 3.6: Effective degrees of freedom for various choice of c and n

	Weighting factor c								
Window Size n	0	0.5	1	1.5	2	2.5	3	3.5	4
100	100	89	75	64	56	49	44	39	36
200	200	178	150	128	111	98	88	79	72

It is worth noting that the use of an exponentially weighted sum of squares type

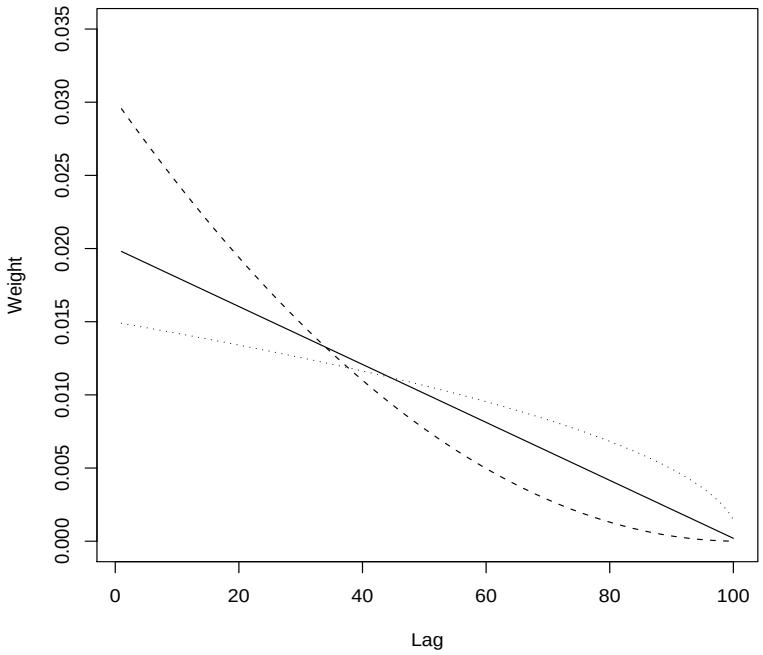

Figure 3.10: Weight vs. lag for various c. Solid line: $c = 0.2$. Dashed line: $c = 2$. Dotted line: $c = 1$. Window Size 100

of estimator is no stranger to statistical control of process dispersion. Montgomery and Mastrangelo [30] used exponentially weighted mean squares (EWMS) as the basis for a control scheme for monitoring a univariate process. They suggested using a χ^2 approximation for the distribution of the EWMS, and the degrees of freedom for such an approximation is $(2-\lambda)/\lambda$, which agrees with the result shown above. They even went further to discuss the distribution of EWMS when the observations involved are not independently distributed. Their results suggest that the "effective degrees of freedom" is a function of both the weighting parameter λ and the correlation structure among these dependent observations. In this dissertation, we expand their results to the multivariate case and look at the generalized weighting schemes instead of just exponential weighting. However our results are restricted

to the independent case only. To the author's knowledge, the expansion to the approximation of the distribution of a linear combination of multivariate correlated observations has yet to be done.

As for the measure of the closeness of the Satterthwaite approximation to the true distribution of the linearly weighted sum of squares, there appears to be only a very limited number of publications about this topic. Tan and Gupta [52] compared the theoretical quantiles (from the approximate distribution formula) with quantiles from Monte Carlo simulation. They concluded that adding more terms (considering higher moments) does not improve the approximation significantly in the general situation. Khuri ([22], [23]) looked at this issue from the perspective of matching the characteristic functions. He suggested that the Satterthwaite approximation does not hold well when some of the weights are negative.

3.6 Numerical Study

We have studied the asymptotic distributions of Nagao's test T_{NT} and the modified likelihood ratio test T_{LR} under both local and distant alternatives, but the sample variance covariance matrix $\mathbf{S}$ used in the derivations follows an exact Wishart distribution. However, if we use a weighted average type estimator, like EWSS or PWSS, in the test statistics, the accuracy of such a substitution needs to investigated. There are two levels of uncertainty involved here. The first layer is the approximation of the distribution of test statistics and the second layer is the use of an unequally weighted sum of squares estimator which is assumed to be

approximately Wishart distributed but in fact is not. A closed form expression for the asymptotic distribution seems to be rather hard to get, so we will do a numerical study of selected cases comparing the power and level of the tests.

We consider a three dimensional observation vector $\mathbf{X} = [x_1\ x_2\ x_3]^T$ which has variance covariance matrix $\mathbf{\Sigma}$. Under the null hypothesis, $\mathbf{\Sigma} = \mathbf{I}_3$. Under the alternative hypothesis

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 + a/\sqrt{k} \end{bmatrix}$$

where a indicates the direction and size of shifts and k is the "effective degrees of freedom" of the sample variance covariance estimator. For simplicity, we only consider EWSS with subgroup size $m = 1$.

Nagao [34] demonstrated that as the degrees of freedom $k \to \infty$, the asymptotic expansion of the null distribution of T_{NT}, in terms of χ^2 distributions, is

$$\begin{aligned} \Pr(\mathrm{T}_{NT} \le x) = \mathrm{P}_d + \frac{1}{k}\Big\{ & -\frac{p(2p^2+3p-1)}{24}\mathrm{P}_d + \frac{p(p+1)^2}{2}\mathrm{P}_{d+2} \\ & - \frac{p(6p^2+13p+9)}{8}\mathrm{P}_{d+4} + \frac{p(4p^2+9p+7)}{12}\mathrm{P}_{d+6}\Big\} \\ & + \mathrm{O}(k^{-1}). \end{aligned} \tag{3.22}$$

where $d = p(p+1)/2$ and $\mathrm{P}_d = \mathrm{P}(\chi^2_f \le x)$.

Figures 3.11-14 display the comparison between the level of T_{NT}, $\mathrm{P}(\mathrm{T}_{NT} > c)$, using the traditional equally weighted sample variance covariance estimator (exactly Wishart distributed, with accuracy up to $\mathrm{O}(1/k)$ order of expansion, from (3.22)), and the actual level of T_{NT}, as estimated using EWSS by 5000 Monte Carlo replica-

tions. The λ values considered are $0.01, 0.02, 0.03, 0.05$ which have effective degrees of freedom $199, 99, 67, 39$ respectively. Overall from the simulations, we see that for the λ values considered, the asymptotic value agrees pretty well with the Monte Carlo value and as λ increases (effective degrees of freedom decreases), the match tends to get worse.

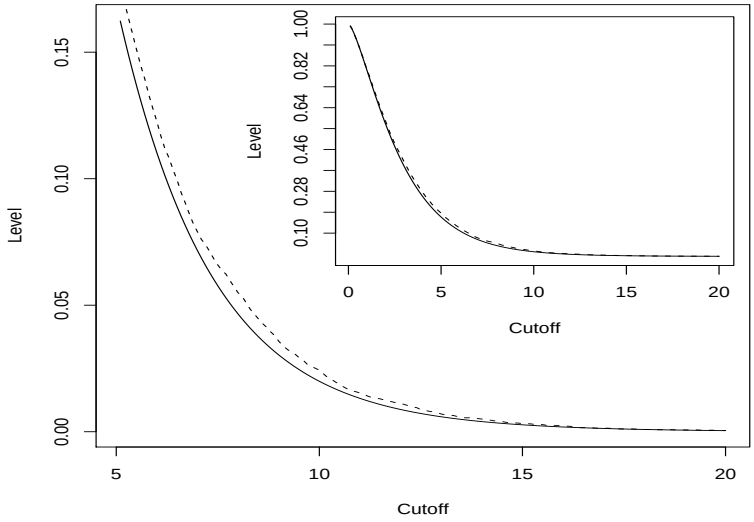

Figure 3.11: Level graph for $\lambda = 0.01$. Solid line : asymptotic formula (3.22), Dashed line : Monte Carlo simulation of true probability. Inset shows all levels from 0 to 1.

Figures 3.15-18 compare the rate of rejecting the null hypothesis for various shift a when $\lambda = 0.01$. We use the asymptotic expansion of the power function of T_{NT} under local alternatives in Theorem 4 to calculate the theoretical rate of rejection. Figure 3.18 suggests a large discrepancy between the theoretical result and Monte Carlo simulation. The source of this large discrepancy when the cutoff is large could be due to that $a = 10$ falls out of the local alternative category. It could be explained as that the $\mathrm{O}(k^{-1})$ terms in the asymptotic expansion formula

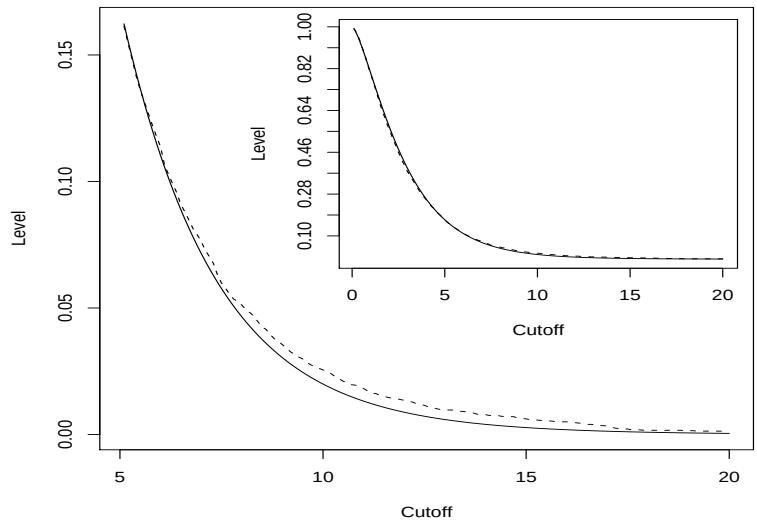

Figure 3.12: Level graph for $\lambda = 0.02$. Solid line : asymptotic formula (3.22), Dashed line : Monte Carlo simulation of true probability. Inset shows all levels from 0 to 1.

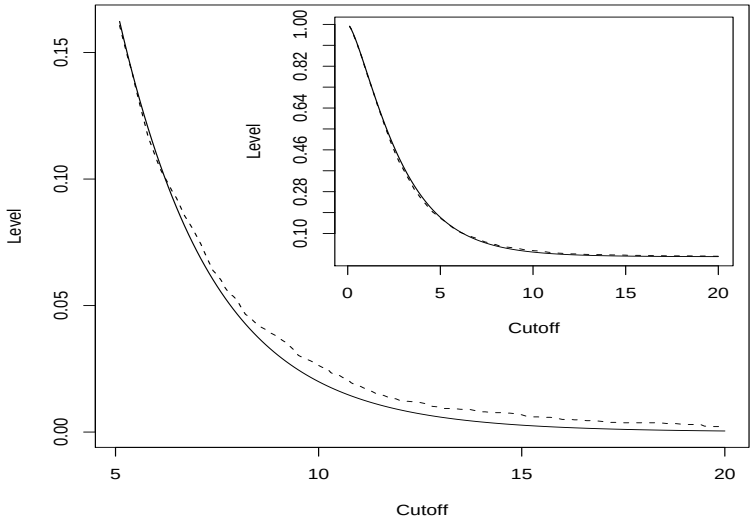

Figure 3.13: Level graph for $\lambda = 0.03$. Solid line : asymptotic formula (3.22), Dashed line : Monte Carlo simulation of true probability. Inset shows all levels from 0 to 1.

