

5-1-2010

Topics in Multivariate Time Series Analysis: Statistical Control, Dimension Reduction Visualization and Their Business Applications

Xuan Huang

University of Massachusetts - Amherst

Follow this and additional works at: http://scholarworks.umass.edu/open_access_dissertations



Part of the [Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

Huang, Xuan, "Topics in Multivariate Time Series Analysis: Statistical Control, Dimension Reduction Visualization and Their Business Applications" (2010). *Open Access Dissertations*. Paper 204.

This Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**TOPICS IN MULTIVARIATE TIME SERIES ANALYSIS:
STATISTICAL CONTROL, DIMENSION REDUCTION, VISUALIZATION
AND THEIR BUSINESS APPLICATIONS**

A Dissertation Presented

by

XUAN HUANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Management
Management Science

© Copyright by Xuan Huang 2010

All Rights Reserved

**TOPICS IN MULTIVARIATE TIME SERIES ANALYSIS:
STATISTICAL CONTROL, DIMENSION REDUCTION, VISUALIZATION
AND THEIR BUSINESS APPLICATIONS**

A Dissertation Presented

by

XUAN HUANG

Approved as to style and content by:

Iqbal Agha, Chair

John Buonaccorsi, Member

Robert Nakosteen, Member

Iqbal Agha, Department Chair
Department of Finance & Operations
Management

DEDICATION

In memory of my doctoral advisor

Søren Bisgaard

ACKNOWLEDGEMENTS

First and foremost, I have to thank Professor Søren Bisgaard, whose close guidance accompanied every single step of my progress. Without him the completion of this thesis wouldn't have been possible.

Professor Bisgaard had always been there – my first conference presentation, my first journal acceptance, qualification exams, and dissertation proposal – to steer me in the correct direction or simply cheer me on. Unfortunately he won't see the finished version of my thesis, but I can almost picture it: he would give it another careful and thorough proofreading, for the, say, fifteenth time. He would find a misplaced punctuation, a misused article or a misaligned formula, and he would point them out in the most gentle and kind way. For the papers we coauthored, we always iterated proofreading – our record was “version 17,” and we had the “final of final version.” He always taught me to be careful, accurate, and precise. With this thought in mind, I endlessly proofread, in fear that the version would not meet his standard.

I am grateful for what he taught me: to never lose sight of real world applications, to not easily give up ideas in doubt of how other researchers would receive them, and that research is a passion rather than a profession. I am grateful that he respected me as a friend and placed my interest before my obligation as his research assistant. He funded me to finish an MSc in Statistics. He encouraged me to learn beyond our research scope. And there were many dinners, barbeques, and afternoon teas at his house. We talked about economics, science, technologies, as well as music, jokes, and trips to Disney World.

And these fun memories wouldn't have been so sweet without Sue Ellen, his wife. She welcomes and loves us graduate students as family members; she prepares wonderful food and provides warm conversations for our every visit. Her unbreakable strength, in accompanying Professor Bisgaard battling with cancer, is admirable and inspiring.

Special thanks also go to my dissertation committee. Thanks to Professor Iqbal Agha, who, in his sabbatical, stepped up and shouldered the responsibility of being the chair of my committee soon after Professor Bisgaard's passing. Thanks to Professor John Buonaccorsi, whose teaching shaped my understanding of statistics and inspired my interest in statistical research. Thanks to Professor Robert Nakosteen, who has always been so gracious to take time out of his busy schedule and flexible with my often short-notice requests. Of course, I thank my committee for reading through drafts of this thesis and for their valuable insights and comments.

My parents, who are probably the most earnest parents in educating their child, have been so supportive during my graduate study and to them I am forever grateful. My father would still worry about my grades. My mother worries about me staying up too late. They both enjoy listening to every trivial detail of my every Friday Skype "report" – what I eat, what's on the news or simply a bad-luck day. I hope that I will soon be able to see them more often, maybe cook meals for them as I have always boasted about – and they are very suspicious of – my second biggest achievement in graduate school: becoming a pretty good cook.

I thank my boyfriend Xing Yi, who has always been there to share my happiness, sometimes tears, stress and whatever I am facing. I cherish these years we have been through together and I look forward to the new life waiting for us.

I am thankful to my officemates at ISOM 307, who have made this “fish ball” office so lovely and feel like home. Thanks to Phuong Anh Nguyen, for those early years when we both basically “lived” in the office, and for advice and help she is always first to offer. Thanks to Davit Khachatryan, my academic little “brother,” whose academic discussions, companionship to conferences, and conversation on future plans are most appreciated. And of course I have to thank all my fellow students in the Isenberg Management Science Program – Patrick Qiang, Zugang Liu, Tina Wakolbinger, Trisha Woolley, Deanna Kennedy, Baris Hasdemir, Ronrapee Leelawong, and Milad Ebtehaj, without whom there wouldn’t have been so many parties, movie nights, barbeques, and fun.

I am immensely blessed with a large circle of friends at UMass, each of whom has been special and dear. I thank Qiao Lan for always lending a nonjudgmental ears through good times or bad. I thank Nikhil Malvankar, Danxiang Li and Apurv Mathur for the exciting journey of “Bug Power.” I thank Kwong Chan for countless sweaty badminton sessions. I thank Shaohan Hu, Jing Yang, Shuang Feng, Liping Qiu and Guang Xu, Li Cai and Jian Du, Ming Li and Lili Cheng – and I wish I could go on to mention all the names I feel grateful to.

Honored to be an Isenberg Scholar, I have to thank Mr. & Mrs. Isenberg for their generous support. I have to thank Professor Bisgaard and Vice Chancellor Michael Malone for selecting me. My understanding of how academics could influence society has been profoundly sharpened through activities in the Isenberg Integration Program. And of course I have always enjoyed the luxury to have some extra pocket money to keep my sleepy days caffeinated.

ABSTRACT

TOPICS IN MULTIVARIATE TIME SERIES ANALYSIS: STATISTICAL CONTROL, DIMENSION REDUCTION, VISUALIZATION AND THEIR BUSINESS APPLICATIONS

May 2010

XUAN HUANG

B.ENG., TSINGHUA UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Søren Bisgaard

Most business processes are, by nature, multivariate and autocorrelated. High-dimensionality is rooted in processes where more than one variable is considered simultaneously to provide a more comprehensive picture. Time series models are preferable to an independently identically distributed (I.I.D.) model because they capture the fact that many processes have a memory of their past. Examples of multivariate autocorrelation can be found in processes in the business fields such as Operations Management, Finance and Marketing.

The topic of statistical control is most relevant to Quality Management. While both multivariate I.I.D. processes and univariate autocorrelated processes have received much attention in the *Statistical Process Control* (SPC) literature, little work has been done to simultaneously address high-dimensionality and autocorrelation. In this dissertation, this gap is filled by extending the univariate *special cause chart* and *common*

cause chart to multivariate situations. In addition, two-chart control schemes are extended to nonstationary processes. Further, a class of Markov Chain models is proposed to provide accurate *Average Run Length* (ARL) computation when the process is autocorrelated.

The second part of this dissertation aims to devise a dimension reduction method for autocorrelated processes. High-dimensionality often obscures the true underlying components of a process. In traditional multivariate literature, Principal Components Analysis (PCA) is the standard tool for dimension reduction. For autocorrelated processes, however, PCA fails to take into account the autocorrelation information. Thus, it is doubtful that PCA is the best choice. A two-step dimension reduction procedure is devised for multivariate time series. Comparisons based on both simulated examples and case studies show that the two-step procedure is more efficient in retrieving true underlying factors.

Visualization of multivariate time series assists our understanding of the process. In the last part of this dissertation a simple three-dimensional graph is proposed to assist visualizing the results of PCA. It is intended to complement existing graphical methods for multivariate time series data. The idea is to visualize multivariate data as a surface that in turn can be decomposed with PCA. The developed surface plots are intended for statistical process analysis but may also help visualize economics data and, in particular, co-integration.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
ABSTRACT.....	viii
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF SYMBOLS	xvii
 CHAPTER	
1. INTRODUCTION	1
1.1 Relevance of Multivariate Time Series Models in Business Applications	3
1.1.1 Case 1: the Furnace Data	4
1.1.2 Case 2: the Hog Data	6
1.1.3 Case 3: the Interest Data	7
1.1.4 Case 4: the Fama-French Data	8
1.2 An Overview of This Dissertation	10
2. MODEL-BASED MULTIVARIATE MONITORING CHARTS FOR AUTOCORRELATED PROCESSES	12
2.1 Statistical Process Control	12
2.1.1 Common Causes and Special Causes	13
2.1.2 Control Charts Constituents: Charted Data.....	13
2.1.3 Control Charts Constituents: Control Limits	14
2.1.4 Average Run Length	14
2.1.5 Phase I and Phase II	15
2.1.6 The Family of Control Charts	15
2.2 Motivation: the Ceramic Furnace Example	15
2.3 A Discussion on “in Control in a Broader Sense” and the Chapter Overview ...	17
2.4 Literature Review of Traditional Multivariate SPC.....	19
2.5 Standard Charts Applied to Autocorrelated Data.....	21
2.6 Proposed Method: The Multivariate Special and Common Cause Charts.....	26
2.7 Revisiting the Furnace Data.....	28
2.8 Run Length Comparisons	32
2.9 Conclusion	47

3. A CLASS OF MARKOV CHAIN MODELS FOR AVERAGE RUN LENGTH COMPUTATION FOR AUTOCORRELATED PROCESSES.....	49
3.1 Introduction.....	49
3.2 The Univariate AR(1) process.....	51
3.3 Extension to the Univariate ARMA(p, q) process	55
3.3.1 Extension to the Univariate AR(2) Process	56
3.3.2 Extension to the Univariate ARMA(1, 1) Process	58
3.3.3 Extension to the Univariate ARMA(p, q) Process	61
3.4 Extension to the Multivariate ARMA(p, q) process	62
3.5 Computation Efficiency	64
3.5.1 How Fine Should the Mesh be?	65
3.5.2 Matrix Computation.....	69
3.6 Conclusion and Discussion	70
4. COMMON FACTORS AND LINEAR RELATIONSHIPS OF MULTIVARIATE TIME SERIES	72
4.1 Motivation.....	72
4.2 Literature Review.....	74
4.2.1 The PCA Method	75
4.2.2 The Box-Tiao Method.....	75
4.2.3 PCA on $\hat{\Sigma}_a$	78
4.2.4 The Badcock Method.....	78
4.2.5 Dynamic Factor Model	79
4.3 Reflections on the Literature.....	80
4.3.1 The Box-Tiao Method and Canonical Correlation Analysis.....	80
4.3.2 The Box-Tiao Method, Cointegration and Co-movement	82
4.3.3 The Box-Tiao Method and the Badcock Method.....	83
4.3.4 A Simple Augmented Model	84
4.3.5 A Simulation Study	86
4.3.6 Other Application of the Box-Tiao Method.....	91
4.3.7 Spurious Box-Tiao Components.....	92
4.4 A Combined method of both the PCA and the Box-Tiao Methods.....	93
4.4.1 The Problem of Using the Box-Tiao method Individually	93
4.4.2 A Two-step Combined Method	96
4.4.3 Demonstration of the Combined Method	97

