

Control Charts for Serially Dependent Multivariate Data

Resul Ödük

Thesis submitted to the Department of Informatics and Mathematical Modelling at
Technical University of Denmark in partial fulfillment of the requirements for the degree of
Master of Science in Mathematical Modelling and Computation

TECHNICAL UNIVERSITY OF DENMARK

2012

ABSTRACT

In the literature, traditional univariate and multivariate control charts have been designed to monitor uncorrelated variables. However, in real life the data collected in time often show serial dependency. Since this serial dependency affects the false alarm rate and the shift detection capability, traditional control charts are effected. In this research we use the X-chart for univariate case and Hotelling T-square control chart for the multivariate case. The first objective is to measure the shift detection performance of proposed methods in the combination of different autocorrelation levels and various magnitudes of shifts in the process mean. For the univariate case proposed methods are to use X-chart based on raw data and based on residuals. For the multivariate case, using the Hotelling T-square control chart based on raw data, residuals and reconstructed data with lagged variables are the proposed methods. Raw data is generated based on the univariate first order autoregressive, AR(1), and bivariate first order vector autoregressive, VAR(1), structure. The residuals are considered as an output of perfectly modelled raw data. Reconstructed data is considered as expanded data with two lagged variables. The second objective is to take autocorrelation into account by adjusting the control limits to in control ARL using the Hotelling T-square control chart based on proposed methods for the multivariate case in the combination of different autocorrelation levels and various magnitudes of shifts in the process mean. Finally, the shift detection performances of the proposed methods are compared by using average run length as performance measure.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Murat Kulahci, who provided me various valuable suggestions and comments during my work on this thesis. This thesis is completed through his continuous support, guidance and encouragement on overcoming problems, making corrections for all the time of research and on writing of this thesis. Without his help and patience, this work would not have been possible.

CONTENTS

ABSTRACT	1
ACKNOWLEDGEMENTS	1
LIST OF TABLES	1
Chapter 1	1
Introduction	1
Chapter 2	4
Literature Review	4
Chapter 3	7
Monitoring Univariate Time Series	7
3.1 AR (p) Models	9
3.2 Determination of the number of observations in Phase I	12
3.3 Residuals of AR (1) Models	18
Chapter 4	26
Monitoring Multivariate Time Series	26
4.1 Multivariate Normal Distribution	27
4.2 Hotelling T-square Control charts	29
4.3 Determination of UCL for different number of observations in Phase I	32
4.4 Hotelling T-square Control Charts for Multivariate Autocorrelated Data	34
4.5 VAR(p) Models	35
Chapter 5	49
Hotelling T-square Statistics on Data Matrix with Lagged Variables	49
Chapter 6	55
Comparison of Proposed Methods	55
Chapter 7	62
Conclusion	62
REFERENCES	64
APPENDIX	66
Appendix A	66
Appendix B	72
Appendix C	74
Appendix D	76
D.1 Simulations for Chapter 3	76
D.2 Simulations for Chapter 4	80
D.3 Simulations for Chapter 5	85
D.4 Simulations for Chapter 6	87

LIST OF TABLES

Table 1 Control limits with known parameters for AR (1) process	15
Table 2 ARLs obtained by using X-chart based on the raw data in the combination of different autocorrelation levels and different number of observations in Phase I for AR (1) process	16
Table 3 ARLs obtained by using X-chart based on the	20
Table 4 ARLs obtained by using X-chart based on the residuals with	21
Table 5 The detection capability of first and subsequent residual	22
Table 6 ARLs obtained by using X-chart based on raw data and residual from AR(1) process.....	24
Table 7 ARLs obtained by using Hotelling T-square control chart based on independent data with theoretical UCL and simulation based UCL	33
Table 8 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase I for different autocorrelation levels and various magnitudes of shifts.....	39
Table 9 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase I for different autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$	40
Table 10 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different positive autocorrelation levels and various magnitudes of shifts	41
Table 11 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different positive autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$	45
Table 12 Comparison of the ARL obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase I with different autocorrelation levels.....	50
Table 13 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different positive autocorrelation levels and various magnitudes of shifts.....	51
Table 14 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different positive autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$	53
Table 15a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.1)	56
Table 15b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on propose methods with Φ matrix in (6.1)	56
Table 16a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.1) with $cov(e_{x_1}, e_{x_2}) = 0.9$	57
Table 16b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.1) with $cov(e_{x_1}, e_{x_2}) = 0.9$	57
Table 17a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.2)	57
Table 17b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.2).....	58
Table 18a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.2) with $cov(e_{x_1}, e_{x_2}) = 0.9$	58
Table 18b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.2) with $cov(e_{x_1}, e_{x_2}) = 0.9$	59

Table 19a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.3)	59
Table 19b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.3)	60
Table 20a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.3) with $cov(e_{x_1}, e_{x_2}) = 0.9$	60
Table 20b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.3) with $cov(e_{x_1}, e_{x_2}) = 0.9$	61
Table A.1 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different negative autocorrelation levels and various magnitudes of shifts	57
Table A.2 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different negative autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$	57
Table B.1 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different negative autocorrelation levels and various magnitudes of shifts	70
Table B.2 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different negative autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$	71
Table C.1 Comparison of the ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.1)	72
Table C.2 Comparison of the ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.2)	73
Table C.3 Comparison of the ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.3)	73

Chapter 1

Introduction

Statistical process control (SPC) is a powerful method to increase the product quality and lower the production costs by controlling, monitoring and improving the process. It was originally introduced by Walter Shewhart in the early 1930's. Shewhart called a process that operates under the common causes variation as being in statistical control while the process with assignable causes indicates out of control. Common causes are usual or predictable whereas assignable causes are unusual or unpredictable variations in the system. The process with common causes could be described by a probability distribution. In SPC, it is often assumed that the quality characteristic is normally distributed. The parameters of this normal distribution are used to determine the control limits. For normally distributed statistics, control limits often cover the 99.73% of all statistics, which indicates control limits are at ± 3 standard deviation distance from the mean. In control charts, if the plotted point falls within these control limits, the process is considered as in control process, but if plotted point falls either above or below the control limits, the control chart signals or an alarm is declared.

In many statistical control applications the process would have more than one quality characteristics. Control charts for single variables, i.e. univariate control charts, can only monitor one quality characteristic, which means that the engineer should look at each quality characteristic separately. But by doing this, any correlation among the quality characteristics would be ignored. For that, traditional multivariate statistical process control (MSPS) charts such as Hotelling T-square (multivariate Shewhart), multivariate exponentially weighted moving average (MEWMA), multivariate cumulative sum (MCUSUM) control charts are used. Applications with multivariate statistical techniques contain the correlation information among the quality characteristics. So, considering the multivariate methods in the case of more than one quality characteristics would be better in comparison to univariate methods.

The general assumption for multivariate control procedure is that the observations are uncorrelated or statistically independent over time. In real life, however, the data collected in time often show serial dependency. Many manufacturing and chemical processes yield multivariate data that have correlation between the successive observations and also cross correlation between the quality characteristics. It is expected that the autocorrelation affects the false alarm rate and the shift detection power. Therefore, when the assumption of independence is often violated, the control charts developed under the assumption of independence would be effected by this violation. In this study we investigate the impact of autocorrelation on the performance of univariate and multivariate control charts. We use X-

chart for the univariate case and Hotelling T-square control chart for the multivariate case, which is one of the widely used techniques in multivariate statistical process control.

In the literature there are two general approaches to deal with autocorrelation in the process. For the first method, when the univariate control charts are being used for autocorrelated data, it is suggested to fit univariate time series models such as ARMA to the data and monitor the residuals. For multivariate autocorrelated data, multivariate time series models such as VARMA are used. For the second method, traditional control charts with modified control limits are used to monitor the autocorrelated data to account for autocorrelation. However a problem with multivariate time series model is the number of variables. When the number of variables is large, the model estimation would be difficult. The number of parameters would be estimated increases with the large number of variables. Therefore, the estimation of parameters with large number of variables would be almost impossible even with modern day's computer. Alternatively, we also propose to fit univariate model to individual observations of multivariate data and consider the residuals by using Hotelling T-square control charts. But this would ignore the cross correlation among the variables. Another method we consider for multivariate data is to expand the data by adding lagged variables, and use Hotelling T-square control charts based on the expanded data. Mason and Young (2002) suggest to add lagged variables to dataset and to monitor the process with Hotelling T-square control chart. The problem with that method in the case of large number of variables is how many lagged variables should be added to data matrix.

Although the residuals from a time series model are uncorrelated, they may not be good enough to detect the process mean shift. Harris and Ross (1990), Longnecker and Ryan (1990) and Zhang (1997) recognized that the control charts based on residuals from a first-order autoregressive, AR (1), process may have poor detection power to detect the shift in the process mean. Here we also examine these suggestions for univariate case by using X-chart and extend it to multivariate cases by using Hotelling T-square control charts.

In this study, for the univariate case, we use univariate control charts based on raw data and the residuals of a univariate time series model, and for the multivariate case, we use multivariate control charts based on raw data, based on the residuals of a univariate and a multivariate time series model and expanded data matrix with lagged variables. For simplicity, proposed methods are based on Hotelling T-square control charts on raw data which has bivariate VAR (1) structure, residuals from bivariate VAR (1) and AR (1) model, and expanded data matrix with two lagged variables. The performance comparison of these proposed methods are made based on the combination of different autocorrelation levels and the magnitudes of the shifts in the process mean by calculating the average run lengths. Run length is the time that a process- monitoring scheme first signals. Average run length (ARL) is the average of the run lengths, or the average run length (ARL) is the average number of points that must be plotted before a point indicates an out of control condition (Montgomery, 2009, p. 191), and in the literature it is used to evaluate the performance of the control charts . The fact that run length for good process has exponential distribution. In this study we

calculate the average run length either based on the exponential distribution of run lengths for good process or by simply taking the average of the run lengths.

In chapter 2, literature review is examined on the existing statistical control applications based on autocorrelated data, in chapter 3 we try to compare the shift detection capability of X-chart based on raw data which have first order autoregressive structure and the residuals from AR (1) model in the combination of different autocorrelation levels and the magnitudes of shifts in the process mean. In chapter 4, the performance of Hotelling T-square control chart based on the data which have first order bivariate vector autoregressive structure and the residuals of bivariate VAR(1) model are considered. Shift detection capabilities of these two methods are compared using different autocorrelation levels and the magnitudes of shifts in the process mean. In chapter 5, the performance of Hotelling T-square control chart based on the multivariate autocorrelated data reconstructed with lagged variables is considered. In chapter 6, the shift detection performance of the proposed methods in chapters 4 and 5 with one another method which is to fit AR model to the individuals in the multivariate autocorrelated data matrix is compared by adjusting the control limits in the combination of various magnitudes of shifts with the autocorrelation matrix corresponds to low, moderate and high autocorrelation levels. Finally, in chapter 7, conclusions and future studies are discussed for the proposed methods.

Chapter 2

Literature Review

The main assumption of many traditional univariate process control techniques is that the observations are independent over time. If the variables in the process exhibit correlation over time, this assumption may be violated since the autocorrelation may effect the false alarm rate and the shift detection power. Hence, traditional control charts would be effected by this violation. This problem has been studied by many authors, Vasilopoulos and Stamboulis (1978), Alwan and Roberts (1988), Harris and Ross (1991), Montgomery and Mastrangelo (1991), Maragah and Woodall(1992), Wardell, Moskowitz and Plante (1994), Superville and Adams (1994), Lu and Reynolds (1995), Schmid (1995,1997a,1997b).

In the literature, in order to deal with this problem two general monitoring approaches are recommended. First method is to fit time series model to the data, and then apply traditional control charts such as Shewhart, EWMA (exponentially-weighted moving average) and CUSUM (cumulative sum control) charts to the residuals from the time series model. Second method is to use traditional control charts to monitor autocorrelated observations with modified control limits to account for autocorrelation.

Alwan and Roberts (1988) show that if the correct time series model is known, using residuals from the time series model (ARIMA) may be appropriate to construct the control charts since the residuals of time series model of autocorrelated process are independent and identically distributed with mean 0 and variance σ^2 . Harris and Ross (1991) fit a time series model to the univariate observations, and then investigate the autocorrelation effect on the performance of CUSUM and EWMA chart by using residuals. Montgomery and Mastrangelo (1991) show that the EWMA (exponentially weighted moving average) control charts may be useful for autocorrelated data by applying control charts to the residuals of time series model. Wardell, Moskowitz and Plante (1994) show the ability of EWMA charts to detect the shift more quickly than individual Shewhart charts when the correlation is based on an ARMA (1,1) model. They also suggest that the residual charts are not sensitive to small process shifts. Lu and Reynolds (1995) study the EWMA control charts to monitor the mean of autocorrelated process. They suggest that for the low and moderate level of correlation, a Shewhart control chart of observations will be better at detecting a shift in the process mean than a Shewhart chart of residuals. For low and moderate shifts EWMA chart will be better than Shewhart chart. They also suggest that when there is high autocorrelation in the process, constructing control charts based on estimated parameters should not be used, instead, applying time series model would be appropriate for the construction of control limits. Schmid (1995, 1997a, 1997b) shows that if there is large shift in the process, using Shewhart

chart is appropriate while EWMA and CUSUM charts are sensitive to small and moderate shifts. Maragah and Woodall (1992) adjust the control limits for autocorrelated univariate data by taking autocorrelation into account. But the tables are needed to choose the critical value when the adjustment is necessary. For each structure, the control limits would be different. For the first order autoregressive process such tables are given by Schmid (1995, 1997a, 1997b). But the residual charts need just one joint control limits which are based on independent and identically distributed case. Therefore, residual charts have an advantage on the construction of control limits than adjusting the control limits. Statistical process control applications generally focus on the residuals of univariate autocorrelated chart. However, the autocorrelation problem in univariate case also extends to multivariate cases. Therefore these studies are extended to multivariate cases by various authors. The widely used control charts to detect the mean shift in multivariate processes are Hotelling T-square control charts, MEWMA (multivariate exponentially-weighted moving average) charts and MCUSUM (multivariate cumulative sum control) charts.

Pan and Jarret (2004) propose using vector autoregressive model (VAR) to monitor multivariate process in the presence of serial correlation by using the residuals of the model. They examine the effects of shifts in the process parameters on the VAR residual chart. Kalgonda and Kulkarni (2004) propose a control chart called Z-chart for the first order vector autoregressive (VAR (1)) process. They also suggest using Z-chart to identify the source of the shift. Pan and Jarret (2007) extend Alwan and Roberts's approach to multivariate cases, using the residuals from the vector autoregressive model on the Hotelling T-square control charts to monitor the multivariate process in the presence of serial correlation. They examine the effects of shifts in process parameters on the residuals of VAR model. They mention that using residuals from a VAR model on Hotelling T-square control chart is effective when the small changes occurred in the mean, covariance and autocorrelation coefficient. They use individual univariate Shewhart charts to further identify the variables which is responsible for the shift. H. Brian Hwang and Yu Wang (2010) propose a neural network identifier (NNI) for multivariate autocorrelated process and benchmark the proposed scheme with Hotelling T-square control chart, MEWMA chart and Z chart. Snoussi (2011) study an approach which is a combination of multivariate residual charts for autocorrelated data and the multivariate transformation technique for independent and identically process observations of short length.

However, some authors such as Harris and Ross (1990), Longnecker and Ryan (1990), Zhang (1997) suggest that for the univariate case, using X-chart based on residuals do not have the same properties as the X-charts for an independent process and show that when the process has mean shift, the detection capability of X-chart based on residuals and X-chart for an independent process are not equal. In this dissertation, we investigate whether the study made for univariate autocorrelated data is valid for the multivariate autocorrelated data. In addition, in the relevant literature, although the performances of Hotelling T-square control charts based on residuals from a VAR model have been used for multivariate autocorrelated process, there exists no study that shows the comparison with performances of Hotelling T-

square control charts based on raw data which have VAR structure. Therefore, in this study these charts (Hotelling T-square charts based on residuals and raw data) are evaluated based on the first order vector autoregressive structure by using average run length as the performance measure.

Chapter 3

Monitoring Univariate Time Series

There are generally two phases in statistical process control (SPC) applications. In Phase I, a historical set of data is considered to determine the in control process performance and understand the variation in the process over time. In Phase II, actual process monitoring is performed based on the control chart constructed in Phase I.

The general assumption is that the data are normally and independently distributed with mean μ and standard deviation σ when the process is in control. If this assumption is violated, the control charts are effected by the violation of independence, and may not work well. In this dissertation we deal with two types of data which are univariate and multivariate data. For the univariate case, we use X-chart, and for the multivariate case, Hotelling T-square control chart is considered. In univariate X-chart, there are two important parameters which are mean value and the standard deviation. If we assume that the univariate process is normally distributed with mean μ and standard deviation σ , where μ and σ are known, then the following control limits with a center line can be used on X-chart for individual observations,

$$\begin{aligned}UCL &= \mu + Z_{\alpha/2}\sigma \\Center\ Line &= \mu \\LCL &= \mu - Z_{\alpha/2}\sigma\end{aligned}\tag{3.1}$$

It is usual to replace $Z_{\alpha/2}$ by 3, so that three sigma limits are employed, which means for normally distributed data, in control average run length of 370 with the 0.0027 false alarm rate. If an observation falls outside of these limits, then a signal is declared.

The time which a control chart first signals is called run length. The probability distribution of the run lengths is called run length distribution, and the average value of this distribution is called average run length or in other words, average run length is defined as the measurement of average number of points will be plotted on a control chart before an out of control situation is occurred, and it is a widely used indicator to have an idea about the effectiveness of a control chart. ARL can be expressed as,

$$ARL = \frac{1}{P(\text{one point plots out of control})}$$

or, for in control ARL

$$ARL_o = \frac{1}{\alpha}$$

where α indicates the probability of false alarm. If there is no change in the process or when the process is in control, the probability of false alarm indicates the probability of a sample point plotted outside the control limits, and it is sometimes called probability of a type I error. For univariate control chart, if α value indicates the probability of an observation plotted outside the control limits, it is expected $1/\alpha$ points will be plotted before a false alarm is indicated.

Now assume that the parameters, μ and σ , are unknown, and when the process is in control they should be estimated from the preliminary or Phase I data. By estimating these parameters, control limits can be calculated, and considered control limits are used to monitor the process in Phase II. Estimation of mean value and variance is considered respectively as in the following, estimated mean value or sample mean is,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.2)$$

Estimated variance or sample variance is,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.3)$$

where, n is the number of observations taken from Phase I when the process is in control, and x_i is the i th observation in the process. Now the control limits can be constructed by the estimated parameters which are sample mean and sample variance with 3 sigma limits,

$$UCL = \bar{x} + 3s$$

$$Center\ Line = \bar{x} \quad (3.4)$$

$$LCL = \bar{x} - 3s$$

Until now we get the brief introduction about the construction of standard control limits for univariate X-chart. If the univariate data have some dependency over time such as autocorrelation which indicates the relationship between the observations at two different time points, then the construction of control limits will be dependent on the autocorrelation level. We know that the key parameters of any univariate normally distributed process are the mean and the variance, but if there is a relationship between observations for the different time periods, another parameter plays an important role for the construction of control limits. This new parameter is explained as the time series correlation which is defined as autocorrelation function (ACF). Autocorrelation function shows the autocorrelation coefficient which is the measurement of the correlation between observations at different times. For lag k operator, autocorrelation function is defined as,

$$\rho_k = \frac{\sum_{t=1}^{N-k} (x_t - \mu)(x_{t+k} - \mu)}{\sum_{t=1}^N (x_t - \mu)^2} \quad (3.5)$$

However, if we consider the sample data, then we need to use sample autocorrelation function which is expressed as,

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (3.6)$$

where r_k indicates the sample autocorrelation between observations k lags apart and \bar{x} denotes the sample mean.

3.1 AR (p) Models

When the observations at different time points are correlated, the data is commonly modeled as an ARMA (p, q) process given as,

$$x_t = c + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (3.7)$$

or

$$x_t = c + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.8)$$

where c is the constant value, ϕ_1, \dots, ϕ_p are the autoregressive parameters, $\theta_1, \dots, \theta_p$ are the moving average parameters, p and q are the lag orders of the process, ε_t is the error term which is assumed to be uncorrelated and normally distributed with mean is 0 and variance σ_ε^2 . For simplicity, in this section autoregressive (AR) processes will be investigated since the further studies will be based on autoregressive processes. In autoregressive models, the observed time series x_t depends on a weighted linear sum of the past values of x_t and an error term ε_t . Autoregressive, AR (p), model is expressed as,

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (3.9)$$

or

$$x_t = c + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t \quad (3.10)$$

where x_t denotes the time series observations, c is a constant value, ε_t indicates the error term and ϕ_1, \dots, ϕ_p indicates the autocorrelation coefficient of the model. In that formula the value of p is called order of the AR model. Sometimes autoregressive processes are expressed in the literature by using the lag operator L , which is defined as,

$$Lx_t = x_{t-1} \quad p = 1$$

$$L^2x_t = x_{t-2} \quad p = 2$$

$$L^s x_t = x_{t-s} \quad p = s$$

Now we can write the AR (p) process with lag operator L ,

$$x_t = c + \phi_1 Lx_t + \phi_2 L^2x_t + \dots + \phi_p L^p x_t + \varepsilon_t \quad (3.11)$$

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)x_t = c + \varepsilon_t \quad (3.12)$$

where $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ indicates the polynomial of lag operator which is called lag polynomial. So, $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ represents the polynomial of order p and then

$$\phi(L)x_t = c + \varepsilon_t \quad (3.13)$$

The stationarity of the process is an essential assumption to define a time series process. In stationary time series, it is assumed that the mean, variance and autocorrelation structure do not change over time. Therefore, we should consider the following equations for univariate stationary time series processes,

$$E(x_t) = E(x_{t-1}) = \dots = E(x_{t-k}) = \mu \quad (3.14)$$

$$Var(x_t) = E[(x_t - \mu)^2] = E[(x_{t-1} - \mu)^2] = \dots = E[(x_{t-k} - \mu)^2] = \sigma_x^2 \quad (3.15)$$

$$Cov(x_t, x_{t-k}) = E[(x_t - \mu)(x_{t-k} - \mu)] = \dots = E[(x_{t-j} - \mu)(x_{t-j-k} - \mu)] = \gamma_k \quad (3.16)$$

where γ_k , σ_x^2 and μ respectively denote finite autocovariance, finite variance and finite mean. As it is seen from the equations (3.14) and (3.15) both the mean and the variance are constant while the covariance changes as a function of the k indices in equation (3.16). For the AR (p) process, if the absolute values of roots of the lag polynomial, $\phi(L)$, lie outside the unit circle, then AR(p) process is considered as stationary or stable. Consider the AR (1) process which is first order autoregressive process, and expressed by,

$$x_t = \phi x_{t-1} + \varepsilon_t \quad (3.17)$$

where constant value c is omitted, ε_t is a white noise process with mean zero and unknown but fixed variance σ_ε^2 . For AR (1) process, the stationary condition is computed as in the following,

$$x_t = \phi x_{t-1} + \varepsilon_t \quad (3.18)$$

$$(1 - \phi L)x_t = \varepsilon_t \quad (3.19)$$

The characteristic equation to find the root is $(1 - \phi L) = 0$, then $L = 1/\phi$

$$\text{If } |1/\phi| > 1 \Rightarrow |\phi| < 1 \quad (3.20)$$

For AR (1) process, stationary condition is found as $|\phi| < 1$. After determining the stationary condition for AR (1) process, we can compute the expected mean, variance and covariance of a stationary AR (1) process as in the following since we will use these parameters later for our simulation. Now we assume that the time series are stationary, and the expected mean is the same for all values of time t as in equation (3.14), if the mean is denoted by μ , then the expected value of stationary AR(1) process is found by,

$$\begin{aligned} E(x_t) &= E(c) + E(\phi x_{t-1}) + E(\varepsilon_t) \\ E(x_t) &= E(c) + \phi E(x_{t-1}) + E(\varepsilon_t) \\ \mu &= c + \phi\mu + 0 \\ \mu &= \frac{c}{1 - \phi} \end{aligned} \quad (3.21)$$

If c constant value is considered as 0, then the mean becomes 0. The second moment, variance, for the stationary AR (1) process is computed as,

$$\begin{aligned} var(x_t) &= var(c) + var(\phi x_{t-1}) + var(\varepsilon_t) \\ var(x_t) &= \phi^2 var(x_{t-1}) + \sigma_\varepsilon^2 \\ \sigma_x^2 &= \phi^2 \sigma_x^2 + \sigma_\varepsilon^2 \\ \sigma_x^2 &= \frac{\sigma_\varepsilon^2}{1 - \phi^2} \end{aligned} \quad (3.22)$$

Furthermore, the autocovariance and the autocorrelation coefficients can be computed respectively for the stationary AR (1) process by the following equations,

$$\begin{aligned} cov(x_t, x_{t-k}) &= \frac{\phi^k \sigma_\varepsilon^2}{1 - \phi^2}, \\ cor(x_t, x_{t-k}) &= \phi^k, \end{aligned}$$

For a stationary AR (1) process, autocorrelation function (ACF) is defined as $\rho_k = \phi^k$. Until now we get the brief introduction about the parameters of the stationary first order autoregressive, AR (1), processes. Now we need to construct a control chart for an AR (1) process. Estimation of control limits for the stationary AR (1) time series process is constructed by considering the equations (3.21) and (3.22) as following,

$$\begin{aligned}
UCL &= \frac{c}{1 - \phi} + 3 \sqrt{\frac{\sigma_{\varepsilon}^2}{1 - \phi^2}} \\
Center\ Line &= \frac{c}{1 - \phi} \\
LCL &= \frac{c}{1 - \phi} - 3 \sqrt{\frac{\sigma_{\varepsilon}^2}{1 - \phi^2}}
\end{aligned} \tag{3.23}$$

In the equation (3.23) control limits of a stationary AR (1) process on the raw data is expressed by taking the autocorrelation coefficient into account.