(3.17) are no longer negligible for large shifts.

Last we compare the Monte Carlo simulation of the power of T_{NT} using EWSS with the theoretical power from Equation 3.17 in Figure 3.20 with $\lambda = 0.01$ and significance level=0.05. The maximum difference is 0.04 which happens at $a = 6$.

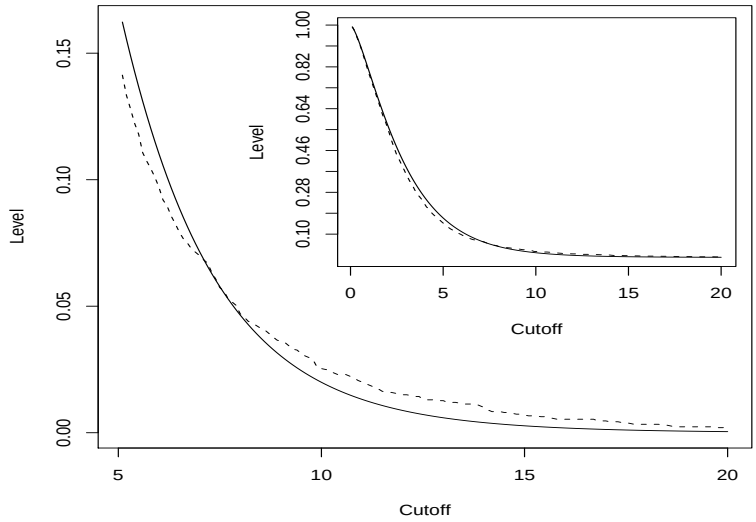

Figure 3.14: Level graph for $\lambda = 0.05$. Solid line : asymptotic formula (3.22), Dashed line : Monte Carlo simulation of true probability. Inset shows all levels from 0 to 1.

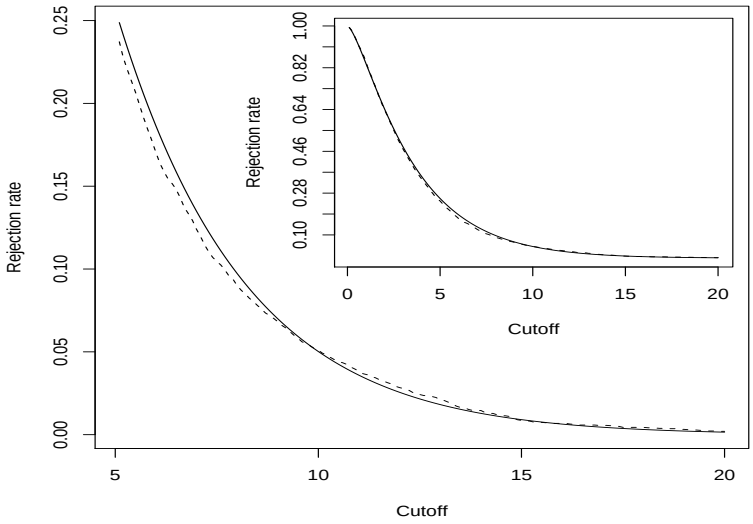

Figure 3.15: Rate of rejection vs. cutoff. $\lambda = 0.01$ and $a = 1$. Solid line : asymptotic formula (3.17), Dashed line : Monte Carlo simulation of the true power. Inset shows all rejection rates.

From the graph we also observe that the Monte Carlo power is consistently below the theoretical power which is expected probably because of the uncertainty added

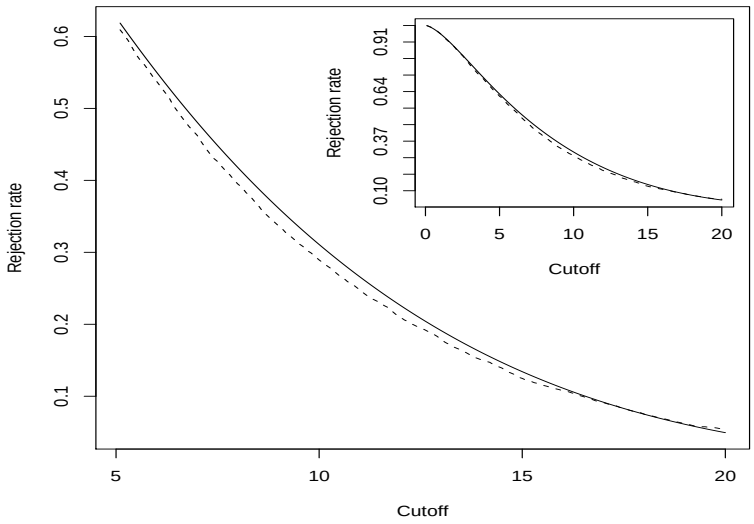

Figure 3.16: Rate of rejection vs. cutoff. $\lambda = 0.01$ and $a = 3$. Solid line : asymptotic formula (3.17), Dashed line : Monte Carlo simulation of the true power. Inset shows all rejection rates.

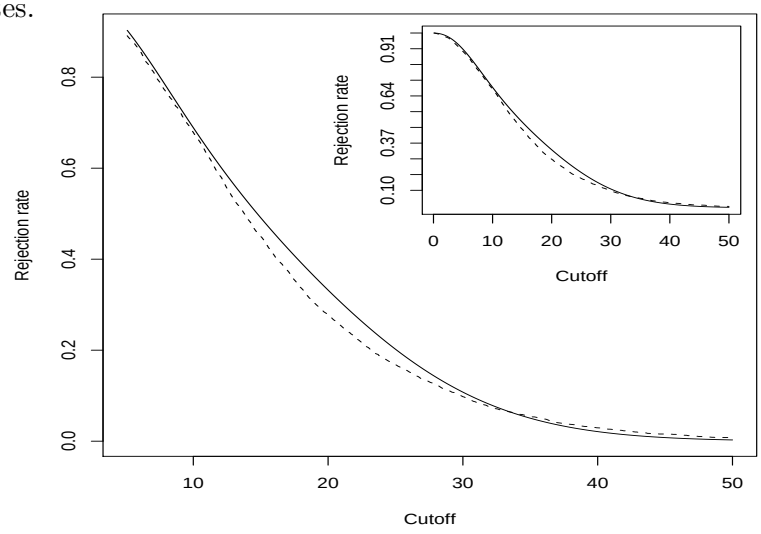

Figure 3.17: Rate of rejection vs. cutoff. $\lambda = 0.01$ and $a = 5$. Solid line : asymptotic formula (3.17), Dashed line : Monte Carlo simulation of the true power. Inset shows all rejection rates.

by the Satterthwaite approximation.

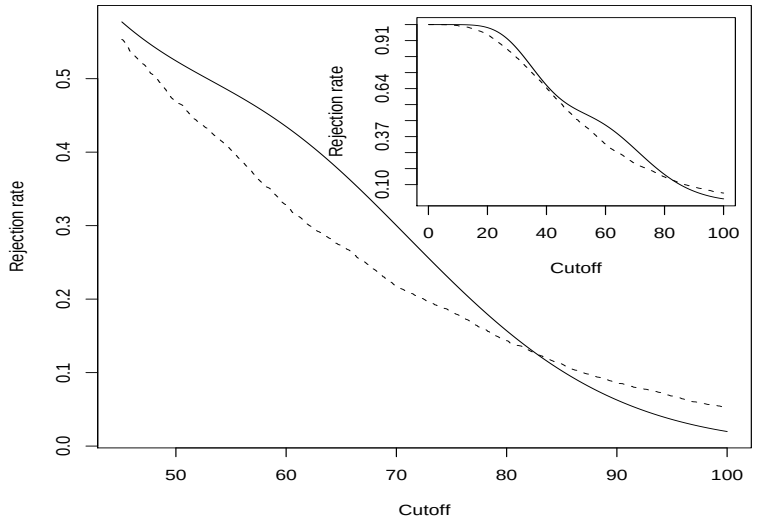

Figure 3.18: Rate of rejection vs. cutoff $\lambda = 0.01$ and $a = 10$. Solid line : asymptotic formula (3.17), Dashed line : Monte Carlo simulation of the true power. Inset shows all rejection rates.

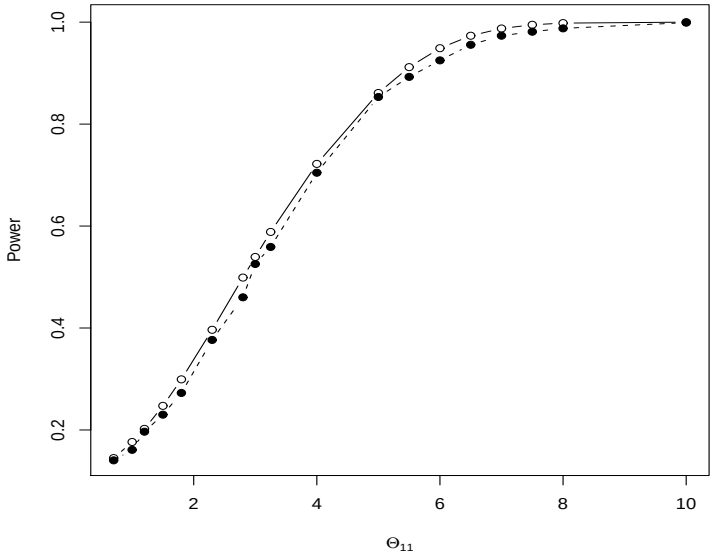

Figure 3.19: Comparison of Power functions. Solid line, circle: asymptotic formula (3.17). dashed line, solid circle: Monte Carlo simulation of the true power. P_1 is consistently below P_2 with maximum difference 0.04

Chapter 4

Proposed Algorithms

4.1 SPC Application

In Chapter 3, we discussed the unequally weighted process dispersion matrix estimators and also studied the asymptotic distribution of the log likelihood ratio test and Nagao's test under both distant alternatives and sequences of local alternatives. Recall that the weighted dispersion matrix estimator is

$$\mathbf{S}_t = \sum_{i=t-n+1}^{t} w_i \mathbf{V}_i = \sum_{i=t-1}^{t} \sum_{j=1}^{m} w_i \mathbf{X}_{ij} \mathbf{X}_{ij}^T / m.$$

When the process is under control

$$\mathrm{E}\mathbf{X}_{ij} \mathbf{X}_{ij}^T = \mathbf{I},$$

and when the process is out of control

$$\mathrm{E}\mathbf{X}_{ij} \mathbf{X}_{ij}^T = \mathbf{\Sigma} \neq \mathbf{I}.$$

So the impact of the shift will slowly enter into $\mathbf{S}_t$, and one can see that $\mathbf{S}_t$ will depart from $\mathbf{I}$ more and more as we bring in more new observations. The test statistic T_{NT} is the Frobenius norm that measures that distance between $\mathbf{S}_t$ and $\mathbf{I}$. The larger the discrepancy between $\mathbf{S}_t$ and $\mathbf{I}$, the bigger the average value of T_{NT}. The power function of T_{NT} is also monotone with respect to the discrepancy between the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$ and $\mathbf{I}$. All things considered, we expect that Nagao's test T_{NT}

combined with the improved variance covariance matrix estimator should be a viable way to monitor multivariate process dispersion. Using the improved estimator in the log likelihood ratio test should also be viable since T_{LR} is monotone with respect to the discrepancy of eigenvalues. Alt *et al.* ([2], [3]) also discussed a Shewhart type scheme based on the log likelihood ratio test. They used the traditional equally weighted variance covariance matrix estimator instead of an unequally weighted one.