4.4.4 Simulated Study Revisit.....	100
4.4.5 Furnace Data Revisited.....	103
4.5 Business Applications	107
4.5.1 Statistical Process Control	107
4.5.2 Applications in Asset Pricing.....	108
5. VISUALIZING PRINCIPAL COMPONENTS ANALYSIS FOR MULTIVARIATE PROCESS DATA.....	112
5.1 Motivation.....	112
5.2 Graphical Representation of Multiple Time Series.....	114
5.3 Principal Components Decomposition	116
5.4 Examples.....	118
5.4.1 Example 1: The Furnace Temperature Data.....	119
5.4.2 Example 2: The Hog Data.....	122
5.5 Further Discussion: Adding Up the Surfaces.....	126
5.6 Visualization of Comovement.....	129
5.7 Conclusion	131
6. CONCLUSION.....	133
6.1 Future Work	134
APPENDICES	
A. EIGENVALUES INVARIANT TO TRANSFORMATION	135
B. ARL CALCULATION FOR RESIDUAL BASED METHOD	136
C. PROOF OF (3.1).....	138
D. DETERMINING THE END STATES IN Figure 3-3	139
BIBLIOGRAPHY.....	140

LIST OF TABLES

Table 2-1. ARL0 With Different set of AR Coefficient Φ for Hotelling's T^2	22
Table 2-2. Examples examined. “++” means both eigenvalues are positive.	35
Table 2-3. ARL's comparison for four cases. NC= Noncentrality. “Z” is the Hotelling's T^2 chart based on the observed data, “A” is the residuals based Hotelling's T^2 chart.....	36
Table 2-4. ARL's comparison for Case 5 and 6. NC= Noncentrality. “Z” stands for observed data based Hotelling's T^2 chart, “A” stands for residuals based Hotelling's T^2 chart.....	41
Table 3-1: Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on AR(1) models.	55
Table 3-2. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on AR(2) models.	58
Table 3-3. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on ARMA(1,1) models.....	60
Table 3-4. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on VAR(1) models.....	64
Table 4-1. Cointegration and Co-movement.....	83
Table 4-2. Correlation between each factor and its matching component.....	102
Table 5-1. Principal Components Analysis Based on the Covariance Matrix of the Furnace-Temperature Data.....	119
Table 5-2. Principal Components Analysis based on the Correlation Matrix of the Hog Data.....	124

LIST OF FIGURES

Figure 1-1. A large ceramic furnace.....	4
Figure 1-2. Hourly temperature measurements at 9 locations of the ceramic furnace.	5
Figure 1-3. A traditional graphical representation of the Hog data.....	7
Figure 1-4. Monthly deposit interest rates in Taiwan: Mar 1961 to Jul 1989.....	8
Figure 1-5. Fama-French factors from Nov 1942 to Jul 2008.....	9
Figure 2-1. A typical control chart.	13
Figure 2-2. Temperature measurements at 2 adjacent locations of a ceramic furnace, (a) Combined display plotted on a common scale. (b) Separate displays on separate scales.	16
Figure 2-3. A T^2 chart using the Holmes and Mergen (1993) estimator for the covariance matrix.	23
Figure 2-4. Control charts on Phase II production: (a) Shewhart chart on z_8 ; (b) Shewhart chart on z_9 ; (c) Hotelling's T^2 on both dimensions. The control limits derived from Phase I data.	25
Figure 2-5. The autocorrelation of the residuals after fitting the IMA model with 5% significance limits for the autocorrelations: (a) residuals from z_8 ; (b) residuals from z_9	29
Figure 2-6. Hotelling's T^2 on residuals after fitting the Vector IMA(1, 1) model.	30
Figure 2-7. Special cause chart with z_8 and fitted values \hat{z}_8 imposed.	31
Figure 2-8. A Time series plot (a) and a box plot (b) of ∇z_{8t}	32
Figure 2-9. Triangular Region of ϕ_1 and ϕ_2	34
Figure 2-10. ARL's Comparison by graphs.	45
Figure 2-11. ϕ_{21} , ϕ_{22} Space when eigenvalues are 0.2 and 0.8 respectively.	47
Figure 3-1. Markov Chain State Space for AR(1) model.	52
Figure 3-2. Markov Chain State Space for AR(2) model.	57

Figure 3-3. Markov Chain State Space for ARMA(1,1) model.	59
Figure 3-4. Markov Chain State Space of VAR(1) model.	63
Figure 3-5. ARL v.s. s for AR(1) model, $\phi = 0$	66
Figure 3-6. ARL v.s. s for AR(1) model, $\phi = 0.5$	66
Figure 3-7. ARL v.s. s for AR(2) model, $\phi_1 = 0.5, \phi_2 = 0.2$	67
Figure 3-8. ARL v.s. s for ARMA(1,1) model, $\phi = 0.8, \theta = 0.5$	68
Figure 3-9. Sparse transition matrices.	69
Figure 4-1. A traditional time series plot of temperatures measured at nine different locations of a large industrial furnace.	73
Figure 4-2. Time Series plot of y_1, y_2 and n_1	87
Figure 4-3. (a) The Time Series plot of y_1, y_2, n_1 and the first two PCs; (b) The Matrix plot of y_1, y_2 and n_1 against two PCs.	89
Figure 4-4. (a) The Time Series plot of y_1, y_2, n_1 and the first three Box-Tiao components; (b) The Matrix plot of y_1, y_2 and n_1 against the first three Box-Tiao components.	90
Figure 4-5. Predictability of Box-Tiao Factors of the Furnace data.	94
Figure 4-6. Correlation of each Box-Tiao factor with each dimension of observed data.	95
Figure 4-7. (a) The Time Series plot of y_1, y_2, n_1 and the two PCA-BT components; (b) The Matrix plot of y_1, y_2 and n_1 against the two PCA-BT components.	101
Figure 4-8. The Four PCA-BT Factors on Furnace data: Time Series plots and Autocorrelation Fuction.	104
Figure 4-9. Correlation between the four BT-PCA factors and each dimension of the observed data.	105
Figure 4-10. The Four PCAs on Furnace data: Time Series plots and Autocorrelation Fuction.	106
Figure 4-11. How well do results from statistical factor retrieving methods match with theoretical proposed pricing factors.	111
Figure 5-1. Surface plot of the industrial furnace temperature profile.	115

Figure 5-2. Contour plot of the furnace temperature profile.	116
Figure 5-3. Plot of the Eigenvector Elements Versus Their Dimension Index Based on the Covariance Matrix of the Furnace Data with (a) for \hat{e}_1 , (b) for \hat{e}_2 and so forth..	120
Figure 5-4. Principal component surfaces for the Furnace Temperature Data: (a) surface of the first principal component C_1 , (b) surface of the second principal component C_2 , (c) surface of the third principal component C_3 , (d) surface of the ninth principal component C_9	121
Figure 5-5. A three-dimensional scatter plot of the furnace data.	122
Figure 5-6. Surface plot of the Hog Data.....	123
Figure 5-7. Plot of the eigenvector elements versus their dimension index based on the correlation matrix of the Hog data with (a) for \hat{e}_1 , (b) for \hat{e}_2 and so forth.	124
Figure 5-8. Principal component surfaces for the Hog Data: (a) surface of the first principal component C_1 , (b) surface of the second principal component C_2 , (c) surface of the third principal component C_3 , (d) surface of the fourth principal component C_4	125
Figure 5-9. Adding up the rank one principal component surfaces for the Hog data: (a) surface of the added-up first two principal components X_2 , (b) surface of the added-up first three principal components X_3	127
Figure 5-10. Adding up the rank one principal component surfaces for the Furnace data: (a) surface of the added-up first two principal components X_2 , (b) surface of the added-up first four principal components X_4 , (c) surface of the added-up first six principal components X_6 , (b) surface of the added-up all principal components X	128
Figure 5-11. Time series plots of the five principal components, PC1, ..., PC5 for the Hog data.	131

LIST OF SYMBOLS

\mathbf{z}_t	The multivariate time series we observe.
\mathbf{a}_t	A vector of white noise, mean of which is $\mathbf{0}$.
$\boldsymbol{\mu}$	The mean of \mathbf{z}_t .
$\boldsymbol{\Sigma}_z$	The variance-covariance matrix of \mathbf{z}_t .
$\boldsymbol{\Sigma}_a$	The variance-covariance matrix of \mathbf{a}_t .
$\Gamma(l)$	$Cov(\mathbf{z}_t, \mathbf{z}_{t+l})$: the cross-covariance matrix of \mathbf{z}_t at lag l .
B	The backshift operator.
t	Time point t .
k	The dimension of \mathbf{z}_t and \mathbf{a}_t .
r	The dimension of common factors (true underlying factors).
p	The autoregressive order.
q	The moving average order.
l	Lag.

CHAPTER 1

INTRODUCTION

Multivariate Time Series models describe relationships among a vector of k time series variables $z_{1t}, z_{2t}, \dots, z_{kt}$. Multivariate processes arise when several related time series are observed simultaneously over time instead of observing just a single series as is the case in univariate time series analysis. In the study of multivariate processes, a framework is needed for describing not only the properties of the individual series but also the possible cross-relationships among the series. The purposes of analyzing and modeling the series jointly are to understand the dynamic relationships over time among the series and to improve the accuracy of forecasts for individual series by utilizing the additional information available from the related series in the forecasts for each series. Among all the models, we base our discussion on the vector autoregressive integrated moving average (VARIMA) time series models.

For a comprehensive but elementary introduction to VARIMA, see Brockwell and Davis (2002). See also Tiao and Box (1981). For a more in-depth discussion, see Reinsel (1997). Here we only state basic facts and define terms.

Let $\{\mathbf{z}_t\}$ be a k -dimensional vector autoregressive moving average ARMA(p, q) time series $\Phi(B)\tilde{\mathbf{z}}_t = \Theta(B)\mathbf{a}_t$ where $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \boldsymbol{\mu}$ is the mean adjusted observation vector (in the non-stationary case I define $\tilde{\mathbf{z}}_t = \mathbf{z}_t$ because the non-stationary process does not have a mean; stationarity is to be defined below), $\Phi(B) = \mathbf{I} - \Phi_1 B - \dots - \Phi_p B^p$, $\Theta(B) = \mathbf{I} - \Theta_1 B - \dots - \Theta_q B^q$ are matrix-valued polynomials, \mathbf{I} is the $k \times k$ identity matrix, B the backshift operator and it is assumed that $\mathbf{a}_t \stackrel{iid}{\sim} N_k(\mathbf{0}, \Sigma_a)$, see e.g. Reinsel (1997).

The ARMA(p, q) process is said to be strictly stationary if the probability distribution of $[\mathbf{z}_{t_1}, \mathbf{z}_{t_2}, \dots, \mathbf{z}_{t_n}]$ and $[\mathbf{z}_{t_1+l}, \mathbf{z}_{t_2+l}, \dots, \mathbf{z}_{t_n+l}]$ are the same for arbitrary times t_1, t_2, \dots, t_n , all n , and all lags and leads $l = \pm 1, \pm 2, \dots$. It is said to be weakly stationary if all the roots of $\det\{\Phi(B)\} = 0$ are greater than one in absolute value and invertible if all the roots of $\det\{\Theta(B)\} = 0$ are greater than one in absolute value. In most of what follows I call a process stationary if it satisfies the weak stationarity criteria. If not specified separately, I also assume that the processes are invertible. Hence, provided the inverse exists, $[\Theta(B)]^{-1}\Phi(B)\tilde{\mathbf{z}}_t = \mathbf{a}_t$. In other words, $[\Theta(B)]^{-1}\Phi(B)$ is a linear filter that when applied to the data $\tilde{\mathbf{z}}_t$ produces a white noise series \mathbf{a}_t .