3.2 Determination of the number of observations in Phase I

Now we have two methods to construct the control limits for a stationary AR (1) process, one of them is calculated by ignoring the autocorrelation effect in the process, other is constructed by taking the autocorrelation into account. Here we will compare these two methods for different number of observations in Phase I. But first we try to investigate how the impact of autocorrelation effects the distribution of the run lengths for these two methods. We generate 5000 datasets with 5000 observations each. For the first method we use the control limits in equation (3.4), and the sample mean \bar{x} and the sample standard deviation s are estimated from the 5000 observations which is considered as good enough to estimate the parameters. For the second method, we use the equation (3.23) in which autocorrelation level is taken into account.

In Figure 1, it can be seen the q-q plot of 5000 run lengths and the histogram of the run lengths which are acquired from 5000 datasets in the case that the parameters are unknown and known when there is no autocorrelation in the process. The case with unknown parameters indicates the calculations based on the control limits with estimated parameters while the case with known parameters indicates the calculations based on the control limits calculated in equation (3.23). Since the observations are normally distributed with mean is 0 and variance is 1, the control limits for the case with known parameters in which $\phi = 0$, will be expressed as,

$$\begin{aligned}
 UCL &= 0 + 3 \sqrt{\frac{1}{1 - 0^2}} = 3 \\
 \text{Center Line} &= 0 \\
 LCL &= 0 - 3 \sqrt{\frac{1}{1 - 0^2}} = -3
 \end{aligned}
 \tag{3.24}$$

The average of 5000 run lengths when there is no autocorrelation is 372.59 for the calculations based on the method in which unknown parameters are considered. The average of the run lengths is 372.64 when the known parameters are considered. In Figure 1, q-q plot is based on the exponential distribution for the run lengths since the fact that run lengths for a good process have exponential distribution. According to the Figure 1 exponential distribution for the run lengths seems valid when the observations are normally distributed but not autocorrelated.

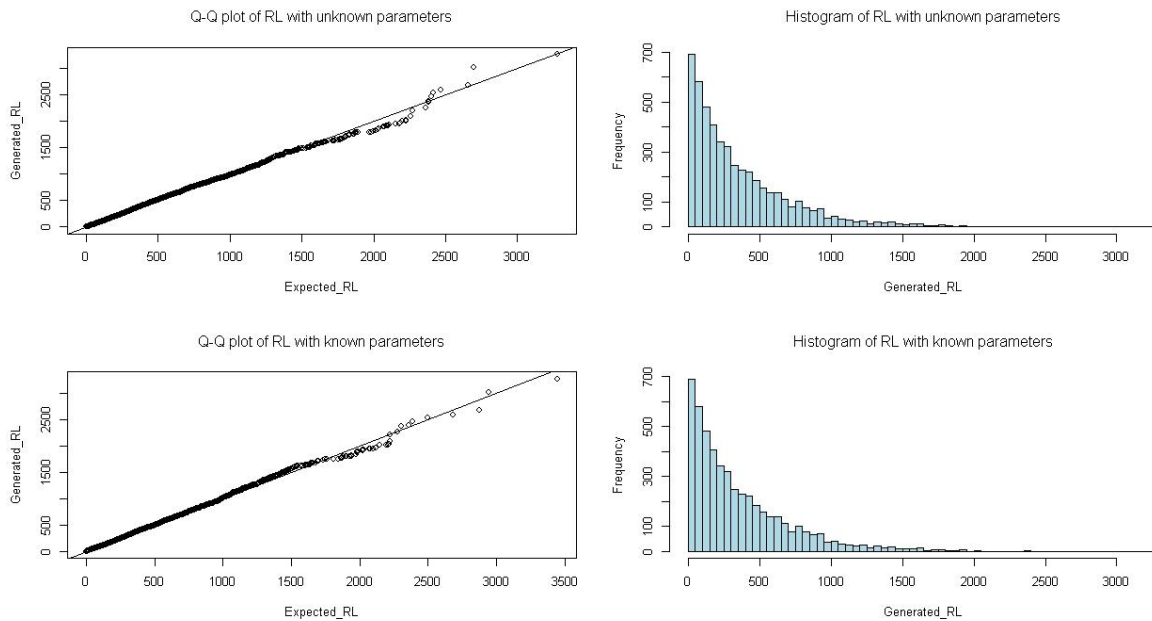


Figure 1 Distribution of the run lengths and histogram of the run lengths with known and unknown parameters when $\phi = 0$

Then we generate the 5000 datasets with autocorrelation level 0.7. Figure 2 shows the q-q plot of 5000 run lengths and the histogram of the run lengths with autocorrelated observations based on the control limits with known and unknown parameters. For the method with unknown parameters, we estimate the sample mean and the sample variance from the autocorrelated observations, and construct the control limits based on these estimated parameters. The average run length is 468.56 for this method. For the method with known

parameters, we use the control limits in equation (3.23) with the autocorrelation level 0.7, and the control limits based on the known parameters for the autocorrelated process (AR(1)) in which error term is normally distributed with mean 0 and variance 1,

$$UCL = 0 + 3 \sqrt{\frac{1}{1 - 0.7^2}} = 4.2$$

$$Center\ Line = 0 \tag{3.25}$$

$$LCL = 0 - 3 \sqrt{\frac{1}{1 - 0.7^2}} = -4.2$$

The average run length is 469.13 in the case of using the control limits in equation (3.25) when the process is autocorrelated with the level of 0.7.

Figure 2 shows the q-q plot of the run lengths and the histogram of the run lengths based on autocorrelated observations with known and unknown parameters. According to the q-q plots of the run lengths, exponential distribution for the run lengths seems valid when the observations are autocorrelated. However the average run length changes with the autocorrelation level.

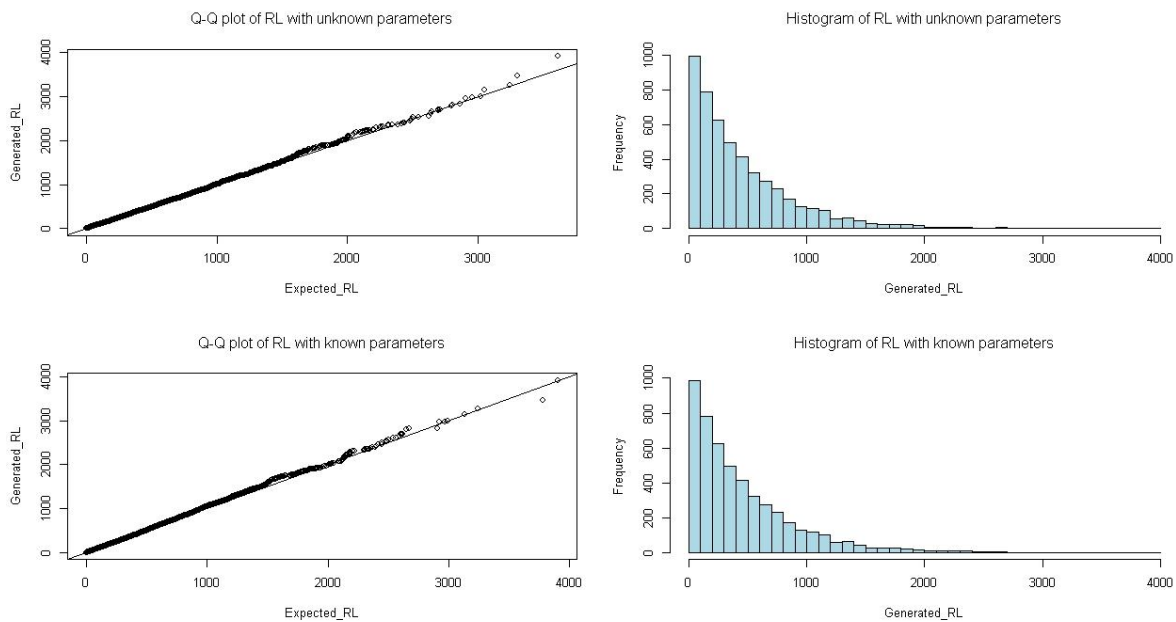


Figure 2 Distribution of the run lengths and histogram of the run lengths with known and unknown parameters when $\phi = 0.7$

Until now we consider 5000 observations so that at least one of the observations gives signal in each dataset. But now we will try to calculate the average run lengths for different number of observations in Phase I to see whether we can use exponential distribution for the run

lengths in the case of small number of observations in Phase I. To calculate the average run lengths for small number of observations by using exponential distribution, we calculate the number of datasets for which we have a signal. The ratio of this number to total N number of datasets is used as an estimate for the probability of run lengths is less than n ($\Pr(RL < n)$) where n is the dataset size and run lengths are exponentially distributed with certain λ ($RL \sim \text{EXP}(\lambda)$). Hence we can estimate $1/\lambda$ which is used for ARL. Also note that this method fails if all datasets signal. However what we look for is when not all datasets signal anyway since sample average of the run lengths will not be appropriate as some run lengths are capped at n. Since we consider that the exponential distribution for the run lengths seems valid when we use 5000 observations in the case of known and unknown parameters, now we will try to compare the average run lengths which are acquired by the use of control limits based on equations (3.4) and (3.23) for small number of observations. Here we generate different number of observations based on the first order autoregressive process (AR (1)) in which correlation coefficients are considered as,

$$\phi = \{0, 0.3, 0.5, 0.7, 0.9, -0.3, -0.5, -0.7, -0.9\}.$$

For the method in which we use the known parameters, the mean of the data generated with first order autoregressive structure is assumed to be 0, error term is normally distributed with mean 0 and standard deviation 1, and the control limits based on the considered autocorrelation levels by using the equation (3.23) are,

ϕ	UCL	LCL
0	3	-3
0.3	3.14	-3.14
0.5	3.46	-3.46
0.7	4.20	-4.20
0.9	6.88	-6.88
-0.3	3.14	-3.14
-0.5	3.46	-3.46
-0.7	4.20	-4.20
-0.9	6.88	-6.88

Table 1 Control limits with known parameters for AR (1) process

When we are taking autocorrelation into account, the control limits above are considered to calculate the average run length based on X-chart for the data which has first order autoregressive structure. Table 2 shows the average run lengths in the combination of different autocorrelation levels and the different number of dataset size for the AR(1) process. The ARLs under the ‘known parameters’ column is calculated in terms of the control limits considered in Table 1 while the ARLs under the column of ‘unknown parameters’ is calculated by the use of control limits constructed with estimated parameters as in equation (3.4) by ignoring autocorrelation.

n	ϕ	Known parameters		Unknown parameters	
		Exponential	Average	Exponential	Average
50	0	373	24	352	24
	0.3	395	25	337	24
	0.5	395	25	299	25
	0.7	477	24	281	24
	0.9	817	25	179	23
100	0	375	47	364	47
	0.3	372	48	346	48
	0.5	407	49	352	48
	0.7	475	47	362	47
	0.9	833	48	361	47
200	0	366	92	364	91
	0.3	376	90	361	90
	0.5	399	91	370	90
	0.7	459	93	405	91
	0.9	808	95	517	92
300	0	377	129	373	129
	0.3	372	131	364	131
	0.5	392	132	375	131
	0.7	459	134	415	132
	0.9	832	138	604	135
400	0	364	165	363	165
	0.3	377	164	373	163
	0.5	397	169	381	168
	0.7	459	168	431	166
	0.9	855	183	666	178
500	0	363	194	361	194
	0.3	379	195	375	194
	0.5	395	199	385	197
	0.7	472	205	441	202
	0.9	810	223	672	217
600	0	367	219	364	219
	0.3	376	222	370	221
	0.5	392	228	382	226
	0.7	452	236	434	234
	0.9	832	266	701	259
700	0	365	247	363	246
	0.3	379	248	374	247
	0.5	395	251	388	250
	0.7	454	267	434	263
	0.9	812	301	703	294

Table 2 ARLs obtained by using X-chart based on the raw data in the combination of different autocorrelation levels and different number of observations in Phase I for AR (1) process

n	ϕ	Known parameters		Unknown parameters	
		Exponential	Average	Exponential	Average
800	0	369	265	368	265
	0.3	370	268	367	268
	0.5	400	273	393	271
	0.7	461	288	442	285
	0.9	829	338	726	328
900	0	373	281	372	281
	0.3	384	287	381	285
	0.5	406	295	402	293
	0.7	455	315	439	310
	0.9	839	368	748	360
1000	0	367	303	367	303
	0.3	375	303	373	302
	0.5	400	310	395	307
	0.7	457	332	445	328
	0.9	823	397	755	387
2000	0	374	356	374	356
	0.3	384	373	384	372
	0.5	390	380	388	377
	0.7	441	441	434	436
	0.9	823	639	786	623
3000	0	369	367	369	366
	0.3	427	373	427	372
	0.5	383	390	383	388
	0.7	440	458	440	454
	0.9	826	755	804	739
4000	0	NA	369	NA	368
	0.3	NA	374	NA	374
	0.5	NA	397	NA	395
	0.7	469	464	469	460
	0.9	838	798	817	783
5000	0	NA	371	NA	370
	0.3	NA	373	NA	373
	0.5	NA	396	NA	395
	0.7	NA	458	NA	456
	0.9	810	819	797	807
6000	0	NA	371	NA	371
	0.3	NA	383	NA	383
	0.5	NA	397	NA	396
	0.7	NA	465	NA	463
	0.9	892	818	868	801

Table 2 Continued

In Table 2, ‘Exponential’ indicates the ARLs which are calculated according to exponential distribution of the run lengths, and ‘Average’ indicates the simple average of the run lengths.

For the method in which parameters are estimated from the generated datasets, if the number of observation is less than 200, the impact of the autocorrelation may not be detected by considering exponential distribution of the run lengths. As it is seen, when the number of observation is 50, the average run length decreases if the level of autocorrelation increases. Also if the number of observation is 100, it is not easy to see the impact of the autocorrelation since the calculations of the average run lengths based on exponential distribution for the run lengths are around 360 in the case of different autocorrelation levels. Another result for the method in which parameters are estimated to construct the control limits is that when the

number of observations is increasing, the average run length values which are calculated based on exponential distribution of the run lengths are approaching the average run length values that we found in the case of exponential distribution of run lengths with the use of control limits based on equation 3.23 in Table 1 (Known parameters). But, if the number of observations are higher than 3000, since all datasets signal for some autocorrelation levels, consideration of ARL may not be possible by using the exponential distribution of run lengths based on the control limits constructed with known and estimated parameters. For example, when the number of observations is equal or higher than 4000, and the autocorrelation level is 0.5, NA indicates that the calculation of exponential distribution of run lengths based on the ratio of the datasets for which we have a signal to total number of datasets does not give meaningful result since each dataset shows a false alarm. But if it is considered to take high number of observations such as 4000 and above, taking the average of the run lengths with known and unknown parameters gives more meaningful results. Also there is no significant difference between average values of the run lengths based on known parameters and the average values of the run lengths based on estimated parameters for all different number of observations. They are small if the number of observations is small, since we consider the average of the run lengths by ignoring the data which do not signal.

As a result, from Table 2, we can say that for the small number of observations in the dataset which has AR (1) structure, to calculate the average run lengths it is possible to use exponential distribution of the run lengths based on the control limits constructed with known parameters by taking autocorrelation into account, and also it is possible to calculate the average run length by taking the average of the run lengths based on the control limits with known parameters in which autocorrelation is taken into account and unknown parameters in which parameters are estimated when the number of observation is higher than 4000.

3.3 Residuals of AR (1) Models

To fit an ARMA (p,q) model, we need to determine the order p and q . To do this the plots of autocorrelation (ACF) and partial autocorrelation functions (PACF) are required. ACF shows the coefficients of correlation between x_t and x_{t-k} for $k=1, 2, \dots$. PACF is the autocorrelation between x_t and x_{t-k} after removing any linear dependency on other lags. The orders p and q are determined by the behaviors of ACF and PACF. After identifying the order of time series model, parameter estimation should be considered based on the model. In our simulations we used maximum likelihood estimation method to estimate the parameters of model. By using these estimated parameter residuals of the model are calculated to assess the adequacy of the model. Residuals are the differences between actual observation value and the fitted value. Since the assumption is that the residuals are independent and identically distributed, then it should be checked whether the residuals behave like white noise by applying the traditional control charts.

Suppose that $\hat{\phi}$ is an estimate of ϕ , \hat{c} and $\hat{\sigma}_\varepsilon^2$ are the estimates of c and σ_ε^2 obtained from the preliminary data of the AR process where error term $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and \hat{x}_t is the fitted value of x_t . Then the residuals can be calculated for AR (1) process as

$$x_t = c + \phi x_{t-1} + \varepsilon_t \quad (3.26)$$

$$e_t = x_t - \hat{x}_t \quad (3.27)$$

$$e_t = x_t - [\hat{c} + \hat{\phi}_1 x_{t-1}]$$

$$e_t = X_t - [(1 - \hat{\phi}_1)\hat{\mu} + \hat{\phi}_1 x_{t-1}]$$

$$e_t = X_t - [\hat{\mu} + \hat{\phi}_1(x_{t-1} - \hat{\mu})]$$

where e_t indicates the residual at time t, and these residuals are assumed to be approximately normally distributed with mean is zero and constant variance σ_ε^2 for stationary process.

For simplicity, first we generate 1000 datasets which have first order autoregressive (AR(1)) structure with no change in the mean. Since we use 100 observations in Phase I, it is expected to use exponential distribution of the run lengths to calculate the in control ARL based on the control limits constructed with known parameters. However, we show that if the sample size is large such as 4000 and above in Phase I, it is also expected to get reasonable results by using the control limits constructed with estimated parameters since the uncertainty for the estimation of parameters will be low. In Phase II, we use 5000 observations so that we have at least one false alarm for each dataset. When the each dataset signals, the total number of run lengths would be 1000. Taking the average of these run lengths is considered as the ARL of the process.

In our simulation, when we are constructing the control limits we use known parameters such as,

$$\mu = 0, \quad c = (1 - \phi)\mu = 0,$$

$$\sigma_\varepsilon = 1 \quad \text{and} \quad \phi = \{0, 0.25, 0.5, 0.75, 0.95, -0.25, -0.5, -0.75, -0.95\}$$

For the X-chart (individuals chart) of the observations with the parameters assumed to be known, the control limits are constructed by taking the autocorrelation into account for the AR(1) process as following,

$$\begin{aligned}
 UCL &= 0 + 3 \sqrt{\frac{1}{1 - \phi^2}} \\
 \text{Center Line} &= 0 \\
 LCL &= 0 - 3 \sqrt{\frac{1}{1 - \phi^2}}
 \end{aligned} \quad (3.28)$$

As we consider before, we can use the exponential distribution of run lengths to calculate the average run lengths for small number of observations in Phase I since there is no significant difference if we consider the average of the run lengths in the case of the number of observation higher than 4000 observations in Phase I where almost at least one observation signals for the each data simulation.

Table 3 shows the in control ARL under the column of ‘Average’, which is the average number of observations before an out of control signal generated with corresponding autocorrelation levels using X-chart with 3 sigma control limits based on known parameters in which autocorrelation level is taken into account when the number of observations is 5000 for AR(1) process. Also under the column of ‘Exponential’ we can see the in control ARLs calculated by the use of exponential distribution of run lengths based on X-chart with known parameters when the number of observation is 100.

There is no significant difference between taking the average of the run lengths of 1000 datasets in which each dataset has 5000 observations and ARL based on the exponential distribution of the run lengths when the number of observation is 100 in Phase I in the case of different autocorrelation levels. The increase in the average run length is explained by the increase of autocorrelation level, or in other words, when the autoregressive parameter ϕ is getting larger, the in control ARLs increase when the X-chart for AR(1) process is constructed with known parameters by taking the autocorrelation into account.

	Average	Exponential
ϕ	N=5000	N=100
0	369	372
0.25	374	375
0.5	397	392
0.75	503	498
0.95	1205	1192

Table 3 ARLs obtained by using X-chart based on the raw data with exponential distribution and taking the average of run lengths in Phase I for AR(1) process

Since we consider the control limits constructed with known parameters, corresponding residuals are calculated with these known parameters such as

$$e_t = x_t - \phi x_{t-1} - c$$

As we mention before, residuals are assumed to be independent and identically distributed with mean is zero and variance is one, i.e. $\varepsilon_t \sim N(0,1)$, the construction of the control limits for residuals with 3 sigma limits are made as following,

$$UCL = 0 + 3 \times 1 = 3$$

$$Center\ Line = 0 \tag{3.29}$$

$$LCL = 0 - 3 \times 1 = -3$$

where, expected value of residuals based on AR(1) model is assumed to be zero and standard deviation is one. Now we can use these control limits (3.28) and (3.29) to monitor the process. Until now we assume that all the parameters that we need are known. Control limits of X-chart based on raw data which have AR(1) structure and the residuals of AR(1) model are calculated in terms of these known parameters.

Then we consider the residuals of AR(1) model which is fitted to the datasets in which each dataset has 100 observations in Phase I. To calculate the ARLs based on these residuals we use exponential distribution of run lengths. Table 4 shows the average run lengths acquired by using X-chart based on residuals with different autocorrelations, in which control limits of residuals are considered as in equation (3.29). Each scenario has approximately the same in control ARLs, around 370.

	Exponential		Exponential
ϕ	N=100	ϕ	N=100
0	372	0	373
0.25	375	-0.25	377
0.5	371	-0.5	373
0.75	376	-0.75	375
0.95	374	-0.95	372

Table 4 ARLs obtained by using X-chart based on the residuals with the exponential distribution of run lengths in Phase I for AR(1) process

Many authors suggest that the control charts based on residual should be used to monitor to process. However, Harris and Ross (1990), Longnecker and Ryan (1990) discuss that the control charts based on residuals from a first-order autoregressive (AR (1)) process may have poor detection power to detect the process mean shift. Longnecker and Ryan (1990) discuss that control charts based on residuals may have high detection power to detect a shift in the process mean when the first residual is plotted, but if the control chart based on residuals fails to detect the shift when the first residual is plotted, then the subsequent residuals would have low probability to detect the shift for an AR(1) process with positive autocorrelations. Zhang (1997) studies detection capability of X-chart based on residuals for general stationary univariate autoregressive process such as AR (1) and AR (2), furthermore, compares detection capability of X-chart based on residuals with the traditional X-chart based on raw data and shows that when the process has a mean shift, the detection capability of X-chart based on residuals for which observations are perfectly modeled and the traditional X-chart based on raw data for an independent process are not equal. Here, we also show when the X-chart based on residuals from AR (1) process will have poor performance to detect the shifts in the process mean. If there is a shift in the process mean given as

$$\mu^* = \mu + \delta\sigma_x \tag{3.30}$$

Then the mean of the residual at time t=T is,

$$E [e_T] = \mu^* - \mu = \delta\sigma_x$$

$$\text{for } t = T + 1, \quad E [e_{T+1}] = \mu^* - [\mu + \phi(\mu^* - \mu)] = (\mu^* - \mu)(1 - \phi) = \delta\sigma_x(1 - \phi)$$

As it is seen, since the expected value of residuals at $t = T$ is bigger than the expected value of residuals at $t = T + 1$, ($\delta\sigma_x > \delta\sigma_x(1 - \phi)$), most of the shift proportion is captured by the first residual, subsequent residuals capture just a proportion of first residual, which depends on the autocorrelation level. Since standardized residuals are related to residual control charts, we have

$$Z_T = \frac{e_t - 0}{\sigma_e}$$

$$E(Z_T) = \frac{\mu^* - \mu}{\sigma_x\sqrt{1 - \phi^2}} = \frac{\delta\sigma_x}{\sigma_x\sqrt{1 - \phi^2}} = \frac{\delta}{\sqrt{1 - \phi^2}} \tag{3.31}$$

$$E(Z_{T+1}) = \frac{(\mu^* - \mu)(1 - \phi)}{\sigma_x\sqrt{1 - \phi^2}} = \frac{\delta\sigma_x(1 - \phi)}{\sigma_x\sqrt{1 - \phi^2}} = \frac{\delta(1 - \phi)}{\sqrt{1 - \phi^2}} \tag{3.32}$$

From the equations above, for AR (1) process, it is seen that $\frac{1}{\sqrt{1 - \phi^2}}$ of the shift is

captured by first residual (3.31), and $(1 - \phi) / \sqrt{1 - \phi^2}$ of the shift is captured by

subsequent residuals (3.32). The problem is that, if the shift is not detected by the first residual, then it will take more time to detect the shift with subsequent residuals when the autocorrelation is positive. But the situation will change when the autocorrelation is negative, subsequent residuals would have higher probability of detecting the shift than the first residual.

ϕ	First	Subsequent	ϕ	First	Subsequent
0	1	1	0	1	1
0.25	1.032	0.774	-0.25	1.032	1.291
0.5	1.154	0.577	-0.5	1.154	1.732
0.75	1.511	0.378	-0.75	1.511	2.645
0.95	3.202	0.160	-0.95	3.202	6.244

Table 5 The detection capability of first and subsequent residual based on X-chart for AR(1) process

Table 5 shows the detection capability rate of the first and the subsequent residuals for different autocorrelation levels. As it is seen, for positively autocorrelated dataset which has AR(1) structure, first residual have high probability to detect the shift, but if the shift could not be captured with first residual, then the subsequent residuals have less probability to detect the shift than it would do with independent data. Also if the positive autocorrelation level is getting higher, then the first residual detection probability increases while the detection probability of subsequent residuals decreases, for different negative autocorrelation levels, subsequent residuals have higher detection probability than the detection probability

of first residual, and also the detection probability of subsequent and first residual increases with the higher negative autocorrelation.

Now suppose that different magnitudes of shifts based on standard deviation unit (3.30) in the process mean is produced, and resulting average run lengths obtained by the use of X-chart constructed based on the control limits with known parameter by taking different autocorrelation level into account are calculated. For this, we generate 1000 datasets which have AR (1) structure with the dataset size of 100 observations in Phase I. To be able to calculate the more reasonable ARLs in Phase II, we consider the number of observation to be generated in Phase II as 5000 so that each dataset shows at least one false alarm. By this way, we will have 1000 run lengths and taking the average of these run lengths would be satisfactory. Here we show how the in control average run length changes in the the combination of different magnitudes of shift and autocorrelation level. In Table 6, we can see the performance of X-chart based on raw data comparison with the X-chart based on residuals from AR (1) process by considering the average run lengths in the combination of various amounts of shifts with different autocorrelation levels. In Table 6, ϕ and δ indicate respectively autocorrelation level and the amount of standard deviation unit shift in the process mean, and the values under the column of 'RESIDUAL' shows the ARLs of X-chart based on residuals of AR(1) model in which observations are perfectly modelled while the values under the column of 'RAW' express the ARLs of X-chart based on raw data which has AR(1) structure.