We denote these two schemes based on Nagao's test and log likelihood ratio test MTNT and MTLR respectively. The stopping rules for them are

$$\mathrm{L}_{MTNT} = \inf\Big\{t \in \mathbb{N} : (k/2)\mathrm{tr}(\mathbf{S}_t\boldsymbol{\Sigma}_0^{-1} - \mathbf{I})^2 > h\Big\}$$
$$\mathrm{L}_{MTLR} = \inf\Big\{t \in \mathbb{N} : -2\ln\big(\mathrm{LR}(\mathbf{S}_t)\big) > h\Big\}$$

where h is the control limit.

In the multivariate setting, $\boldsymbol{\Sigma}$ has $p(p+1)/2$ parameters. There are many ways in which these parameters can change. Here we do not have any specific interest in any particular direction or size of change. Unlike some other schemes that are only concerned with increased variability, the MTNT scheme has power against both the increase and decrease of process dispersion, which means that situations involving the increase or decrease of variance of a single characteristic variable, a subset of characteristic variables, covariances, etc., are all considered. Hence the proposed SPC procedure has the flavor of being "Omnibus".

Knoth ([20], [21]) discussed the EWMA-S^2 chart for monitoring univariate

process variability. The stopping rule of the EWMA-S^2 chart is characterized by

$$\mathrm{L} = \inf\Big\{ t \in \mathbb{N} : S_t > \sigma_0^2 + h\sqrt{\frac{\lambda}{2-\lambda}}\sqrt{\frac{2}{m}}\sigma_0^2 \Big\} \tag{4.1}$$

where m is the sample subgroup size, λ is the EWMA weighting parameter. Some simple change of expression of (4.1) gives

$$\mathrm{L} = \inf\Big\{ t \in \mathbb{N} : m(2-\lambda)(S_t/\sigma_0^2 - 1)^2/2\lambda > h \Big\} \tag{4.2}$$

Hence the EWMA-S^2 chart is a univariate version of the proposed MTNT scheme.

It is to be noted that in the works by Yeh *et al.* [58] and Chang *et al.* [9], a similar EWSS estimator for the process dispersion matrix was used. They pointed out that the asymptotic variance covariance is

$$(\lambda/(2-\lambda))\mathbf{\Sigma}$$

This agrees with our proposal except that they didn't have a distributional assumption.

The optimal design of the proposed algorithm relies upon the choice of the parameters, λ in EWSS or n and c in PWSS. Here we mainly focus upon discussing the EWSS scenario. There are mainly three factors that drive the choice of λ. First, the tests derived in Chapter 3 depend on the asymptotic properties as the "effective degrees of freedom" k approaches ∞ and in the approximation of degrees of freedom

$$k = m(2-\lambda)/\lambda \qquad \text{or} \qquad \lambda = 2m/(k+m)$$

The higher order terms in the approximation of the distribution of T_{NT} have order $\mathrm{O}(1/k)$ or $\mathrm{O}(\lambda)$, hence a small value of λ means small error in the approximation.

Also given a fixed "effective degrees of freedom", if we increase the subgroup sample size m, λ can be made bigger as well, making the scheme more flexible. Finally the choice of λ also depends on the dimension of sample observation $\mathbf{X}_{tj}$. The larger p, the smaller the value λ should be since ideally we want a full rank estimation of dispersion matrix and $m(2/\lambda - 1) \geq p$.

The RiskMetrics database created by J.P.Morgan [16] uses a EWSS type of estimator to estimate and update daily financial market volatility. The company found that $\lambda = 0.04$ gives forecasts of the variance rate that come closest to the realized variance rate. Lucas and Saccucci [25] examined properties of the EWMA control scheme to monitor the mean of a univariate normally distributed process. They pointed out that

> for a fixed in-control ARL, the optimal value of λ increases as the shift in the process increases.

For an in-control average run length of 300, they recommenced using $0.05 \leq \lambda \leq 0.06$ for a mean change from 0 to 0.5, $0.14 \leq \lambda \leq 0.15$ for a mean change from 0 to 1 and $0.38 \leq \lambda \leq 0.42$ for a mean change from 0 to 2. However as pointed out before, the proposed algorithm is "omnibus" and has no particular sensitivity to any size or direction of change. Our repeated simulations suggest that for the newly proposed control scheme, a small value of λ is desirable. When the subgroup sample size is one, we recommend using λ between 0.01 and 0.03. Within this range, the approximate degrees of freedom range from 199 to 66. On the other hand, a too small λ value will cause numerical instability. For the EWMA-S^2 control scheme, the univariate

version of the proposed testing criterion T_{NT}, Knoth [20], in referring to Mittag [29], even discussed λ as small as 0.000042 with $m = 4$ and in control average run length 250. For this value of λ, the corresponding "effective degrees of freedom" is 190472 and the h in (4.1) is 0.000064375308. It is quite clear that in practical situations, SPC practitioners will have a hard time achieving this level of numerical accuracy and a small misspecification of control limit h could be disastrous. Another reason for this is that if λ is too small, it will take a long time until the test statistic process stabilizes. Very few practitioners can afford a 190472 observations long burn in process.

We now examine a numerical example for the proposed testing criteria. We look at a three dimensional process $\mathbf{X} = [x_1, x_2, x_3]^T$ which has in control variance covariance matrix $\mathbf{\Sigma}_0 = \mathbf{I}_3$. Shifts in the variance of the third dimension is introduced at time $t = 300$. The new variance covariance matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 + a/\sqrt{k} \end{bmatrix}$$

where a indicates the shift size and k is the "effective degrees of freedom" of sample variance covariance estimator. Here we use the EWSS with subgroup sample size one and $\lambda = 0.01$, hence $k = 199$.

Figure 4.1 plots the original data of one realization of $\mathbf{X}$ process from time $t = 0$ to $t = 700$. Figure 4.2 shows the corresponding sequence of T_{NT}. Since T_{NT} is based on a weighted average of past and current sum of squares, we expect that there exists autocorrelation in the sequence of test statistics. Figure 4.3 and 4.4 display

the autocorrelation and partial autocorrelation patterns of T_{NT} for $\lambda = 0.01$, from which we can see that T_{NT} for this value of λ resembles an AR(1) process. In fact, if we let T_t be T_{NT} at time t, $\mathbf{S}_t$ be the variance covariance matrix estimator using data from the whole time period, $[0, t]$. Let $\mathbf{V}_t$ be the variance covariance estimator using only data observed at time t, then

$$\begin{aligned}
\mathrm{T}_t &= \frac{k}{2}\mathrm{tr}(\mathbf{S}_t - \mathbf{I})^2 = \frac{k}{2}\mathrm{tr}\{(1-\lambda)\mathbf{S}_{t-1} + \lambda\mathbf{V}_t - \mathbf{I}\}^2 \\
&= \frac{k}{2}\mathrm{tr}\{(1-\lambda)(\mathbf{S}_{t-1} - \mathbf{I}) + \lambda(\mathbf{V}_t - \mathbf{I})\}^2 \\
&= \frac{k}{2}(1-\lambda)^2\mathrm{tr}(\mathbf{S}_{t-1} - \mathbf{I})^2 + k\lambda(1-\lambda)\mathrm{tr}(\mathbf{S}_{t-1} - \mathbf{I})(\mathbf{V}_t - \mathbf{I}) + \frac{k}{2}\lambda^2\mathrm{tr}(\mathbf{V}_t - \mathbf{I})^2 \\
&= (1-\lambda)^2\mathbf{T}_{t-1} + (2-\lambda)(1-\lambda)\mathrm{tr}(\mathbf{S}_{t-1} - \mathbf{I})(\mathbf{V}_t - \mathbf{I}) + \frac{\lambda(2-\lambda)}{2}\mathrm{tr}(\mathbf{V}_t - \mathbf{I})^2.
\end{aligned}$$

Note that $\mathbf{S}_{t-1} - \mathbf{I}$ and $\mathbf{V}_t - \mathbf{I}$ are statistically independent with means both equal to $\mathbf{0}$. So when λ is small, the observed T_{NT} sequence is approximately an AR(1) process with autoregressive coefficient $(1-\lambda)^2$ and the white noise part played by $\frac{\lambda(2-\lambda)}{2}\mathrm{tr}(\mathbf{V}_t - \mathbf{I})^2$.

Figure 4.5 shows the average sample path of 1000 realizations of T_{NT} when $a = 5$, from which we observe four phases of T_{NT}. The first phase is the wiggle at the start. This is the burn-in period when T_{NT} tries to stabilize itself and the impact of the starting value $\mathbf{S}_0$ is phased out. The length of this phase depends upon the choice of starting value for $\mathbf{S}_0$. Choosing $\mathbf{S}_0$ close to the in control variance covariance matrix typically means a short burn-in phase. After the burn-in, T_{NT} stabilizes and enters into a steady state in-control phase during which it oscillates around the in-control mean $p(p+1)/2 = 6$. After the variance of the third dimension shifts from 1 to $1 + a/\sqrt{k}$ at time $t = 300$, T_{NT} displays an nearly linear upward

trend. The slope of the trend depends on the size of the shift; the greater the shift the bigger the slope. After roughly twice the "effective degrees of freedom", T_{NT} enters into the fourth phase, a new stationary state when it oscillates around the new mean $p(p+1)/2 + k\text{tr}(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1} - \mathbf{I})^2/2$ [$=6 + a^2/2$ for the numerical example discussed here].

Figure 4.6 plots the behavior of T_{NT} around the time when the shifts take place (same data from Figure 4.4). We can see that almost immediately after the shifts are introduced, T_{NT} begin its upward trend. This is true even for small size shifts. This quick response to shift is ideal for SPC since we aim for the fastest detection.