When the process is stationary, it possesses constant mean and variance. In this dissertation, I use $\boldsymbol{\mu}$ to denote the mean of \mathbf{z}_t , use $\boldsymbol{\Sigma}_z$ to denote the contemporaneous variance-covariance matrix of \mathbf{z}_t . Further I use $\Gamma(l)$ to denote the cross-covariance matrix at lag l , between \mathbf{z}_t and \mathbf{z}_{t+l} . Note

$$\Gamma(l) = \text{Cov}(\mathbf{z}_t, \mathbf{z}_{t+l}) = [\text{Cov}(\mathbf{z}_{t+l}, \mathbf{z}_t)]' = [\text{Cov}(\mathbf{z}_{t+l}, \mathbf{z}_{(t+l)-l})]' = \Gamma(-l)'.$$

In modeling multivariate time series processes, we generally want to describe, predict, and control the processes. A simple example where a multivariate time series model is useful is the daily inventory level of ice cream and waffle cone in a local grocery store. To maintain a healthy inventory level, demand data of past years could be examined and a time series model is fitted. It will not be surprising that the demand data would show patterns and seasonality with peaks during hot seasons and lows during cold seasons. Also the demands of these two complementary products would resemble each

other and are fairly cross-correlated. A two-dimension Seasonal Vector ARIMA model would be capable of capturing these patterns and provides a fairly accurate short term prediction of the demands. Actions could be taken in advance upon a predicted extreme demand and it helps prevent supply falling short or a glut of supply eventually incurring high inventory cost. This is only a simple example of how multivariate time series modeling can be used to describe, predict and control a business process. More real cases are provided in the next section.

1.1 Relevance of Multivariate Time Series Models in Business Applications

Most, if not all business applications are by nature multivariate and autocorrelated. Multivariability is rooted in processes where more than one variable is considered simultaneously to provide a more comprehensive picture of the processes. Time series models are preferable to an I.I.D. (independently identically distributed) model because they capture the fact that many processes have a memory of their past, i.e., the current state of the processes depend on their past states to some extent. In this section, four business cases are introduced to provide insights into why multivariate time series models are relevant in business applications. These cases include operations management process, econometric data and financial data. However it should be acknowledged that multivariate time series models also can be applied in many more fields, e.g., marketing, accounting. The cases provided here will be referred to from time to time throughout the dissertation.

1.1.1 Case 1: the Furnace Data

Figure 1-1 shows a large ceramic furnace. The raw materials are poured in at the left hand side. The materials then get heated; reactions take place along the furnace and finally the finished materials flow out at the opposite end. One important responsibility of operations management is to supervise the quality. Maintaining high yield is critically important to the economy of the firm. Modern quality control has shifted from dependence on final inspection to closely monitoring the process. For this reason, thermocouples are installed in both the top and bottom of the furnace at equal distance along the length of the furnace and the temperatures are sampled hourly and monitored closely. Clearly this process is multivariate.

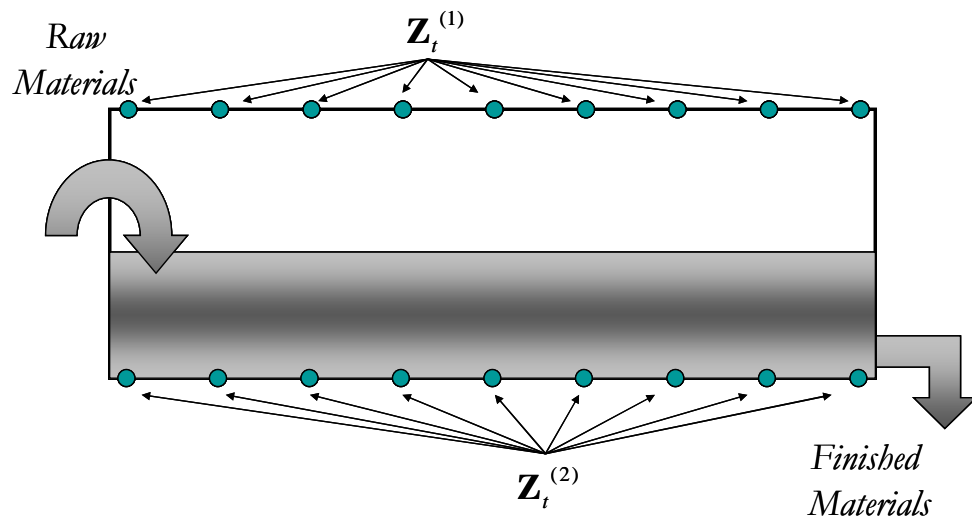


Figure 1-1. A large ceramic furnace.

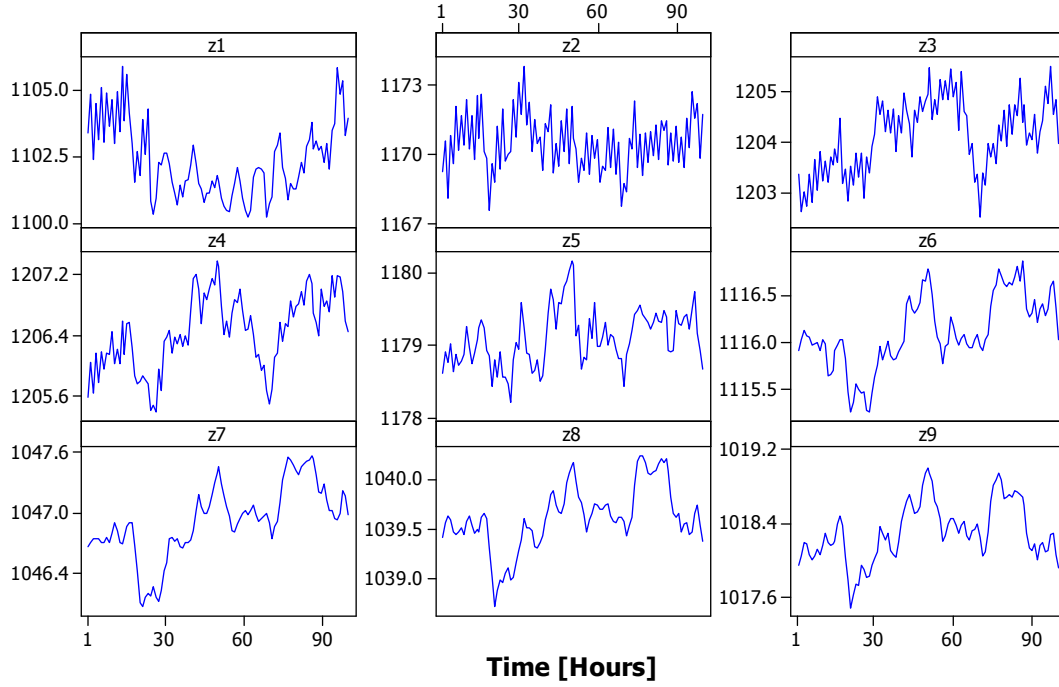


Figure 1-2. Hourly temperature measurements at 9 locations of the ceramic furnace.

This process is autocorrelated as well. By eyeballing each individual series in Figure 1-2 we find none of them resemble an I.I.D process. They all exhibit certain dynamic and even some non-stationarity. More interestingly, z_{6t} through z_{9t} all look alike; they reach peaks and troughs around the same time.

In this dissertation, I answer the following questions: how do we implement control charts on multivariate time series processes? How do we judge a process is in-control if the process is non-stationary? What is the latent factor that make z_{6t} through z_{9t} all look alike? Can we reduce the dimension of the time series to a number smaller than nine?

Case 1 is a major motivation of this dissertation. Every chapter of this dissertation aims at one perspective of understanding and controlling such a process. At the conclusion of this dissertation, building blocks of results from each chapter will be put together to show how the goal is achieved.

1.1.2 Case 2: the Hog Data

The second example is an econometrics example. Figure 1-3 shows the traditional graphical representation of the so-called “hog data” discussed extensively in the statistical time series literature first used by Quenouille (1957) and later Box and Tiao (1977) and many others. It is a five-dimensional multiple time series consisting of 82 yearly observations from 1867-1948 of quantities relevant to the US hog trade. The five dimensions are Hog supply, Hog price, Corn price, Corn supply and Farm wages respectively. The data was originally provided in “Historical Statistics of the United States 1789-1945” and “Statistical Abstract of the United States, 1950” published by the US Department of Commerce. Quenouille (1957) and Box and Tiao (1977) made a few adjustments where necessary, log transformed each series, lagged some and linearly coded the logs. I use the data as adjusted by Box and Tiao (1977).

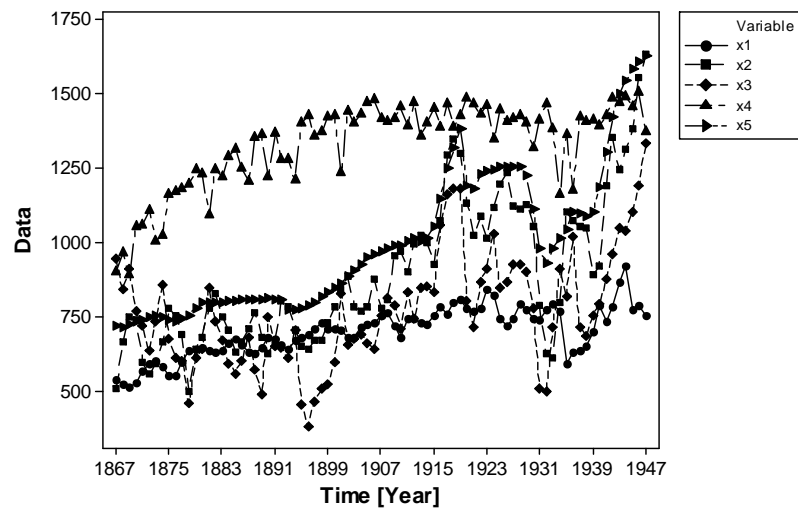


Figure 1-3. A traditional graphical representation of the Hog data.

This data is again multivariate and autocorrelated. One question of interest to the economists is whether there is any equilibrium or long term relationship among the five series, that is, whether co-integration exists among the five series. Further more, how does co-integration imply about underlying factors? I answer these questions in Chapter 4 and extend the idea of co-integration to co-movement. Another problem I consider is whether there is a better way to display multivariate time series than the traditional graph of Figure 1-3, which superimposes all the series together and is hard to read. In Chapter 5 I propose a solution – a 3D surface plot complementary to the traditional displays. I demonstrate how this plot can assist our perception of PCA and co-integration.

1.1.3 Case 3: the Interest Data

Figure 1-4 is the time series plot of the one-, three-, and six-month deposit interest rates in Taiwan from March 1961 to July 1989. The observations represent the averages of the deposit rates offered by nine major banks. The two big swings are associated with the two recessions caused partially by the oil crises around 1973 and 1981. Not

surprisingly, the three interest rates tend to move together. Although this is more or less “man-made” data because the interest rate is controlled by the government, it is still interesting to ask: are the three interest rates co-integrated? How many true common factors are there?