ϕ	δ	RAW	RESIDUAL
0.25	0	381.4	373.2
	0.5	168.9	217.3
	1	48.8	77.1
	2	8.2	12.6
	3	3.1	2.8
0.5	0	383.3	376.3
	0.5	178.3	258.6
	1	54.9	119.8
	2	9.9	23.1
	3	3.5	4.5
0.75	0	508.4	370.5
	0.5	246.1	304.4
	1	88.9	196.3
	2	15.4	38.4
	3	4.4	3.1
0.95	0	1314.3	365.7
	0.5	724.3	360.3
	1	277.4	145.8
	2	56.9	1
	3	7.2	1
-0.25	0	368.8	382.1
	0.5	157.5	105.4
	1	44.7	22.8
	2	7.1	3.5
	3	2.8	1.6
-0.5	0	404.2	376.3
	0.5	165.1	62.2
	1	47.4	10.5
	2	7.1	2.1
	3	2.8	1.3
-0.75	0	519.6	378.2
	0.5	188.8	22.4
	1	58.1	3.5
	2	9.1	1.5
	3	2.7	1.1
-0.95	0	1365.4	379.3
	0.5	445.1	2.7
	1	142.4	1.4
	2	22.1	1
	3	2.7	1

Table 6 ARLs obtained by using X-chart based on raw data and residual from AR(1) process in the combination of different autocorrelation levels and various magnitudes of shifts

ARLs obtained by using X-chart based on residuals and the X-chart based on raw data are calculated for AR (1) process with autocorrelation levels, 0.25, 0.5, 0.75, 0.95, -0.25, -0.5, -0.75,-0.95. The comparison is made in the combination of different amounts of shifts and the autocorrelation levels. Throughout the simulated examples, it is shown that when the autocorrelation level is 0.95, since the first residual detection capability is 3.20 while subsequent residual detection capability is 0.16, the X-chart based on residuals can detect the shift earlier than the X-chart based on raw data for all combination of shifts considered. Also

if the amount of shift is 3, then X-chart based on residuals can detect the shift earlier when the autocorrelation level is 0.75 and 0.95. For negative autocorrelation levels, since the detection capability of the subsequent residual is higher than the detection capability of first residual, X-chart based on residuals detects the shift earlier than the X-chart based on raw data. From the Table 6, ARLs obtained by using the X-chart based on residuals for negative autocorrelation levels are lower than the ARLs obtained by using the X-chart based on raw data for all combination of autocorrelation levels and the magnitudes of shifts.

Chapter 4

Monitoring Multivariate Time Series

In many statistical process control (SPC) applications, it is often the case that we have more than one quality characteristic to monitor. Monitoring these quality characteristics simultaneously is important since the correlation among the variables should be taken into account, or in other words, individual or univariate monitoring of variables will ignore the correlation among the variables. In real life, it is also reasonable to observe serial dependency for data collected in time. Therefore, in many SPC applications, it is assumed that the observations in the data matrix are correlated over time and the variables would have some correlation with each other. Ignoring these dependencies may cause incorrect interpretations when monitoring the data. In multivariate statistical process control applications, since several variables are of interest, multivariate control charts should be used. In the literature, there are three main multivariate control charts which are Hotelling T-square control chart, multivariate exponentially-weighted moving average (MEWMA) and multivariate cumulative sum (MCUSUM) control charts. Here, we use Hotelling T-square multivariate process control technique for monitoring simultaneously several correlated and autocorrelated quality characteristics. Hotelling T-square control chart is a multivariate extension of univariate control chart.

At the first part of this chapter, we apply Hotelling T-square control charts to the bivariate autocorrelated data, and in the second part, we use a bivariate time series model which is vector autoregressive model to take into account the autocorrelation, which is the multivariate extension of the univariate autoregressive model used in the previous chapter. Then we monitor the residuals of the vector autoregressive model by using Hotelling T-square control chart. These applications are made in the case of different autocorrelation levels with the first order vector autoregressive model (VAR (1)) as the reference model. We then as in the case from the previous chapter study the performance of the two Hotelling T-square control charts (one based on raw data and other based on the residuals from a VAR (1) model) in detecting a shift in the mean. Hotelling T-square control charts for raw data and the residuals from the first order vector autoregressive model are compared in terms of average run length performance measures. In Hotelling T-square calculations, it is expected that, in Phase I, the process parameters, sample mean and sample variance-covariance matrix are estimated. In that phase it is aimed to get the in control sample mean vector and sample variance covariance matrix, then they are used to obtain the Hotelling T-square statistics. But here we assume that the mean vector and the variance covariance matrix of the process are known.

4.1 Multivariate Normal Distribution

In univariate case, generally the process is assumed to be normally distributed. This assumption can be used for the multivariate case. Multivariate normal distribution is an extension of the univariate normal distribution to multiple variables. The generation of the random values from a p -dimensional multivariate normal distribution will have the following probability density function,

$$\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (4.1)$$

where, $\boldsymbol{\mu}$ represents $p \times 1$ mean vector of p variables $\boldsymbol{\Sigma}$ represents the $p \times p$ variance-covariance matrix of p variables, the diagonal elements of that matrix are the variances of each variables and off-diagonal elements are the covariances. \mathbf{x} represents the vector of $p \times 1$ random variable, $\mathbf{x} = (x_1, x_2, \dots, x_p)$. Hence we have $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, in our simulations, the mean vector $\boldsymbol{\mu}$ was set as zero vector,

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p) = (0, 0, 0, \dots, 0) \quad (4.2)$$

and variance-covariance matrix generally was unless otherwise specified set to $p \times p$ diagonal matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

where σ_{ij} , ($i = 1, \dots, p$ and $j = 1, \dots, p$, $i \neq j$) represents the correlations among the quality characteristics. When the diagonal elements are set as one, this matrix is equivalent to correlation matrix. If any off-diagonal element is different from zero, then it means corresponding variables are correlated. In addition, the covariance matrix $\boldsymbol{\Sigma}$ should be symmetric positive definite for multivariate normality assumption. It means that all eigenvalues of covariance matrix should be positive and $\boldsymbol{\Sigma}$ must be symmetric. In our simulations, after defining the mean vector and the correlation matrix which is symmetric positive definite, we generate the matrix of a dataset which is multivariate normally distributed as,

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

The above matrix represents the p dimensional multivariate normal distribution in which each dimension has n observations. In our simulations, first we assume that the variables are uncorrelated and the dataset has p dimensional multivariate normal distribution with mean

zero and identity covariance matrix Σ . But in real life the variables would often be correlated. Therefore, we made some changes in the covariance matrix for the further monitoring simulations to make comparison in the manner of different correlations among the variables. Here we assume that we know the mean vector and the covariance matrix of the multivariate normal distribution. If the parameters are unknown, we should calculate the sample mean vector and sample covariance matrix of the multivariate distribution. If we have a sample from a multivariate normal distribution such as,

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$$

where n is the number of observations in which i th sample vector contains observations on each of the p variables

$$x_{i1}, x_{i2}, \dots, x_{ip}$$

Then the sample mean vector is

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad (4.3)$$

and the sample covariance matrix of multivariate normal distribution is

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (4.4)$$

where diagonal elements are the sample variances and off-diagonal elements are sample covariances which is expressed,

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (4.5)$$

where $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, p$ for $j \neq k$. For the dataset which has multivariate normality, sample mean vector and sample covariance matrix are unbiased estimators of the population, which are expressed

$$E(\bar{\mathbf{x}}) = \boldsymbol{\mu} \quad (4.6)$$

$$E(\mathbf{S}) = \Sigma \quad (4.7)$$

where $\boldsymbol{\mu}$ and Σ are population mean vector and population covariance matrix respectively.

4.2 Hotelling T-square Control charts

As we mention before, if there is no correlation among the quality characteristics, then univariate control charts may be applied to determine whether the process is control or not. However, applying individual control chart to the each quality characteristics which has correlation with others may give erroneous conclusions. Therefore, in this study, Hotelling T-square control charts are used.

Hotelling T-square multivariate control charts are the multivariate extension of univariate Shewhart control charts. The main parameters for Hotelling T-square calculations are mean value of each variable and the variance-covariance matrix. These parameters are estimated from preliminary samples when the process is assumed to be in control. In the literature, Hotelling T-square control charts are expressed for subgroup data and for individual observations.

For the subgrouped data, consider a multivariate normally distributed data with p variables, and the population mean and population covariance are known, then χ^2 distribution would be used to monitor the mean and determine the control limits, the statistics of χ^2 control charts is expressed as,

$$\chi_t^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad t = 1, 2, \dots, m \quad (4.8)$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_p]$ is the vector of population mean which is known, $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$ is the vector of sample mean, n is the sample size, $\boldsymbol{\Sigma}$ is the covariance matrix of population. Suppose that we have m samples. The sample means are calculated from each sample as,

$$\bar{x}_{jk} = \frac{1}{n} \sum_{i=1}^n x_{ijk} \quad (4.9)$$

where $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, m$, x_{ijk} is the i th observation on the j th quality characteristic in the k th sample.

For only two variables x_1 and x_2 , the equation (4.8) becomes

$$\chi^2 = \frac{n}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} [\sigma_2^2 (\bar{x}_1 - \mu_1)^2 + \sigma_1^2 (\bar{x}_2 - \mu_2)^2 - 2\sigma_{12} (\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)] \quad (4.10)$$

where \bar{x}_1 and \bar{x}_2 are the sample averages of two quality characteristics computed from a sample size n , μ_1 and μ_2 are the population means of these variables, σ_1 and σ_2 are the standard deviations of x_1 and x_2 and σ_{12} is the covariance between x_1 and x_2 .

The lower control limit for the statistic in (4.8) is zero and the upper control limit with known parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ has the following equation with the specified false alarm rate and the number of variables,

$$UCL = \chi_{\alpha,p}^2 \quad (4.11)$$

If the mean vector and the variance covariance matrix of the multivariate data are unknown, they should be estimated from the in control data which is called Phase I with enough number of observations. These estimates are used in Phase II. In addition, when the true population values are not known, Hotelling T-square statistics are used instead of χ^2 statistics. When the parameters are unknown and the samples have subgroups, the Hotelling T-square calculation is expressed as following,

$$T^2 = n(\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \bar{\bar{\mathbf{x}}}) \quad (4.12)$$

where $\bar{\mathbf{x}}$ is the vector of sample mean, $\bar{\bar{\mathbf{x}}}$ is the vector of grand averages of variables. Estimated mean value and covariance matrix for subgrouped data are calculated as in the following equation,

$$\bar{\bar{x}}_j = \frac{1}{m} \sum_{t=1}^m \bar{x}_{jk} \quad (4.13)$$

$\bar{\bar{x}}_j$ is the grand average of variable j where $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, m$. The $\bar{\bar{x}}_j$ are the elements of vector $\bar{\bar{\mathbf{x}}}$ which is an unbiased estimate of $\boldsymbol{\mu}$ when the process is in control. Sample variance of variable j ,

$$\bar{s}_j^2 = \frac{1}{m} \sum_{t=1}^m s_{jk}^2 \quad (4.14)$$

where s_{jk}^2 is the sample variance on the j th variable in the k th sample, and \bar{s}_j^2 is the j th diagonal element of covariance matrix.

$$\bar{s}_{jh} = \frac{1}{m} \sum_{t=1}^m s_{jhk} \quad \text{where } j \neq h \text{ and } h = 1, 2, \dots, p \quad (4.15)$$

where \bar{s}_{jh} are the off-diagonal elements of sample covariance matrix \mathbf{S} . Then sample covariance matrix for subgrouped data is expressed with matrix notation

$$\mathbf{S} = \begin{bmatrix} \bar{s}_1^2 & \bar{s}_{12} & \bar{s}_{13} & \dots & \bar{s}_{1p} \\ & \bar{s}_2^2 & \bar{s}_{23} & \dots & \bar{s}_{2p} \\ & & \bar{s}_3^2 & \dots & \bar{s}_{3p} \\ & & & \ddots & \vdots \\ & & & & \bar{s}_p^2 \end{bmatrix}$$

This sample covariance matrix \mathbf{S} is an unbiased estimate of $\boldsymbol{\Sigma}$ when the process is in control.

The control limits for Hotelling T-square control chart is dependent on which phase we are monitoring. For in control process, Phase I with subgrouped data, where sample mean and sample covariance matrix is estimated, the control limits are,

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad (4.16)$$

$$LCL = 0 \quad (4.17)$$

To monitor the process in Phase II, control limits are

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \quad (4.18)$$

$$LCL = 0 \quad (4.19)$$

where p is the number of variables, m denotes the number of samples, n shows the sample group size, α is the false alarm rate.

When the sample group size $n=1$ in the multivariate simultaneous monitoring applications, and the parameters are known than χ^2 statistic with upper and lower control limit is calculated as following

$$\chi^2 = n(\mathbf{x}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}) \quad t = 1, 2, \dots, m \quad (4.20)$$

$$UCL = \chi_{\alpha,p}^2 \quad (4.21)$$

$$LCL = 0 \quad (4.22)$$

But if they are unknown and must be estimated from the preliminary data, the sample mean of the variable j to be estimated is

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_j \quad (4.23)$$

where m is the number of observations. Sample covariance matrix of individual multivariate data,

$$\mathbf{S} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (4.24)$$

where \mathbf{x}_i is the vector of i th observations of each variable and $\bar{\mathbf{x}}$ is the vector of sample means of variables. Then the Hotelling T-square equation which is dependent on sample mean vector and sample variance covariance matrix is,

$$T_t^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \quad (4.25)$$

Tracy, Young, and Mason (1992) consider that if the observations are individual observations, Hotelling T-square statistics would be based the beta distribution as in the following,

$$T_t^2 \sim \frac{(m-1)^2}{m} \beta_{\frac{p}{2}, \frac{m-p-1}{2}} \quad (4.26)$$

where, $\beta_{\frac{p}{2}, \frac{m-p-1}{2}}$ is a beta distribution with parameters $p/2$ and $m - p - 1/2$.

Hence, the Phase I limits should be based on beta distribution and the phase I limits with α rate calculated as,

$$UCL \sim \frac{(m-1)^2}{m} \beta_{\alpha, \frac{p}{2}, \frac{m-p-1}{2}} \quad (4.27)$$

$$LCL = 0 \quad (4.28)$$

where, $\beta_{\alpha, \frac{p}{2}, (m-p-1)/2}$ is the upper α percentage point of beta distribution with parameters $p/2$ and $(m - p - 1)/2$. The Phase II control limits are based on F distribution, which are

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p} \quad (4.29)$$

$$LCL = 0 \quad (4.30)$$

4.3 Determination of UCL for different number of observations in Phase I

To see the effect of sample size used in Phase I on the performance of the T-square control chart, we consider two methods to obtain more accurate upper control limit for Hotelling T-square statistics. We compare the performance of Hotelling T-square control charts based on the theoretical value of upper control limit as described above and the simulation based method in which upper control limit is found by the simulation in the case of different sample sizes.

For the simulation based method, we generate 1000 multivariate normal datasets with two variables and calculate the Hotelling T-square control statistics for each datasets. Then we take the maximum values of Hotelling T-square values of each datasets. Now we have 1000 maximum values, then we sort these in a descending order of maximum values. According to individual false alarm rate considered as 0.0027 which is equal to average run length 370, we found the overall α value for different size of samples as,

$$\alpha_{overall} = 1 - (1 - 0.0027)^n$$

Then, on the descending order of maximum values, we found the ($\alpha_{overall} \times 1000$)th value and we considered that value as our upper control limit. This is the UCL found by the simulation based method that we use below. We compare upper control limit which is found by simulation based method with the theoretical value of UCL for different sample sizes by considering average run lengths. Table 7 shows the calculated upper control limits and average run lengths for different number of observations. In Table 7, the ARLs under the column of 'Exponential' is calculated by using the exponential distribution of run lengths and

the ARLs under the column of ‘Average ’ is calculated by taking the average of the run lengths from the 1000 datasets. To calculate the average run lengths for different number of observations by using exponential distribution, the same procedure which is considered in chapter 3 for the calculations of ARLs based on X-chart is used. But here, the calculations of ARLs are based on Hotelling T-square chart. ‘Theoretical UCL’ values are the upper control limits found by using the beta distribution in equation (4.27).

Observation number n	Theoretical UCL	Simulation Based UCL	Exponential		Average Value	
			Theoretical	Simulation Based	Theoretical	Simulation Based
100	11.25	11.33	354	371	49	49
200	11.54	11.62	351	368	92	92
300	11.64	11.72	354	370	131	132
400	11.68	11.78	350	369	166	168
500	11.71	11.81	351	367	196	198
600	11.73	11.84	351	372	224	228
700	11.74	11.85	353	374	247	252
800	11.75	11.86	353	372	266	273
900	11.76	11.86	352	371	282	289
1000	11.77	11.86	351	369	297	305
1100	11.77	11.88	350	370	309	319
1200	11.78	11.89	348	370	320	331
1300	11.78	11.91	350	373	327	341
1400	11.78	11.92	353	378	329	344
1500	11.79	11.92	353	376	339	355
1600	11.79	11.93	350	377	344	365
1700	11.79	11.95	347	379	348	371
2000	11.79	11.97	339	380	357	386
2500	11.80	12.15	181	381	363	427
3000	11.80	12.33	88	365	367	471
4000	11.81	12.91	NA	468	365	628

Table 7 ARLs obtained by using Hotelling T-square control chart based on independent data with theoretical UCL and simulation based UCL

As it is seen in the Table 7, upper control limits found by the simulation based method is bigger than the theoretical UCL, and also they increase as the number of sample size increases. Average run lengths calculated with the theoretical UCL by using exponential distribution of run lengths express the average run length around 350 until the number of observation of 1700. For the sample size 2500 and above, this method is not working well to calculate the average run lengths. Exponential distribution of the run lengths for the small number of observations by using UCL obtained by simulation based method is giving more reasonable results. ARLs are almost around 370 which is expected for 0.0027 false alarm rate. But it gives high average run length, 468, for the large sample size such as 4000.

For the large sample sizes, calculating the average value of the run lengths by using theoretical UCL is giving more accurate results. Also calculating the average value of run length by using simulation based method is not giving expected ARL. Taking the average value of run lengths for small number of observations is giving incorrect results since it is ignoring the datasets which do not have a false alarm, so small average run length values are obtained. As a result, unless the sample size is high (>2000), we recommend using

exponential distribution of run lengths in Phase I with the upper control limits which are calculated based on simulation based methods.

4.4 Hotelling T-square Control Charts for Multivariate Autocorrelated Data

To control the multivariate processes, the traditional control charts require the assumption of the process data should be independently distributed. When the observations in the multivariate datasets are autocorrelated, the common approach is to reduce or remove the autocorrelation from the process by using appropriate time series models and use the residuals from the time series model to monitor the process. Since the residuals would be uncorrelated after using appropriate time series model, then the assumption of traditional control charts will be satisfied. But for the multivariate case, model selection would not be easy as the univariate case since the number of variables to be monitored simultaneously can be large. If the number of variables is large, then the number of parameters to be estimated would also be large. In multivariate cases, the commonly used multivariate time series model is the vector autoregressive moving average model (VARMA). Let's see how many parameters we need to estimate if we consider VARMA (p,q) models for multivariate autocorrelated process with k dimensional multivariate time series. VARMA (p,q) model is given as,

$$\mathbf{x}_t = \mathbf{c} + \Phi_1 \mathbf{x}_{t-1} + \dots + \Phi_p \mathbf{x}_{t-p} + \boldsymbol{\epsilon}_t + \theta_1 \boldsymbol{\epsilon}_{t-1} + \dots + \theta_q \boldsymbol{\epsilon}_{t-q} \quad (4.31)$$

where Φ_1, \dots, Φ_p are $k \times k$ autoregressive parameter matrices while $\theta_1, \dots, \theta_p$ are moving average parameters with also $k \times k$ matrices, \mathbf{c} is a $k \times 1$ vector of constants, $\boldsymbol{\epsilon}_t$ is a $k \times 1$ vector of error terms which are multivariate normally distributed with mean vector is 0 and variance covariance matrix $\Sigma_{k \times k}$. If we write the equation (4.31) in matrix notation,

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{k,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & \phi_{12}^1 & \dots & \phi_{1k}^1 \\ \phi_{21}^1 & \phi_{22}^1 & \dots & \phi_{2k}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1}^1 & \phi_{k2}^1 & \dots & \phi_{kk}^1 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \vdots \\ x_{k,t-1} \end{pmatrix} + \dots + \begin{pmatrix} \phi_{11}^p & \phi_{12}^p & \dots & \phi_{1k}^p \\ \phi_{21}^p & \phi_{22}^p & \dots & \phi_{2k}^p \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1}^p & \phi_{k2}^p & \dots & \phi_{kk}^p \end{pmatrix} \begin{pmatrix} x_{1,t-p} \\ x_{2,t-p} \\ \vdots \\ x_{k,t-p} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \vdots \\ \epsilon_{k,t} \end{pmatrix} \\ + \begin{pmatrix} \theta_{11}^1 & \theta_{12}^1 & \dots & \theta_{1k}^1 \\ \theta_{21}^1 & \theta_{22}^1 & \dots & \theta_{2k}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k1}^1 & \theta_{k2}^1 & \dots & \theta_{kk}^1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ \vdots \\ \epsilon_{k,t-1} \end{pmatrix} + \dots + \begin{pmatrix} \theta_{11}^q & \theta_{12}^q & \dots & \theta_{1k}^q \\ \theta_{21}^q & \theta_{22}^q & \dots & \theta_{2k}^q \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k1}^q & \theta_{k2}^q & \dots & \theta_{kk}^q \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-q} \\ \epsilon_{2,t-q} \\ \vdots \\ \epsilon_{k,t-q} \end{pmatrix}$$

As it is seen there are many parameters to be estimated in a VARMA process. For example if the number of variables in the process is 5, and VARMA(1,1) model is used, then the number of parameters to be estimated would be; for constant term vector (5x1) 5, autoregressive parameter matrix (5x5) 25, moving average parameter matrix (5x5) 25, error term matrix

which is Σ , 5(diagonal)+10(off-diagonal)=15. Since the total number of parameters to be estimated is 70, fitting an appropriate time series model for multivariate autocorrelated data would not be easy with even 5 variables. Still if we assume that the number of variables is 2, $k=2$, the total number of the parameters to be estimated is 13 (2 constant terms, 4 parameters for autoregression matrix, 4 parameters for moving average matrix and 3 parameters for error term matrix). Therefore, for simplicity we will consider VAR (1) process with 2 variables, then the number of parameters to be estimated would be 9 (2 constant terms, 4 autoregressive parameters, 3 error term parameters). The advantage of using VAR model is that it can be estimated easily compare to the VARMA models, but we should remember that we need large sample sizes to have good estimations of parameters.

4.5 VAR(p) Models

In time series analysis, it is generally assumed that the observed values of a variable is dependent on some immediate past value. The vector autoregressive model (VAR) is a popular multivariate time series model as it is easy to apply and interpret. It is an extension form of univariate autoregressive model to multivariate data. The basic p lag vector autoregressive (VAR (p)) model with k variables can be expressed with matrix notation,

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{k,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & \phi_{12}^1 & \cdots & \phi_{1k}^1 \\ \phi_{21}^1 & \phi_{22}^1 & \cdots & \phi_{2k}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1}^1 & \phi_{k2}^1 & \cdots & \phi_{kk}^1 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \\ \vdots \\ x_{k,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \phi_{11}^p & \phi_{12}^p & \cdots & \phi_{1k}^p \\ \phi_{21}^p & \phi_{22}^p & \cdots & \phi_{2k}^p \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k1}^p & \phi_{k2}^p & \cdots & \phi_{kk}^p \end{pmatrix} \begin{pmatrix} x_{1,t-p} \\ x_{2,t-p} \\ \vdots \\ x_{k,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{k,t} \end{pmatrix}$$

or

$$\mathbf{x}_t = \mathbf{c} + \Phi_1 \mathbf{x}_{t-1} + \cdots + \Phi_p \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t \quad (4.32)$$

where, \mathbf{x}_t denotes an $k \times 1$ vector of time series variables at time t , \mathbf{c} is a $k \times 1$ vector of constants, Φ_1, \dots, Φ_p are $k \times k$ autoregressive parameter matrices, $\boldsymbol{\varepsilon}_t$ is a $k \times 1$ vector of error terms which are multivariate normally distributed with zero mean vector and variance covariance matrix $\Sigma_{k \times k}$.