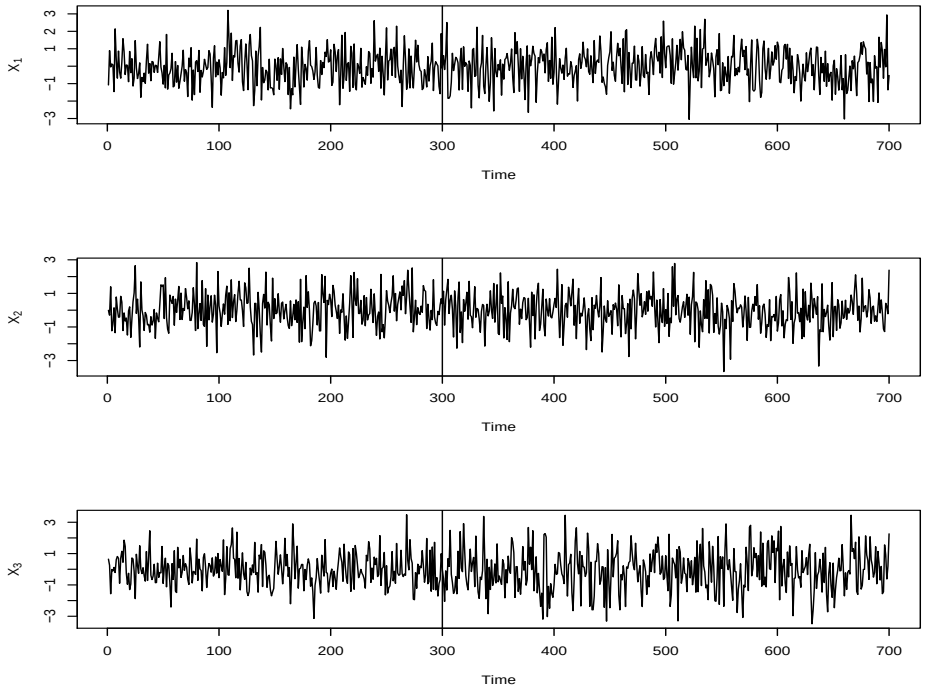

Figure 4.1: Original data for three response variables from $t = 0$ to $t = 700$. In control variance covariance matrix $\mathbf{\Sigma}_0 = \mathbf{I}$, $\lambda = 0.01$, effective degrees of freedom=199, Variance of the third dimension, $\mathbf{\Sigma}_{33}$, shifts to $1 + 10/\sqrt{199}$ at $t = 300$. The vertical line indicates the time when the change happens.

4.2 ARL Study

The numerical studies in Chapter 3 give us some first hand knowledge about the performance of proposed control schemes. The test statistics are highly correlated because of the structure of the unequally weighted variance covariance matrix estimators, so that we can not use the Type I error rate of the distribution derived in Chapter 3 to set up the control limits. In that chapter, we reviewed how to use integral equation methods to calculate the ARL of EWMA/CUSUM for a univariate process. Extension of this methodology to the proposed control scheme seems to be

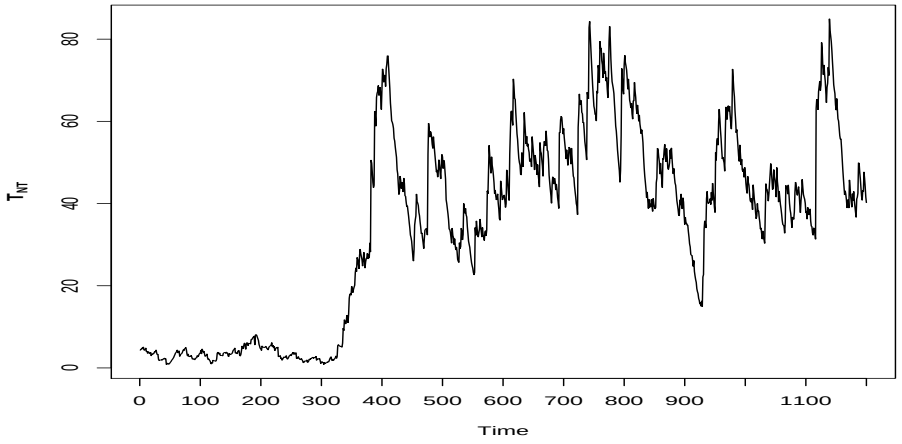

Figure 4.2: One realization of T_{NT} for a three-dimensional observation process. In control variance covariance matrix $\mathbf{\Sigma}_0 = \mathbf{I}$, $\lambda = 0.01$, effective degrees of freedom=199, Variance of the third dimension, $\mathbf{\Sigma}_{33}$, shifts to $1 + 10/\sqrt{199}$ at $t = 300$.

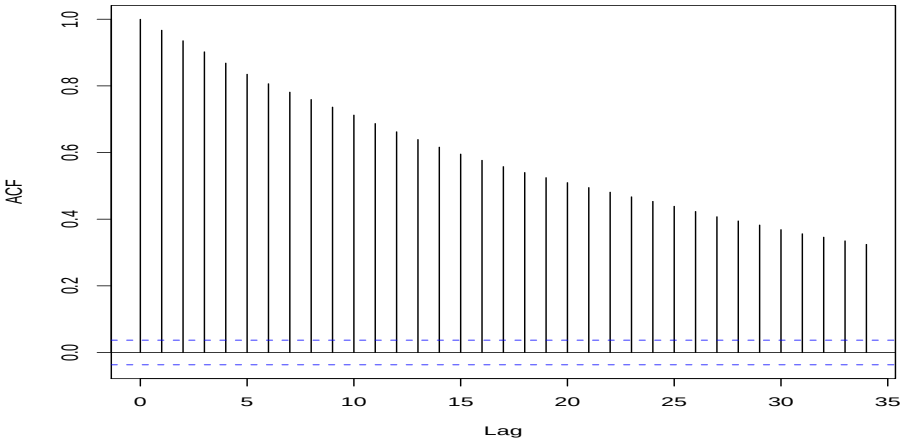

Figure 4.3: ACF plot for T_{NT}. $\lambda = 0.01$. For this small value of λ, T_{NT} displays significant autocorrelation structure. Even at lag 34, the autocorrelation is still 0.31.

rather complicated.

Let $\mathbf{S}$ be initialized at $\mathbf{S}_0 = \mathbf{A}$, where $\mathbf{A}$ is symmetric and positive definite,

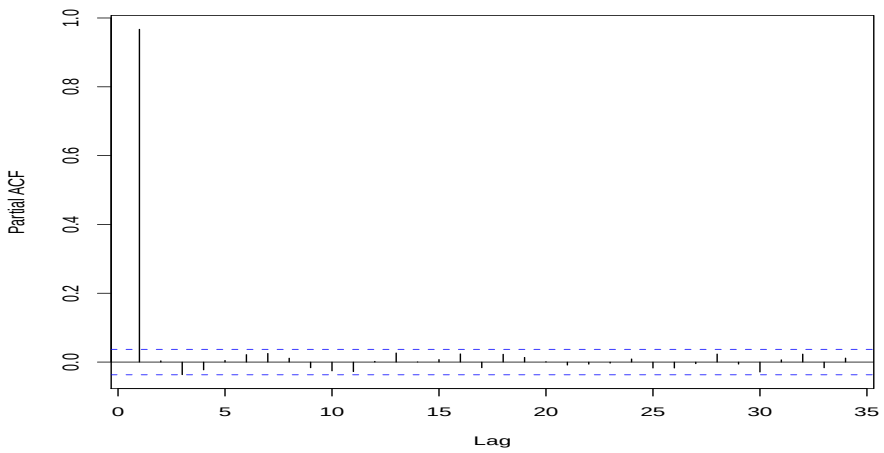

Figure 4.4: PACF plot for T_{NT}. $\lambda = 0.01$. For this small value of λ, the first partial autocorrelation is nearly 1, but the others are not significantly different from 0.

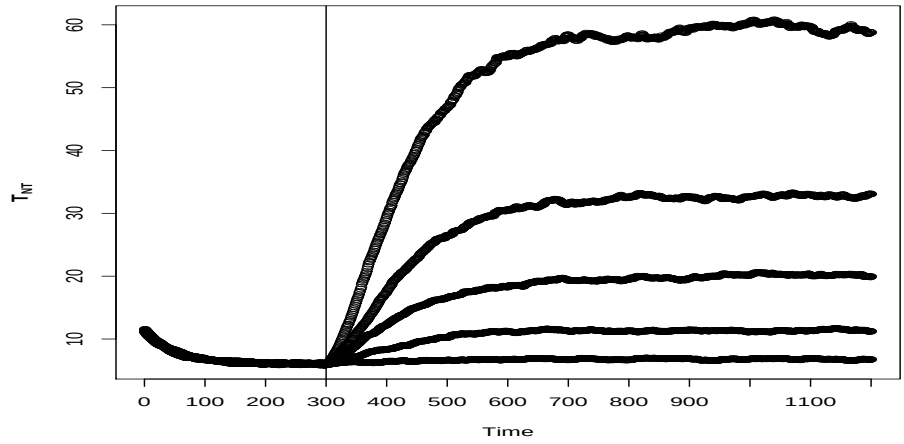

Figure 4.5: Mean path of T_{NT} for a three-dimensional observation process for a variety of shift scenarios. $\mathbf{\Sigma}_0 = \mathbf{I}$, $\lambda = 0.01$, $k = 199$, Variance of the third dimension, $\mathbf{\Sigma}_{33}$, shifts to $1 + a/\sqrt{k}$ at $t = 300$. The vertical line indicates the time when the change happens. From top to bottom is the average path of 1000 realizations for $a = 10, 7, 5, 3, 0$ respectively. $a = 0$ corresponds to the in-control case. The wiggle at the start is the burn-in phase. A larger shift corresponds to a larger slope, which is desirable for SPC.