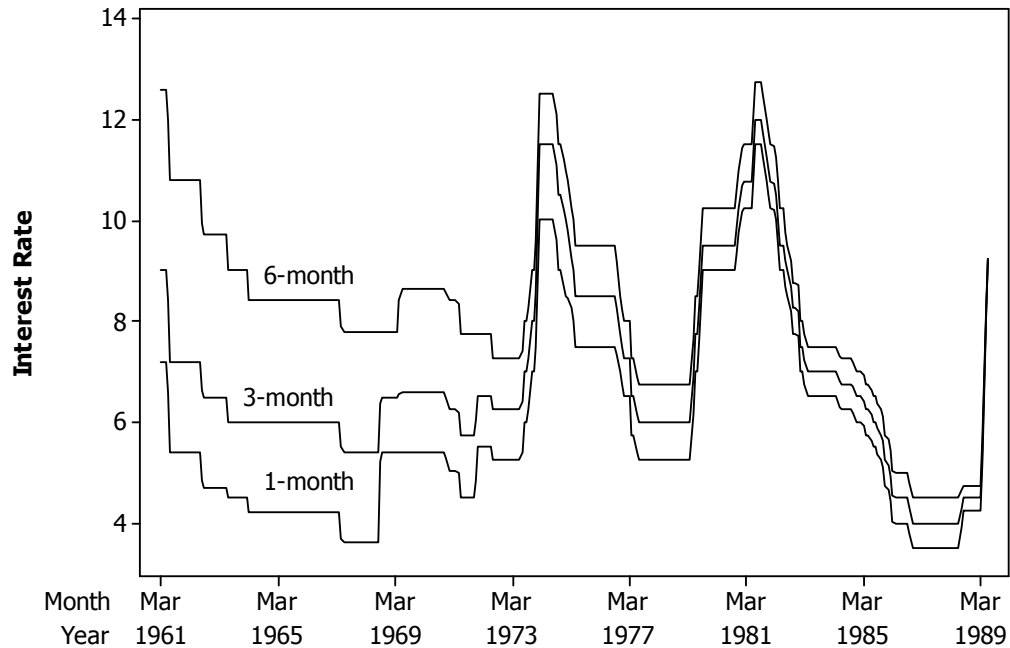


Figure 1-4. Monthly deposit interest rates in Taiwan: Mar 1961 to Jul 1989.

1.1.4 Case 4: the Fama-French Data

Fama and French (1992) investigated important factors that explain the cross-sectional variability of expected stock returns. They find the overall market factor, the size factor and the book-to-market equity factor are significantly priced.

These factors were constructed based upon financial economics theories. The size factor, SMB (small minus big), is the difference between the average returns on small-stock portfolio and the average return on big-stock portfolio. The book-to-market equity factor, HML (high minus low), is the difference between the average of the return on

high-BE/ME (book equity to market equity) portfolio and the average return on low-BE/ME portfolio.

Since the paper was published, Dr. Kenneth French continuously updates the factor data upon newly available stock returns and makes them available through his website. Figure 1-5 shows the three Fama-French factors from Nov 1942 to Jul 2008. Also available is the return on portfolios formed on each size decile (stocks are ranked according to size – market capitalization, and then divided into ten groups; portfolios are formed for each group).

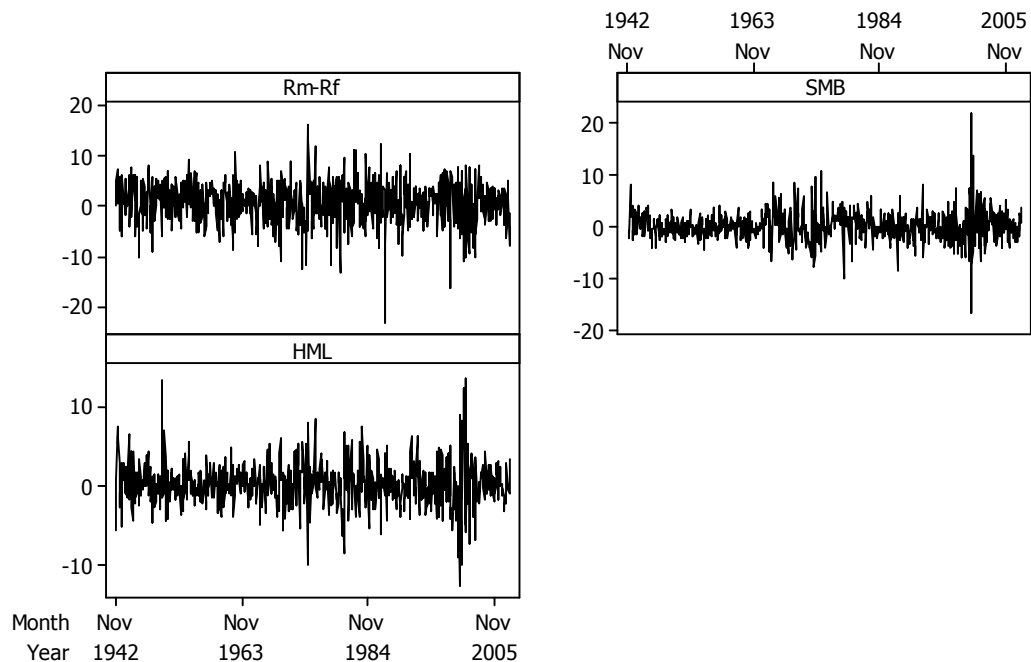


Figure 1-5. Fama-French factors from Nov 1942 to Jul 2008.

Stock market price/return is a commonly analyzed multivariate autocorrelated data set. In this context, what I am interested is how these theory-based factors possibly match empirically determined latent factors through some kind of dimension reduction method. I look into this question in Chapter 4.

1.2 An Overview of This Dissertation

Chapter 2 through Chapter 5 each treat an individual problem yet fit in the same scope of the dissertation topic. Chapter 2 deals with process monitoring of multivariate processes and in particular with the autocorrelation and non-stationarity issues that appears to be incompatible with traditional multivariate control charts and thus may seriously impact their performance. We suggest modeling processes with multivariate ARIMA time series models and propose two model-based monitoring charts. One monitors the predicted value and provides information about the need for mean adjustments. The other is a Hotelling's T^2 control chart applied to the residuals. The ARL performance of the residual based Hotelling's T^2 chart is compared to the observed data based Hotelling's T^2 chart for a group of first order vector autoregressive models. We show that the new chart in most cases performs well.

Chapter 3 deals with *Average Run Length* computation problem for autocorrelated processes. The ARL of conventional control charts is typically computed assuming temporal independence. However, this assumption is frequently violated in practical applications. Alternative ARL computations have often been conducted via time consuming and not necessarily very accurate simulations. In Chapter 3, we develop a class of Markov chain models for evaluating the run length performance of traditional control charts for autocorrelated processes. We show extensions from the simplest univariate AR(1) model to the general univariate ARMA(p, q) model, and the multivariate VARMA(p, q) time series case. The results of the proposed method are compared with simulation results and a discussion on computational accuracy and efficiency is provided.

In Chapter 4 I take on the question of how to improve the current dimension reduction methods for multivariate time series. In traditional multivariate literature, PCA is the standard tool for dimension reduction and factor retrieving. For autocorrelated process, however, PCA fails to take into account the time structure information. Thus it is doubtful that PCA is the best choice. I review several existing reduction methods and compare them with PCA. I propose to combine the PCA method and Box-Tiao method (Box and Tiao, 1977) and therefore get a two-step procedure which is demonstrated to be more efficient in retrieving underlying factors for autocorrelated multivariate processes. I also investigate its application in econometrics and financial asset pricing context.

In Chapter 5 we suggest a simple method for visualizing the results of PCA intended to complement existing graphical methods for multivariate time series data applicable for process analysis and control. The idea is to visualize multivariate data as a surface that in turn can be decomposed with PCA. The surface plots developed in this research are intended for statistical process analysis but may also help visualize economic data and in particular co-integration.

Chapter 2 is based on Huang and Bisgaard (2010); Chapter 3 is based on Huang and Bisgaard (2009); Chapter 4 is a working paper; Chapter 5 is based on Bisgaard and Huang (2008). This dissertation is largely problem motivated and emphasis is on how the proposed methodologies can solve the problems.

CHAPTER 2

MODEL-BASED MULTIVARIATE MONITORING CHARTS FOR AUTOCORRELATED PROCESSES

For more than half a century, control charts were designed for processes that were temporally independent. It was only until recent decades (Alwan and Roberts, 1988) that autocorrelation started to be recognized to greatly impact the performance of control charts. Complementary to existing literature, this chapter addresses the issue of autocorrelation for multivariate processes.

Autocorrelation or nonstationarity may seriously impact the performance of conventional Hotelling's T^2 charts. We suggest modeling processes with multivariate ARIMA time series models and propose two model-based monitoring charts. One monitors the predicted value and provides information about the need for mean adjustments. The other is a Hotelling's T^2 control chart applied to the residuals. The ARL performance of the residual based Hotelling's T^2 chart is compared to the observed data based Hotelling's T^2 chart for a group of first order vector autoregressive models. We show that the new chart in most cases performs well.

2.1 Statistical Process Control

It was a milestone for quality management when Shewhart (1931) introduced control charts in the early twentieth century. This technique has been proven a simple yet effective means of understanding data from real-world processes. The profound idea of Shewhart's control charts reside in two key concepts, common causes and special causes.

2.1.1 Common Causes and Special Causes

Common causes refer to routine variation which was called “controlled variation” by Shewhart. A process under control only displays this consistent pattern of variation. Shewhart attributed such variation to “chance” causes. Special causes refer to exceptional variation (uncontrolled variation) which is characterized by a pattern of variation that changes over time in an unpredicted manner. Shewhart attributed these unpredictable changes in the pattern of variation to “assignable” causes. See Alwan and Roberts (1988) for more discussion.

2.1.2 Control Charts Constituents: Charted Data

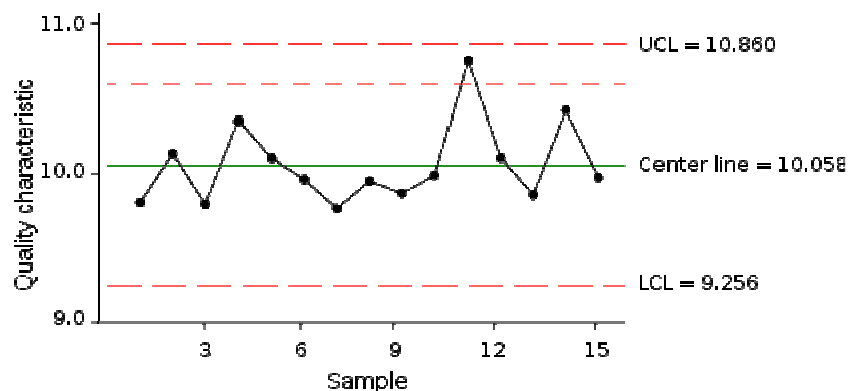


Figure 2-1. A typical control chart.

A typical control chart is shown in Figure 2-1. Resembling a time series chart, it charts the observations directly or a summary statistics if sub-grouping is used, e.g., subgroup mean, subgroup range. Sub-grouping refers to grouping the process outcomes into small samples. It could happen naturally, e.g., taking products from the same batch as a subgroup; or it could be done purposely because sub-grouping and taking average make “chance” variation smaller and thus a same-size mean shift is more easily

detectable. In multivariate cases, charted data could also be a vector based statistics, like the Hotelling's T^2 .

2.1.3 Control Charts Constituents: Control Limits

No matter what is being charted, a set of control limits will be imposed, as the Upper Control Limit (UCL) and Lower Control Limit (LCL) shown in Figure 2-1. The limits function to help judge whether the process is in control. In some sense, the charting process works like a series of hypothesis tests. When a new observation or a new charting statistics becomes available, it will be compared against the limits. Any statistic outside the limits is considered a sign of out-of-control and process check would be called. Tight control limits are preferable so they are sensitive to anything abnormal; however, too tight control limits have a high probability of making false alarms. Production frequently interrupted by false alarms will result in unnecessary operating cost. A tradeoff has to be made by choosing a set of effective and economical limits.