Here we consider the number of variables $k=2$. Bivariate vector autoregressive model for the p -lag is expressed by the following form,

$$x_{1,t} = c_1 + \phi_{11}^1 x_{1,t-1} + \phi_{12}^1 x_{2,t-1} + \phi_{11}^2 x_{1,t-2} + \phi_{12}^2 x_{2,t-2} + \cdots + \phi_{11}^p x_{1,t-p} + \phi_{12}^p x_{2,t-p} + \varepsilon_{1,t}$$

$$x_{2,t} = c_2 + \phi_{21}^1 x_{1,t-1} + \phi_{22}^1 x_{2,t-1} + \phi_{21}^2 x_{1,t-2} + \phi_{22}^2 x_{2,t-2} + \cdots + \phi_{21}^p x_{1,t-p} + \phi_{22}^p x_{2,t-p} + \varepsilon_{2,t}$$

For simplicity, the easiest applicable model in multivariate time series models is bivariate first order vector autoregressive model which has two quality characteristics. The bivariate VAR (1) model can be written as,

$$\mathbf{x}_t = \mathbf{c} + \Phi \mathbf{x}_{t-1} + \epsilon_t \quad (4.33)$$

or in matrix form,

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

or

$$\begin{aligned} x_{1,t} &= c_1 + \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \epsilon_{1,t} \\ x_{2,t} &= c_2 + \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \epsilon_{2,t} \end{aligned}$$

where the autocorrelation coefficient matrix is,

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

constant vector \mathbf{c} is,

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

and the error term vector is

$$\epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

ϵ_t has the multivariate normal distribution with mean vector is zero and covariance matrix is $\Sigma_{k \times k}$

For the stationarity of the process, all eigenvalues (λ_i) of autocorrelation coefficient matrix Φ in a VAR (1) model should be within the unit circle or absolute value of λ_i should be less than one, ($|\lambda_i| < 1$). Now we assume that the all absolute eigenvalues of autocorrelation coefficient matrix Φ less than one, and process variables have finite mean and finite variance. So we can compute the expected value and the covariance matrix of a stationary first order vector autoregressive model (VAR(1)) as in the following.

$$\mathbf{x}_t = \mathbf{c} + \Phi \mathbf{x}_{t-1} + \epsilon_t \quad (4.34)$$

$$E(\mathbf{x}_t) = E(\mathbf{c}) + \Phi E(\mathbf{x}_{t-1}) + E(\epsilon_t) \quad (4.35)$$

$$\boldsymbol{\mu} = \mathbf{c} + \Phi \boldsymbol{\mu} \quad (4.36)$$

$$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1} \mathbf{c} \quad (4.37)$$

where, $\boldsymbol{\mu}$ is the vector of expected values of each variable, \mathbf{I} is the identity matrix, $\boldsymbol{\Phi}$ is the matrix of autocorrelation coefficients, \mathbf{c} is the vector of constant terms. In this study, the mean vector of the multivariate time series is assumed to be zero. Then the covariance matrix of a stationary first order vector autoregressive model (VAR (1)) is computed by using the following equation,

$$\begin{aligned} cov(\mathbf{x}_t) &= cov(\mathbf{c}) + cov(\boldsymbol{\Phi} \mathbf{x}_{t-1}) + cov(\boldsymbol{\epsilon}_t) \\ \boldsymbol{\Gamma}(0) &= \boldsymbol{\Phi}' \boldsymbol{\Gamma}(0) \boldsymbol{\Phi} + \boldsymbol{\Sigma} \end{aligned} \quad (4.38)$$

where, $\boldsymbol{\Gamma}(0)$ is the covariance matrix of the data which have first order vector autoregressive structure, $\boldsymbol{\Phi}$ is the matrix of autocorrelation coefficients, and $\boldsymbol{\Sigma}$ is the covariance matrix of errors. As it is seen from the equation above, covariance of the first order vector autoregressive process is dependent on the autocorrelation coefficients and the covariance matrix of the error terms. Therefore, in this chapter we will see how the changes in these parameters effect the process by using average run length as the performance measure. In the previous chapter, since we assume that the univariate autocorrelated time series is perfectly modeled and control limits are constructed by taking autocorrelation into account, here for the multivariate autocorrelated time series we also assume that the multivariate time series are perfectly modeled and theoretical control limits are used. In addition, for multivariate time series, we know that if the parameters are unknown, Hotelling T-square statistics is dependent on sample mean vector and sample variance covariance matrix, but here we will use the true values instead of sample mean vector and sample covariance matrix by taking autocorrelation into account, which are considered in equations (4.37) and (4.38).

In this chapter, we discuss the effect of autocorrelation in Hotelling T-square control chart based on multivariate autocorrelated raw data which is generated in terms of bivariate first order vector autoregressive structure. The same procedure will be applied for the residuals of bivariate first order vector autoregressive model. Since we consider that the time series will be perfectly modeled, instead of sample mean vector of residuals and sample variance covariance matrix of residuals, respectively we will use zero vector and true covariance matrix of error terms in the calculation of Hotelling T-square statistics for the residuals of first order vector autoregressive model. We will see how the effect of autocorrelation changes for different levels of autocorrelation. Then we will add various levels of shifts in the means of the variables. Shifts will be based on standard deviation unit. The comparison between the ARLs which are obtained by using the Hotelling T-square chart based on raw data and ARLs obtained by using the Hotelling T-square chart based on residuals of first order vector autoregressive model will depend on the combination of autocorrelation level of each variable and the amount of shift in the mean of each variable. Lastly, we will add correlation between the errors, and then we will see how the correlation in the errors affects the autocorrelated process in the case of various amount of shifts in the process mean.

At first, we will look at the Phase I data in which the process is assumed to be in control by considering different levels of autocorrelations in the variables and the correlation between the error terms. The following VAR (1) model is used,

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \begin{pmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{pmatrix}$$

The eigenvalues of Φ matrix should be within the unit circle or absolute value of eigenvalues should be less than one, and the error terms are generated as multivariate normally distributed with mean vector is zero, and covariance matrix of error terms is

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

As it is seen correlation between the error terms, $cov(e_{x_1}, e_{x_2}) = 0$.

The Table 8 shows the averages of 1000 run lengths obtained by the use of Hotelling T-square control charts based on generated bivariate data under the column 'RAW' and the averages of 1000 run lengths obtained by the use of Hotelling T-square control charts based on residuals from VAR (1) model under the column 'RESIDUAL' when the $cov(e_{x_1}, e_{x_2}) = 0$. The ARLs based on raw data and ARLs based on residuals with $\alpha = 0.0027$ false alarm rate are considered with different autocorrelation levels,

$$\phi_{11} = \pm 0.25, \pm 0.5, \pm 0.75, \pm 0.95$$

and

$$\phi_{22} = \pm 0.25, \pm 0.5, \pm 0.75, \pm 0.95$$

ϕ_{11}	$\delta_{x_1} = 0$ $\delta_{x_2} = 0$			ϕ_{11}	$\delta_{x_1} = 0$ $\delta_{x_2} = 0$		
	ϕ_{22}	RAW	RESIDUAL		ϕ_{22}	RAW	RESIDUAL
0	0	368.68	368.68	0	0	364.49	364.49
	0.25	385.76	381.29		-0.25	364.07	349.34
	0.5	363.57	351.99		-0.5	376.21	373.45
	0.75	405.61	361.57		-0.75	393.14	377.26
	0.95	569.25	372.35		-0.95	530.19	371.17
0.25	0	360.70	362.23	-0.25	0	374.85	370.57
	0.25	378.86	372.68		-0.25	388.63	366.66
	0.5	397.50	366.35		-0.5	394.80	366.90
	0.75	401.05	371.12		-0.75	404.25	363.63
	0.95	542.44	370.12		-0.95	537.02	376.70
0.5	0	362.19	362.81	-0.5	0	384.76	372.28
	0.25	383.23	374.94		-0.25	355.44	359.78
	0.5	392.01	373.61		-0.5	410.63	373.95
	0.75	430.59	360.52		-0.75	427.87	357.88
	0.95	570.53	354.54		-0.95	597.93	383.69
0.75	0	396.91	366.55	-0.75	0	409.01	351.91
	0.25	425.15	367.28		-0.25	401.58	386.63
	0.5	412.15	364.29		-0.5	433.69	354.77
	0.75	486.20	373.16		-0.75	477.89	355.98
	0.95	677.14	380.79		-0.95	695.39	371.20
0.95	0	553.34	374.77	-0.95	0	537.86	361.41
	0.25	549.86	383.66		-0.25	556.60	368.79
	0.5	585.05	363.55		-0.5	587.20	358.18
	0.75	712.71	360.16		-0.75	677.72	383.43
	0.95	1234.46	358.40		-0.95	1236.99	373.02

Table 8 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase I for different autocorrelation levels and various magnitudes of shifts

As it is seen from the Table 8, ARLs obtained by the use of the Hotelling T-square control charts based on raw data increases while the absolute value of autocorrelation level in any variable increase when the autocorrelation level of other variable is fixed. The ARLs obtained by the use of Hotelling T-square control charts based on residuals for different autocorrelation levels are close to 370 which is ARL value of in control process when the false alarm rate is 0.0027. In Phase I studies, using the Hotelling T-square control chart based on residuals is effective when reducing or removing time dependency from the process.

Now we will check whether the correlation between the error terms effects the average run lengths in Phase I. To do this we just change the off-diagonal element in variance covariance matrix of error term, $cov(e_{x_1}, e_{x_2}) = 0.9$ which is a high level of correlation between the error terms.

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

	$\delta_{x_1} = 0$ $\delta_{x_2} = 0$				$\delta_{x_1} = 0$ $\delta_{x_2} = 0$		
ϕ_{11}	ϕ_{22}	RAW	RESIDUAL	ϕ_{11}	ϕ_{22}	RAW	RESIDUAL
0	0	356.80	356.80	0	0	393.67	393.67
	0.25	368.15	360.30		-0.25	379.36	365.41
	0.5	393.16	367.78		-0.5	426.99	375.21
	0.75	485.57	376.79		-0.75	462.97	379.25
	0.95	608.90	377.37		-0.95	588.41	370.35
0.25	0	377.55	361.74	-0.25	0	368.79	359.11
	0.25	378.21	376.52		-0.25	387.51	377.39
	0.5	401.37	382.97		-0.5	391.59	367.01
	0.75	452.36	362.85		-0.75	423.80	374.53
	0.95	601.96	354.23		-0.95	592.51	349.05
0.5	0	420.39	367.09	-0.5	0	414.47	368.20
	0.25	385.06	354.46		-0.25	398.70	373.38
	0.5	397.70	355.67		-0.5	384.47	347.45
	0.75	435.11	351.50		-0.75	473.30	362.67
	0.95	601.08	368.95		-0.95	644.51	376.06
0.75	0	461.03	363.49	-0.75	0	465.64	360.40
	0.25	491.86	370.35		-0.25	462.94	369.84
	0.5	455.92	357.35		-0.5	452.54	372.69
	0.75	485.87	380.54		-0.75	471.18	388.43
	0.95	739.39	357.98		-0.95	773.12	374.92
0.95	0	560.02	366.76	-0.95	0	588.82	343.38
	0.25	586.66	387.45		-0.25	636.32	369.09
	0.5	658.71	371.18		-0.5	632.30	377.76
	0.75	769.38	385.35		-0.75	766.97	369.11
	0.95	1172.95	389.74		-0.95	1224.63	374.06

Table 9 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase I for different autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$

Table 9 shows the ARLs obtained by the use of Hotelling T-square control charts based on raw data and the residuals with $cov(e_{x_1}, e_{x_2}) = 0.9$. Although we consider $cov(e_{x_1}, e_{x_2}) = 0.9$, the trend in the average run lengths given in Table 9 for different autocorrelation levels within the variables are similar to the average run lengths values in Table 8. Here we see that in the case of considering autocorrelation only within the variables or in other words when the off-diagonal elements are zero in the autocorrelation coefficient matrix, the correlation among the error terms does not effect the average run lengths significantly in terms of the use of Hotelling T-square control charts applied to raw data or residuals from VAR (1) model.

To make our study comparable to previous chapter in which univariate time series, AR (1), and residuals of AR(1) model with a change in the mean is considered, we add different amounts of shifts to each variable of bivariate first order autoregressive process.

$$\mu^* = \mu + \delta\sigma_x \tag{4.39}$$

$$\begin{pmatrix} \mu_{x_1}^* \\ \mu_{x_2}^* \end{pmatrix} = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix} + \begin{pmatrix} \delta_{x_1} \sigma_{x_1} \\ \delta_{x_2} \sigma_{x_2} \end{pmatrix}$$

where μ_{x_1} and μ_{x_2} are equal zero since we assume that the expected value of variables in phase I is zero, δ_{x_1} and δ_{x_2} are the amounts of shifts for each variable, σ_{x_1} and σ_{x_2} are the

standard deviations of each variable. As it is seen from the equation (4.39), shift is considered in standard deviation unit.

Here we will show how the shifts in the means are caught by the Hotelling T-square control charts. In the literature there is not enough theoretical analysis of how the shifted mean effects ARLs obtained by the use of Hotelling T-square control charts based on raw data and the residuals from VAR models. The amount of shifts in standard deviation unit considered for each variable are,

$$\delta_{x_1} = 0.5, 1, 2, 3 \text{ and } \delta_{x_2} = 0.5, 1, 2, 3$$

Table 10 shows the ARLs obtained by the use of Hotelling T-square control chart based on raw data and the residuals from the VAR (1) model when at least one of mean of the variable is shifted to a new value.

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	204.61	204.61	127.37	127.37	65.20	65.20
	0.25	198.04	196.58	136.14	157.84	67.23	67.12
	0.5	207.94	205.48	139.78	172.95	69.39	65.14
	0.75	207.55	192.45	151.58	191.80	71.43	66.09
	0.95	253.66	201.09	217.47	185.81	77.94	68.80
0.25	0	203.31	247.01	126.78	151.34	70.77	106.55
	0.25	210.01	245.46	130.75	186.05	73.62	110.49
	0.5	205.41	244.32	139.46	199.90	67.92	106.94
	0.75	229.09	265.21	159.39	219.36	75.01	116.18
	0.95	278.20	245.15	208.41	235.16	86.12	114.58
0.5	0	206.60	273.43	136.27	166.25	80.01	159.39
	0.25	218.23	299.39	136.30	202.78	75.81	165.26
	0.5	225.53	295.49	146.98	249.69	76.86	167.55
	0.75	251.09	290.06	158.98	263.48	81.34	160.51
	0.95	282.07	294.94	226.45	283.34	93.53	169.73
0.75	0	262.03	342.03	152.06	187.03	104.56	245.41
	0.25	249.87	339.92	164.41	235.18	99.67	242.94
	0.5	268.35	348.12	172.10	267.58	97.45	240.37
	0.75	274.86	326.87	184.68	289.61	108.38	249.04
	0.95	362.41	327.68	296.39	298.21	128.27	232.25
0.95	0	395.95	343.50	219.27	206.13	207.90	172.21
	0.25	412.78	337.71	213.08	229.25	206.83	160.54
	0.5	426.25	351.96	238.98	281.11	205.16	180.22
	0.75	482.63	341.15	272.75	319.46	239.99	187.31
	0.95	826.36	349.54	588.26	303.71	339.04	186.15

Table 10 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different positive autocorrelation levels and various magnitudes of shifts

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	51.41	51.41	27.02	27.02	9.07	9.07
	0.25	54.03	57.63	27.74	36.29	9.49	9.38
	0.5	53.64	64.25	29.92	46.77	9.61	9.51
	0.75	57.66	64.70	35.49	54.62	9.87	9.57
	0.95	72.01	65.82	50.02	29.06	10.27	9.41
0.25	0	53.17	77.47	28.34	37.06	10.49	20.63
	0.25	54.20	90.53	29.20	52.23	10.53	19.78
	0.5	56.37	101.58	31.18	68.77	10.71	19.75
	0.75	59.45	108.32	37.74	88.97	10.68	20.41
	0.95	74.14	103.73	55.06	50.85	11.22	18.24
0.5	0	56.17	104.37	29.35	45.94	12.46	43.50
	0.25	59.74	118.55	31.32	69.24	12.55	39.41
	0.5	62.08	143.25	33.68	95.27	13.14	41.40
	0.75	69.32	163.17	42.81	128.86	13.12	40.28
	0.95	81.97	154.43	60.16	73.95	14.74	39.93
0.75	0	74.67	145.55	36.07	52.15	17.78	67.79
	0.25	78.83	169.41	39.02	84.19	17.64	68.95
	0.5	77.62	192.46	40.42	123.91	18.66	74.47
	0.75	89.04	222.96	52.66	160.13	18.99	67.94
	0.95	123.25	210.64	74.99	97.26	21.44	69.30
0.95	0	132.07	114.96	52.70	32.71	47.60	1.15
	0.25	124.29	119.46	57.81	53.60	48.48	1.32
	0.5	145.40	148.37	63.01	76.17	52.28	1
	0.75	174.68	142.36	83.14	92.22	52.23	1
	0.95	280.61	131.12	173.13	46.75	64.50	1.48

Table 10 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	9.29	8.29	6.50	6.50	2.96	2.95
	0.25	9.24	8.34	6.59	7.18	3.29	3.91
	0.5	9.56	8.73	6.57	7.47	3.71	4.67
	0.75	10.67	9.22	7.25	7.66	4.04	3.96
	0.95	11.11	7.72	7.99	3.87	4.93	1
0.25	0	10.40	17.25	7.04	11.03	3.14	3.78
	0.25	10.15	16.98	7.40	14.66	3.40	5.32
	0.5	10.27	16.47	7.64	14.98	3.60	6.62
	0.75	11.27	19.36	8.07	16.06	4.18	6.54
	0.95	12.81	17.61	9.40	7.87	5.28	1.01
0.5	0	12.70	30.51	8.26	18.01	3.45	4.25
	0.25	12.32	32.94	8.10	22.23	3.76	6.99
	0.5	12.45	35.74	8.72	27.77	4.06	10.02
	0.75	13.08	37.95	10.10	31.41	4.80	10.41
	0.95	16.42	33.32	11.55	12.39	6.27	1
0.75	0	18.26	47.28	11.31	21.58	4.06	3.81
	0.25	17.14	52.86	11.10	30.89	4.32	6.44
	0.5	18.80	61.03	12.48	43.08	5.06	10.03
	0.75	20.09	68.02	13.43	45.68	5.95	10.56
	0.95	23.48	49.56	16.96	13.21	8.70	1
0.95	0	40.04	1	19.69	1.03	4.89	1
	0.25	42.69	1.03	20.35	1	5.49	1
	0.5	46.44	1.75	24.81	1.23	6.57	1.01
	0.75	54.86	1.06	29.61	1	9.19	1
	0.95	77.06	1	49.71	1	19.96	1

Table 10 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$		$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	2.53	2.53	2.43	2.43	2.29	2.29
	0.25	2.55	2.57	2.52	2.55	2.24	2.24
	0.5	2.63	2.64	2.35	2.33	2.31	2.31
	0.75	2.61	2.54	2.55	2.51	2.16	2.07
	0.95	2.47	2.41	2.56	2.10	2.33	1.33
0.25	0	2.87	4.23	2.77	3.96	2.32	3.04
	0.25	2.76	4.09	2.76	4.07	2.41	3.34
	0.5	2.91	4.40	2.81	4.13	2.36	3.33
	0.75	2.94	4.50	2.68	3.76	2.52	3.42
	0.95	2.94	4.38	2.78	3.37	2.75	1.83
0.5	0	3.31	7.20	3.31	6.31	2.68	4.19
	0.25	3.43	7.69	3.12	6.32	2.84	5.04
	0.5	3.33	7.32	3.36	6.22	2.85	5.27
	0.75	3.55	7.11	3.38	7.47	2.90	5.34
	0.95	3.50	7.55	3.33	5.62	3.01	2.16
0.75	0	4.68	6.02	4.34	4.13	3.71	3.32
	0.25	5.02	5.84	4.59	5.81	3.74	3.58
	0.5	4.95	6.48	4.84	6.28	3.44	4.41
	0.75	4.70	6.29	4.76	5.62	3.97	3.74
	0.95	5.23	6.63	4.98	4.05	4.31	1.31
0.95	0	12.15	1	10.78	1	6.33	1
	0.25	12.43	1	9.76	1	7.15	1
	0.5	11.95	1	10.83	1	8.05	1
	0.75	14.72	1	12.99	1	8.99	1
	0.95	17.06	1	15.02	1	11.99	1

Table 10 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$		$\delta_{x_1} = 3$ $\delta_{x_2} = 3$	
		RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.59	1.59	1.19	1.19
	0.25	1.57	1.58	1.24	1.25
	0.5	1.63	1.61	1.28	1.21
	0.75	1.71	1.34	1.25	1.03
	0.95	1.80	1	1.29	1
0.25	0	1.69	1.85	1.24	1.23
	0.25	1.68	2.01	1.21	1.26
	0.5	1.75	1.95	1.27	1.26
	0.75	1.78	1.6	1.29	1.04
	0.95	1.79	1.01	1.32	1
0.5	0	1.8	1.81	1.19	1.11
	0.25	1.79	2.25	1.24	1.27
	0.5	1.91	2.68	1.28	1.25
	0.75	1.95	2.27	1.31	1.04
	0.95	2.18	1	1.44	1
0.75	0	1.90	1.29	1.21	1.04
	0.25	2.04	1.62	1.29	1.08
	0.5	2.23	1.70	1.37	1.06
	0.75	2.53	1.19	1.42	1.01
	0.95	2.77	1	1.66	1
0.95	0	2.42	1	1.28	1
	0.25	2.61	1	1.35	1
	0.5	3.08	1	1.45	1
	0.75	3.99	1	1.61	1
	0.95	6.50	1	2.41	1

Table 10 Continued

In Table 10 we show how the average run lengths change in the combination of different autocorrelation level and different magnitudes of the shifts which are considered in standard deviation unit in the process mean. In the comparison of ARLs obtained by the use of Hotelling T-square control chart based on raw data and the Hotelling T-square control chart based on residuals from the VAR (1) model, the lower ARLs obtained by the use of Hotelling T-square control chart based on residuals than ARLs obtained by the use of Hotelling T-square control charts based on raw data are marked with red color.

As it is seen from the Table 10, Hotelling T-square control charts with the residuals shows better performance when the ϕ_{11} and ϕ_{22} are larger than 0.75 for all magnitudes of shifts. We can see the same interpretation was valid for the univariate autocorrelated chart in the previous chapter. In Table 7 in chapter 3, when the autocorrelation level of variable is larger than 0.75 for all magnitudes of shifts, X-chart based on the residuals from the first order autoregressive model detect the shift earlier than the X-chart based on the raw data. Therefore, we can say that if both of the variables have high autocorrelation level such as 0.95, or the first and the second eigenvalues of autocorrelation matrix is 0.95, then the Hotelling T-square control charts based on residual statistics can detect the shift earlier than Hotelling T-square charts based on raw data, or in other words, out of control ARLs obtained by the use Hotelling T-square control chart based on residuals is less than that of ARLs obtained by the use of Hotelling T-square control chart based on raw data when at least one of the process variable has standard deviation unit shift in the process mean.

Another result from Table 10 is that if any autocorrelation level of any variable is as high as 0.95 and any of the variable has at least 2 standard deviation unit shift in the process mean, then the Hotelling T-square chart based on residual statistics performs better than the Hotelling T-square charts based on raw data. If one of the variables has no autocorrelation, then the Hotelling T-square control charts based on residual statistics shows better performance for all combination of autocorrelation level and the amount of shift for the second variable. When both variables have at least 2 standard deviation unit shift, or one of the variable has at least 3 standard deviation unit shift and the other has at least 0.5 standard deviation unit shift in the process mean, then the Hotelling T-square control chart based on the residual performs well if the one of the variable has not autocorrelation and the other has high autocorrelation level such as 0.75 and 0.95.

Finally we observe that if the shift in standard deviation unit for both variables is as high as 3, then the Hotelling T-square control chart based on residual performs well to detect the shift in the process mean almost for all combinations of moderate and high autocorrelation levels. This result may also be seen for the univariate autocorrelated process in chapter 3 in which when the process shift is 3 standard deviation unit, residual chart performs well.

Now we will see how the correlation between the error terms effects the average run lengths in Phase II when the process mean shifted to a new value. As we consider in Phase I, we assume the correlation level between the variables as 0.9, i.e. $cov(e_{x_1}, e_{x_2}) = 0.9$

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

Table 11 with the correlation effect between the error terms can be seen with different autocorrelation levels and different amounts of shift in the process mean.