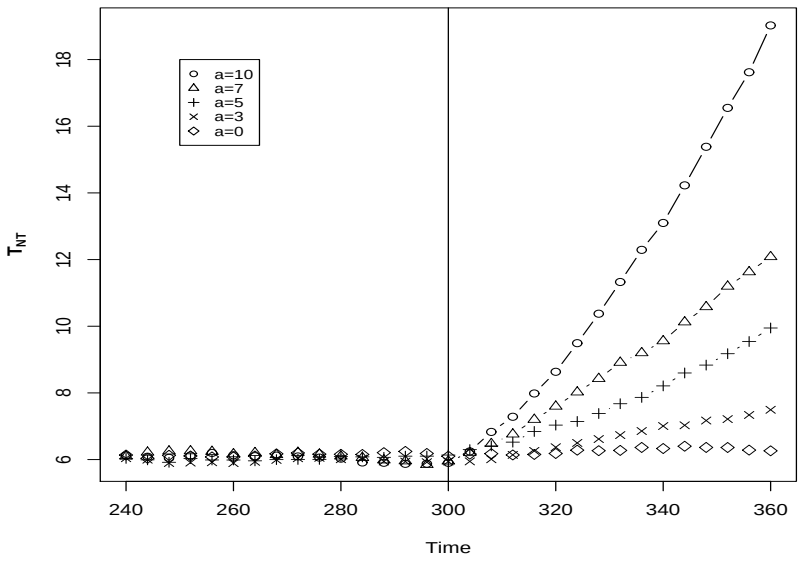

Figure 4.6: Replot of data from Fig 4.3 for $t = 240$ to $t = 360$. $\mathbf{\Sigma}_0 = \mathbf{I}$, $\lambda = 0.01$, $k = 199$, Variance of the third dimension, $\mathbf{\Sigma}_{33}$, shifts to $1 + a/\sqrt{199}$ at $t = 300$. The vertical line indicates the time when the change happens. From top to bottom is the average path of 1000 realizations for $a = 10, 7, 5, 3, 0$ respectively. $a = 0$ corresponds to the in-control case. Graph indicates that the proposed algorithm is also responsive to the change even if the shift size is small.

and let the corresponding run length be $\mathrm{L}(\mathbf{A})$. Define $\mathbf{Z}_i = \mathbf{X}_i\mathbf{X}_i^T$. Then

$$\mathbf{Z}_i \backsim \mathrm{Wishart}(1, \mathbf{\Sigma}).$$

With the first observation $\mathbf{X}_1$

$$\mathbf{S}_1 = (1 - \lambda)\mathbf{A} + \lambda\mathbf{Z}_1.$$

Conditional on the outcome of the first step, we have the following

$$\begin{aligned}
\mathrm{L}(\mathbf{A}) &= 1 \times \mathrm{P}(\mathrm{T}(\mathbf{S}_1) \geq h) + \int_{\mathrm{T}(\mathbf{S}_1)<h} (1 + \mathrm{L}((1-\lambda)\mathbf{A} + \lambda\mathbf{Z}))f(\mathbf{Z})\, d\mathbf{Z} \\
&= 1 + \int_{(k/2)tr((1-\lambda)\mathbf{A}+\lambda\mathbf{Z}-\mathbf{I})^2<h} \mathrm{L}((1-\lambda)A + \lambda\mathbf{Z}) f(\mathbf{Z})\, d\mathbf{Z} \\
&= 1 + (1/\lambda) \int_{(k/2)tr(\mathbf{B}-\mathbf{I})^2<h} \mathrm{L}(\mathbf{B}) f\Big(\frac{\mathbf{B} - (1-\lambda)A}{\lambda}\Big)\, d\mathbf{B} \qquad (4.3)
\end{aligned}$$

where f($\mathbf{Z}$) is the distribution function of Wishart$(1, \mathbf{\Sigma})$. (4.3) is also a Fredholm integral equation of the second kind with matrix inputs.

However to numerically approximate L($\mathbf{A}$) is a rather challenging task. This is because unlike in the univariate case when we monitor process mean, the numerical approximation can be done because the support of integration variable is on a subset of the real line. However, in the case of monitoring multivariate process dispersion, the input for the kernel function $f()$ should be a positive definite matrix; that is

$$\mathbf{B} > (1-\lambda)\mathbf{A}.$$

Also to numerically evaluate $f(\mathbf{A})$ by applying the Nystrom method [37] and Quadrature method [5], we need to subdivide the integration region of (4.3) into small regions of dimension p. The "Curse of Dimensionality" will cause the computational burden to grow exponentially as p increases. Hence in this dissertation we will only look at numerical simulation for a limited number of cases.

We examine the following out of control scenarios. For a p dimensional process $\mathbf{X} = [x_1, x_2, \ldots, x_p]^T$ which has in control variance covariance matrix $\mathbf{\Sigma}_0 = \mathbf{I}_p$. Shifts to the variance of the last dimension x_p is introduced at some time after the statistic

has stabilized. The new variance covariance matrix is

$$\boldsymbol{\Sigma}^{\star} = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} \\ \mathbf{0} & 1+a \end{bmatrix}$$

where a indicates the direction and size of the shift and $a = 0$ corresponds to the in control scenario. Unlike in previous chapters, the small shifts considered are not measured in units of $1/\sqrt{k}$, but rather are absolute amounts $a = 0.2, 0.4, 0.6, 0.8$. The large shifts considered are $d = 1, 2, 3, 4, 5$. Since SPC practitioners are usually more concerned with process variability increases than decreases, we put emphasis on the increase in variance cases. However as a way of illustrating the properties of proposed schemes, we also considered two cases with decreasing variance, $a = -0.2, -0.4$. For the purpose of comparison, two weighting parameters, $\lambda = 0.01, 0.03$, are considered. Also due to limitation of computer capability, we only consider multivariate processes of up to 5 dimensions, that is $p = 2, 3, 4, 5$.

In the numerical simulation, the process will run until the test statistics reach steady state before shifts take place. We will compare the MTNT scheme, MTLR scheme and the scheme based on the sample generalized variance test using unequally weighted variance covariance matrix estimators, which in this paper we call MTGV. The MTGV chart is usually called generalized variance chart elsewhere in the literatures. It calculates the ratio of the determinant of the sample variance covariance matrix against the true process generalized variance and compares the ratio with an upper control limit. This scheme is designed to monitor variability increases. Interested readers can refer to Aparisi *et al.* [4].

We used 7000 Monte Carlo trials for all in control simulations The control

limits are set such that the in control ARL is 200 for all combinations of test, p and λ. Repeated simulations suggest that the standard errors for these in control run lengths are around 1.3 times the simulated in control ARL, hence the standard errors of the in control ARLs are about 1.5% of the ARL with 95% confidence region being $(0.97 \times ARL, 1.03 \times ARL)$. Table 4.1 to 4.4 show the in control average run length of MTNT and MTLR Schemes for all combinations of $p = 2, 3, 4, 5$ and $\lambda = 0.01, 0.03$. Test statistics after the burn-in period are compared with the control limits taken from quantiles of $\chi^2_{p(p+1)/2}$ distributions. The schemes stop and give a signal when the statistic is out of the in control region. The control limits only depend on the dimension of the process and the desired in control average run length, which is an important feature of the MTNT and MTLR schemes. In general for the same quantile (different degrees of freedom), the in control ARL decreases as the dimension p increases.

We used 2000 Monte Carlo trials to compute the out of control ARLs. The simulation shows that as the shift size a increases, the standard deviation of out of control ARL decreases. At $a = 0.2$, the standard errors of these out of control run lengths are about 1.1 times the corresponding out of control ARL, hence the standard errors for their means, the out of control ARLS, are about 2.5% of the corresponding ARL value with 95% confidence region being $(0.95 \times ARL, 1.05 \times ARL)$. For larger shifts, for example when $a = 4$, the standard errors of these out of control run lengths are about 0.7 times the corresponding out of control ARLs, hence the standard errors are about 1.5% of the ARLs with 95% confidence region being $(0.97 \times ARL, 1.03 \times ARL)$. For each fixed combination of p and λ, the three

Table 4.1: In control average run length for MTNT for various cutoffs, $\lambda = 0.01$

Control Limit	p=2	p=3	p=4	p=5
$\chi^2_{p(p+1)/2,0.945}$	365.69	332.8	313.84	289.13
$\chi^2_{p(p+1)/2,0.94}$	341.02	311.56	293.66	270.34
$\chi^2_{p(p+1)/2,0.935}$	317.5	291.46	27.73	252.56
$\chi^2_{p(p+1)/2,0.93}$	297.28	274.36	260.25	239.71
$\chi^2_{p(p+1)/2,0.925}$	279.02	259.51	245.52	228.61
$\chi^2_{p(p+1)/2,0.92}$	261.2	245.07	231.56	215.99
$\chi^2_{p(p+1)/2,0.915}$	247.19	232.07	219.62	205.19
$\chi^2_{p(p+1)/2,0.91}$	234.6	219.55	209.35	195.5
$\chi^2_{p(p+1)/2,0.905}$	221.32	208.51	199.15	187.01
$\chi^2_{p(p+1)/2,0.9}$	211.89	198.71	188.94	178.16
$\chi^2_{p(p+1)/2,0.895}$	201.5	188.96	180.77	170.14
$\chi^2_{p(p+1)/2,0.89}$	192.89	180.3	173.33	163.67
$\chi^2_{p(p+1)/2,0.885}$	183.61	172.89	166.06	157.13
$\chi^2_{p(p+1)/2,0.88}$	175.93	165.81	158.85	151.86
$\chi^2_{p(p+1)/2,0.875}$	168.2	159.98	152.38	147.02
$\chi^2_{p(p+1)/2,0.87}$	161.49	153.64	146.89	141.83
$\chi^2_{p(p+1)/2,0.865}$	155.18	147.26	140.91	137.52
$\chi^2_{p(p+1)/2,0.86}$	149.15	142.02	135.96	132.74

test statistics are calculated using the same simulation data, hence there exists high amount of correlation among them.

Table 4.5-4.8 list the out of control ARLs of these three schemes for a variety of shifts when $\lambda = 0.01, 0.03$. The entries for MTGV when $a < 0$ are omitted because

Table 4.2: In control average run length for MTLR for various cutoffs. $\lambda = 0.01$

Control Limit	p=2	p=3	p=4	p=5
$\chi^2_{p(p+1)/2,0.945}$	361.45	346.03	331.88	326.07
$\chi^2_{p(p+1)/2,0.94}$	334.83	321.98	309.68	301.45
$\chi^2_{p(p+1)/2,0.935}$	310.43	299.99	288.59	283.44
$\chi^2_{p(p+1)/2,0.93}$	290.84	280.44	270.62	265.73
$\chi^2_{p(p+1)/2,0.925}$	271	264.05	254.56	249.09
$\chi^2_{p(p+1)/2,0.92}$	254.77	247.26	239.46	234.72
$\chi^2_{p(p+1)/2,0.915}$	241.5	233.42	225.51	223.48
$\chi^2_{p(p+1)/2,0.91}$	228.86	220.74	214.83	211.28
$\chi^2_{p(p+1)/2,0.905}$	218.35	209.7	204.18	198.74
$\chi^2_{p(p+1)/2,0.9}$	209.34	201.34	194.46	190.43
$\chi^2_{p(p+1)/2,0.895}$	199.25	190.31	184.84	182.08
$\chi^2_{p(p+1)/2,0.89}$	189.04	183.1	177.15	174
$\chi^2_{p(p+1)/2,0.885}$	181.31	174.80	170.63	166.61
$\chi^2_{p(p+1)/2,0.88}$	172.87	167.21	163.93	158.8
$\chi^2_{p(p+1)/2,0.875}$	164.54	161.39	156.62	151.61
$\chi^2_{p(p+1)/2,0.87}$	156.93	155.22	150.86	145.71
$\chi^2_{p(p+1)/2,0.865}$	149.79	148.57	144.14	140.34
$\chi^2_{p(p+1)/2,0.86}$	143.46	142.71	137.74	135.38

the MTGV scheme is designed to monitor variability increases and has almost no power against decreasing scenarios.