2.1.4 Average Run Length

When developing a control chart algorithm, there are two important questions concerning the performance. First, how often will there be false alarms while nothing actually has changed? Second, how quickly will we detect systematic change?

The answers to these two questions are similar to “Type I Error” and “power” respectively in hypothesis testing context. In control charts context, *Average Run Length* is the measure designated to answer them. By definition, ARL is how long on the average we will plot successive control charts points before we detect a point beyond the control

limits. Often, the answer to the first question is called “in-control ARL,” or denoted simply by ARL_0 while the answer to the second question is called “out-of-control ARL.”

2.1.5 Phase I and Phase II

There are two phases for control charts application. The retrospective Phase I study is a careful scrutiny of a sequence of past data. If based on a period when the process is deemed to be “well behaved,” Phase I data provides a reference for how an “in-control” process behaves. Statistical characteristics of the process are estimated in this phase and are later used to determine UCL and LCL for Phase II. Normally Phase I does not last long in real production. After the process is studied and understanding is gained, we switch to Phase II where the purpose is to monitor the process to maintain control.

2.1.6 The Family of Control Charts

Basic control charts include Individual Moving Range (IMR) chart, $\bar{x} - R$ chart, $\bar{x} - s$ chart, p chart, np chart, c chart and u chart. More recently there are *Cusum* and *EWMA* charts. For a more detailed discussion, see Montgomery (2008). This chapter is dedicated to a multivariate chart, Hotelling’s T^2 chart; more details are provided in section 2.4.

2.2 Motivation: the Ceramic Furnace Example

With the proliferation of computers, automatic sensor technology, communications networks and sophisticated software, data used for *statistical process control* of industrial processes are increasingly multivariate (vector-valued) and sampled

individually at high sampling rates. Typically data vectors arrive one by one at equal time intervals, e.g., every hour or every minute. Sub-grouping, as for example done with the traditional Shewhart $\bar{X} - R$ control chart, is often not convenient, useful or desirable. For such applications, the data will typically not only be cross correlated, but also autocorrelated especially if sampled quickly relative to the dynamics of the system being monitored. Indeed, the data stream used as input to process monitoring algorithms are often high dimensional vector time series.

We consider Case 1 from Chapter 1 as a motivational example. The original dimension of this problem was nine readings from nine locations, but for now, we consider only two adjacent time series, z_8 and z_9 . A sample of 100 hourly temperature observations from these two different locations of the furnace is shown in Figure 2-2.

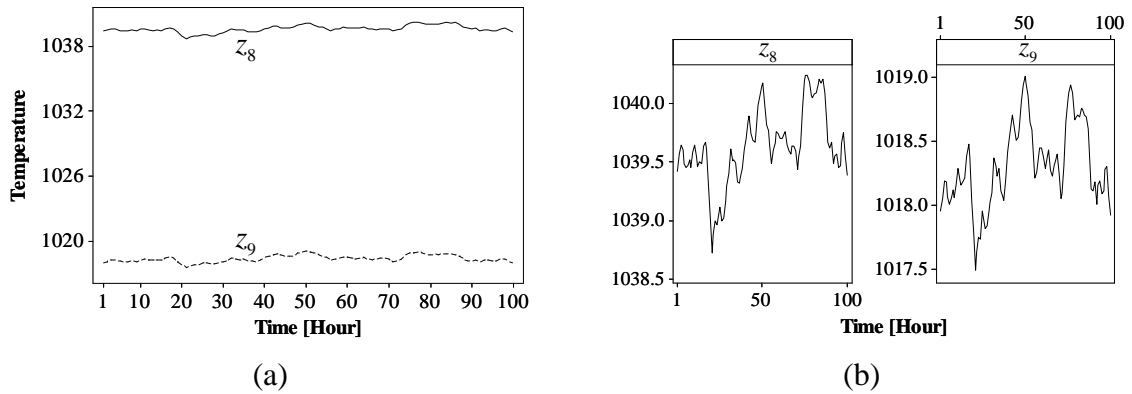


Figure 2-2. Temperature measurements at 2 adjacent locations of a ceramic furnace, (a) Combined display plotted on a common scale. (b) Separate displays on separate scales.

From a practical point of view, the most noticeable observation from Figure 2-2(a) is that the temperature variation is extremely small given the high level of the temperatures and the large scale of the furnace. We further see from Figure 2-2(b) that although the process seems to maintain a relatively stable level and show limited variability, it exhibits patterns and signs of non-randomness. Indeed, a furnace system as

large as this, has natural inertia and will exhibit high autocorrelation; so do most chemical plants. If we used the usual control charts and imposed Western Electric Rules we would find that this process appears to lack statistical control. Yet this furnace and in particular during this time period had very well behaved judged by the team of control engineers who monitored it. That raises the question: is this process in control? Can non-stationary process possibly be in control as well? Or more fundamentally, what do we mean by statistical control? Further, if this process could be considered in control, as was insisted by the control engineers, are the traditional control charts immediately applicable? Or could we devise a control chart more suitable?

2.3 A Discussion on “in Control in a Broader Sense” and the Chapter Overview

In the past it has often been assumed or implied that processes in a state of statistical control should be modeled as independently and identically distributed (I.I.D.) random variables. More recently this view has been challenged, for example, by Alwan and Roberts (1988) and Box and Paniagua-Quiñones (2007). By quoting Shewhart who pointed out that a state of statistical control implies predictable processes, Alwan and Roberts (1988) argued for a more flexible view whereby a process to which a time series model can successfully be fitted, implying predictability, should be considered “in control in a broader sense.”

Traditional control charts are outdated for autocorrelated processes and new charts are therefore sought after. For univariate cases, three approaches have been proposed. This first approach focuses on adjusting parameters and control limits to accommodate the autocorrelation. See, e.g., VanBrackle and Reynolds (1997). The second approach introduces a moving window of observations from the univariate

process and treats it as a multivariate problem using multivariate control charts. See, e.g., Krieger, Champ and Alwan (1992), Alwan and Alwan (1994), Apley and Tsung (2002), Jiang (2004). The third and the most popular approach, is a residual-based method, which is, by its name, to apply traditional control charts (Shewhart chart, EWMA chart, CUSUM chart) to residuals after fitting a time series model. This approach has been discussed by Alwan and Roberts (1988), Harris and Ross (1991), Montgomery and Mastrangelo (1991), Lin and Adams (1996), Vander Weil (1996), Box and Luceño (1997), Apley and Shi (1999), Rosolowski and Schmid (2003), and Box and Paniagua-Quñones (2007). Evaluation of the residual-based method has been studied by Wardell, Moskowitz and Plante (1992, 1994), Superville and Adams (1994), Runger, Willemain and Prabhu (1995), Kramer and Schmid (1997), Lu and Reynolds (1999a, 1999b, 2001). There were a number of other methods which do not fall into any of the abovementioned categories, such as a modified observation-based CUSUM scheme by Atienza and Ang (2002), a CUSUM-triggered CUSCORE chart that utilized the information contained in the dynamics of the fault signature as proposed by Shu et al. (2002), a multivariate control scheme on univariate feedback-controlled process by Apley and Tsung (2002), and a regression model-based control chart by Loredó et al. (2002).

Despite the large research effort spawned by the problem of autocorrelation, much of the previous work was based on stationary process and very little literature has explicitly treated nonstationarity. Moreover, although an extension of the univariate residual-based control chart to multivariate residual-based control chart has been proposed (see, e.g., Pan 2005), it remains unclear how practical this method is and how well it performs. This part of my dissertation thus attempts to address these two

remaining questions: first, how to extend the residual-based method to time-structured multivariate process, including the nonstationary cases; second, how the proposed control chart performs. Further, we demonstrate the practical value of our method with a case study.

The rest of this chapter is organized as following. In Section 2.4 we briefly review the relevant multivariate quality control literature, and in Section 2.5 we discuss problems with the standard Hotelling's T^2 chart applied to autocorrelated process. In Section 2.6 we draw a parallel to the univariate counterpart of the special cause chart and propose a residual based Hotelling's T^2 control chart for multivariate process; as a complementary chart, we also introduce a common cause chart which provides information for process adjustment. A revisit of the case of Furnace Data follows in Section 2.7. To further validate the generalized *special cause chart*, we evaluate in Section 2.8 the proposed special cause chart using ARL comparisons based on a comprehensive discussion of a first order vector autoregressive time series model with varying parameters. Lastly we provide conclusions and some final thoughts in Section 2.9.

2.4 Literature Review of Traditional Multivariate SPC

Hotelling's T^2 chart is a popular tool for monitoring multidimensional cross-correlated processes for continued process stability originally introduced by Hotelling, (1947); see also Hald (1951), Alt (1984) , Jackson (1985), Johnson and Wichern (2002), Mason and Young (2002), Kourti and MacGregor (1996); see Montgomery (2008) for a recent summary of previous work. Herein I limit my discussion to the situation where the T^2 chart is applied to individual observation, i.e., subgroup sample size equals 1. In

general, suppose $\mathbf{z}_t = (z_{t1}, \dots, z_{tk})'$ is a k -dimensional data vector. Suppose further that $\mathbf{z}_t \stackrel{iid}{\sim} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Hotelling T^2 at time t , is given by $T_t^2 = (\mathbf{z}_t - \bar{\mathbf{z}})' \mathbf{S}^{-1} (\mathbf{z}_t - \bar{\mathbf{z}})$ where $\bar{\mathbf{z}}$ is an estimate of the mean and \mathbf{S} is an estimate of the covariance matrix. In retrospective Phase I applications $t = 1, \dots, n$ and in prospective Phase II applications, $t > n$; see Woodall (2000) for a discussion of phases in control chart applications. Typically the estimates $\bar{\mathbf{z}}$ and \mathbf{S} are based on data from Phase I. In Phase I, \mathbf{z}_t is neither independent of the sample covariance matrix \mathbf{S} nor the sample mean $\bar{\mathbf{z}}$. Tracy et al. (1992) suggests $nT_t^2 / (n-1)^2$ is Beta distributed and showed that the upper control limit (UCL) for the T^2 statistics should be computed as $(n-1)^2 n^{-1} B_{1-\alpha, k/2, (n-k-1)/2}$. For Phase II application, Tracy et al. (1992) suggests $n(n-k)T_t^2 / k(n+1)(n-1)$ is F distributed and showed that the UCL for the T^2 statistics is $k(n+1)(n-1)n^{-1}(n-k)^{-1} F_{1-\alpha, k, n-k}$.

The estimation of the covariance matrix for individuals is an issue. In the past, it was standard to use the unbiased Maximum Likelihood estimator (MLE) \mathbf{S} . However, Sullivan and Woodall (1996) discussed this and compared several other estimators, including one suggested by Holmes and Mergen (1993) based on successive differences, which provides a more robust estimate against a sudden mean shift in the process. We discuss this issue below.

Besides the Hotelling's T^2 control chart, cumulative sum (CUSUM) control chart and exponentially weighted moving average (EWMA) control chart both have multivariate versions. Woodall and Ncube (1985), Crosier (1988) and Pignatiello and Runger (1990) proposed various multivariate CUSUM control charts. Lowry et al. (1992)

presented a multivariate EWMA control chart following the same scheme as the univariate EWMA. Wierda (1994) provides a useful overview of various methods. See also Sullivan and Woodall (1996) for an overview and comparison of several methods. Montgomery (2008) provides an up-to-date overview of the state of the art of multivariate control chart methods.

It is acknowledged that all these multivariate control methods are intended for data that comply with an assumption of temporal independence. However, if the data is autocorrelated, most of them either do not work well, or provide misleading results.