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	50.09	50.090	186.50	186.50	5.60	5.60
	0.25	65.29	48.64	192.86	194.22	9.29	6.25
	0.5	113.25	49.30	193.86	157.07	20.21	6.06
	0.75	176.11	47.44	223.05	110.52	43.86	5.68
	0.95	266.69	51.52	242.08	49.12	69.64	6.00
0.25	0	69.79	85.17	193.41	181.54	9.71	11.82
	0.25	49.37	82.78	202.31	250.22	6.64	11.66
	0.5	78.29	90.48	211.20	251.88	11.72	11.68
	0.75	146.74	85.56	242.18	189.20	32.05	11.50
	0.95	262.05	87.06	283.03	93.02	68.59	11.95
0.5	0	125.71	140.46	206.42	146.96	26.77	25.66
	0.25	81.71	135.35	209.86	227.12	13.55	25.41
	0.5	63.21	143.41	208.04	303.82	8.66	26.62
	0.75	104.40	139.32	263.27	265.65	19.19	23.98
	0.95	268.59	143.66	322.85	156.29	69.80	24.00
0.75	0	235.74	219.60	209.28	106.51	71.87	36.42
	0.25	191.89	219.65	239.22	184.69	54.90	44.99
	0.5	131.85	202.24	273.92	266.44	28.17	44.73
	0.75	84.74	203.45	275.67	325.23	11.74	39.04
	0.95	276.25	214.92	404.86	244.96	67.85	39.71
0.95	0	411.83	118.96	243.89	50.87	200.62	1.00
	0.25	451.10	131.94	273.93	90.32	198.15	1.00
	0.5	425.47	115.89	313.60	148.61	183.89	1.00
	0.75	428.22	128.57	420.92	237.97	157.42	1.00
	0.95	277.13	118.87	760.32	318.87	45.31	1.00

Table 11 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different positive autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	32.02	32.02	67.93	67.93	1.10	1.10
	0.25	38.20	21.72	68.30	65.03	1.26	1.10
	0.5	56.59	15.49	69.10	40.64	2.09	1.09
	0.75	67.90	11.97	65.24	19.15	4.55	1.11
	0.95	79.24	7.72	73.66	1.18	8.80	1.10
0.25	0	41.42	82.02	61.60	59.89	1.33	1.11
	0.25	32.20	57.76	65.77	104.22	1.10	1.09
	0.5	44.83	41.88	72.22	99.59	1.36	1.12
	0.75	67.78	27.89	77.38	52.94	3.18	1.12
	0.95	85.11	19.03	78.71	1.23	8.75	1.12
0.5	0	74.66	171.74	68.54	38.64	2.69	1.08
	0.25	53.96	133.37	69.69	98.69	1.51	1.08
	0.5	37.67	100.31	71.93	162.97	1.13	1.05
	0.75	58.24	75.34	84.46	129.75	2.11	1.05
	0.95	93.72	46.61	96.13	3.68	9.16	1.07
0.75	0	116.87	161.78	67.42	20.19	9.20	1.00
	0.25	113.59	185.06	76.56	50.82	5.58	1.00
	0.5	92.28	190.76	86.77	138.58	2.61	1.00
	0.75	55.24	173.38	102.27	237.08	1.21	1.00
	0.95	118.06	119.93	150.11	19.70	7.65	1.00
0.95	0	182.67	1.26	68.78	1.10	43.09	1.00
	0.25	194.24	1.14	81.43	1.46	42.73	1.00
	0.5	224.09	1.03	96.97	3.46	33.83	1.00
	0.75	273.37	2.35	140.46	20.85	22.63	1.00
	0.95	182.49	54.52	334.50	168.21	1.67	1.00

Table 11 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.64	1.64	3.69	3.69	8.52	8.52
	0.25	2.04	1.56	4.25	2.74	8.98	8.25
	0.5	3.71	1.58	6.63	2.38	9.40	4.92
	0.75	6.94	1.68	8.81	2.23	8.96	2.01
	0.95	9.62	2.02	10.27	1.40	7.71	1.00
0.25	0	2.38	2.33	5.07	9.40	9.17	8.04
	0.25	1.79	2.10	3.92	6.17	9.47	18.29
	0.5	2.37	1.96	4.64	4.47	10.37	15.74
	0.75	5.57	2.14	8.08	3.69	10.03	4.30
	0.95	10.14	2.49	11.90	1.77	10.29	1.00
0.5	0	5.88	3.85	11.00	20.52	8.88	4.54
	0.25	2.85	2.99	7.63	17.77	10.30	15.09
	0.5	1.90	2.70	5.00	12.81	11.13	35.50
	0.75	4.07	3.07	7.14	8.21	13.29	17.71
	0.95	11.43	4.66	14.02	3.52	12.93	1.00
0.75	0	18.40	1.74	20.92	3.81	8.66	1.92
	0.25	12.08	1.29	19.72	8.15	10.67	4.50
	0.5	6.08	1.25	15.29	11.23	13.34	19.00
	0.75	2.48	1.63	7.26	14.70	17.82	65.16
	0.95	12.19	6.06	18.19	14.85	21.71	1.00
0.95	0	54.19	1.00	35.14	1.00	7.97	1.00
	0.25	56.40	1.00	36.94	1.00	9.24	1.00
	0.5	56.57	1.00	48.66	1.00	11.66	1.00
	0.75	46.35	1.00	61.66	1.00	19.61	1.00
	0.95	7.09	1.00	23.99	1.00	64.24	1.00

Table 11 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$		$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.00	1	1.00	1.004	1.05	1.05
	0.25	1.00	1	1.03	1.00	1.15	1.06
	0.5	1.08	1	1.24	1.00	1.44	1.09
	0.75	1.48	1	1.89	1.01	2.09	1.23
	0.95	2.32	1	2.60	1.27	2.48	1.50
0.25	0	1.00	1	1.03	1.00	1.17	1.06
	0.25	1.00	1	1.00	1.00	1.06	1.06
	0.5	1.00	1	1.03	1.00	1.20	1.07
	0.75	1.22	1	1.49	1.01	1.96	1.17
	0.95	2.29	1	2.63	1.21	2.70	1.51
0.5	0	1.08	1	1.33	1.00	2.01	1.00
	0.25	1.00	1	1.04	1.00	1.27	1.01
	0.5	1.00	1	1.00	1.00	1.07	1.02
	0.75	1.03	1	1.14	1.00	1.43	1.10
	0.95	2.35	1	2.87	1.09	3.13	1.65
0.75	0	2.09	1	3.35	1.00	4.55	1.00
	0.25	1.43	1	2.12	1.00	3.44	1.00
	0.5	1.05	1	1.18	1.00	1.77	1.00
	0.75	1.00	1	1.00	1.00	1.09	1.00
	0.95	1.91	1	2.37	1.00	3.37	1.23
0.95	0	10.42	1	12.73	1.00	11.30	1.00
	0.25	8.76	1	12.80	1.00	12.39	1.00
	0.5	6.62	1	11.48	1.00	13.43	1.00
	0.75	2.93	1	6.69	1.00	11.96	1.00
	0.95	1.00	1	1.04	1.00	1.25	1.00

Table 11 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$		$\delta_{x_1} = 3$ $\delta_{x_2} = 3$	
		RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.78	1.78	2.40	2.40
	0.25	1.92	1.61	2.49	2.26
	0.5	2.26	1.59	2.41	1.59
	0.75	2.61	1.56	2.14	1.02
	0.95	2.32	1.00	1.70	1.00
0.25	0	2.25	2.92	2.46	2.25
	0.25	1.96	2.38	2.52	3.46
	0.5	2.20	2.15	2.74	3.04
	0.75	2.70	1.77	2.62	1.06
	0.95	2.82	1.00	2.07	1.00
0.5	0	3.35	2.28	2.34	1.60
	0.25	2.94	3.94	2.67	2.91
	0.5	2.26	3.47	3.10	6.00
	0.75	2.69	2.77	3.31	1.72
	0.95	3.52	1.00	2.76	1.00
0.75	0	4.19	1.00	2.23	1.03
	0.25	4.64	1.00	2.60	1.07
	0.5	4.80	1.03	3.24	1.55
	0.75	3.04	2.39	4.24	5.52
	0.95	4.70	1.00	4.39	1.00
0.95	0	4.00	1.00	1.69	1.00
	0.25	4.85	1.00	2.03	1.00
	0.5	7.12	1.00	2.63	1.00
	0.75	11.01	1.00	4.74	1.00
	0.95	8.82	1.00	13.48	1.00

Table 11 Continued

In Table 11, firstly, we see that the number of ARLs in red color increases when we add the correlation between the error terms, so we can say that if the error terms in one variable are highly correlated with the error terms of other variable, the detection capability of Hotelling T-square control chart based on residual statistics increases. When there is a some amount of difference between the shift such as at least 1.5 standard deviation unit difference, and the error terms are highly correlated, $cov(e_{x_1}, e_{x_2}) = 0.9$, then we can say that the Hotelling T-square control chart based on the residual statistics works well to detect the shift than Hotelling T-square charts based on raw data.

In the previous chapter, when the autocorrelation level is negative, the detection capability of X-chart based residuals was better than the X-chart based on raw data. Here we can see the same interpretation for all combination of negative autocorrelation levels and the amount of shifts. The out of control ARL of by the use of Hotelling T-square control chart based on residual statistics for the first order vector autoregressive process with $\phi_{11} = -0.25, -0.5, -0.75, -0.95$ and $\phi_{22} = -0.25, -0.5, -0.75, -0.95$ is smaller than the ARLs of Hotelling T-square control chart based on raw data. The tables with negative autocorrelation levels can be seen appendix in Tables A.1 and A.2. When the amount of shift is low and the negative autocorrelation level is high, the detection capability of Hotelling T-square control chart based on residuals is significantly better than the detection capability of Hotelling T-square control charts based on raw data. For example, $\delta_{x_1} = 0.5$ and $\delta_{x_2} = 0$, the Hotelling T-square control chart based on raw data detect the shift at 540 while Hotelling T-square control chart based on residual detect the shift at 3.204. ARLs in each combination of autocorrelation level and shift show that the residual chart is better. If the correlation coefficient between the error terms is high and the variables are negatively autocorrelated, when the variables have same amount of shift, the ARLs for each combination of autocorrelation level in the variables increases, compared to the tables without any correlation between the error terms. For example when $\delta_{x_1} = 0.5$ and $\delta_{x_2} = 0.5$ the ARLs of each combination of autocorrelation level in Table 12 is lower than the ARLs of Table 11 in which $\delta_{x_1} = 0.5$ and $\delta_{x_2} = 0.5$ conversely, when the variables have different amounts of shifts, ARLs decrease.

Chapter 5

Hotelling T-square Statistics on Data Matrix with Lagged Variables

Mason and Young (2002) suggest that the relationship between the process variables requires adding additional lag variables to the historical data since the observation of one variable at time t may be dependent to previous observations of other variables. For example suppose that the process has two variables $x_{1,t}$ and $x_{2,t}$ where $t=1,2,\dots,n$, and the process has first order vector autoregressive procedure in which $x_{1,t}$ and $x_{2,t}$ are,

$$x_{1,t} = c_1 + \phi_{11}x_{1,t-1} + \phi_{12}x_{2,t-1} + \varepsilon_{1,t}$$

$$x_{2,t} = c_2 + \phi_{21}x_{1,t-1} + \phi_{22}x_{2,t-1} + \varepsilon_{2,t}$$

As it is seen, variable $x_{1,t}$ is dependent on the previous value of itself, $x_{1,t-1}$, and previous value of other variable $x_{2,t-1}$. Similarly $x_{2,t}$ has relationship with $x_{1,t-1}$ and $x_{2,t-1}$. Therefore, according to Mason and Young (2002) the dataset should be reconstructed in the form of

$$\begin{bmatrix} x_{1,t} & x_{1,t-1} & x_{2,t} & x_{2,t-1} \end{bmatrix}$$

where $t=2,\dots,n$. For the higher order autoregressive relationships, more lag variables can be added to the dataset.

To see the effect of these time-lagged variables on the Hotelling T-square control chart, Mason and Young (2002) compare the T-square statistics without and with lag variables in the example of Reactor data. They decide the T-square statistics with lagged variables are more sensitive than the T-square statistics without lagged variables. They show that the T-square statistics with lagged variables perform well in signal detection.

In this chapter, we will work on the performance of reconstructed data with lagged variables. We will see the effect of these time lagged variables on T-square control procedure by using the average run length performance tool. Same combinations of autocorrelation levels and the amount of shifts will be considered for each variable. First we consider the Phase I in which each variables has first order autoregressive procedure. The data vector is reconstructed with one lagged variables as following,

$$\begin{bmatrix} x_{1,t} & x_{1,t-1} & x_{2,t} & x_{2,t-1} \end{bmatrix}$$

The Table 12 shows the average run lengths for each combination of autocorrelation level for two variables in Phase I in which the process is in control.

$\delta_{x_1} = 0$ $\delta_{x_2} = 0$			$\delta_{x_1} = 0$ $\delta_{x_2} = 0$		
ϕ_{11}	ϕ_{22}	LAGGED	ϕ_{11}	ϕ_{22}	LAGGED
0	0	460.48	0	0	462.87
	0.25	483.09		-0.25	481.16
	0.5	438.77		-0.5	448.08
	0.75	459.08		-0.75	467.09
	0.95	454.40		-0.95	492.32
0.25	0	455.42	-0.25	0	465.41
	0.25	440.91		-0.25	466.39
	0.5	464.28		-0.5	449.19
	0.75	449.44		-0.75	457.05
	0.95	503.96		-0.95	506.79
0.5	0	452.72	-0.5	0	449.01
	0.25	461.61		-0.25	454.95
	0.5	483.30		-0.5	464.65
	0.75	483.94		-0.75	463.36
	0.95	492.69		-0.95	462.39
0.75	0	467.25	-0.75	0	457.25
	0.25	461.24		-0.25	486.82
	0.5	446.66		-0.5	468.22
	0.75	460.41		-0.75	489.94
	0.95	515.36		-0.95	504.49
0.95	0	461.59	-0.95	0	492.48
	0.25	485.05		-0.25	509.41
	0.5	488.31		-0.5	511.94
	0.75	506.58		-0.75	498.96
	0.95	501.17		-0.95	510.26

Table 12 Comparison of the ARL obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase I with different autocorrelation levels

As it is seen from the Table 12, when the autocorrelation level increases, the average run length increases or in other words, false alarm rate of process decreases. While standard average run length values are 370 with 0.0027 false alarm rate, here with the lagged variables, this value increases. The reason of having these higher ARLs when the process is in control is the correlation which is occurred in the calculation of T-square statistics with lagged variables. If we focus on T-square calculation with lagged variables;

$$T_t^2 = (\mathbf{x}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}) \quad (5.1)$$

For identity matrix, we have

$$T_t^2 = (\mathbf{x}_t - \boldsymbol{\mu})^T (\mathbf{x}_t - \boldsymbol{\mu}) \quad (5.2)$$

where, $t=2, \dots, n$

$$T_t^2 = \begin{bmatrix} x_{1,t} - \mu_{x_1} & x_{1,t-1} - \mu_{x_1} & x_{2,t} - \mu_{x_2} & x_{2,t-1} \end{bmatrix} \begin{bmatrix} x_{1,t} - \mu_{x_1} \\ x_{1,t-1} - \mu_{x_1} \\ x_{2,t} - \mu_{x_2} \\ x_{2,t-1} - \mu_{x_2} \end{bmatrix}$$

$$T_2^2 = (x_{1,2} - \mu_{x_1})^2 + (x_{1,1} - \mu_{x_1})^2 + (x_{2,2} - \mu_{x_2})^2 + (x_{2,1} - \mu_{x_2})^2$$

$$T_3^2 = (x_{1,3} - \mu_{x_1})^2 + (x_{1,2} - \mu_{x_1})^2 + (x_{2,3} - \mu_{x_2})^2 + (x_{2,2} - \mu_{x_2})^2$$

$$T_4^2 = (x_{1,4} - \mu_{x_1})^2 + (x_{1,3} - \mu_{x_1})^2 + (x_{2,4} - \mu_{x_2})^2 + (x_{2,3} - \mu_{x_2})^2$$

⋮

$$T_{n-1}^2 = (x_{1,t-1} - \mu_{x_1})^2 + (x_{1,t-2} - \mu_{x_1})^2 + (x_{2,t-1} - \mu_{x_2})^2 + (x_{2,t-2} - \mu_{x_2})^2 \tag{5.3}$$

$$T_n^2 = (x_{1,t} - \mu_{x_1})^2 + (x_{1,t-1} - \mu_{x_1})^2 + (x_{2,t} - \mu_{x_2})^2 + (x_{2,t-1} - \mu_{x_2})^2 \tag{5.4}$$

As it seen from the equation 5.3 and 5.4, T_{n-1}^2 and T_n^2 has the same terms which are $(x_{1,t-1} - \mu_{x_1})^2$ and $(x_{2,t-1} - \mu_{x_2})^2$. Each calculation uses the two terms which are used in previous calculation of T-square statistic. So, this is causing the high average run length values in the calculations. To observe how the performance of Hotelling T-square control chart of data with lagged variables in the case of standard deviation unit shift in the process mean with positive autocorrelation level, the results when the process is out of control can be seen in Table 13.

		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	230.79	125.67	59.13	44.18	21.25	5.78	5.56
	0.25	212.35	143.85	58.13	45.93	25.97	6.22	5.16
	0.5	219.05	159.05	57.48	51.13	31.37	6.15	5.63
	0.75	231.49	176.32	58.42	49.31	36.74	6.19	5.92
	0.95	226.45	206.47	63.02	57.00	44.18	6.49	5.70
0.25	0	254.33	152.95	81.75	57.58	25.96	10.35	8.55
	0.25	261.12	165.42	84.69	63.10	32.73	10.29	8.84
	0.5	263.21	179.16	83.11	67.04	39.73	10.05	9.79
	0.75	260.63	206.21	85.70	70.31	47.86	10.28	9.48
	0.95	262.53	248.82	95.07	77.27	64.71	10.11	9.71
0.5	0	301.97	158.83	116.56	74.88	31.30	16.83	13.51
	0.25	299.53	168.44	113.51	80.02	35.86	16.18	15.25
	0.5	294.50	209.42	109.06	93.77	48.14	16.39	15.35
	0.75	296.84	225.95	115.27	97.21	59.04	15.84	15.49
	0.95	316.29	251.77	121.08	107.52	73.34	16.83	16.41
0.75	0	346.34	177.61	147.70	87.96	36.67	27.58	22.97
	0.25	323.85	202.67	151.92	101.09	47.39	29.50	23.34
	0.5	339.06	215.88	147.03	115.85	60.19	26.35	24.72
	0.75	329.08	247.02	154.62	121.46	75.95	28.68	25.81
	0.95	349.95	299.06	166.38	126.85	98.17	28.23	28.25
0.95	0	402.85	219.05	222.85	131.07	44.55	63.36	41.94
	0.25	387.30	241.68	215.72	148.97	59.25	64.44	46.11
	0.5	359.44	251.88	230.44	157.89	79.54	64.19	53.92
	0.75	404.36	291.95	214.79	176.31	99.72	62.84	53.28
	0.95	407.63	333.02	245.04	199.99	141.84	63.01	59.46

Table 13 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different positive autocorrelation levels and various magnitudes of shifts

		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 3$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	4.32	1.83	1.53	1.47	1.34	1.10	1.01
	0.25	4.37	2.27	1.58	1.47	1.42	1.15	1.03
	0.5	4.69	2.60	1.57	1.45	1.35	1.20	1.03
	0.75	5.17	3.18	1.44	1.47	1.43	1.23	1.06
	0.95	5.32	3.74	1.53	1.50	1.43	1.29	1.08
0.25	0	6.32	2.25	2.34	2.14	1.83	1.24	1.03
	0.25	7.09	2.78	2.33	2.02	1.92	1.35	1.05
	0.5	7.58	3.73	2.30	2.03	2.03	1.44	1.07
	0.75	8.05	4.46	2.26	2.21	2.06	1.58	1.11
	0.95	8.33	5.36	2.47	2.29	2.04	1.61	1.26
0.5	0	8.98	2.58	3.38	3.16	2.74	1.43	1.04
	0.25	10.23	3.63	3.71	3.26	2.99	1.63	1.11
	0.5	11.32	4.51	3.63	3.25	2.96	1.71	1.15
	0.75	12.94	5.92	3.67	3.38	3.09	2.11	1.34
	0.95	14.30	8.58	3.70	3.47	3.30	2.12	1.51
0.75	0	12.57	3.03	6.79	5.11	3.95	1.76	1.06
	0.25	15.29	4.43	6.91	5.78	4.27	2.07	1.17
	0.5	17.43	6.01	6.82	6.02	4.47	2.48	1.40
	0.75	20.13	8.76	7.01	6.07	5.40	3.24	1.52
	0.95	23.77	12.72	7.37	6.18	5.77	3.72	1.83
0.95	0	20.32	3.66	18.17	12.12	7.04	1.99	1.08
	0.25	25.92	5.61	17.66	11.43	9.27	2.597	1.28
	0.5	33.49	7.83	18.72	12.83	11.50	3.82	1.42
	0.75	39.57	12.66	16.83	14.09	11.19	5.37	1.94
	0.95	46.74	23.56	17.19	16.06	14.21	8.11	2.70

Table 13 Continued

From the Table 13, the average run lengths with higher autocorrelation levels indicate higher ARLs than the ARLs of lower autocorrelation levels. Also increasing amount of shift leads to decrease in the average run lengths for all different combinations of autocorrelation levels. Now we will consider that there is correlation between the variables such a 0.9 level. Table 15 shows the Hotelling T-square control chart performance for the dataset which have lagged variables with $cov(e_{x_1}, e_{x_2}) = 0.9$. When there is at least 1.5 standard deviation unit difference between the process means, and the variable which has the biggest shift has low autocorrelation level such as 0.25, then the Hotelling T-square chart based on the data matrix with lagged variables may be an alternative to the Hotelling T-square control chart based on raw data to detect the shift early. For example when $\delta_{x_1} = 2$ and $\delta_{x_2} = 0$ and the $\phi_{11} = 0$ or $\phi_{11} = 0.25$, the detection capability of Hotelling T-square control chart based on data matrix with lagged variables is better than the Hotelling T-square control charts based on raw data. Also when the amount of shift for each variable increases, the detection capability of Hotelling T-square control charts based on data matrix with lagged variables gets better compare to the raw data applications. For moderate and low autocorrelation level when $cov(e_{x_1}, e_{x_2}) = 0.9$, the detection capability of Hotelling T-square control chart based on data matrix with lagged variables is better than the Hotelling T-square control chart based on residuals from the VAR (1) model even ARLs are not set to 370.

Hotelling T-square Statistics on Data Matrix with Lagged Variables

		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	42.04	225.91	3.65	22.94	52.22	1.00	1.08
	0.25	45.84	205.08	4.33	19.27	56.11	1.00	1.14
	0.5	60.04	175.82	5.67	15.76	38.98	1.04	1.15
	0.75	62.76	140.73	7.05	14.10	22.45	1.08	1.16
	0.95	73.89	98.50	7.94	9.48	11.30	1.10	1.13
0.25	0	67.17	212.53	7.82	49.58	48.84	1.05	1.80
	0.25	58.94	259.95	6.33	38.01	77.94	1.03	1.298
	0.5	73.41	264.86	7.73	34.18	72.09	1.07	1.49
	0.75	98.75	215.34	12.95	29.90	46.63	1.29	1.77
	0.95	121.90	160.37	16.96	23.53	27.86	1.58	1.71
0.5	0	133.16	184.73	20.17	96.80	36.85	1.74	5.44
	0.25	107.51	245.09	14.54	73.83	77.54	1.26	3.07
	0.5	85.39	296.46	9.85	55.20	105.54	1.08	1.90
	0.75	114.96	264.58	17.57	51.09	98.54	1.46	2.59
	0.95	174.56	228.54	33.42	48.95	57.25	2.74	3.79
0.75	0	235.27	140.97	61.61	131.32	22.93	6.54	21.26
	0.25	211.95	220.32	51.33	136.37	49.15	4.86	14.60
	0.5	169.68	281.68	34.92	118.85	100.85	2.50	8.15
	0.75	118.85	316.75	19.20	82.73	150.26	1.29	2.95
	0.95	235.11	315.14	57.40	96.21	128.56	5.73	8.70
0.95	0	362.72	93.11	192.25	98.73	11.57	42.48	50.31
	0.25	372.79	157.07	198.22	145.30	26.16	40.96	64.50
	0.5	378.45	230.10	175.64	190.30	60.08	33.53	58.70
	0.75	343.95	330.95	145.06	219.24	129.53	21.95	48.87
	0.95	197.47	438.50	45.14	141.76	213.10	2.27	7.87

Table 14 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different positive autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$

		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 3$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	2.00	5.31	1.00	1.00	1.00	1.17	1.37
	0.25	1.72	5.42	1.00	1.00	1.00	1.09	1.38
	0.5	1.53	3.54	1.00	1.00	1.00	1.03	1.15
	0.75	1.41	2.12	1.00	1.00	1.00	1.01	1.01
	0.95	1.19	1.21	1.00	1.00	1.00	1.00	1.00
0.25	0	5.13	5.28	1.00	1.00	1.04	1.92	1.38
	0.25	3.28	9.12	1.00	1.00	1.00	1.50	1.95
	0.5	2.91	8.73	1.00	1.00	1.00	1.32	1.92
	0.75	2.69	4.80	1.00	1.00	1.02	1.15	1.29
	0.95	1.94	2.08	1.00	1.01	1.01	1.02	1.02
0.5	0	13.69	3.41	1.01	1.10	1.74	2.87	1.14
	0.25	9.25	9.01	1.00	1.01	1.17	3.19	1.85
	0.5	5.28	15.10	1.00	1.00	1.03	2.20	3.14
	0.75	5.25	12.47	1.00	1.02	1.10	1.79	2.57
	0.95	4.38	5.14	1.06	1.15	1.18	1.30	1.35
0.75	0	21.35	1.97	1.39	3.07	5.83	2.07	1.02
	0.25	28.19	5.32	1.14	1.91	4.03	4.49	1.29
	0.5	19.84	12.78	1.03	1.15	2.21	5.63	2.75
	0.75	9.95	26.04	1.00	1.00	1.14	4.05	5.72
	0.95	14.00	18.06	1.34	1.53	2.13	3.31	3.23
0.95	0	11.28	1.22	9.36	15.61	7.17	1.14	1.00
	0.25	23.12	2.14	8.77	15.23	11.64	1.76	1.02
	0.5	42.84	5.39	6.35	13.92	14.31	3.94	1.27
	0.75	63.45	17.81	3.35	6.88	13.41	10.71	3.46
	0.95	26.43	61.16	1.00	1.04	1.59	9.73	15.39

Table 14 Continued

When there is high correlation between the variables, generally Hotelling T-square control charts based on residuals is better when the autocorrelation level of one variable is at high level, for the low autocorrelation level Hotelling T-square control chart based on data matrix with lagged variables, even in control ARLs are not set to 370, can be an alternative to the Hotelling T-square control chart based on raw data.

Chapter 6

Comparison of Proposed Methods

The ARL of each combination of autocorrelation level may be deceptive since the in control ARL is not all equal to 370. However, it is important to adjust the control limits to in control ARL to interpret the out of control performance of the proposed charts. In this chapter, in order to obtain in control ARL of 370, the upper control limit is adjusted 370. Datasets are generated based on three different autocorrelation matrices. These datasets are used to compare all the methods we propose until now. Comparison results are presented different combination of the amounts of shifts in the process mean with the covariance of error terms,

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The tables below indicate the performance of proposed methods with 3 different autocorrelation matrices. The first matrix that we consider is

$$\Phi = \begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \quad (6.1)$$

Adjusted limits for that matrix in terms of proposed methods with covariance matrix of error term Σ can be seen in the Table 15a below. In Table 15a, Σ indicates the covariance matrix of error terms of the generated datasets. In this matrix, if the off-diagonal element is different from zero, then it means the error terms of the generated variables are correlated. ‘RAW’ indicates the adjusted upper control limit in order to obtain in control ARL of 370 for the Hotelling T-square control chart based on raw data which have VAR (1) structure. ‘VAR’ shows the theoretical upper control limit for Hotelling T-square control chart based on residuals from VAR (1) model. ‘AR’ indicates the adjusted UCL for Hotelling T-square control chart based on residuals from an AR (1) model fitted to each variable individually. Since fitting AR (1) model to the variables ignores the cross correlation among th variables, UCL limit should be adjusted to 370. ‘LAG’ gives the adjusted upper control limit for Hotelling T-square control chart based on the data matrix with lagged variables. These adjusted upper control limits are considered for the autocorrelation matrix that we considered in (6.1). The results without adjustments in which theoretical UCL is used for all Hotelling T-square control charts mentioned above can be seen in appendix C.