From the tables we observe that when $a > 0$, that is when the variability increases, the control scheme MTNT outperforms MTLR for all out of control sce-

Table 4.3: In control average run length for MTNT for various cutoffs, $\lambda = 0.03$

Control Limit	p=2	p=3	p=4	p=5
$\chi^2_{p(p+1)/2,0.99}$	416.01	325.38	275.33	238.64
$\chi^2_{p(p+1)/2,0.9875}$	372.38	289.82	245.96	213.51
$\chi^2_{p(p+1)/2,0.985}$	336.09	265.08	220.65	195.48
$\chi^2_{p(p+1)/2,0.9825}$	306.53	242.9	202.04	180.34
$\chi^2_{p(p+1)/2,0.98}$	283.23	226.34	188.43	168.94
$\chi^2_{p(p+1)/2,0.9775}$	262.65	212.41	176.43	158.98
$\chi^2_{p(p+1)/2,0.975}$	248.45	199.32	166.88	150.82
$\chi^2_{p(p+1)/2,0.9725}$	232.07	187.94	159.50	143.81
$\chi^2_{p(p+1)/2,0.97}$	221.75	179.15	150.04	136.06
$\chi^2_{p(p+1)/2,0.9675}$	210.13	171.13	143.65	129.02
$\chi^2_{p(p+1)/2,0.965}$	199.13	164.11	138.59	123.39
$\chi^2_{p(p+1)/2,0.9625}$	191.54	156.16	133.42	119.07
$\chi^2_{p(p+1)/2,0.96}$	183.96	149.94	127.11	114.62
$\chi^2_{p(p+1)/2,0.9575}$	176.18	143.56	122.76	110.66
$\chi^2_{p(p+1)/2,0.955}$	167.86	138.81	118.02	107.34
$\chi^2_{p(p+1)/2,0.9525}$	162.15	134.68	114.34	103.97
$\chi^2_{p(p+1)/2,0.95}$	156.28	130.54	111.21	101.51
$\chi^2_{p(p+1)/2,0.9475}$	151.39	126.56	108.4	98.69

narios, while when $d < 0$, that is when the variability decreases, the control scheme MTLR outperforms MTNT. This is true because T_{NT} and T_{LR} have the same asymptotic distribution when the process is under control. Hence for the same in control ARL, they have nearly equal control limits; while under the alternatives, the

Table 4.4: In control average run length for MTLR for various cutoffs. $\lambda = 0.03$

Control Limit	p=2	p=3	p=4	p=5
$\chi^2_{p(p+1)/2,0.99}$	589.65	559.39	534.37	513.34
$\chi^2_{p(p+1)/2,0.9875}$	509.32	485.30	461.31	436.63
$\chi^2_{p(p+1)/2,0.985}$	456.1	421.82	403.42	381.64
$\chi^2_{p(p+1)/2,0.9825}$	406.17	377.72	358.49	342.83
$\chi^2_{p(p+1)/2,0.98}$	365.13	337.24	322.18	309.08
$\chi^2_{p(p+1)/2,0.9775}$	331.78	307.88	292.28	283.23
$\chi^2_{p(p+1)/2,0.975}$	306.12	282.55	267.16	258.51
$\chi^2_{p(p+1)/2,0.9725}$	283.25	259.37	241.99	237.4
$\chi^2_{p(p+1)/2,0.97}$	262.03	239.63	225.33	219.79
$\chi^2_{p(p+1)/2,0.9675}$	242.57	222.37	209.98	204.89
$\chi^2_{p(p+1)/2,0.965}$	224.94	208.31	197.03	193.60
$\chi^2_{p(p+1)/2,0.9625}$	212.47	197.73	187.12	183.88
$\chi^2_{p(p+1)/2,0.96}$	198.37	185.09	176.24	171.73
$\chi^2_{p(p+1)/2,0.9575}$	187.63	177.06	167.72	162.28
$\chi^2_{p(p+1)/2,0.955}$	176.98	168.09	160.84	156.1
$\chi^2_{p(p+1)/2,0.9525}$	168.03	160.92	153.24	147.83
$\chi^2_{p(p+1)/2,0.95}$	160.69	152.93	146.8	141.75
$\chi^2_{p(p+1)/2,0.9475}$	153.36	147.13	141.15	135.36

asymptotic means for T_{NT} and T_{LR} are

$$\mathrm{E}(\mathrm{T}_{NT}) = k\sum_{i=1}^{p}(l_i - 1)^2/2 + p(p+1)/2$$

$$\mathrm{E}(\mathrm{T}_{LR}) = k\{\sum_{i=1}^{p}(l_i - 1) - \textstyle\prod_{i=1}^{p}\ln l_i\} + p(p+1)/2$$

Table 4.5: Out of control average run length with In control ARL=200. $\lambda = 0.01$.

Variance of the last dimension shifts from 1 to $1 + a$

Shift size	p=2			p=3		
	MTNT	MTLR	MTGV	MTNT	MTLR	MTGV
-0.4	56.13	48.92		70.14	57.77	
-0.2	119	100.85		131.71	111.1	
0	201.5	199.25	201.84	198.71	201.34	203.09
0.2	93.41	98.81	73.59	100.1	109.85	87.40
0.4	53.71	56.58	44.33	56.17	63.34	54.65
0.6	37.26	38.87	32.4	40.72	44.73	38.99
0.8	27.66	28.28	25.44	30.71	33.29	29.77
1	21.85	23.88	18.95	24.58	27.08	25.99
2	11.9	12.6	10.9	13.55	14.61	14.02
3	8.23	8.72	7.42	8.79	9.59	9.04
4	6.59	7.18	6.18	7.34	7.87	7.65
5	5.67	5.97	5.17	5.97	6.47	6.22

where $l_1, l_2, \ldots, l_p$ are the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$. Then

$$\begin{aligned} \mathrm{E}(\mathrm{T}_{NT} - \mathrm{T}_{LR}) &= k\{\sum_{i=1}^{p}(l_i - 1)^2 - 2\sum_{i=1}^{p}(l_i - 1) + 2\prod_{i=1}^{p}\ln l_i\}/2 \\ &= k\sum_{i=1}^{p}\{(l_i - 1)(l_i - 3) + 2\ln l_i\}/2 \end{aligned}$$

Table 4.6: Out of control average run length with In control ARL=200. $\lambda = 0.01$. Variance of the last dimension shifts from 1 to $1 + a$

Shift size s	p=4			p=5		
	MTNT	MTLR	MTGV	MTNT	MTLR	MTGV
-0.4	82.04	68.22		97.65	75.84	
-0.2	156.04	131.04		166.84	139.81	
0	199.15	204.18	200.59	195.5	198.74	203.8
0.2	99.92	116.2	93.14	114.25	129.18	112.8
0.4	56.44	66.46	59.38	64.76	77.50	64.18
0.6	43.21	48.4	44.33	45.34	52.44	47.98
0.8	32.81	37.77	34.79	33.67	38.55	38.83
1	26.27	30.13	28.31	29.5	34.22	33.37
2	13.59	15.04	16.29	15.18	17.15	17.68
3	9.48	10.39	11.57	10.42	11.68	12.56
4	7.78	8.38	8.73	8.17	9.04	10.18
5	6.28	6.93	7.15	7.08	7.93	8.53

Also

$$(l_i - 1)(l_i - 3) + 2\ln l_i = \begin{cases} > 0 & , l_i > 1 (a > 0) \\ = 0 & , l_i = 1 (a = 0) \\ < 0 & , 0 < l_i < 1 (-1 < a < 0) \end{cases}$$

that is to say, these eigenvalues that are greater than one make a positive contribution to $\mathrm{E}(\mathrm{T}_{NT} - \mathrm{T}_{LR})$ while these eigenvalues between zero and one make a negative contribution to $\mathrm{E}(\mathrm{T}_{NT} - \mathrm{T}_{LR})$. If the overall positive contribution outweighs the

Table 4.7: Out of control average run length with In control ARL=200. $\lambda = 0.03$. Variance of the last dimension shifts from 1 to $1 + a$

Shift size	p=2			p=3		
	MTNT	MTLR	MTGV	MTNT	MTLR	MTGV
-0.4	82.3	48.57		127.83	60.75	
-0.2	186.33	114.45		224.66	134.89	
0	199.13	198.37	202.39	199.32	197.73	198.65
0.2	85.72	104.51	84.1	97.5	115.15	93.82
0.4	46.39	56.21	49.2	52.77	62.57	60.46
0.6	31.9	36.52	34.99	35.32	41.98	42.17
0.8	23.63	27.58	26.02	26.72	30.54	32.01
1	18.04	19.64	21.06	20.03	22.56	27.22
2	9.22	10.58	9.89	10.71	12.08	13.51
3	7.02	7.63	7.56	7.47	8.28	9.47
4	5.34	5.84	5.98	6.28	6.95	7.39
5	4.42	4.89	4.8	4.76	5.32	5.93

negative contribution then T_{NT} will stabilize at a higher level than T_{LR} after the shift. This combined with the fact that it takes almost equal number of observations for both schemes to stabilize, result in a greater slope for T_{NT} than T_{LR}; and vice versa. This explains the phenomena we observe in Table 4.5-8 since in the simulation $l_1 = l_2 = 1$ and $l_3 = 1 + a$. Overall the ARL improvement of MTNT over MTLR when $a > 0$ is between 5% and 15%, while when $a < 0$, MTLR outperforms MTNT by 10% to 20%. Figure 4.7 compares the two power functions of T_{LR} and

Table 4.8: Out of control average run length with In control ARL=200. $\lambda = 0.03$. Variance of the last dimension shifts from 1 to $1 + a$

Shift size	p=4			p=5		
	MTNT	MTLR	MTGV	MTNT	MTLR	MTGV
-0.4	165.75	74.51		172.48	94.29	
-0.2	233.35	147.94		249.41	163.77	
0	202.04	197.03	199.47	195.48	204.89	199.72
0.2	114.07	128.48	107.48	110.5	132.94	109.75
0.4	61.87	72.4	67.95	66.61	80.85	71.04
0.6	39.48	46.95	47.33	43.86	53.81	54.81
0.8	29.21	33.66	38.43	33.02	39.10	43.91
1	23.29	26.71	32.81	25.94	31.32	37.61
2	11.58	13.26	15.86	12.53	14.57	19.56
3	8.15	8.97	11.52	8.74	10.2	13.29
4	6.76	7.39	9.67	6.79	7.82	10.21
5	5.53	6.08	7.75	6.09	6.48	8.27

T_{NT}.