2.5 Standard Charts Applied to Autocorrelated Data

Blindly applying traditional control charts to autocorrelated processes can be harmful. The impact on the performance of classical SPC procedures on univariate processes has been studied by Johnson and Bagshaw (1974), Bagshaw and Johnson (1975), and Harris and Ross (1991). In this section we discuss briefly the impact of autocorrelation on Hotelling's T^2 control chart.

As discussed in Section 2.1, the ARL is typically used for evaluating the performance of a control chart method. Under the assumption of temporal independence, the T_t^2 at Phase II is independent. Thus the run length has a geometric distribution with the mean equal to $1/p$, where p is the probability that an individual T_t^2 falls outside of the control limits. However, if the data \mathbf{z}_t are autocorrelated, the individual's T_t^2 will also be autocorrelated. Thus the logical argument that leads to the geometric distribution is no longer valid.

To illustrate what may happen, consider a two dimensional first order vector autoregressive time series, VAR(1). Specifically, let Φ be a 2×2 coefficient matrix and $\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \mathbf{a}_t$ where it is assumed that $\mathbf{a}_t \stackrel{iid}{\sim} N(\mathbf{0}, \mathbf{I})$. Note that $\Phi = \mathbf{0}$ is equivalent to no autocorrelation. To simplify we assume that Σ_z is known even though only its estimate will be so in a Phase II study; this difference will not change our overall conclusion below. In Table 2-1 I compute the ARL0 (ARL when the process is in control) of Hotelling's T^2 on observed data for different values of Φ corresponding to different scenarios of autocorrelation. The UCL is fixed across all cases, chosen to give an ARL0 of 300 when the process is i.i.d.. The Markov chain method introduced in Chapter 3 is used for ARL computation.

Table 2-1. ARL0 With Different set of AR Coefficient Φ for Hotelling's T^2 .

Φ	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0.3 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0 \\ 0 & 0.3 \end{bmatrix}$	$\begin{bmatrix} 0.7 & 0 \\ 0 & 0.8 \end{bmatrix}$	$\begin{bmatrix} -0.7 & 0 \\ 0 & -0.5 \end{bmatrix}$	$\begin{bmatrix} -0.9 & 0 \\ 0 & -0.9 \end{bmatrix}$
ARL0	300.18	367.51	338.51	406.08	422.37	534.51

Note that the Φ matrices shown in Table 2-1 are all diagonal. However, this is not essential for our conclusion. Thus Table 2-1 indicates that it is difficult to control the ARL0 for the Hotelling's T^2 chart without knowing the underlying model.

The ARL is not the only problem; many useful techniques developed based on assuming temporal independence produce unreliable result for autocorrelated processes. In particular practitioners are cautioned against blindly applying standard statistical software; they may use parameter estimators that are seriously non-robust to temporal dependence. For example, Figure 2-3 shows a standard Hotelling's T^2 applied to the furnace data using the Holmes and Mergen (1993) covariance estimator, an estimator

robust to sudden step shift. Figure 2-3 shows numerous out-of-control signals. The UCL is computed through Phase I formula $UCL = (n-1)^2 n^{-1} B_{1-\alpha, k/2, (n-k-1)/2}$, with $n = 100$, $k = 2$ and $\alpha = 1/500$ so that ARL0 is 500.

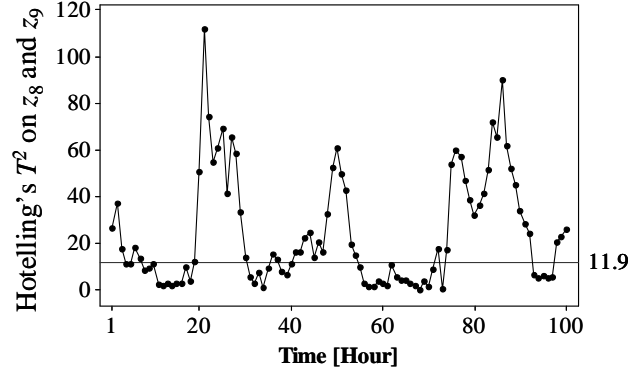


Figure 2-3. A T^2 chart using the Holmes and Mergen (1993) estimator for the covariance matrix.

The Holmes-Mergen's estimator of the variance-covariance matrix is,

$$\mathbf{S}_d = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \mathbf{d}_i \mathbf{d}_i' \quad (2.1)$$

where $\mathbf{d}_t = \mathbf{z}_{t+1} - \mathbf{z}_t$, $t = 1, \dots, n-1$ are successive differences, and n is the sample size. As indicated \mathbf{S}_d was intended to provide a robust estimate of Σ_z against sudden mean step changes, similar to that sometimes advocated in the univariate case where the moving range is a popular estimator of the variance; see Box and Luceño (1997), page 44 for comments. We acknowledge that this estimator never was intended to be used for autocorrelated processes. However, in the present case of positive autocorrelation, the most common type for continuously monitored processes, \mathbf{S}_d seriously underestimates the variance-covariance matrix, resulting in the excessive number of false alarms.

A simple ad hoc solution to the problem of positive serial correlation is to inflate the variance-covariance matrix by an appropriate scalar factor using an empirical reference distribution for benchmarking. For example, a test data set of appropriate length from a time segment where the process is deemed stable by knowledgeable process engineers can be used as a reference distribution. The process engineers can then either adjust the control limits to make sure no false alarms are recorded during this period or inflate the variance-covariance matrix by a scalar factor to accomplish the same. Both approaches are simple and work in practice; see also Kourti and MacGregor (1996). However, this remains an ad hoc approach.

Another way to fix the problem of autocorrelation is to use the MLE, \mathbf{S} directly. That may work reasonably well for Hotelling's T^2 when we have a sufficiently long time series for the Phase I estimation of the covariance matrix; see Johnson and Langeland (1991).

It is even more dangerous to use Hotelling's T^2 when the time series data is non-stationary. Box and Luceño (1997) pointed out that most real world processes are non-stationary because of the second law of thermodynamics. The furnace data, a representative of an industry process, provides good evidence for their argument. Analyzing the Phase I data, we find that the furnace temperature exhibit patterns modeled well by a Vector IMA(1,1) model. If the data from the initial 100 observations shown in Figure 2-2 are assumed to be i.i.d. normal and used as a Phase I period, we would have found the process hopelessly out of control in the following three month, as showed in Figure 2-4. The UCL of Individual chart (Figure 2-4a and Figure 2-4b) is determined by the traditional formula of $\hat{\mu} + 3\hat{\sigma}$ where $\hat{\mu}$ and $\hat{\sigma}$ are computed from Phase I, the first

100 observations; the LCL is determined by $\hat{\mu} - 3\hat{\sigma}$. The UCL of Hotelling's T^2 (Figure 2-4c) is computed through Phase II formula $k(n+1)(n-1)n^{-1}(n-k)^{-1}F_{1-\alpha,k,n-k}$, with $n = 1274$, $k = 2$ and $\alpha = 1/500$ so that ARL0 is 500.

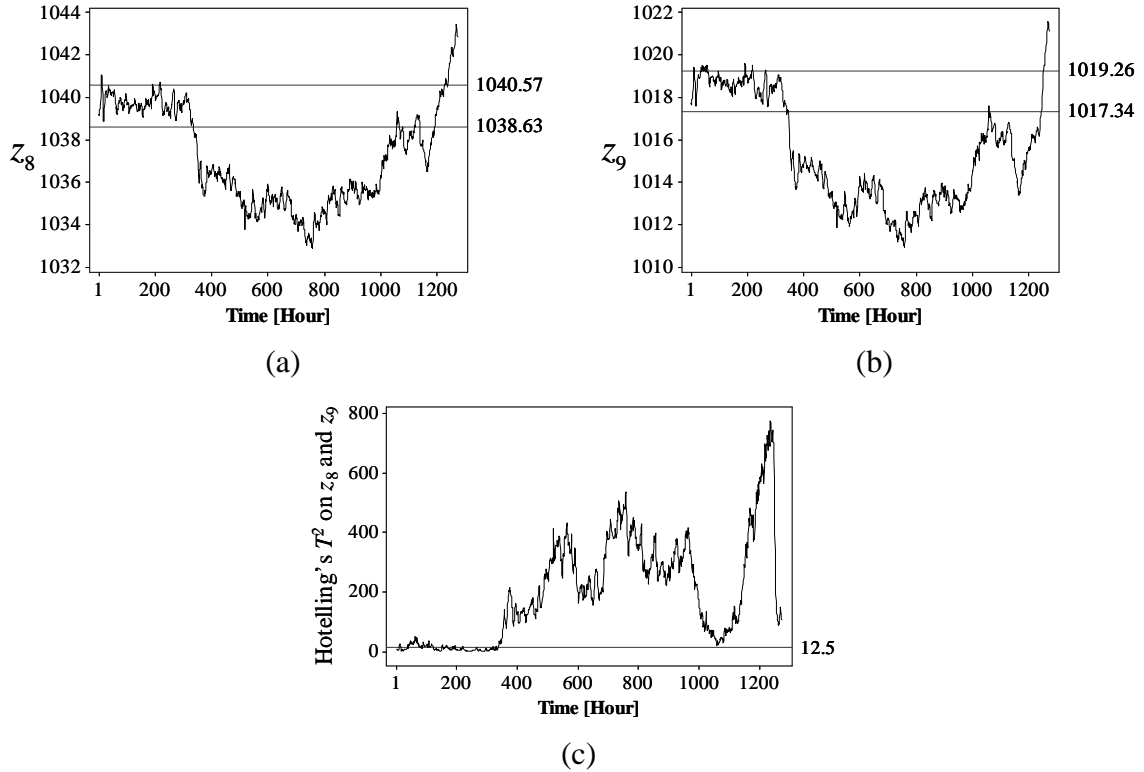


Figure 2-4. Control charts on Phase II production: (a) Shewhart chart on z_8 ; (b) Shewhart chart on z_9 ; (c) Hotelling's T^2 on both dimensions. The control limits derived from Phase I data.

Specifically, we see from Figure 2-4 (a) and (b) that both z_8 and z_9 exhibit a step decline around observation 320 and then climb back up around observation 1170. Most likely both process shifts are the effect of process adjustments by the process engineers. Indeed temperature changes of an inherently non-stationary process is normal and should not necessarily be a cause for alarm. Rather it is a cause for a routine adjustment, for example, production switch from making clear glass to making tinted glass. Alarms

should be reserved for something extraordinary, a systems change shown up in the residual chart.

2.6 Proposed Method: The Multivariate Special and Common Cause Charts

Echoing earlier work by Box and Jenkins (1968), Box, Jenkins and MacGregor (1974), and Berthouex, Hunter and Pallesen (1978), Alwan and Roberts (1988) suggested that we consider monitoring two charts, a “common cause” chart and a “special cause” chart. The common cause chart is a chart of the fitted values or one-step-ahead forecasts from a time series model. The special cause control chart is a time series plot on the residuals. The common cause chart provides an estimate of the current mean assuming the process continues to follow the same stationary or non-stationary time series model as in Phase I, whereas the special cause chart of the residuals triggers an alarm whenever a deviation from our expectation is too large, indicating that a special cause or system change may have occurred. The same philosophy has been expressed by Box (1991), Box and Luceño (1997), and Box and Paniagua-Quiñones (2007).

Based on fitting a multivariate time series model, we extend the univariate special cause and common cause chart to multivariate horizon.