In Table 15b, in the case of different combination of the amounts of the shifts, the values under the column ‘RAW’ indicates the ARLs obtained by using Hotelling T-square control chart based on raw data which have VAR(1) structure with the corresponding adjusted upper control limit, 10.35. ‘VAR’ shows the ARLs obtained by using Hotelling T-square control

chart based on residuals from VAR (1) model while the ‘AR’ shows the ARLs obtained by using Hotelling T-square control chart based on residuals from the individual AR (1) model fitted to each variable. ‘LAG’ mentions the ARLs of Hotelling T-square control chart based on the data matrix with lagged variables. For the above mentioned four cases of Hotelling T-square control charts, upper control limits are adjusted in order to obtain the in control ARL of 370 when the process has no change in the mean. In this chapter the rest of the tables would have the same structure, but the autocorrelation matrix Φ and covariance matrix of error terms Σ would be different.

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	10.35	11.81	12.75	15.56

Table 15a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.1)

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	373.08	362.12	367.20	375.77
	0.5	165.83	309.23	345.20	220.08
	1	47.64	127.37	177.80	75.77
	2	6.18	1.16	1.56	10.58
	3	1.52	1.00	1.00	2.05
0.5	0	147.11	293.28	330.01	209.40
	0.5	260.67	328.43	284.02	292.11
	1	122.09	152.24	138.48	168.85
	2	14.82	1.00	1.14	22.49
	3	2.26	1.00	1.00	3.52
1	0	41.95	158.40	203.64	66.85
	0.5	96.93	183.89	167.20	142.90
	1	144.02	79.19	57.78	177.19
	2	28.01	1.00	1.00	41.08
	3	3.80	1.00	1.00	6.49
2	0	6.32	7.12	18.80	9.48
	0.5	10.54	6.61	10.24	18.21
	1	18.57	2.79	4.22	30.03
	2	33.49	1.00	1.00	46.17
	3	9.61	1.00	1.00	14.87
3	0	1.48	1.00	1.01	1.79
	0.5	1.85	1.00	1.00	2.57
	1	3.07	1.00	1.00	4.89
	2	6.43	1.00	1.00	9.99
	3	7.39	1.00	1.00	10.89

Table 15b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.1)

The Φ matrix used in this case corresponds to high autocorrelation since the first eigenvalue is 0.96 and second eigenvalue is 0.74. When the level of shift is at low level, the shift detection capability of Hotelling T-square control chart based on raw data is better than other proposed methods while Hotelling T-square control chart based on residuals from VAR (1) model is better in catching the shift when the amount of shift is large such as at least 1σ for any of the two variables. The Table 16a shows the adjusted control limits for the proposed methods with $cov(e_{x_1}, e_{x_2}) = 0.9$ and the Table 16b shows the ARLs obtained by the use of

Hotelling T-square control chart based on the four different proposed methods according to the considered Φ matrix in (6.1) with $cov(e_{x_1}, e_{x_2}) = 0.9$.

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$	10.32	11.81	13.14	15.55

Table 16a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.1) with $cov(e_{x_1}, e_{x_2}) = 0.9$

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	388.46	372.2	367.17	385.77
	0.5	15.78	11.03	35.70	23.10
	1	1.48	1.00	1.00	1.66
	2	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00
0.5	0	13.88	37.42	96.41	19.48
	0.5	292.59	325.17	274.78	320.39
	1	14.68	1.00	1.12	22.02
	2	1.00	1.00	1.00	1.00
	3	1.00	1.00	1.00	1.00
1	0	1.55	1.00	1.13	1.68
	0.5	11.71	36.18	65.62	16.70
	1	153.16	42.21	30.14	178.74
	2	1.41	1.00	1.00	1.56
	3	1.00	1.00	1.00	1.00
2	0	1.00	1.00	1.00	1.00
	0.5	1.00	1.00	1.00	1.00
	1	1.24	1.02	1.32	1.27
	2	37.39	1.00	1.00	55.52
	3	1.15	1.00	1.00	1.26
3	0	1.00	1.00	1.00	1.00
	0.5	1.00	1.00	1.00	1.00
	1	1.00	1.00	1.00	1.00
	2	1.11	1.00	1.00	1.11
	3	9.98	1.00	1.00	14.97

Table 16b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.1) with $cov(e_{x_1}, e_{x_2}) = 0.9$

When the autocorrelation is at high level and the correlation between the variables is 0.9, residual based approach is generally better than the other proposed methods.

The second matrix we considered is,

$$\Phi = \begin{pmatrix} 0.3 & 0.2 \\ 0.1 & 0.7 \end{pmatrix} \quad (6.2)$$

This matrix is used in the case corresponds to moderate autocorrelation. Adjusted upper control limits for that matrix in terms of proposed methods are shown in the Table 17a,

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	11.57	11.81	12.45	15.68

Table 17a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.2)

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	376.99	361.12	368.02	371.26
	0.5	215.68	298.38	324.62	245.70
	1	80.07	180.75	244.51	105.03
	2	13.14	40.39	65.41	17.12
0.5	3	3.23	4.03	7.47	3.54
	0	200.36	252.51	256.62	205.48
	0.5	174.37	276.06	227.16	206.47
	1	86.67	241.22	178.60	119.15
1	2	17.54	64.93	57.00	23.38
	3	4.20	6.21	7.20	5.05
	0	65.25	106.01	110.45	68.75
	0.5	73.87	146.59	103.94	88.24
2	1	53.52	155.59	86.41	70.96
	2	16.12	66.79	32.93	23.05
	3	4.23	6.03	5.04	5.60
	0	9.20	18.71	20.37	8.27
3	0.5	11.57	24.57	20.74	11.04
	1	11.47	26.49	17.03	11.99
	2	7.44	15.87	7.57	8.90
	3	3.34	3.11	1.87	4.08
3	0	2.60	3.75	4.53	2.03
	0.5	2.90	3.88	4.16	2.36
	1	2.96	4.26	3.79	2.60
	2	2.78	2.83	2.08	2.61
	3	1.87	1.22	1.14	1.91

Table 17b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.2)

The first eigenvalue is 0.75 and second is 0.26 for the new Φ matrix which means moderate autocorrelation. From the Table 17b Hotelling T-square chart based on raw data is better than other methods to detect the shift. But when the amount of shift for both variable increases to 3σ , then the model based approach seems to be better. As it is seen when $\delta_{x_1} = 3$ and $\delta_{x_2} = 3$, VAR and AR based methods detect the shift early than other methods.

We then consider the correlation between error terms, $cov(e_{x_1}, e_{x_2}) = 0.9$ with the Φ matrix in (6.2). The Table 18a below shows the adjusted upper control limits for proposed methods in order to obtain the in control ARL of 370. The Table 18b shows the ARLs obtained by the use of Hotelling T-square control charts based on proposed methods with $cov(e_{x_1}, e_{x_2}) = 0.9$ in the case of shift in the process mean.

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$	11.47	11.81	13.6	15.67

Table 18a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.2) with $cov(e_{x_1}, e_{x_2}) = 0.9$

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	389.13	375.48	380.52	378.92
	0.5	62.90	84.14	188.93	58.97
	1	8.92	6.34	34.13	6.52
	2	1.14	1.00	1.00	1.02
0.5	3	1.00	1.00	1.00	1.00
	0	51.80	53.60	78.32	38.02
	0.5	225.35	299.54	204.78	245.04
	1	48.32	103.98	218.62	56.55
1	2	2.06	1.01	1.78	1.49
	3	1.00	1.00	1.00	1.00
	0	6.80	6.03	12.82	4.07
	0.5	33.23	32.28	31.99	22.68
2	1	83.46	152.33	66.42	93.66
	2	6.13	2.92	13.69	5.88
	3	1.09	1.00	1.00	1.02
	0	1.16	1.02	1.10	1.00
3	0.5	1.74	1.40	1.93	1.12
	1	3.84	3.60	4.19	2.11
	2	14.61	21.86	8.26	13.99
	3	3.19	1.00	1.00	3.48
3	0	1.00	1.00	1.00	1.00
	0.5	1.01	1.00	1.00	1.00
	1	1.07	1.03	1.08	1.00
	2	1.90	1.80	1.90	1.21
	3	3.68	1.66	1.36	3.31

Table 18b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.2) with $cov(e_{x_1}, e_{x_2}) = 0.9$

If the correlation between the variables is 0.9 with Φ matrix with less autocorrelation, then the detection capability of Hotelling T-square chart based on data matrix with lagged variables seems to be better if the amount of shift is low. For the large amounts of shifts, model based approach seems better than the other methods to detect the shift in the process mean.

The third Φ autocorrelation matrix that we considered is

$$\Phi = \begin{pmatrix} 0.4 & 0.1 \\ 0.3 & 0.4 \end{pmatrix} \tag{6.3}$$

Adjusted upper control limits for that matrix in terms of proposed methods are given in Table 19a without any correlation among the error terms. In the Table 19b, in the case of shift in the process mean, the ARLs calculated by the use of Hotelling T-square control charts based on the four different proposed methods with the Φ matrix in (6.3) can be seen.

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	11.67	11.81	12.55	15.72

Table 19a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.3)

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	359.34	367.01	385.02	358.42
	0.5	208.73	269.05	265.90	218.28
	1	69.44	121.49	127.29	79.44
	2	10.37	21.95	25.59	10.32
0.5	3	3.06	4.00	5.44	2.41
	0	197.02	279.61	286.08	218.55
	0.5	157.37	265.63	212.21	203.11
	1	66.82	163.40	107.69	90.88
1	2	11.90	29.80	21.60	13.06
	3	3.25	4.92	4.92	2.85
	0	69.69	116.90	154.10	77.11
	0.5	71.14	156.00	129.72	88.09
2	1	43.00	140.60	76.97	59.80
	2	11.22	32.70	19.01	13.11
	3	3.09	4.63	3.84	2.97
	0	10.42	20.79	33.74	9.61
3	0.5	11.73	28.24	30.66	12.31
	1	11.18	35.02	23.31	12.64
	2	5.96	18.57	8.67	7.15
	3	2.45	3.44	2.46	2.61
	0	2.75	4.10	6.78	2.19
	0.5	3.22	5.17	6.56	2.75
	1	3.13	5.24	5.12	2.78
	2	2.54	3.90	2.81	2.55
	3	1.59	1.53	1.28	1.55

Table 19b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.3)

This Φ matrix in (6.3) corresponds to low autocorrelation level since the eigenvalues are at low level, first eigenvalue is 0.57 and second is 0.23. To detect the shift earlier in the process mean, the Hotelling T-square control chart based on data matrix with lagged variables can be an alternative to the Hotelling T-square control chart based on raw data when one of the variable has large amount of shift in the process mean, such as $\delta_{x_1} = 3$ or $\delta_{x_2} = 3$. But generally Hotelling T-square control chart based on raw data is better than other methods to detect the shift earlier when the autocorrelation level is low.

The Table 20a below shows the adjusted upper control limits based proposed methods with Φ autocorrelation matrix in (6.3) when $cov(e_{x_1}, e_{x_2}) = 0.9$. The table shows the ARLs obtained by the use of Hotelling T-square control chart based on the proposed methods in the case of shift in the process mean with $cov(e_{x_1}, e_{x_2}) = 0.9$

Σ	RAW	VAR	AR	LAG
$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$	11.68	11.81	13.85	15.68

Table 20a Adjusted upper control limits for Hotelling T-square control charts based on proposed methods with Φ matrix in (6.3) with $cov(e_{x_1}, e_{x_2}) = 0.9$

δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	378.78	369.30	363.80	376.57
	0.5	44.06	63.31	83.04	37.57
	1	5.35	6.11	12.33	3.61
	2	1.05	1.00	1.03	1.00
0.5	3	1.00	1.00	1.00	1.00
	0	39.95	44.96	112.27	30.98
	0.5	211.55	281.62	242.29	227.21
	1	32.61	59.46	50.33	30.82
1	2	1.55	1.16	1.55	1.15
	3	1.00	1.00	1.00	1.00
	0	4.93	5.17	19.15	2.91
	0.5	26.17	29.70	106.01	18.97
2	1	80.43	177.48	135.14	100.03
	2	3.68	4.03	4.44	3.00
	3	1.03	1.00	1.00	1.00
	0	1.05	1.02	1.13	1.00
3	0.5	1.47	1.37	3.03	1.05
	1	3.12	3.40	15.69	1.90
	2	12.72	34.94	24.11	14.77
	3	2.25	1.50	1.72	2.03
3	0	1.00	1.00	1.00	1.00
	0.5	1.00	1.00	1.00	1.00
	1	1.03	1.02	1.07	1.00
	2	1.76	1.75	5.65	1.20
	3	3.32	4.73	3.90	3.25

Table 20b Comparison of the ARLs obtained by using Hotelling T-square control charts with adjusted upper control limits based on proposed methods with Φ matrix in (6.3) with $cov(e_{x_1}, e_{x_2}) = 0.9$

When there is high correlation between the error terms with the matrix corresponds to low autocorrelation, detection capability of Hotelling T-square control chart based on data matrix with lagged variable is better than other methods when the amount of shift for one variable is different from the amount of shift of other variable.

Chapter 7

Conclusion

The main reason for using control charts in statistical quality control is to detect any out of control situation quickly. As we did in chapter 3, when there is autocorrelation between the successive observations of process, generally time series models are fit to remove or reduce the autocorrelation, and then residuals are monitored to detect the shift in process mean. However control charts based residuals do not have the same properties as the traditional control chart. We show that the X-chart based on residuals from the first autoregressive AR (1) process have poor detection capability to detect the shift in the process mean, since the first residual has the high probability to detect the shift than the subsequent residuals when the autocorrelation is positive. Conversely, when the autocorrelation is negative, the subsequent residuals have higher probability to detect the shift in process mean. Therefore, the use of X-chart based on residual is suggested when the process is negatively autocorrelated in the univariate case. They can be used to detect the larger shift such as 3 standard deviation unit when the positive autocorrelation is high. But in this study we focus on the multivariate data since multivariate control charts consider also the cross correlation among the variables. We use Hotelling T-square control charts to observe the relation between the observations. Then we attempt to extend the case in univariate time series to the multivariate time series. For multivariate time series for simplicity, we use the data which have bivariate VAR (1) structure. In chapter 4 we compare the shift detection performance of Hotelling T-square control charts based on raw data with the Hotelling T-square control charts based on the residuals from the VAR (1) model. The results show that the if the autocorrelation level which is considered in terms of eigenvalue of autocorrelation matrix is at high level Hotelling T-square control chart based on residuals performs well if the model estimation is assumed to be perfect. Also if the amount of shift increases, such as 3σ for two variables, then the detection capability of Hotelling T-square control chart based on residuals increases even the autocorrelation level decreases to 0.75.

Then in chapter 5, we consider another method which is suggested by Mason and Young (2002). They propose to reconstruct the data with lagged variables, and to apply the Hotelling T-square control chart to the reconstructed data. But we show that even we assume that the data are independent, it still shows autocorrelation effect since in each calculation of Hotelling T-square statistic, previous observation is used.

Last we fit individual AR models to the multivariate data, but since it ignores the correlation and cross correlation between the variables, we expect that using AR model for multivariate data may give misleading results. Finally to compare these proposed methods, we adjust the

control limits to give the same false alarm rate, then we consider the shift in the process mean. The results show that if the estimation of Φ matrix corresponds to high autocorrelation with $cov(e_{x_1}, e_{x_2}) = 0$, generally residual based approach is better to detect the shift when the amount of shift is large. Also when $cov(e_{x_1}, e_{x_2}) = 0.9$, the shift detection power of Hotelling T-square control chart based on residuals will increase. If the autocorrelation matrix indicates moderate autocorrelation with $cov(e_{x_1}, e_{x_2}) = 0$, it is suggested to use Hotelling T-square control chart based on raw data when the magnitudes of shifts are low or moderate. When the autocorrelation matrix corresponds to low autocorrelation with $cov(e_{x_1}, e_{x_2}) = 0$, Hotelling T-square control chart based on raw data generally detects the shift earlier than other proposed methods. But Hotelling T-square control chart based on data matrix with lagged variables can be an alternative to Hotelling T-square control charts based on raw data when the $cov(e_{x_1}, e_{x_2}) = 0.9$. In addition if the off-diagonal elements in the autocorrelation matrix do not show high autocorrelation, Hotelling T-square based on residuals of individual AR model can be used as an alternative to the Hotelling T-square control chart based on residuals from the VAR model to detect the shift in the process mean. Also using Hotelling T-square control chart based on the residuals of individual AR model for each variable may be advantageous if the data have large number of variables compare to the Hotelling T-square control chart based on multivariate autoregressive model such as VAR model.

In this study we focus on the performance of Hotelling T-square control charts based on bivariate data which has VAR structure, residuals from VAR model, residuals from AR models on the individual time series and the data matrix with lagged variables. Statistical performance of these proposed methods could be extended for more than two variables with different level of correlations and autocorrelations. In addition, other control charts such as MCUSUM and MEWMA may be used to see the detection capability when the process mean shift.

REFERENCES

- Nien Fan Zhang, Detection capability of residual control charts for stationary process data. Statistical Engineering Division, National Institute of Standard and Technology, Gaithersburg, USA
- Alwan LC., Roberts, H.V., 1988. Time series modeling for detecting level shifts of autocorrelated processes. *Journal of Business and Economics Statistics* 6, 87-96
- Harris, T.J., Ross, W.H., 1991. Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering* 69, 48-57.
- Hotelling, H. 1947. Multivariate quality control. In: Eisenhart, C., Hastay, M.W., Wallis, W.A. *Techniques of Statistical Analysis*. McGraw-Hill, New York.
- Jackson, J.E., 1985. Multivariate quality control. *Communications in Statistics- Theory and Methods* 14, 2657-2688
- Jarrett, J.E., Pan, X., 2006. The quality control chart for monitoring multivariate autocorrelated processes. *Computational Statistics and Data Analysis*.
- Lu, C.W., Reynolds, M.R., 1999. Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology* 31, 259-274
- Mason, R.L., Tracy, N.D., Young, J.C., 1995. Decomposition of T-square for multivariate control charts interpretation. *Journal of Quality Technology* 27, 99-108
- Montgomery, D.C., Mastrangelo, C.M., 1991. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* 23, 179-193
- Wardell, D., Moskowitz, H., Plante, R., 1992. Control charts in presence of data correlation. *Management Science* 38, 1084-1105
- Alwan, L.C 1992, Effects of autocorrelation on control chart performance. *Communications in statistics: Theory and Methods*, 21, 1025-1049
- Longnecker, M.T, Ryan T.P. 1990. A deficiency in residuals charts for correlated data. Technical Report 131 (Department of Statistics, Texas University)
- Wardell, D.G., Moskowitz, H., Plante, R 1994. Run length distribution of special cause control charts for correlated process.
- Hwarg, H.B., 2005. Simultaneous identification of mean shift and correlation change in AR (1) processes. *International Journal of Production Research*. 43, 1761-1783.
- Jarrett, J.E., Pan, X., 2007. Using vector autoregressive residuals to monitor multivariate processes in the presence of serial correlation. *International Journal of Production Economics*, 106, 204-216
- Montgomery, D.C., 2009. Introduction to statistical quality control. 6th edition. New York: Wiley
- Mason, R.L., Tracy, N.D., and Young, J.C., 1995. Decomposition of T-square for multivariate control chart interpretation. *Journal of Quality Technology*, 27 (2), 99-108

- Mason, R.L., Tracy, N.D., and Young, J.C., 1997. A practical approach for interpreting multivariate T-square control chart signals. *Journal of Quality Technology*, 29(4), 396-406
- M. Kulahci, S. Bisgaard. 2011. *Time series analysis and forecasting by example*, Wiley
- Mason, R.L., Young, J.C., *Multivariate statistical process control with industrial application*.
- H.Madsen. *Time series analysis*, Chapman & Hall
- J.D Williams, W.H. Woodall, J.B. Birch, J.H. Sullivan. On the distribution of Hotelling T-square statistics based on successive differences covariance matrix estimator.
- M.Kulahci and S. Bisgaard, 2005. The effect of autocorrelation on statistical process control procedures, *Quality engineering*.
- Mastrangelo C.M., and Forrest, D.R. 2002 Multivariate autocorrelated processes: Data shift generation. *Journal of Quality Technology* 34, 216-220
- Kalgonda A.A. and Kulkarni S.R. 2004 Multivariate quality control chart for autocorrelated processes. *Journal of applied statistics*, 31, 317-327
- Lu, C.W and Reynolds, M.R Jr 1999 EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology*, 31, 166-188
- T.C Cheng, P. H. Hsieh, S.F. Yang 2011. On the Hotelling T-square control chart for vector autoregressive process. National Chengchi University, Department of statistics
- Karin Kandanond, 2010. The detection of shifts in autocorrelated processes with moving range and exponentially weighted moving average charts. Rajabhat University, Faculty of Industrial Technology
- Zhang N.F., A statistical control chart for stationary process data. *Technometrics*, 1998, 40(1), 24-38
- Alwan, L.C. & Roberts H.V 1988. Time Series modeling for statistical process control. *Business and economic statistics*, 6, 87-95
- Hotelling H. 1947. *Multivariate Quality control: Techniques of statistical analysis*. New York: McGraw Hill.
- Montgomery D.C & Wadsworth H.M 1972. *Some techniques for multivariate quality control applications*.

APPENDIX

Here we add some supplementary tables which contain supporting information to understand the analysis we did.

Appendix A

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	193.50	193.50	125.36	125.36	68.16	68.16
	-0.25	195.12	192.03	124.70	99.00	66.99	65.56
	-0.5	208.74	205.09	132.22	65.76	68.86	67.77
	-0.75	227.48	208.54	143.22	29.13	72.40	70.80
	-0.95	259.99	197.35	194.37	3.04	82.35	67.78
-0.25	0	211.03	148.22	132.40	105.03	67.99	36.75
	-0.25	200.50	145.53	122.58	81.41	65.60	34.78
	-0.5	209.05	150.87	138.45	57.17	65.63	35.55
	-0.75	208.39	149.40	141.37	24.66	66.99	35.89
	-0.95	272.88	148.43	196.30	3.02	79.95	35.15
-0.5	0	203.19	93.58	137.53	70.98	69.98	16.23
	-0.25	199.18	89.64	136.29	56.29	67.69	15.91
	-0.5	205.08	92.36	136.75	39.98	70.85	15.60
	-0.75	219.97	92.62	146.51	19.73	70.75	16.08
	-0.95	277.39	99.20	212.73	2.92	78.69	15.78
-0.75	0	219.46	31.73	142.86	27.95	77.92	4.74
	-0.25	232.45	34.41	144.38	26.37	76.90	4.62
	-0.5	227.69	35.65	144.55	20.90	84.84	4.68
	-0.75	237.25	34.92	160.01	12.54	81.57	4.54
	-0.95	297.05	34.39	232.89	2.73	92.53	4.56
-0.95	0	346.86	3.24	193.12	2.93	137.24	1.54
	-0.25	339.18	3.26	209.83	3.11	127.01	1.54
	-0.5	365.21	3.17	196.84	2.89	143.32	1.54
	-0.75	375.85	3.29	211.42	2.73	151.53	1.54
	-0.95	540.62	3.20	374.14	1.97	202.60	1.50

Table A.1 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different negative autocorrelation levels and various magnitudes of shifts

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	54.30	54.30	27.78	27.78	9.11	9.11
	-0.25	52.61	44.54	29.50	19.02	9.52	9.45
	-0.5	50.01	33.05	28.23	9.44	9.54	9.46
	-0.75	52.80	17.33	27.94	3.65	10.09	9.83
	-0.95	62.14	2.79	37.53	1.48	9.98	9.24
-0.25	0	52.26	30.67	25.66	17.46	8.94	4.46
	-0.25	49.32	25.37	26.71	12.78	8.56	4.52
	-0.5	50.78	20.96	26.55	7.85	8.87	4.41
	-0.75	50.23	12.81	28.69	3.40	9.09	4.50
	-0.95	70.87	2.66	41.95	1.44	9.81	4.54
-0.5	0	50.83	14.37	27.00	9.36	8.76	2.37
	-0.25	55.98	12.59	26.72	8.04	8.80	2.47
	-0.5	53.93	11.10	28.23	5.63	9.41	2.48
	-0.75	55.98	7.77	31.07	3.06	9.61	2.43
	-0.95	66.46	2.46	40.92	1.41	9.98	2.44
-0.75	0	56.58	4.46	29.71	3.81	10.86	1.64
	-0.25	59.21	4.14	28.71	3.44	10.16	1.61
	-0.5	59.93	4.17	30.40	2.98	10.01	1.62
	-0.75	63.46	3.51	33.28	2.28	11.32	1.60
	-0.95	80.67	2.12	47.40	1.40	11.69	1.62
-0.95	0	89.03	1.49	38.57	1.47	19.20	1.00
	-0.25	93.39	1.52	39.78	1.46	20.16	1.00
	-0.5	100.48	1.52	42.70	1.46	19.04	1.00
	-0.75	101.01	1.50	47.12	1.40	23.37	1.00
	-0.95	166.06	1.40	84.15	1.11	23.79	1.00