Next we compare MTGV with the other two schemes. In general, when shift size is small, for example $d = 0.2, 0.4$, it appears that MTGV has the shortest ARL among these three scheme. When $p = 3, 4, 5$, in general, the MTNT scheme has the shortest ARLs among these three with some exceptions happen when the shift size a is really small. The relation between ARL_{MTLR} and ARL_{MTGV} depends on a and p, for example when $p = 4$ and $a = 3$, $ARL_{MTNT} = 9.48 < ARL_{MTLR} =$

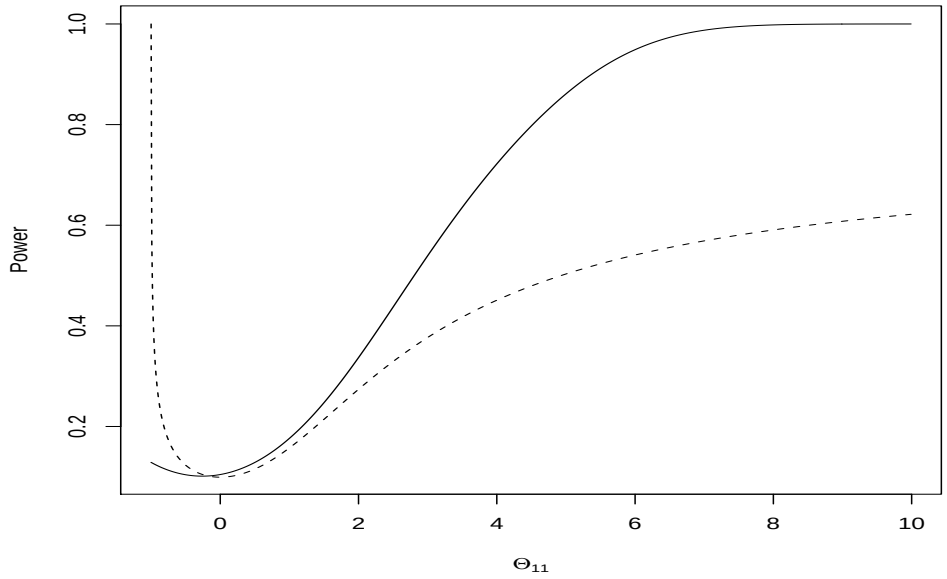

Figure 4.7: Power function comparison between T_{NT} and T_{LR}. Dotted line: T_{LR}; Solid Line: T_{NT}

$10.39 < ARL_{MTGV} = 11.57$, while when $p = 3$ and $a = 0.4$, $ARL_{MTNT} = 55.89 < ARL_{MTLR} = 57.15 < ARL_{MTGV} = 60.12$. In the variability increase scenarios, the MTLR scheme appears to be always worse than the better of the other two.

For the comparison between MTNT and MTGV, it appears that for each fixed combination of λ and p, there exists an equilibrium point (shift size a_0) where if $d < d_0$ then MTGV has shorter out of control ARL and when $a > a_0$ MTNT has the shorter ARL. For example, when $p = 2$, $\lambda = 0.01$, for all the 9 out of control cases considered, MTNT has longer ARL than MTLR, which means the equilibrium point is beyond 5, $a_0 > 5$. When $p = 3$, $\lambda = 0.01$, at $a = 0.2, 0.4, 0.6, 0.8$ MTNT has longer ARL than MTGV, while when $a = 1, 2, 3, 4, 5$, MTNT has shorter ARL than MTGV, which suggests that the equilibrium point is between 0.8 and 1. While for all other p and λ combinations, the equilibrium point seems to be between 0.2 and

0.4.

It is widely perceived that small λ values should be used in EWMA to monitor small shifts and large λ's for large shifts. By cross comparing Table 4.5, 4.6 and Table 4.7, 4.8, we see that this is largely true in the three schemes considered here. For the same shift, dimension, use of a smaller λ grants a shorter ARL when a is small and a larger λ gets a longer ARL, for example, when $a = 5$,$p = 2$, $\lambda = 0.03$, the ARLs are $ARL_{MTNT} = 4.42, ARL_{MTLR} = 4.89, ARL_{MTGV} = 4.8$, they are all smaller than the corresponding ARL when $\lambda = 0.01$, $ARL_{MTNT} = 5.67, ARL_{MTLR} = 5.97, ARL_{MTGV} = 5.17$.

The above simulations focused on the change of one dimension (eigenvalue) only, therefore they do not project a complete picture about the performance comparison among the three control schemes. In fact both MTNT and MTLR are sensitive to more out of control scenarios than MTGV. All the three schemes considered here are based on the sample eigenvalues of $\mathbf{S}\boldsymbol{\Sigma}_0^{-1}$. As pointed out earlier, the impacts of these p eigenvalues on MTNT and MTLR are additive, while this is not true for MTGV since MTGV is based on the product of sample eigenvalues, hence an increase of one eigenvalue can be negated by the decrease of another which results in the overall product staying the same. SPC practitioners should take precautions about this.

Chapter 5

Other Considerations and Conclusion

5.0.1 Review

In this dissertation, we consider the problem of monitoring multivariate process variability. We derive the asymptotic distributions of the log likelihood ratio test and Nagao's test under both local and distant alternatives, and investigate the properties of unequally weighted sum of squares type estimators for the process variance covariance matrix. Satterthwaite's approximation is used to estimate the degrees of freedom of the proposed weighted sum of squares type estimator for the underlying process variance-covariance matrix. By combining the log likelihood ratio test and Nagao's test with the weighted sum of squares estimator, we come up with two new control schemes, MTLR and MTNT.

The performance of the proposed new schemes are compared with the existing methods using Monte Carlo method for various scenarios. The new scheme MTNT is easy to implement and the numerical simulations presented have shown that MTNT outperforms the schemes based on the likelihood ratio and sample generalized variance under medium to large variance increase situations. However this new scheme has its own limitations too. It has longer out of control ARLs than the MTLR scheme when the process variability decreases. Also this test is based upon the discrepancies of the characteristic roots of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$ from 1, but $\mathbf{\Sigma}$ has $p(p+1)/2$

parameters while $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$ only has p eigenvalues, so it is always possible that the process variability shifts in such a way that the eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}$ remain 1 while the internal structure of $\mathbf{\Sigma}$ differs from that of $\mathbf{\Sigma}_0$.

5.0.2 Issues to be answered

In the future, we plan to answer the following challenges..

The first issue is the validity and accuracy of Satterthwaite's approximation. Satterthwaite's approximation of the distribution of the EWSS estimator is a method of moments estimator. It is based upon the assumption that a weighted moving average of independent Wishart distributed matrices is also Wishart. The validity of this assumption needs to be checked and the accuracy of the approximation needs to be further investigated. We may examine this from the perspective of matching characteristic functions.

Let the batch size m be 1 and moving window size be n. The characteristic function of the weighted moving average of independent Wishart distributed matrices

$$\mathbf{S} = \sum_{j=1}^{n} w_j X_j X_j^T$$

is

$$\begin{aligned}\text{Char}(\mathbf{S}) &= \prod_{j=1}^{n} \left|\mathbf{I} - 2i\mathbf{\Theta} w_j \mathbf{\Sigma}\right|^{-1/2} \\ &= \Bigg|\mathbf{I} - 2i\mathbf{\Theta}\mathbf{\Sigma} + (-2i\mathbf{\Theta}\mathbf{\Sigma})^2 \sum_{j=1}^{n}\sum_{l>j} w_j w_l + (-2i\mathbf{\Theta}\mathbf{\Sigma})^3 \sum_{j=1}^{n}\sum_{l>j}\sum_{o>l} w_j w_l w_o \\ &\quad + \cdots + (-2i\mathbf{\Theta}\mathbf{\Sigma})^n \prod_{j=1}^{n} w_j \Bigg|^{-1/2}.\end{aligned}$$

The characteristic function of a Wishart distributed matrix $\mathbf{W}$ with k degrees of freedom and scale matrix $\mathbf{\Sigma}/k$ is

$$\begin{aligned}\text{Char}(\mathbf{W}) &= \left|\mathbf{I} - 2i\mathbf{\Theta\Sigma}\right|^{-k/2} \\ &= \Big|\mathbf{I} - 2i\mathbf{\Theta\Sigma} + (-2i\mathbf{\Theta\Sigma})^2(k-1)/2k + (-2i\mathbf{\Theta\Sigma})^3(k-1)(k-2)/(6k^2) \\ &\quad + \cdots + (-2i\mathbf{\Theta\Sigma})^k/(k^k)\Big|^{-1/2}\end{aligned}$$

The objective is to find a measure of the difference between two characteristic functions of $\mathbf{S}$ and $\mathbf{W}$ with

$$k = 1/\sum_{j=1}^{n} w_j^2$$

as the optimization solution. The natural restriction on the optimal value k should always be that it is no more than the window size n. If we attempt to match the orders of $\mathbf{\Theta\Sigma}$, k should be such that the higher order terms (higher than k) will be neglected, while this neglect can be compensated by the lower or equal order terms.

Second, in the above algorithm, we assume that the mean of the process is $\mathbf{0}$ and remains so throughout the whole period. However if this assumption fails and the process mean does shift to another level, without or without shift in the process dispersion, we may see a signal. This is true because if at time t, the mean shifts to μ instead of the assumed $\mathbf{0}$, then

$$\begin{aligned}\text{E}(\mathbf{X}_t\mathbf{X}_t^T) &= \text{Var}(\mathbf{X}_t) + \text{E}(\mathbf{X}_t)\text{E}(\mathbf{X}_t^T) \\ &= \mathbf{\Sigma} + \mu\mu^T.\end{aligned}$$

hence a shift in the process mean from $\mathbf{0}$ to μ resembles a shift in the process variance-covariance matrix from $\mathbf{\Sigma}$ to $\mathbf{\Sigma} + \mu\mu^T$. At time t, the mean of the EWSS

estimator of the process variance covariance matrix is

$$\begin{aligned} \mathrm{E}(\mathbf{S}_t) &= \mathrm{E}\{(1-\lambda)\mathbf{S}_{t-1}\} + \lambda \mathrm{E}\{\mathbf{X}_t \mathbf{X}_t^T\} \\ &= (1-\lambda)\boldsymbol{\Sigma} + \lambda(\boldsymbol{\Sigma} + \mu\mu^T) \\ &= \boldsymbol{\Sigma} + \lambda\mu\mu^T, \end{aligned}$$

so a change in the process mean is transformed into the shift in the process dispersion from $\boldsymbol{\Sigma}$. This is one side effect of this control scheme.

In order to identify the correct source of variation, we may run two control schemes simultaneously, one for the process mean and one for the process variability. The results of these two schemes can be compared. If both schemes issue signals, then the source of variation could be only the process mean vector or both mean vector and variance-covariance matrix; if only the proposed scheme for monitoring process variability issues a signal, then the source of variation is the process variance-covariance matrix. The change in the mean vector will cause signals with or without change in the variance-covariance matrix.

The third issue is the assumption that the observations are independent. However, this independence assumption is often violated in many applications, and the existence of autocorrelation has large impact on the performance of control charts. The autocorrelation among the observations could happen if we take samples within a very short time interval, or the underlying process itself is a time-dependent time series. The classical time series models for autocorrelated data include $AR(p)$, $ARMA(p,q)$, etc. The first step may be to investigate the performance of the proposed schemes when we have a simple autoregressive process $AR(p)$. The robustness

against the autocorrelation is the main concern here.

Finally, an issue that has been persistent in monitoring multivariate process is that even though the scheme can correctly issue a signal for an overall change, it can not help identify the original source of the variation. Which dimension or which subset of dimensions has changed? The proposed schemes can not help answer this question. We may investigate the individual dimensions separately after the scheme(s) have issued signals.