To construct a special cause chart we suggest computing the following scalar quantity, $\hat{T}_t^2 = \hat{\mathbf{a}}_t' \mathbf{S}_a^{-1} \hat{\mathbf{a}}_t, t = 1, K$ where $\mathbf{S}_a = \hat{\Sigma}_a$ obtained by fitting an appropriate time series model to $\{\mathbf{z}_t\}$ and $\hat{\mathbf{a}}_t = [\hat{\Theta}(B)]^{-1} \hat{\Phi}(B) \tilde{\mathbf{z}}_t$ is the filtered data. This residual-based T^2 chart is generalization of the univariate special-cause chart proposed by Alwan and Roberts (1988) and Box and Paniagua-Quiñones (2007).

Our proposed procedure for monitoring a multivariate process is as follows:

Step 1. (Phase I): Select a set of data from a time period that by the operators familiar with the process is deemed “well behaved.” The data size n should typically be larger than what is usually required for temporally independent data. How large depends on how autocorrelated the process is. For moderate positive autocorrelation a minimum of 100 seems a good rule of thumb but larger sample size is often needed. This data set should then be plotted in all possible ways to check for outliers and other unusual patterns. If needed we may also apply transformations to obtain approximate multivariate normality; see Johnson and Wichern (2002).

Step 2. (Phase I): After the preliminary screening of the data we proceed with the iterative time series modeling scheme of identification, estimation and checking advocated by Box and Jenkins (1976). Tiao and Box (1981) and Peña, Tiao and Tsay (2001, Chapter 11) provide good overviews of the model building process for multivariate time series.

Step 3. (Phase I): If a time series model fits the data well, the Phase I data can be considered to represent a predictable process and we proceed to Step 4. If not, it would be an indication that the process is not in statistical control according the extended definition of being predictable. Thus we need to investigate why the process is not predictable.

Step 4. (Phase II): Apply the filter $\hat{\mathbf{a}}_t = [\hat{\boldsymbol{\Theta}}(B)]^{-1} \hat{\boldsymbol{\Phi}}(B) \tilde{\mathbf{z}}_t$ to new incoming Phase II data. Construct a chart for tracking $\hat{T}_t^2, t = n + 1, \dots$. This is the special cause chart. Points falling above the UCL indicate the advent of special causes resulting in large deviations from the expected value. A special cause is a fundamental change to the system generating the time series. Hence it may either be a change in the distribution of \mathbf{a}_t , or to

the structure of time series model specified by $\Phi(B)$ and $\Theta(B)$ and the parameters in those polynomials. χ^2 distribution is suggested for control limit computation.

Step 5. (Phase II): In addition to the special cause chart, we suggest charting each dimension of the fitted value $\hat{\mathbf{z}}_{t+1}$ either individually or jointly. These are common cause charts. Note that for these charts we do not superimpose control limits. Rather we may impose *action limits* to indicate when the predicted vector has moved too far from the target and the process is in need of an adjustment.

With this procedure aside, let us go back to our previous question: can a nonstationary process possibly in control or not? This question is critically important because many real-world processes, such as our Furnace, exhibit nonstationary patterns. Keeping in mind Alwan and Roberts's (1988) argument on "in control in a broader sense," we answer this question with a resounding Yes. As long as the nonstationary process can be fitted by a time series model and follows the same model throughout Phase II, the only difference of controlling this process is to have a common cause chart that behaves nonstationary. The special cause chart on the residuals will not sound an alarm as long as the process does not meander away from what the nonstationary model could explain.

2.7 Revisiting the Furnace Data

We now use the furnace data introduced in Section 1.1.1 and plotted in Figure 2-2, as our Phase I data. The primary purpose of a retrospective study of the data, is to understand the process and test whether the process can be considered in control here interpreted in the wider Shewhart sense of being predictable. For the current data we

found that a Vector IMA(1,1) model fit the data well and that the residuals are well behaved. The fitted model is,

$$(1-B)\mathbf{z}_t = \left(I - \begin{bmatrix} -0.123 & -0.358 \\ -0.457 & -0.097 \end{bmatrix} B \right) \mathbf{a}_t$$

The autocorrelation functions (ACF) of the residuals after fitting a vector IMA(1,1) model to the first 100 observation shown in Figure 2-5. The ACF's indicates that this model fits well for the Phase 1 data. Thus we proceed to Phase II.

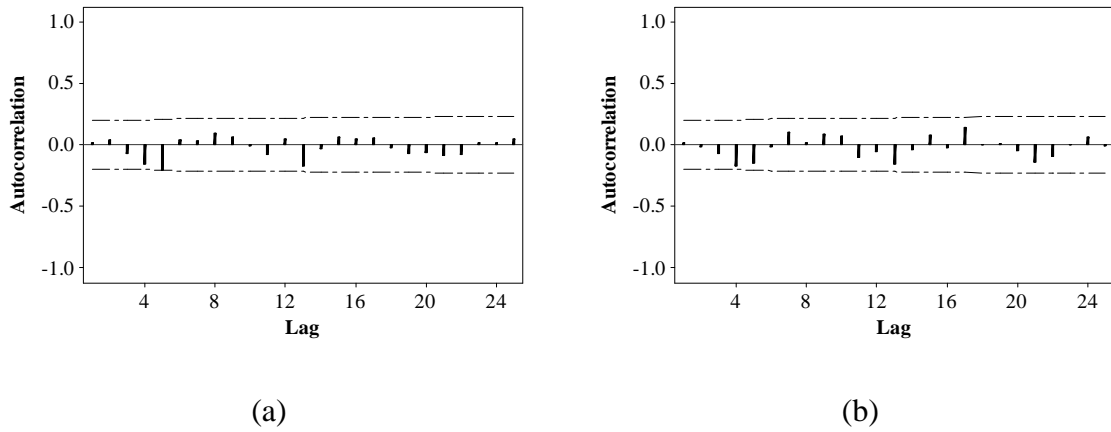


Figure 2-5. The autocorrelation of the residuals after fitting the IMA model with 5% significance limits for the autocorrelations: (a) residuals from z_8 ; (b) residuals from z_9

In Phase II, residuals and fitted values need to be computed recursively as time passes using the model and fixed parameter values as estimated in Phase I. Figure 2-6 shows the residual based special cause chart for the Phase II data. The UCL is determined through $UCL = \chi^2(1-\alpha, k)$, where $k=2$ and $\alpha=1/500$ so that ARL0 is 500. Note the $\chi^2(k)$ approximation is very rough because none of $\hat{\Phi}(B)$, $\hat{\Theta}(B)$ and $\hat{\Sigma}_a$ is perfect estimate. We suspect this is a major reason that causes some false alarms and do consider a more accurate UCL part of our future work. For now we would recommend ignoring slightly out-of-control points because UCL has been underestimated. Although not

perfect, compared with Figure 2-4(c), this chart provides an impression more consistent with how the process engineers considered this period of time.

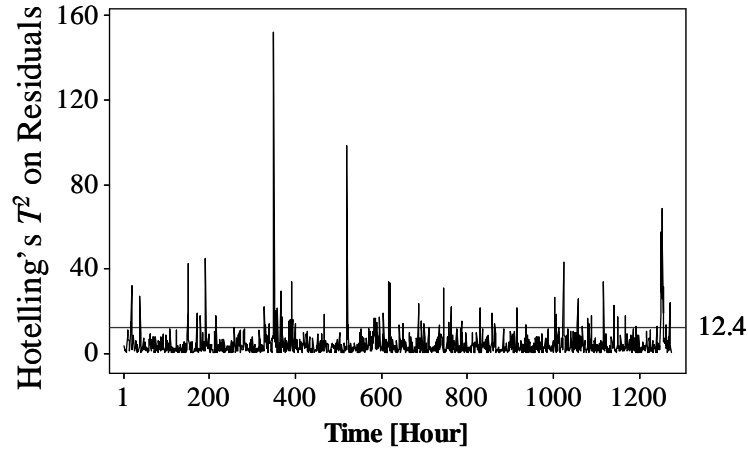


Figure 2-6. Hotelling's T^2 on residuals after fitting the Vector IMA(1, 1) model.

As already indicated, the engineers from time to time adjusted the process during the Phase II period, mostly, we surmise, by eyeballing something akin to the fitted values. Such adjustments are natural and sometimes necessitated because of materials changes, environmental changes, and occasional momentary power failures. Figure 2-7 overlays z_8 , the common cause charts of the fitted values \hat{z}_8 (\hat{z}_9 would be similar and thus omitted) and the special cause chart of Hotelling's T^2 on the residuals. Both z_8 and \hat{z}_8 are scaled so that they can be shown and compared to the special cause chart. On the special cause chart we see three large spikes at time 349, 520 and 1254 indicating out-of-control situations. Around these three time points, as indicated by the three bold windows, the process experienced a sharp plunge, a noise spike and a sudden increase. The first and third significant alarms most likely occurred when the engineers decided to adjust the process to a new target temperature, evidenced by the steep trends around those points. Such adjustments are a normal part of operating of the process. With open

communication between the process control engineers and the quality control engineers, each party should know the reason and not be alarmed. Only the second alarm around observation 520 looks like a special cause. The fitted value at this point is consistent with neither of its neighbors. Given the large inertia of the furnace, a more than one-degree temperature drop in one hour and an immediate return in the next hour is unusual. Thus this is more likely an outlier possibly caused by a thermo couple misfire. In fact if we plot ∇z_g , the first difference of z_g , as shown in Figure 2-8, the spike captured in second window of Figure 2-7 results in a first difference significantly outside its empirical reference distribution, as circled in both charts of Figure 2-8. This further indicates that the second special cause might be due to an outlier.

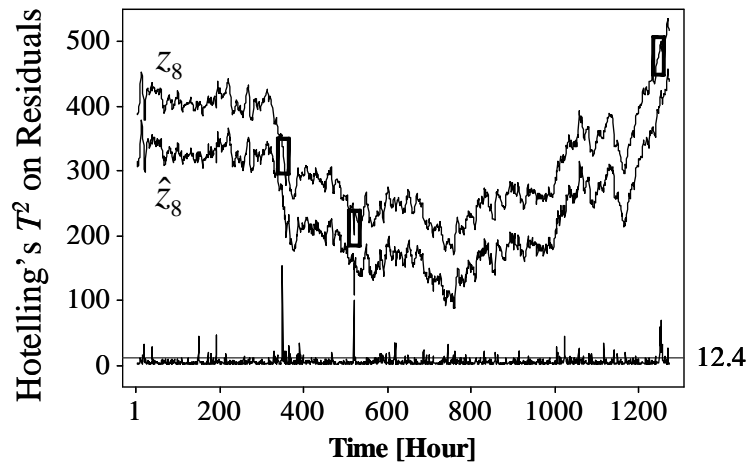


Figure 2-7. Special cause chart with z_g and fitted values \hat{z}_g imposed.