Table A.1 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	8.50	8.50	6.56	6.56	2.92	2.92
	-0.25	8.08	7.80	6.11	5.28	2.94	2.36
	-0.5	8.41	7.28	6.64	4.11	2.83	1.77
	-0.75	8.87	5.69	6.76	2.60	2.85	1.38
	-0.95	9.20	2.13	7.12	1.31	3.09	1.00
-0.25	0	8.26	4.30	6.00	3.59	2.94	2.36
	-0.25	8.43	4.33	5.98	3.37	2.79	2.01
	-0.5	7.80	3.86	6.27	2.91	2.68	1.67
	-0.75	8.21	3.47	6.36	2.17	2.72	1.36
	-0.95	8.64	1.95	6.39	1.31	2.83	1.00
-0.5	0	8.14	2.39	6.36	2.25	2.84	1.79
	-0.25	7.97	2.36	5.96	2.12	2.80	1.71
	-0.5	8.39	2.34	6.22	2.06	2.67	1.51
	-0.75	8.56	2.23	6.32	1.86	2.86	1.31
	-0.95	9.29	1.72	6.65	1.25	2.92	1.00
-0.75	0	9.39	1.61	6.26	1.55	2.85	1.36
	-0.25	8.88	1.59	6.57	1.54	2.84	1.36
	-0.5	9.76	1.58	6.51	1.50	2.96	1.32
	-0.75	9.98	1.56	7.09	1.47	2.69	1.17
	-0.95	10.52	1.46	8.35	1.12	2.73	1.00
-0.95	0	16.09	1.00	9.69	1.00	3.08	1.00
	-0.25	14.81	1.00	9.69	1.00	2.83	1.00
	-0.5	16.22	1.00	9.22	1.00	2.75	1.00
	-0.75	18.60	1.00	11.02	1.00	2.98	1.00
	-0.95	21.54	1.00	13.98	1.00	3.63	1.00

Table A.1 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$		$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	2.47	2.47	2.41	2.41	2.25	2.25
	-0.25	2.49	2.50	2.57	2.52	2.17	2.07
	-0.5	2.61	2.59	2.46	2.31	2.17	1.90
	-0.75	2.49	2.49	2.42	2.09	2.19	1.61
	-0.95	2.70	2.61	2.57	1.55	2.37	1.14
-0.25	0	2.39	1.80	2.32	1.75	2.04	1.63
	-0.25	2.44	1.82	2.43	1.75	2.04	1.65
	-0.5	2.48	1.82	2.28	1.70	2.01	1.55
	-0.75	2.42	1.81	2.29	1.67	2.09	1.46
	-0.95	2.51	1.83	2.31	1.44	2.05	1.11
-0.5	0	2.18	1.44	2.19	1.42	1.95	1.38
	-0.25	2.27	1.42	2.16	1.40	1.98	1.39
	-0.5	2.31	1.48	2.17	1.42	1.96	1.37
	-0.75	2.26	1.45	2.24	1.41	1.98	1.32
	-0.95	2.42	1.44	2.30	1.33	1.98	1.07
-0.75	0	2.09	1.10	2.09	1.11	1.91	1.08
	-0.25	2.21	1.10	2.05	1.12	1.81	1.07
	-0.5	2.16	1.11	2.01	1.11	1.81	1.08
	-0.75	2.26	1.11	2.14	1.09	1.90	1.07
	-0.95	2.29	1.09	2.11	1.06	1.88	1.01
-0.95	0	2.40	1.00	2.23	1.00	1.94	1.00
	-0.25	2.35	1.00	2.21	1.00	1.88	1.00
	-0.5	2.42	1.00	2.18	1.00	1.84	1.00
	-0.75	2.32	1.00	2.18	1.00	1.87	1.00
	-0.95	2.54	1.00	2.54	1.00	1.82	1.00

Table A.1 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$		$\delta_{x_1} = 3$ $\delta_{x_2} = 3$	
		RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.61	1.61	1.20	1.20
	-0.25	1.63	1.51	1.18	1.16
	-0.5	1.54	1.34	1.25	1.11
	-0.75	1.54	1.18	1.17	1.01
	-0.95	1.51	1.00	1.17	1.00
-0.25	0	1.52	1.38	1.19	1.16
	-0.25	1.53	1.35	1.20	1.16
	-0.5	1.53	1.29	1.17	1.08
	-0.75	1.49	1.17	1.17	1.01
	-0.95	1.48	1.00	1.16	1.00
-0.5	0	1.46	1.24	1.17	1.09
	-0.25	1.46	1.24	1.20	1.09
	-0.5	1.46	1.18	1.19	1.05
	-0.75	1.46	1.11	1.17	1.00
	-0.95	1.42	1.00	1.17	1.00
-0.75	0	1.50	1.04	1.18	1.01
	-0.25	1.46	1.05	1.17	1.01
	-0.5	1.44	1.04	1.17	1.01
	-0.75	1.38	1.01	1.16	1.00
	-0.95	1.41	1.00	1.18	1.00
-0.95	0	1.44	1.00	1.16	1.00
	-0.25	1.40	1.00	1.16	1.00
	-0.5	1.39	1.00	1.16	1.00
	-0.75	1.40	1.00	1.15	1.00
	-0.95	1.35	1.00	1.15	1.00

Table A.1 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	48.42	48.42	193.31	193.31	6.13	6.13
	-0.25	65.49	49.60	199.54	145.07	9.05	5.92
	-0.5	108.22	44.93	207.81	58.09	21.39	6.06
	-0.75	177.31	49.87	216.67	9.47	49.61	5.79
	-0.95	273.25	50.11	234.79	1.75	74.79	6.16
-0.25	0	67.27	23.79	195.31	133.24	8.73	3.05
	-0.25	48.05	25.79	204.96	145.77	5.49	3.18
	-0.5	68.56	25.71	206.39	82.50	9.48	3.13
	-0.75	145.10	24.51	211.79	13.54	29.56	3.18
	-0.95	263.25	23.48	246.43	1.74	72.03	3.21
-0.5	0	113.28	10.27	205.09	56.76	20.21	1.98
	-0.25	72.95	10.35	196.36	77.95	10.26	1.95
	-0.5	50.42	10.12	204.05	81.69	5.84	1.95
	-0.75	102.59	10.67	217.51	22.80	15.11	1.97
	-0.95	241.45	10.67	269.21	1.78	62.29	1.94
-0.75	0	179.61	3.41	196.83	9.54	47.80	1.43
	-0.25	166.96	3.32	223.02	13.54	36.76	1.40
	-0.5	106.59	3.36	226.68	23.39	17.30	1.42
	-0.75	57.34	3.35	231.40	31.58	6.78	1.43
	-0.95	229.91	3.23	332.18	1.94	48.75	1.43
-0.95	0	343.72	1.36	230.09	1.75	128.98	1.00
	-0.25	334.59	1.36	259.21	1.74	123.71	1.00
	-0.5	352.31	1.34	276.28	1.77	125.72	1.00
	-0.75	345.04	1.33	368.96	1.95	88.39	1.00
	-0.95	150.671	1.363	505.6999	3.112	13.303	1.000

Table A.2 Comparison of the ARLs obtained by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process in Phase II for different negative autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 1$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	30.75	30.75	61.80	61.84	1.12	1.12
	-0.25	39.58	46.57	60.96	32.74	1.26	1.11
	-0.5	57.57	66.94	57.96	8.28	2.22	1.11
	-0.75	75.01	31.01	60.67	2.15	4.85	1.10
	-0.95	81.50	1.94	59.68	1.01	8.90	1.10
-0.25	0	36.13	9.84	63.15	32.56	1.25	1.07
	-0.25	31.48	15.42	61.38	33.28	1.11	1.07
	-0.5	41.68	24.97	65.63	13.44	1.28	1.07
	-0.75	65.07	30.96	65.27	2.42	2.83	1.08
	-0.95	81.48	2.02	61.00	1.02	8.20	1.09
-0.5	0	57.16	3.43	62.23	8.10	1.91	1.02
	-0.25	42.21	4.21	62.14	13.53	1.28	1.02
	-0.5	31.04	6.10	63.87	14.42	1.10	1.03
	-0.75	47.25	13.58	70.93	3.42	1.50	1.02
	-0.95	86.96	2.20	72.36	1.02	6.47	1.02
-0.75	0	77.07	1.81	58.60	2.16	4.80	1.00
	-0.25	74.31	1.84	66.67	2.54	2.64	1.00
	-0.5	63.17	1.97	72.16	3.36	1.42	1.00
	-0.75	37.05	2.38	80.32	4.13	1.10	1.00
	-0.95	93.73	3.06	97.85	1.10	3.83	1.00
-0.95	0	122.35	1.00	58.30	1.01	16.85	1.00
	-0.25	129.07	1.00	64.15	1.02	16.76	1.00
	-0.5	145.35	1.00	72.67	1.03	11.91	1.00
	-0.75	176.32	1.00	101.27	1.11	6.22	1.00
	-0.95	91.14	1.15	175.13	1.50	1.11	1.00

Table A.2 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$		$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.63	1.63	3.48	3.48	8.42	8.42
	-0.25	2.18	1.78	4.53	5.08	8.42	4.27
	-0.5	4.12	2.14	7.44	8.23	8.09	2.01
	-0.75	8.05	4.60	9.71	3.67	7.09	1.43
	-0.95	9.71	2.36	9.50	1.31	5.15	1.00
-0.25	0	1.91	1.35	4.04	1.85	8.54	4.23
	-0.25	1.52	1.37	3.17	2.17	8.54	4.28
	-0.5	2.09	1.45	4.58	2.95	8.14	2.29
	-0.75	5.32	1.87	8.11	4.07	7.89	1.46
	-0.95	10.31	2.89	10.23	1.33	5.77	1.00
-0.5	0	3.44	1.16	6.67	1.48	8.17	1.95
	-0.25	1.96	1.16	4.29	1.48	8.59	2.32
	-0.5	1.48	1.20	3.07	1.60	8.40	2.33
	-0.75	2.72	1.31	5.69	2.18	8.85	1.57
	-0.95	9.13	2.35	10.19	1.40	6.90	1.00
-0.75	0	8.14	1.00	10.36	1.05	6.48	1.41
	-0.25	5.81	1.00	10.24	1.05	7.34	1.51
	-0.5	2.75	1.00	6.28	1.09	8.69	1.52
	-0.75	1.40	1.01	3.23	1.22	10.45	1.60
	-0.95	7.49	1.25	10.75	1.49	10.08	1.00
-0.95	0	19.60	1.00	14.88	1.00	5.15	1.00
	-0.25	22.16	1.00	18.09	1.00	5.94	1.00
	-0.5	22.34	1.00	20.87	1.00	6.51	1.00
	-0.75	16.63	1.00	25.94	1.00	10.50	1.00
	-0.95	1.40	1.00	4.75	1.00	22.36	1.00

Table A.2 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$		$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$		$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	
		RAW	RESIDUAL	RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.00	1	1.00	1.00	1.05	1.05
	-0.25	1.02	1	1.03	1.00	1.13	1.07
	-0.5	1.07	1	1.22	1.00	1.56	1.11
	-0.75	1.44	1	1.90	1.02	2.39	1.57
	-0.95	2.34	1	2.74	1.70	2.59	1.50
-0.25	0	1.00	1	1.02	1.00	1.11	1.02
	-0.25	1.00	1	1.00	1.00	1.04	1.03
	-0.5	1.00	1	1.03	1.00	1.18	1.06
	-0.75	1.17	1	1.39	1.01	1.68	1.19
	-0.95	2.08	1	2.37	1.27	2.40	1.50
-0.5	0	1.05	1	1.17	1.00	1.39	1.00
	-0.25	1.00	1	1.03	1.00	1.15	1.00
	-0.5	1.00	1	1.00	1.00	1.03	1.00
	-0.75	1.02	1	1.11	1.00	1.27	1.05
	-0.95	1.61	1	2.01	1.04	2.31	1.45
-0.75	0	1.34	1	1.56	1.00	1.82	1.000
	-0.25	1.16	1	1.31	1.00	1.45	1.00
	-0.5	1.02	1	1.10	1.00	1.27	1.000
	-0.75	1.00	1	1.00	1.00	1.04	1.00
	-0.95	1.26	1	1.47	1.00	1.76	1.04
-0.95	0	1.97	1	2.23	1.00	2.38	1.00
	-0.25	1.77	1	2.25	1.00	2.45	1.00
	-0.5	1.48	1	1.92	1.00	2.21	1.00
	-0.75	1.23	1	1.39	1.00	1.65	1.00
	-0.95	1.00	1	1.00	1.00	1.04	1.00

Table A.2 Continued

ϕ_{11}	ϕ_{22}	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$		$\delta_{x_1} = 3$ $\delta_{x_2} = 3$	
		RAW	RESIDUAL	RAW	RESIDUAL
0	0	1.79	1.79	2.36	2.36
	-0.25	2.10	2.18	2.26	1.73
	-0.5	2.47	1.93	2.07	1.43
	-0.75	2.52	1.58	1.73	1.03
	-0.95	1.99	1.00	1.39	1.00
-0.25	0	1.85	1.38	2.25	1.72
	-0.25	1.69	1.45	2.19	1.72
	-0.5	1.93	1.67	2.13	1.42
	-0.75	2.45	1.57	1.85	1.03
	-0.95	2.06	1.00	1.42	1.00
-0.5	0	2.04	1.16	1.97	1.39
	-0.25	1.77	1.17	2.03	1.41
	-0.5	1.68	1.25	2.14	1.41
	-0.75	1.99	1.40	1.92	1.07
	-0.95	2.10	1.00	1.47	1.00
-0.75	0	2.23	1.00	1.74	1.02
	-0.25	2.20	1.00	1.86	1.03
	-0.5	1.90	1.00	1.89	1.07
	-0.75	1.57	1.03	2.00	1.08
	-0.95	2.22	1.00	1.67	1.00
-0.95	0	1.78	1.00	1.39	1.00
	-0.25	1.89	1.00	1.43	1.00
	-0.5	2.10	1.00	1.45	1.00
	-0.75	2.48	1.00	1.60	1.00
	-0.95	1.53	1.00	2.20	1.00

Table A.2 Continued

Appendix B

		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	230.48	124.98	56.72	42.43	20.89	6.30	5.58
	-0.25	214.81	105.35	59.44	38.07	15.15	6.28	5.15
	-0.5	219.14	83.66	57.85	32.56	8.83	6.08	4.72
	-0.75	220.67	38.43	59.04	18.77	3.51	6.20	4.02
	-0.95	232.49	3.15	62.10	2.48	1.00	6.14	1.47
-0.25	0	179.49	113.57	36.73	27.80	14.39	3.31	3.09
	-0.25	179.87	89.49	37.84	25.45	11.38	3.18	2.92
	-0.5	177.31	69.92	35.91	21.75	7.50	3.26	2.85
	-0.75	172.00	34.84	36.03	14.02	3.17	3.26	2.48
	-0.95	183.22	3.03	39.55	2.16	1.00	3.26	1.27
-0.5	0	125.17	73.42	17.92	15.26	9.29	1.69	1.55
	-0.25	118.56	66.90	18.14	13.55	7.37	1.64	1.57
	-0.5	122.50	53.57	18.07	12.46	5.16	1.66	1.55
	-0.75	121.80	28.77	18.28	8.75	2.56	1.67	1.44
	-0.95	133.35	2.80	18.17	2.01	1.00	1.71	1.08
-0.75	0	52.20	36.26	4.81	4.62	3.51	1.04	1.02
	-0.25	48.35	32.60	5.00	4.34	3.03	1.02	1.03
	-0.5	52.09	28.29	5.20	3.64	2.48	1.02	1.02
	-0.75	54.41	17.28	5.01	3.25	1.65	1.01	1.01
	-0.95	52.38	2.60	5.14	1.42	1.00	1.04	1.00
-0.95	0	3.27	3.05	1.01	1.00	1.00	1.00	1.00
	-0.25	3.38	3.10	1.00	1.00	1.00	1.00	1.00
	-0.5	3.34	2.86	1.00	1.00	1.00	1.00	1.00
	-0.75	3.21	2.40	1.01	1.00	1.00	1.00	1.00
	-0.95	3.38	1.29	1.01	1.00	1.00	1.00	1.00

Table B.1 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different negative autocorrelation levels and various magnitudes of shifts

		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 3$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	4.03	1.75	1.46	1.40	1.30	1.10	1.00
	-0.25	3.46	1.47	1.56	1.47	1.31	1.05	1.00
	-0.5	2.77	1.13	1.52	1.44	1.22	1.02	1.00
	-0.75	1.74	1.00	1.54	1.37	1.09	1.00	1.00
	-0.95	1.00	1.00	1.51	1.06	1.00	1.00	1.00
-0.25	0	2.53	1.45	1.16	1.13	1.11	1.03	1.00
	-0.25	2.27	1.23	1.12	1.11	1.07	1.01	1.00
	-0.5	1.97	1.09	1.14	1.13	1.06	1.00	1.00
	-0.75	1.41	1.00	1.13	1.09	1.04	1.00	1.00
	-0.95	1.00	1.00	1.12	1.02	1.00	1.00	1.00
-0.5	0	1.44	1.15	1.00	1.01	1.00	1.00	1.00
	-0.25	1.39	1.09	1.01	1.01	1.00	1.00	1.00
	-0.5	1.32	1.03	1.01	1.01	1.00	1.00	1.00
	-0.75	1.14	1.00	1.01	1.00	1.00	1.00	1.00
	-0.95	1.00	1.00	1.01	1.00	1.00	1.00	1.00
-0.75	0	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	-0.25	1.02	1.00	1.00	1.00	1.00	1.00	1.00
	-0.5	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	-0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-0.95	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	-0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table B.1 Continued

		$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0$	$\delta_{x_1} = 0.5$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0$	$\delta_{x_1} = 1$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 1$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0$	$\delta_{x_1} = 2$ $\delta_{x_2} = 0.5$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	40.85	222.54	3.53	23.49	54.07	1.00	1.12
	-0.25	44.16	167.90	4.37	35.87	31.28	1.00	1.26
	-0.5	59.63	78.82	5.49	52.23	9.79	1.03	1.88
	-0.75	64.73	14.04	7.15	30.99	1.39	1.08	4.13
	-0.95	76.62	1.01	7.30	1.06	1.00	1.11	1.58
-0.25	0	26.63	176.49	2.33	10.53	31.27	1.00	1.01
	-0.25	23.91	177.70	2.10	12.71	31.61	1.00	1.01
	-0.5	26.18	108.67	2.43	23.19	14.73	1.00	1.05
	-0.75	34.14	20.59	2.99	32.43	1.86	1.00	1.52
	-0.95	38.59	1.01	3.31	1.19	1.00	1.00	2.14
-0.5	0	13.17	77.40	1.40	3.19	8.80	1.00	1.00
	-0.25	12.75	107.23	1.34	3.99	14.86	1.00	1.00
	-0.5	11.77	115.25	1.28	6.13	16.11	1.00	1.00
	-0.75	12.68	35.60	1.36	14.86	3.05	1.00	1.01
	-0.95	15.65	1.05	1.44	1.62	1.00	1.00	1.63
-0.75	0	3.49	13.55	1.00	1.07	1.41	1.00	1.00
	-0.25	3.19	20.32	1.00	1.13	1.87	1.00	1.00
	-0.5	3.35	35.80	1.00	1.28	3.00	1.00	1.00
	-0.75	3.12	44.67	1.00	1.83	4.65	1.00	1.00
	-0.95	3.39	1.24	1.00	2.86	1.00	1.00	1.00
-0.95	0	1.00	1.01	1.00	1.00	1.00	1.00	1.00
	-0.25	1.00	1.02	1.00	1.00	1.00	1.00	1.00
	-0.5	1.00	1.05	1.00	1.00	1.00	1.00	1.00
	-0.75	1.00	1.21	1.00	1.00	1.00	1.00	1.00
	-0.95	1.00	2.99	1.00	1.00	1.00	1.00	1.00

Table B.2 Comparison of the ARLs obtained by using Hotelling T-square control charts based on data matrix with lagged variables in Phase II for different negative autocorrelation levels and various magnitudes of shifts with $cov(e_{x_1}, e_{x_2}) = 0.9$

		$\delta_{x_1} = 2$ $\delta_{x_2} = 1$	$\delta_{x_1} = 2$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0$	$\delta_{x_1} = 3$ $\delta_{x_2} = 0.5$	$\delta_{x_1} = 3$ $\delta_{x_2} = 1$	$\delta_{x_1} = 3$ $\delta_{x_2} = 2$	$\delta_{x_1} = 3$ $\delta_{x_2} = 3$
ϕ_{11}	ϕ_{22}	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED	LAGGED
0	0	1.98	5.29	1	1.00	1.00	1.20	1.30
	-0.25	3.06	2.81	1	1.00	1.00	1.40	1.07
	-0.5	5.56	1.15	1	1.00	1.08	1.31	1.00
	-0.75	2.73	1.00	1	1.03	1.42	1.00	1.00
	-0.95	1.00	1.00	1	1.35	1.00	1.00	1.00
-0.25	0	1.19	2.77	1	1.00	1.00	1.00	1.08
	-0.25	1.33	2.90	1	1.00	1.00	1.03	1.07
	-0.5	1.96	1.42	1	1.00	1.00	1.11	1.00
	-0.75	2.89	1.00	1	1.00	1.03	1.00	1.00
	-0.95	1.00	1.00	1	1.08	1.00	1.00	1.00
-0.5	0	1.00	1.17	1	1.00	1.00	1.00	1.00
	-0.25	1.008	1.478	1	1.000	1.000	1.000	1.002
	-0.5	1.042	1.507	1	1.000	1.000	1.001	1.006
	-0.75	1.430	1.004	1	1.000	1.000	1.006	1.000
	-0.95	1.000	1.000	1	1.000	1.001	1.000	1.000
-0.75	0	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.25	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.5	1.000	1.003	1	1.000	1.000	1.000	1.000
	-0.75	1.000	1.013	1	1.000	1.000	1.000	1.000
	-0.95	1.004	1.000	1	1.000	1.000	1.000	1.000
-0.95	0	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.25	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.5	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.75	1.000	1.000	1	1.000	1.000	1.000	1.000
	-0.95	1.000	1.000	1	1.000	1.000	1.000	1.000

Table B.2 Continued

Appendix C

ARLs obtained by using Hotelling T-square control chart with theoretical control limits based on proposed methods for the Φ matrix corresponds to high, moderate and low autocorrelation levels can be see in this appendix.

$\Phi = \begin{pmatrix} 0.8 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$					
δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	702.04	375.74	256.44	525.53
	0.5	296.08	314.61	219.93	302.61
	1	84.73	120.37	107.46	100.30
	2	9.17	1.02	1.47	12.56
	3	1.67	1.00	1.00	2.05
0.5	0	268.39	302.74	223.37	289.28
	0.5	459.09	298.33	167.90	377.62
	1	200.58	148.76	82.68	205.85
	2	19.74	1.30	1.24	27.32
	3	3.18	1.00	1.00	3.89
1	0	66.48	163.30	128.60	84.66
	0.5	155.01	195.20	109.93	175.18
	1	237.08	99.99	45.51	233.33
	2	42.33	1.12	1.07	53.58
	3	5.15	1.00	1.00	7.75
2	0	7.71	6.09	8.93	9.64
	0.5	15.88	8.25	7.17	20.82
	1	27.44	3.14	2.17	37.90
	2	51.64	1.00	1.00	58.43
	3	13.38	1.00	1.00	16.96
3	0	1.57	1.00	1.00	1.84
	0.5	2.35	1.00	1.00	2.91
	1	3.76	1.00	1.00	5.22
	2	8.51	1.00	1.00	12.30
	3	9.24	1.00	1.00	12.34

Table C.1 Comparison of the ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.1)

$\Phi = \begin{pmatrix} 0.3 & 0.2 \\ 0.1 & 0.7 \end{pmatrix}$					
δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	402.10	383.89	284.52	471.56
	0.5	241.54	289.63	251.82	312.91
	1	79.02	183.33	173.69	110.69
	2	13.84	39.26	52.78	18.31
	3	3.62	4.20	6.57	4.20
0.5	0	222.00	239.34	193.20	248.88
	0.5	184.90	298.50	175.51	256.83
	1	93.85	243.52	134.85	145.14
	2	17.51	62.86	45.33	26.92
	3	4.38	5.17	4.90	5.40
1	0	68.05	106.64	87.79	82.81
	0.5	79.14	146.10	86.12	104.14
	1	54.49	143.75	66.44	84.52
	2	17.67	62.92	24.86	26.12
	3	4.37	5.14	3.35	6.02
2	0	9.80	17.88	16.87	9.67
	0.5	12.05	23.89	16.53	12.58
	1	12.88	26.46	14.60	14.49
	2	8.22	18.13	6.30	10.35
	3	3.41	2.60	1.64	4.26
3	0	2.64	3.83	3.96	2.12
	0.5	2.98	3.85	3.55	2.56
	1	3.24	4.01	3.08	2.80
	2	2.97	2.90	1.97	3.00
	3	1.94	1.19	1.07	1.97

Table C.2 Comparison of ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.2)

$\Phi = \begin{pmatrix} 0.4 & 0.1 \\ 0.3 & 0.4 \end{pmatrix}$					
δ_{x_1}	δ_{x_2}	RAW	VAR	AR	LAG
0	0	386.96	364.10	258.89	443.36
	0.5	211.79	270.71	192.01	260.78
	1	72.85	122.87	98.07	90.61
	2	11.02	22.82	19.77	11.90
	3	2.94	3.69	3.99	2.53
0.5	0	201.50	249.84	192.48	245.46
	0.5	163.61	263.10	145.54	228.37
	1	72.45	162.57	80.99	104.33
	2	12.39	30.63	18.01	14.25
	3	3.19	4.68	3.65	2.87
1	0	72.18	117.96	111.28	86.00
	0.5	72.42	153.85	91.92	105.40
	1	48.80	137.63	60.69	73.23
	2	11.90	37.58	15.57	16.29
	3	3.13	5.14	3.23	3.27
2	0	10.84	21.20	26.48	11.06
	0.5	12.96	29.40	24.85	14.79
	1	11.28	32.50	17.42	14.31
	2	6.04	17.20	6.65	7.99
	3	2.64	3.69	2.03	2.91
3	0	2.96	4.19	5.58	2.45
	0.5	3.15	4.77	5.18	2.90
	1	3.23	5.27	4.26	3.20
	2	2.63	3.94	2.44	2.77
	3	1.61	1.62	1.28	1.65

Table C.3 Comparison of ARLs obtained by using Hotelling T-square control charts with theoretical upper control limits based on proposed methods with Φ matrix in (6.3)