Bibliography

[1] Anderson, T. W. (1958),"An Introduction to Multivariate Statistical Analysis," *Wiley, New York.*

[2] Alt, F. B. (1985),"Multivariate Quality Control," *Encyclopedia of Statistical Sciences, 6. John Wiley, New York.*

[3] Alt, F. B., and Smith, N. D. (1988), "Multivariate Process Control," *Quality Control and Reliability, 7. Handbooks of Statistics, Amsterdam, North-Holland.*

[4] Aparisi, F., Jabaloyes, J. and Carrion, A. (2001), "Generalized Variance Chart Design with Adaptive Sample Sizes. The Bivariate Case," *Communications in Statistics, Simulation and Computation, 30, 4, pp. 931-948.*

[5] Atkinson K. (1989),"An Introduction to Numerical Analysis," *John Wiley. 2nd ed.*

[6] Brook, D and Evans, D. A. (1972), "An Approach to the Probability Distribution of CUSUM run Length," *Biometrika, 59, pp. 539-548.*

[7] Calzada, M. E., and Scariano, S. M. (2003), "Reconciling the Integral Equation and Markov Chain Approaches for Computing EWMA Average Run Lengths," *Communications in Statistics, Simulation and Computation, 32, No.2, pp. 591-604.*

[8] Champ C. W. and Rigdon S. E. 1991, "A Comparison of the Markov Chain and the Integral Equation Approaches for Evaluating the Run Length Distribution of Quality Control Charts," *Communications in Statistics, Ser.B, Simulation and Computation, 20, No. 1, pp. 191-204..*

[9] Chang, S. I. and Zhang, K. (2007), "Statistical Process Control for Variance Shift Detections of Multivariate Autocorrelated Processes," *Quality Technology and Quantitative Management, 4, No.3, pp. 413-435.*

[10] Crowder, S. C. (1987), "A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts," *Technometrics, 29, No.4, pp. 401-407.*

[11] Dick, J., Kritzer, P., Kuo, F. Y. and Sloan, I. H. (2006), "Lattice-Nystorm method for Fredholm Integral Equations of the Second Kind," *Journal of Complexity, 23, pp. 752-772.*

[12] Fan, F. F. (1995), "Jointly Monitoring Process Mean and Variability Using Exponentially Weighted Moving Average Control Charts," *Technometrics, 37(4), pp. 446-453.*

[13] Gupta, S. D. (1969), "Properties of Power Functions of Some Tests Concerning Dispersion Matrices of Multivariate Normal Distributions," *Annals of Mathematical Statistics, 40, No. 2, pp. 697-701.*

[14] Hawkins, D. M. (1991), "Multivariate quality control based on regression-adjusted variables," *Technometrics, 33, No.4, pp. 61-75.*

[15] Chen, F. L. and Huang, H. J. (2005), "A Synthetic control chart for monitoring process dispersion with sample standard deviation," *Computers and Industrial Engineering, 49, pp. 221-240.*

[16] Hull, J. C. (2006), "Options, Futures and Other Derivatives," *6th Edition. Prentice Hall.*

[17] Hunter, J. S. (1986), "The Exponentially Weighted Moving Average," *Journal of Quality Technology, 19, pp, 239-250.*

[18] Jiang, W. (2001), "Average Run Length Computation of ARMA Charts for Stationary Processes," *Communications in Statistics, Ser.B, Simulation and Computation, 30, No. 3, pp. 699-716.*

[19] Kemp, K. W. (1971), "Formal Expressions Which Can Be Applied to CUSUM Charts," *Journal of the Royal Statistical Society, Ser. B. 33, pp. 331-360* .

[20] Knoth, S. (2004), "The Art of Evaluating Monitoring Schemes-How to Measure the Performance of Control Charts?," *Preceedings of the VIIIth International Workshop on Intelligent Statistical Quality Control, pp. 117-143.*

[21] Knoth, S. (2005), "Accurate ARL Computation for $EWMA - S^2$ Control Charts," *Statistics and Computing, 15, pp. 341-352.*

[22] Khuri, A.I., Mathew, T., and Nel, D. G. (1994), "A Test to Determine Closeness of Multivariate Satterthwaite's Approximations," *Journal of Multivariate Analysis, 51, pp. 201-209.*

[23] Khuri, A. I. (1995), "A Measure to Evaluate the Closeness of Satterthwaite's Approximation," *Biometrical Journal, 37, pp. 547-563.*

[24] Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992), "A Multivariate Exponentially Weighted Moving Average Chart," *Technometrics, 34, 1, pp. 46-53.*

[25] Lucas, J. M. and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements" *Technometrics, 32, pp. 1-12.*

[26] Macgregor, J. F. and Harris, T. J. (1993), "The Exponentially Weighted Moving Variance," *Journal of Quality Technology, 25, pp. 106-118.*

[27] Masaro, J., and Wong, C. S. (2003), "Wishart Distributions Associated with Matrix Quadratic Forms," *Journal of Multivariate Analysis, 85, pp. 1-9.*

[28] Mathew, T., and Nordstrom, K. (1997), "Wishart and Chi-square Distributions Associated with Matrix Quadratic Forms," *Journal of Multivariate Analysis, 61, No.1, pp. 129-143.*

[29] Mittag, H. J., Stemann, D. and Tewes, B. (1998), "EWMA-Karten zur Uberwachung der Streuung von Qualitatsmerkmalen," *Allemeines Statistische Archiv, 82, pp. 327-338.*

[30] Montgomery, D. C. and Mastrangelo, C. M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology, 23, pp. 179-193.*

[31] Mortarino, C. (2005), "A Decomposition for A Stochastic Matrix with An Application to MANOVA," *Journal of Multivariate Analysis, 92, pp. 134-144.*

[32] Nagao, H. (1967), "Monotonicity of the Modified Likelihood Ratio Test for a Covariance Matrix," Journal of Science of Hiroshima University, Ser. A-I Math, 31, pp 147-150.

[33] Nagao, H. (1970), "Asymptotic Expansions of Some Test Criteria for Homogeneity of Variances and Covariance Matrices from Normal Populations," *Journal of Science of Hiroshima University, Ser. A-I Math, 34, pp. 153-247.*

[34] Nagao, H. (1973), "On Some Test Criteria For Covariance Matrix," *Annals of Statistics, Vol.1, No.4, pp. 700-709.*

[35] Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika, 41, pp. 100-114.*

[36] Rao, B. V., Disney, R. L., and Pignatiello, J. J. (2001), "Uniqueness and Convergence of Solutions to Average Run Length Integral Equations for Cumulative Sum and Other Control Charts," *IIE Transactions, 33, pp. 463-469.*

[37] Reinhardt H. J. (1985), "Analysis of Approximation Methods for Differential and Integral Equations," *Springer-Verlag New York, LLC.*

[38] Reynolds, M. R. and Cho, G. Y. (2006), "Multivariate Control Chart for Monitoring the Mean Vector and Covariance Matrix," *Journal of Quality Technology, 38, pp. 230-253.*

[39] Roberts, S. W. (1959), "Control Chart Tests Based on Geometric Moving averages," *Technometrics, 1, pp. 239-250.*

[40] Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance components," *Biometrics Bulletin, 2, pp. 110-114.*

[41] Robinson, P. B., and Ho, T. Y. (1978), "Average Run Lengths of Geometric Moving Average Charts by Numerical Methods," *Technometrics, 20, No.1, pp. 85-93.*

[42] Runger, G. C., and Prabhu, S. S. (1996), "A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart," *Journal of the American Statistical Association, 91, pp. 1701-1706.*

[43] Shewhart, W. A. (1931), "Economic Control of Quality of Manufactured Product," *Princeton, NJ: Van Nostrand Reinhold Co., 1931. (Republished in 1981 by the American Society for Quality Control, Milwaukee, WI).*

[44] Srebro, N. (2004), "Learning with Matrix Factorizations," *Phd Dissertation, MIT.*

[45] Stoer, J., and Bulirsch, R. (1992), "Introduction to Numerical Analysis," Springer, N.Y.

[46] Sugiura, N., and Nagao, H. (1968), "Unbiasedness of Some Test Criteria for the Equality of One or Two Covariance Matrices," *Annals of Mathematical Statistics, 39, pp. 1686-1692.*

[47] Sugiura, N. (1969), "Asymptotic Expansions of The Distributions of The Likelihood Ratio Criteria For Covariance Matrix," *Annals of Mathematical Statistics, Vol. 40, No.6, pp. 2051-2063*

[48] Sugiura, N., and Fujikoshi, Y. (1969), "Asymptotic Expansions of the Non-null Distribution of the Likelihood Ratio Criteria for Multivariate Linear Hypothesis and Independence," *Annals of Mathematical Statistics, 40, pp. 942-952* .

[49] Sugiura, N. (1969), "Asymptotic Expansions of the Distributions of the Likelihood Ratio Criteria for Covariance Matrix," *Annals of Mathematical Statistics, 40, pp. 942-952* .

[50] Sugiura, N. (1973), "Asymptotic Non-null Distributions Of The Likelihood Ratio Criteria For Covariance Matrix Under Local Alternatives," *Annals of Statistics, Vol.1, No.4, pp. 718-728*

[51] Surtihadi, J., Raghavachari, M. and Runger, G. (2004), "Multivariate Control Charts for Process Dispersion," *International Journal of Production Researcj, 42, No. 15, pp. 2993C3009.*

[52] Tan W. Y., and Gupta, R. P. (1983), "On Approximating a Linear Combination of Central Wishart Matrices with Positive Coefficientss," *Communications in Statistics, Theory and Methods, 12, No. 22, pp. 2589-2600.*

[53] Vanbrackle, L. and Williamson, G. D. (1999), "A Study of the Average Run Length Characteristics of the National Notifiable Diseases Surveillence System," *Statistics In Medicine, 19, pp. 3309-3319.*

[54] Wilks S. S. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis," *Annals of Mathematical Statistics, 9, pp. 60-62.*

[55] Wong, C. S., Masaro, J., and Wang, T. (1991), "Multivariate Versions Of Cochran's Theorems," *Journal of Multivariate Analysis, 39, pp. 157-174.*

[56] Wong, C. S., Masaro, J., and Wang, T. (1993), "Multivariate Versions Of Cochran's Theorems II," *Journal of Multivariate Analysis, 44, pp. 146-159.*

[57] Yeh, A. B., Lin, D. K. J., Zhou, H. H. and Venkataramani, C. (2003), "A Multivariate Exponentially Weighted Moving Average Control Chart for Monitoring Process Variability," *Journal of Applied Statistics, 30, No. 5, pp. 507C536.*

[58] Yeh, A. B., L.H and Wu, C. W. (2005), "A Multivariate EWMA Control Chart for Monitoring Process Variability with Individual Observations," *IIE Transactions, 37, pp. 1023-1035.*