Appendix D

D.1 Simulations for Chapter 3

R codes for the calculation of ARLs based on X-chart in the combination of different autocorrelation levels and different number of observations in Phase I for AR (1) process.

```

rm(list=ls())
sq<-seq(100,1000,100)
sq2<-seq(2000,6000,1000)
n1=c(50,sq,sq2)
dd<-5000
phi1<-c(0,0.3,0.5,0.7,0.9)
mu1<-0
Ar11x <- matrix(rep(0),dd,1)
Ar111x <- matrix(rep(0),dd,1)
Ar11x2 <- matrix(rep(0),dd,1)
Ar111x2<- matrix(rep(0),dd,1)
cc <-matrix(rep(0),length(phi1),4)
ccs <-matrix(rep(0),length(phi1)*length(n1),4)
ucllcl<-matrix(rep(0),dd,2)

for (s in 1:length(n1))
{
dat<- matrix(rep(0),n1[s],dd)
for (q in 1:length(phi1))
{
for (i in 1:dd)
{
c1<-(1-phi1[q])*mu1
a1<-arima.sim(n=n1[s],list(ar=c(phi1[q])),mean=c1)

##### Control Limits with estimated parameters #####
ybarx<-mean(a1)
ysdx<-sd(a1)
UCLx<- ybarx+3*ysdx
LCLx<-ybarx-3*ysdx
ucllcl[i,]<-c(UCLx,LCLx)
dat[,i]<-a1
}
m<-colMeans(ucllcl)

##### Control Limits with known parameters by taking autocorrelation into account #####
ybar<-mu1
ysd<-(1/sqrt(1-phi1[q]^2))
UCL<-ybar+3*ysd
LCL<-ybar-3*ysd

##### Calculation of Run lengths #####
for (j in 1:dd)

```

Appendix D

```
{
  Arl1x[j,]<-min(which(dat[,j]>UCL))
  Arl11x[j,]<-min(which(dat[,j]<LCL))
  Arl1x[is.infinite(Arl1x)]<-NA
  Arl11x[is.infinite(Arl11x)]<-NA

  Arl1x2[j,]<-min(which(dat[,j]>m[1]))
  Arl11x2[j,]<-min(which(dat[,j]<m[2]))
  Arl1x2[is.infinite(Arl1x2)]<-NA
  Arl11x2[is.infinite(Arl11x2)]<-NA
}

Va1<-cbind(Arl1x,Arl11x)
av1a<-matrix(0,dd,1)
av1a[(which(Va1[,1]&Va1[,2]))]<- apply(Va1[(which(Va1[,1]&Va1[,2]))],1,min)
av1a[which(is.na(Va1[,1])&is.na(Va1[,2]))]=NA
av1a[which(Va1[,1]&is.na(Va1[,2]))<-Va1[which(Va1[,1]&is.na(Va1[,2])),1]
av1a[which(is.na(Va1[,1])&Va1[,2])<-Va1[which(is.na(Va1[,1])&Va1[,2]),2]

Va2<-cbind(Arl1x2,Arl11x2)
av1b<-matrix(0,dd,1)
av1b[(which(Va2[,1]&Va2[,2]))]<- apply(Va2[(which(Va2[,1]&Va2[,2]))],1,min)
av1b[which(is.na(Va2[,1])&is.na(Va2[,2]))]=NA
av1b[which(Va2[,1]&is.na(Va2[,2]))<-Va2[which(Va2[,1]&is.na(Va2[,2])),1]
av1b[which(is.na(Va2[,1])&Va2[,2])<-Va2[which(is.na(Va2[,1])&Va2[,2]),2]

exp1<-dim(matrix(which(av1a<=(n1[s])))[1]/dd)
exp2<-dim(matrix(which(av1b<=(n1[s])))[1]/dd)
arl1m<-mean(av1a,na.rm=TRUE)
arl2m<-mean(av1b,na.rm=TRUE)

rate<-seq(0,1,0.0001)
d1<-matrix(rep(0),length(rate),1)
d2<-matrix(rep(0),length(rate),1)

for (r in 1:length(rate))
{
  d1[r]<-qexp(exp1,rate[r])
  d2[r]<-qexp(exp2,rate[r])
}
k1<-min(which(round(d1)<=n1[s]))
k2<-min(which(round(d2)<=n1[s]))
arl1<-1/rate[k1]
arl2<-1/rate[k2]

cc[q,]<-c(arl1,arl1m,arl2,arl2m)
}

ccs[(length(phi1)*(s-1)+1):(length(phi1)*s),]<-cc
}
```

ccs

For the figures of exponential distribution when the number of observation is 5000 and the number of dataset is 5000 with autocorrelation levels 0 and 0.7

```
##### Exponential Distribution of run lengths #####
par(mfrow=c(2,2))
Generated_RL<-av1b
Expected_RL<-rexp(dd,1/cc[4] )
qqplot(Expected_RL,Generated_RL,main=expression( "Q-Q plot of RL with unknown parameters"))
abline(0,1)
hist(Generated_RL,breaks=50,col="lightblue",main =expression( "Histogram of RL with unknown
parameters"))

Generated_RL<-av1a
Expected_RL<-rexp(dd,1/cc[2] )
qqplot(Expected_RL,Generated_RL,main=expression( "Q-Q plot of RL with known parameters"))
abline(0,1)
hist(Generated_RL,breaks=50,col="lightblue",main =expression( "Histogram of RL with known parameters"))
```

R codes for the calculation of ARLs by using X-chart based on raw data and residuals from AR(1) process in the combination of different autocorrelation levels and various magnitudes of shifts in the process mean.

```
rm(list=ls())
n1=100
n2<-5000
sn2<-0
shf<-c(0,0.5,1,2,3)
dd<-1000
phi1<-c(0.25,0.5,0.75,0.95,-0.25,-0.5,-0.75,-0.95)
mu1<-0

Arl1xr <- matrix(rep(0),dd,1)
Arl11xr <- matrix(rep(0),dd,1)
Arl1x <- matrix(rep(0),dd,1)
Arl11x <- matrix(rep(0),dd,1)
cc<-matrix(rep(0),length(phi1),2)
ccs <-matrix(rep(0),length(phi1)*length(shf),2)
datb1<- matrix(rep(0),n2-sn2,dd)

for (sh in 1:length(shf))
{
for (q in 1:length(phi1))
{
c1<-(1-phi1[q])*mu1
```

Appendix D

```
ybar<-mu1
ysd<-(1/sqrt(1-phi1[q]^2))

#### Control limits for raw data ####
UCL<-ybar+3*ysd
LCL<-ybar-3*ysd

#### Control limits of residuals ####
UCLr<- 0+3*1
LCLr<- 0-3*1

mu2<- mu1+shf[sh]*(1/sqrt(1-phi1[q]^2))
c2<-(1-phi1[q])*mu2

for (j in 1:dd)
{

a1<-arima.sim(n=n1,list(ar=c(phi1[q]),mean=c1)
mushf<-a1[n1]+shf[sh]*(1/sqrt(1-phi1[q]^2))

z = matrix(0,n2,1)
e = rnorm(n2,0,1)
z[1] =mushf
  for (jf in 2:n2)
    {
      z[jf] = c2+phi1[q]*z[jf-1]+e[jf]
    }
b1sn<-z[(sn2+1):n2]
datb1[,j]<-b1sn

ra1m<-matrix(0,n2-sn2,1)
ra1m[1]<-b1sn[1]-phi1[q]*a1[n1-1]-c1

for (k in 2:(n2-sn2))
{
ra1m[k]<-b1sn[k]-phi1[q]*b1sn[k-1]-c1
}

Arl1xr[j,]<-min(which(ra1m>UCLr))
Arl11xr[j,]<-min(which(ra1m<LCLr))
Arl1xr[is.infinite(Arl1xr)]<-NA
Arl11xr[is.infinite(Arl11xr)]<-NA

Arl1x[j,]<-min(which(datb1[,j]>UCL))
Arl11x[j,]<-min(which(datb1[,j]<LCL))
Arl1x[is.infinite(Arl1x)]<-NA
Arl11x[is.infinite(Arl11x)]<-NA
}

Valr<-cbind(Arl1xr,Arl11xr)
```

Appendix D

```
av1ar<-matrix(0,dd,1)
av1ar[(which(Va1r[,1]&Va1r[,2]))]<- apply(Va1r[(which(Va1r[,1]&Va1r[,2]))],1,min)
av1ar[which(is.na(Va1r[,1])&is.na(Va1r[,2]))]=NA
av1ar[which(Va1r[,1]&is.na(Va1r[,2]))]<-Va1r[which(Va1r[,1]&is.na(Va1r[,2])),1]
av1ar[which(is.na(Va1r[,1])&Va1r[,2])<-Va1r[which(is.na(Va1r[,1])&Va1r[,2]),2]
ar11mr<-mean(av1ar,na.rm=TRUE)
```

```
Va1<-cbind(Ar11x,Ar11x)
av1a<-matrix(0,dd,1)
av1a[(which(Va1[,1]&Va1[,2]))]<-apply(Va1[(which(Va1[,1]&Va1[,2]))],1,min)
av1a[which(is.na(Va1[,1])&is.na(Va1[,2]))]=NA
av1a[which(Va1[,1]&is.na(Va1[,2]))]<-Va1[which(Va1[,1]&is.na(Va1[,2])),1]
av1a[which(is.na(Va1[,1])&Va1[,2])<-Va1[which(is.na(Va1[,1])&Va1[,2]),2]
```

```
ar11m<-c(mean(av1a,na.rm=TRUE))
ar11mr<-mean(av1ar,na.rm=TRUE)
```

```
cc[q,]<-c(ar11m,ar11mr)
}
ccs[(length(phi1)*(sh-1)+1):(length(phi1)*sh),]<-cc
}
ccs
```

D.2 Simulations for Chapter 4

R codes for the determination of simulation and theoretical based upper control limits for the multivariate normally distributed data by using Hotelling T-square statistics

```
rm(list=ls())
library(qcc)
library(mvtnorm)
library(MASS)
p=2
dd=1000
ds<-100
sigma<-diag(p)
mu = rep(0,p)
sq<-seq(100,1700,100)
k<-c(sq,2000,2500,3000,4000)
nler<-matrix(rep(0),length(k),2)

for (h in 1:length(k))
{
dat<- matrix(rep(0),k[h],dd)
mmax1<- matrix(rep(0),dd,1)
UCL<-limits.T2.single(k[h], 1, p, 0.9973)$control[2]
UCL2<-matrix(rep(0),ds,1)
for (s in 1:ds)
{
for (d in 1:dd)
```

Appendix D

```
{
e = mvrnorm(k[h],mu,sigma)
T<-stats.T2.single(e, center=NULL, cov=NULL )$statistics
dat[,d]<-T
k1=max(dat[,d])
mmax1[d,]<-k1
}
mx<- mmax1[rev(order(mmax1))]
alph1<-1-((0.9973)^(k[h]))
UCL2[s,]<-mx[alph1*dd]
}
nler[h,]<-c(UCL,mean(UCL2))
}
nler
```

R codes for the calculation of average run lengths by using Hotelling T-square control chart based on multivariate normally distributed data in terms of simulation and theoretical UCL. We need to define the number of observation and upper control limit based on the specified number of observation determined by the previous simulation when we are running the code

```
rm(list=ls())
library(qcc)
library(mvtnorm)
library(MASS)

#### Define the number of observation n ####
n = 100

p=2
dd=1000
cvler<-100

sigma<-diag(p)
mu = rep(0,p)

Arl<- matrix(rep(0),dd,1)
Arl2<- matrix(rep(0),dd,1)

cv<- matrix(rep(0),cvler,4)

dat<- matrix(rep(0),n,dd)
mmax1<- matrix(rep(0),dd,1)
UCL<-limits.T2.single(n, 1, p, 0.9973)$control[2]

for (jk in 1:cvler){
for (d in 1:dd)
{
e = mvrnorm(n,mu,sigma)
T<-stats.T2.single(e, center=NULL, cov=NULL )$statistics
```

Appendix D

```
dat[,d]<-T
k1=max(dat[,d])
mmax1[d,]<-k1
Arl[d,]<-min(which(T>UCL))
Arl1<-Arl
Arl1[is.infinite(Arl1)]<-NA
}

### Theoretical UCL ###
exp1<-dim(matrix(which(Arl1<=n)))[1]/dd
rate<-seq(0,1,0.0001)
d1<-matrix(rep(0),length(rate),1)
for (r in 1:length(rate))
{
d1[r]<-qexp(exp1,rate[r])
}
k1<-min(which(round(d1)<=n))
ar1<-1/rate[k1]

### Define UCL2 in terms of the UCL determined previous simulation #####
UCL2<-12.91

### Simulation based UCL #####
for (s in 1:dd)
{
Arl2[s,]<-min(which(dat[,s]>UCL2))
Arl22<-Arl2
Arl22[is.infinite(Arl22)]<-NA
}
exp2<-dim(matrix(which(Arl22<=n)))[1]/dd
rate<-seq(0,1,0.0001)
d2<-matrix(rep(0),length(rate),1)
for (r in 1:length(rate))
{
d2[r]<-qexp(exp2,rate[r])
}
k2<-min(which(round(d2)<=n))

ar2<-1/rate[k2]

ccc<-mean(Arl1,na.rm=TRUE)
cc2<-mean(Arl22,na.rm=TRUE)

cv[jk,]<-c(ar1,ar2,ccc,cc2)

}

colMeans(cv)
```


R codes for the calculation of ARLs by using Hotelling T-square control charts based on raw data and residuals from VAR(1) process with different positive autocorrelation levels and various amount of magnitudes of shifts. We need to define the magnitudes of shift, sigma error term and the autocorrelation level when we are running the code.

```

rm(list=ls())
library(mvtnorm)
library(MASS)
library(qcc)
library(vars)
n = 100
n2= 5000
ne<- 0
p = 2
dd= 1000

#### define the magnitudes of shift #####
shf<-c(0,0)

sigma<-matrix(c(1,0,0,1),2,2)
phi = matrix(rep(0),p,p)

####
a<-c(0,0.25,0.5,0.75,0.95)
mu = rep(0,p)

Arl1z<- matrix(rep(0),dd,1)
Arl2z<- matrix(rep(0),dd,1)
datz1<- matrix(rep(0),n2-ne,dd)
datz2<- matrix(rep(0),n2-ne,dd)

ccxz<-matrix(rep(0),length(a),2)
ccxxz<-matrix(rep(0),length(a)^2,2)

for (i in 1:length(a))
{
phi[1]=a[i]
  for (tt in 1:length(a))
  {

phi[p,p]=a[tt]

c = as.vector((diag(p)-phi)%*%mu)

y = matrix(0,p,n)
y[,1] = mu
z = matrix(0,p,n2)

for (d in 1:dd)

```

Appendix D

```
{
e = mvrnorm(n+n2,mu,sigma)
e = t(e)
e1<-e[,1:n]
  for (j in 2:n)
    {
      y[,j] = c+phi%*%y[,j-1]+e1[,j]
    }
e2<-e[, (n+1):(n+n2)]
u <- t(y)
yy1 <- u[(ne+1):n,]

ecov<-solve(diag(4)-kronecker(phi, phi))
stsigma<-cbind(stack(as.data.frame(sigma))[,1])
dfg<-ecov%*%stsigma
ncov<-matrix(c(dfg),2,2)

mu2<-as.vector(mu+shf*sqrt(diag(ncov)))
c2 = as.vector((diag(p)-phi)%*%mu2)

mushf<-y[,n]+shf*sqrt(diag(ncov))

z[,1] =mushf
  for (jf in 2:n2)
    {
      z[,jf] = c2+phi%*%z[,jf-1]+e2[,jf]
    }

uz <- t(z)
zz1<- uz[(ne+1):n2,]

re1<-matrix(rep(0),p,n2-(ne))
re1[,1]<-z[,1]-(c+phi%*%y[,n-1])

for (gh in 2:(n2-ne))
  {
    re1[,gh]<-z[,gh]-(c+phi%*%z[,gh-1])
  }
re1t<-t(re1)

Tzz1<-stats.T2.single(zz1, center=mu, cov=ncov )$statistics
Tzz2<-stats.T2.single(re1t, center=c(0,0), cov=sigma )$statistics

datz1[,d]<-Tzz1
datz2[,d]<-Tzz2

}

UCL1z<-limits.T2.single(n2-ne, 1, p, 0.9973)$control[2]
```

```

for (s in 1:dd)
  {
  Arl1z[s,]<-min(which(abs(datz1[,s])>UCL1z))
  Arl2z[s,]<-min(which(abs(datz2[,s])>UCL1z))
  Arl11z<-Arl1z
  Arl22z<-Arl2z
  Arl11z[is.infinite(Arl11z)]<-NA
  Arl22z[is.infinite(Arl22z)]<-NA
  }

ccxz[tt,]<-c(mean(Arl11z,na.rm=TRUE),mean(Arl22z,na.rm=TRUE))
}
ccxxz[(length(a)*(i-1)+1):(length(a)*i),]<-ccxz
}

```

D.3 Simulations for Chapter 5

R codes for the calculation of ARLs by using Hotelling T-square control chart based on data matrix with lagged variables with different positive autocorrelation levels and various amount of magnitudes of shifts. We need to define the amount of shift, sigma error term and autocorrelation level when we are running the code.

```

rm(list=ls())
library(mvtnorm)
library(MASS)
library(qcc)
library(vars)
n = 5000
n2= 5000
ne<- 0
p = 2
dd= 1000

#### Define the amount of shift ####
shf<-c(0,0)

#### Define the sigma error term ####
sigma<-matrix(c(1,0,0,1),2,2)
phi = matrix(rep(0),p,p)

#### Define the autocorrelation level ####

a<-c(0,0.25,0.5,0.75,0.95)
mu = rep(0,p)
Arl1zz3<- matrix(rep(0),dd,1)
datzz3<- matrix(rep(0),n2-ne-1,dd)
ccxz<-matrix(rep(0),length(a),1)

```

Appendix D

```
ccxxz<-matrix(rep(0),length(a)^2,1)

for (i in 1:length(a))
{
  phi[1]=a[i]
  for (tt in 1:length(a))
  {

    phi[p,p]=a[tt]
    c = as.vector((diag(p)-phi)%**%mu)
    y = matrix(0,p,n)
    y[,1] = mu
    z = matrix(0,p,n2)

    for (d in 1:dd)
    {
      e = mvrnorm(n+n2,mu,sigma)
      e = t(e)
      e1<-e[,1:n]
      for (j in 2:n)
      {
        y[,j] = c+phi%**y[,j-1]+e1[,j]
      }
      e2<-e[(n+1):(n+n2)]
      u <- t(y)
      yy1 <- u[(ne+1):n,]

      ecov<-solve(diag(4)-kronecker(phi, phi))
      stsigma<-cbind(stack(as.data.frame(sigma))[,1])
      dfg<-ecov%**stsigma
      ncov<-matrix(c(dfg),2,2)

      mu2<-as.vector(mu+shf*sqrt(diag(ncov)))
      c2 = as.vector((diag(p)-phi)%**%mu2)
      mushf<-y[,n]+shf*sqrt(diag(ncov))

      z[,1] =mushf
      for (jf in 2:n2)
      {
        z[,jf] = c2+phi%**z[,jf-1]+e2[,jf]
      }
      uz <- t(z)
      zz1<- uz[(ne+1):n2,]

      yy3<- cbind(yy1[2:(n-ne),1],yy1[1:(n-ne-1),1],yy1[2:(n-ne),2],yy1[1:(n-ne-1),2])
      zz3<- cbind(zz1[2:(n2-ne),1],zz1[1:(n2-ne-1),1],zz1[2:(n2-ne),2],zz1[1:(n2-ne-1),2])

      Tzz3<-stats.T2.single(zz3, center=colMeans(yy3), cov=cov(yy3) )$statistics
      datzz3[,d]<-Tzz3
    }
    UCL2zz3<-limits.T2.single(n2-ne-1, 1, 4 , 0.9973)$control[2]
```

```

for (s in 1:dd)
  {
  Arl1zz3[s,]<-min(which(abs(datzz3[,s])>UCL2zz3))
  Arl1zz3<-Arl1zz3
  Arl1zz3[is.infinite(Arl1zz3)]<-NA
  }
ccxz[tt,]<-c(mean(Arl1zz3,na.rm=TRUE))
}
ccxxz[(length(a)*(i-1)+1):(length(a)*i),]<-ccxz
}

```

D.4 Simulations for Chapter 6

R codes for the calculation of ARLs by using Hotelling T-2 control charts with adjusted control limits based on proposed methods with the autocorrelation matrix corresponds to high, moderate and low autocorrelation. We need to define sigma error term, autocorrelation level and upper control limits when we are running the code.

```

rm(list=ls())
library(mvtnorm)
library(MASS)
library(qcc)
library(vars)
n = 5000
n2= 5000
ne<- 0
p = 2
dd= 1000
shf<-c(0,0)

#### define sigma ####
sigma<-matrix(c(1,0,0,1),2,2)
phi = matrix(rep(0),p,p)

#### define autocorrelation matrix #####
phi[1,1]<-0.4
phi[2,2]<-0.4
phi[1,2]<-0.1
phi[2,1]<-0.3

#### define UCL for each method ####
UCL1<-11.67  ##UCL for raw data
UCL2<-limits.T2.single(n2-ne, 1, p, 0.9973)$control[2]  ## UCL for residual of VAR model
UCL3<-12.55  ## UCL for residual of individual AR model
UCL4<-15.7  ## UCL for data matrix with lagged variables

```

Appendix D

```
mu = rep(0,p)
Arl1<- matrix(rep(0),dd,1)
Arl2<- matrix(rep(0),dd,1)
Arl3<- matrix(rep(0),dd,1)
Arl4<- matrix(rep(0),dd,1)

datz1<- matrix(rep(0),n2-ne,dd)
datz2<- matrix(rep(0),n2-ne,dd)
datz3<- matrix(rep(0),n2-ne,dd)
datz4<- matrix(rep(0),n2-ne-1,dd)

sh<-c(0,0.5,1,2,3)
ccxz<-matrix(rep(0),length(sh),4)
ccxxz<-matrix(rep(0),length(sh)^2,4)
c = as.vector((diag(p)-phi)%**%mu)

y = matrix(0,p,n)
y[,1] = mu
z = matrix(0,p,n2)

for (sh1 in 1:length(sh)){
shf[1]<-sh[sh1]
  for (sh2 in 1:length(sh)){
shf[2]<-sh[sh2]

for (d in 1:dd)
  {
e = mvrnorm(n+n2,mu,sigma)
e = t(e)
e1<-e[,1:n]
  for (j in 2:n)
    {
      y[,j] = c+phi%**y[,j-1]+e1[,j]
    }
e2<-e[(n+1):(n+n2)]
u <- t(y)
yy1 <- u[(ne+1):n,]
yy2<- cbind(yy1[2:(n-ne),1],yy1[1:(n-ne-1),1],yy1[2:(n-ne),2],yy1[1:(n-ne-1),2])

ecov<-solve(diag(4)-kronecker(phi, phi))
stsigma<-cbind(stack(as.data.frame(sigma))[,1])
dfg<-ecov%**stsigma
ncov<-matrix(c(dfg),2,2)

mu2<-as.vector(mu+shf*sqrt(diag(ncov)))
c2 = as.vector((diag(p)-phi)%**%mu2)
mushf<-y[,n]+shf*sqrt(diag(ncov))

z[,1] =mushf
  for (jf in 2:n2)
    {
```

Appendix D

```
z[,jf] = c2+phi%*%z[,jf-1]+e2[,jf]
}

uz <- t(z)
zz1<- uz[(ne+1):n2,]

zz2<- cbind(zz1[2:(n2-ne),1],zz1[1:(n2-ne-1),1],zz1[2:(n2-ne),2],zz1[1:(n2-ne-1),2])

##VAR residual
re1<-matrix(rep(0),p,n2-(ne))
re1[,1]<-z[,1]-(c+phi%*%y[,n-1])
for (gh in 2:(n2-ne))
{
re1[,gh]<-z[,gh]-(c+phi%*%z[,gh-1])
}
re1t<-t(re1)

##AR residual
phihatar<-matrix(rep(0),2,2)
phihatar[1,1]<-phi[1,1]
phihatar[2,2]<-phi[2,2]
re2<-matrix(rep(0),p,n2-(ne))
re2[,1]<-z[,1]-(c+phihatar%*%y[,n-1])
for (gh in 2:(n2))
{
re2[,gh]<-z[,gh]-(c+phihatar%*%z[,gh-1])
}
re2t<-t(re2)

Tzz1<-stats.T2.single(zz1, center=mu, cov=ncov )$statistics #RAW
Tzz2<-stats.T2.single(re1t, center=c(0,0), cov=sigma )$statistics #VAR residual
Tzz3<-stats.T2.single(re2t, center=c(0,0), cov=sigma )$statistics #AR residual
Tzz4<-stats.T2.single(zz2, center=colMeans(yy2), cov=cov(yy2) )$statistics #LAG

datz1[,d]<-Tzz1
datz2[,d]<-Tzz2
datz3[,d]<-Tzz3
datz4[,d]<-Tzz4
}

for (s in 1:dd)
{
Arl1[s,]<-min(which(abs(datz1[,s])>UCL1))
Arl2[s,]<-min(which(abs(datz2[,s])>UCL2))
Arl3[s,]<-min(which(abs(datz3[,s])>UCL3))
Arl4[s,]<-min(which(abs(datz4[,s])>UCL4))
Arl11<-Arl1
Arl22<-Arl2
Arl33<-Arl3
Arl44<-Arl4
Arl11[is.infinite(Arl11)]<-NA
```

Appendix D

```
Arl22[is.infinite(Arl22)]<-NA  
Arl33[is.infinite(Arl33)]<-NA  
Arl44[is.infinite(Arl44)]<-NA  
}
```

```
ccxz[sh2,]<-  
c(mean(Arl11,na.rm=TRUE),mean(Arl22,na.rm=TRUE),mean(Arl33,na.rm=TRUE),mean(Arl44,na.rm=TRUE))  
}  
ccxxz[(length(sh)*(sh1-1)+1):(length(sh)*sh1),]<-ccxz  
}  
ccxxz
```