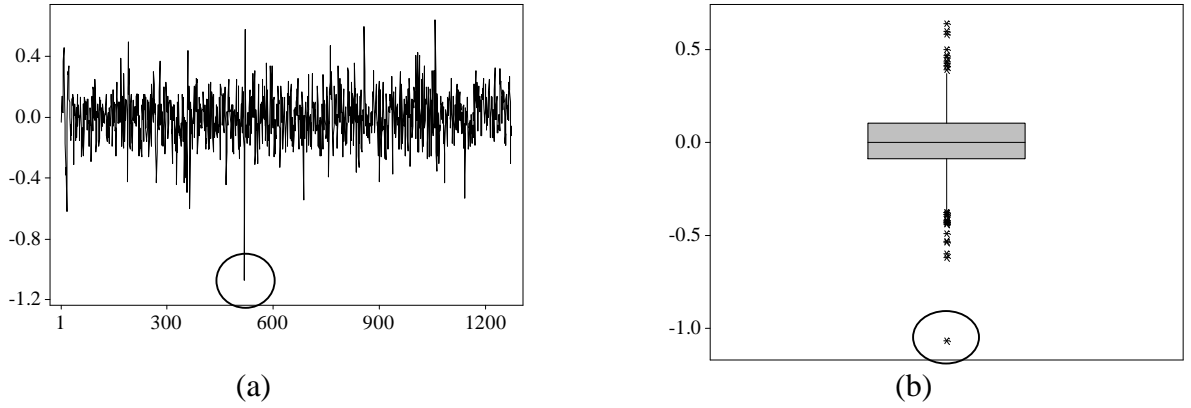


Figure 2-8. A Time series plot (a) and a box plot (b) of ∇z_{8t} .

The common cause chart of \hat{z}_8 in Figure 2-7 serves as a compass for the control engineers. Depending on it, they are able to know where the process is relative to a given target temperature, when it needs to be adjusted, and by how much. The same common cause chart should be made for \hat{z}_9 . However, due to the high correlation between z_8 and z_9 , it is not shown. When the fitted model is stationary and the process is of high dimension, a Hotelling's T^2 like chart (a generalized Euclidean distance measures) of fitted values is recommended. Note the variance-covariance matrix of the fitted process needs to be used, and it is not defined for non-stationary processes.

2.8 Run Length Comparisons

For non-stationary process, the Hotelling's T^2 chart applied directly to the observed data is neither defined nor meaningful. Thus we confine the ARL performance comparison for the Hotelling's T^2 charts to stationary process. The performance of the special-cause chart introduced by Alwan and Roberts (1988), the univariate counterpart to our proposed generalization, has been discussed by Wardell et al. (1992, 1994). For univariate process, they showed that the residual-based special cause chart is not

necessarily the best chart to use for every type of autocorrelation. In general, the residual chart outperforms the Shewhart chart only when the data are negatively autocorrelated. Unfortunately, this is less common in industrial practice, making the residual-based chart less appealing for practical use. Should the same be true for the multivariate extension, there may be little motivation for the use of the proposed residual-based Hotelling's T^2 . However, as we show below, the multivariate case is more complicated. Specifically, the performance of the residual based Hotellings T^2 depends on the parameter matrices as well as the direction and size of a mean shift.

To investigate the complexity of the situation consider a 2-dimension VAR(1) model $\mathbf{z}_t - \boldsymbol{\mu}_t = \boldsymbol{\Phi}(\mathbf{z}_{t-1} - \boldsymbol{\mu}_t) + \mathbf{a}_t$ where $\mathbf{a}_t \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_a)$. Without loss of generality, we assume $\boldsymbol{\Sigma}_a = \mathbf{I}_{2 \times 2}$. The structure of this time series model depends entirely on $\boldsymbol{\Phi}$ with $\boldsymbol{\Sigma}_z$ being determined through $\boldsymbol{\Phi}$ (Reinsel 1997, p. 30). The situation can therefore be analyzed by investigating the performance of the two Hotelling's T^2 charts for different values of $\boldsymbol{\Phi}$ and the combination of different size and direction of the mean shift. We base our discussion on eigenvalues of $\boldsymbol{\Phi}$, which are invariant to transformation on \mathbf{z}_t , see APPENDIX A. Let

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}.$$

The two eigenvalues are

$$\lambda_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

where, $\phi_1 = \phi_{11} + \phi_{22}$ and $\phi_2 = \phi_{12}\phi_{21} - \phi_{11}\phi_{22}$.

For stationarity we require that $|\lambda_i| < 1, i = 1, 2$; see Reinsel, p. 29. Thus ϕ_1 and ϕ_2 must satisfy: $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$ and $-1 < \phi_2 < 1$. This region is shown in Figure 2-9.

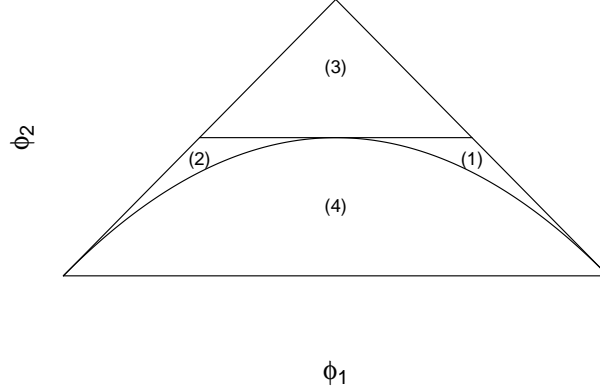


Figure 2-9. Triangular Region of ϕ_1 and ϕ_2 .

The stationary region can be further divided into four sub-regions according to whether the value of the eigenvalues $\lambda_i, i = 1, 2$: (1) both positive, (2) both negative, (3) one positive one negative or (4) two imaginary eigenvalues. Note with eigenvalues fixed, Φ is not fixed. Representative examples of these four cases are presented in Table 2-2 as Case 1 to Case 4. For each of these four cases, we examine examples of the Phase II ARL performance under a set of mean shift scenarios given by

$$\boldsymbol{\mu}_t = \begin{cases} \mathbf{0} & t < T \\ \boldsymbol{\delta} & t \geq T \end{cases}$$

where $\boldsymbol{\delta} = a \cdot (\cos \alpha, \sin \alpha)'$, $\alpha = i\pi / 6$ with $i = 0, 1, \dots, 5$ (due to the symmetric density function of \mathbf{Z}_T , we need not consider cases from $i = 6$ to 11), and the scalar a varies so that the non-centrality parameter $\boldsymbol{\delta}'\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\delta}$ takes values of $0.5j$ with $j = 0, 1, \dots, 9$. In other words, a and α represent the size and direction of the shift vector, respectively. Pan (2005) made a thorough discussion on three types of parameter shift without providing ARL computation.

Table 2-2. Examples examined. “++” means both eigenvalues are positive.

Case	1: ++	2: --	3: - +	4: imaginary	5: ++	6: ++
$\phi_{1,2}$	1, -0.16	-0.5, -0.031	-0.334, 0.334	0.833, -0.667	1, -0.16	1, -0.16
$\lambda_{1,2}$	0.2, 0.8	-0.073, -0.427	-0.768, 0.435	$0.417 \pm 0.702i$	0.2, 0.8	0.2, 0.8
Φ	$\begin{bmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{bmatrix}$	$\begin{bmatrix} -0.25 & 0.25 \\ 0.125 & -0.25 \end{bmatrix}$	$\begin{bmatrix} 0.167 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.667 \\ -0.75 & 0.333 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 1 \\ 0.09 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.692 & 0.048 \\ 1.115 & 0.308 \end{bmatrix}$

Calculating the ARL for Hotelling’s T^2 chart based on observed data is not straight forward because T_t^2 is autocorrelated. Our computation is through the Markov chain method discussed in Chapter 3 of this dissertation. Computing the ARL for the residuals-based Hotelling T^2 chart is discussed in APPENDIX B.

Table 2-3 provides ARL comparisons for Case 1 through Case 4. The control limits are calibrated to make ARL0 approximately 300. For each case and same mean shift, the ARL is presented in bold face whenever the residual-based chart performs better than the observed data chart. It can be seen that in Case 1, the residual-based chart has larger ARL’s for all scenarios. In Case 2, the situation is the opposite with the residual-based chart performing significantly better. The same is true for Case 4. In Case 3, the residual-based chart performs mostly better except for shifts in directions around $\pi/6$.

These four cases obviously do not provide a base for general inference. However, they suggest that the residual based chart has potentially better performance in some realistic scenarios. Indeed this conclusion is consistent with other cases we experimented with but not shown here.

Table 2-3. ARL's comparison for four cases. NC= Noncentrality. "Z" is the Hotelling's T^2 chart based on the observed data, "A" is the residuals based Hotelling's T^2 chart.

(a) Case 1

NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	300.40	300.00	300.40	300.00	300.40	300.00	300.40	300.00	300.40	300.00	300.40	300.00
0.5	113.47	163.34	133.51	226.73	106.81	226.73	113.32	163.34	133.54	144.29	106.51	144.29
1	62.16	103.97	79.29	175.04	56.49	175.04	62.36	103.97	79.05	86.48	56.67	86.48
1.5	39.68	71.80	53.06	136.41	35.69	136.41	39.87	71.80	53.00	57.76	35.62	57.76
2	27.77	52.14	38.51	106.60	24.59	106.60	27.80	52.14	38.33	41.21	24.60	41.21
2.5	20.42	39.17	29.27	83.21	18.08	83.21	20.67	39.17	29.24	30.76	18.04	30.76
3	15.77	30.15	23.12	64.74	13.88	64.74	15.82	30.15	23.09	23.73	13.81	23.73
3.5	12.70	23.64	18.56	50.16	10.96	50.16	12.57	23.64	18.58	18.79	10.99	18.79
4	10.30	18.81	15.27	38.69	8.90	38.69	10.28	18.81	15.29	15.18	8.94	15.18
4.5	8.57	15.14	12.83	29.71	7.47	29.71	8.53	15.14	12.92	12.48	7.46	12.48

Continued on next page

Continued

(b) Case 2

NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	300.90	300.00	300.90	300.00	300.90	300.00	300.90	300.00	300.90	300.00	300.90	300.00
0.5	107.37	64.20	108.61	92.22	107.68	92.66	107.27	65.90	107.46	46.48	107.53	45.49
1	56.72	28.25	56.71	45.67	56.71	45.99	56.57	29.23	56.61	18.87	56.71	18.38
1.5	34.75	16.01	35.13	27.34	35.23	27.56	35.01	16.63	34.95	10.45	34.94	10.17
2	23.57	10.45	23.83	18.24	23.70	18.41	23.68	10.87	23.93	6.86	23.65	6.67
2.5	17.06	7.48	17.19	13.09	17.11	13.22	17.12	7.79	17.15	5.02	17.02	4.89
3	12.91	5.72	13.05	9.90	12.84	10.01	12.85	5.96	13.01	3.95	12.84	3.86
3.5	10.03	4.60	10.23	7.80	9.97	7.89	10.09	4.79	10.20	3.28	9.98	3.21
4	8.09	3.84	8.23	6.34	8.02	6.42	8.04	3.99	8.23	2.84	7.99	2.78
4.5	6.67	3.30	6.84	5.30	6.60	5.36	6.69	3.43	6.81	2.52	6.54	2.47

Continued on next page

Continued

(c) Case 3

NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	300.82	300.00	300.82	300.00	300.82	300.00	300.82	300.00	300.82	300.00	300.82	300.00
0.5	109.73	86.93	107.83	184.11	109.55	73.33	109.65	18.17	108.40	9.98	110.09	21.15
1	59.81	41.99	58.03	125.87	59.43	33.45	59.74	6.66	58.67	3.92	58.99	7.75
1.5	37.82	24.71	36.76	91.56	37.54	19.16	37.76	3.90	36.79	2.61	37.67	4.44
2	25.87	16.26	25.36	69.33	25.83	12.46	25.85	2.85	25.09	2.13	25.80	3.16
2.5	18.76	11.52	18.30	54.00	18.72	8.81	18.75	2.34	18.34	1.89	18.86	2.54
3	14.22	8.61	13.85	42.94	14.14	6.62	14.33	2.05	13.89	1.74	14.25	2.18
3.5	11.12	6.71	10.73	34.70	11.05	5.21	11.01	1.86	10.78	1.64	11.00	1.96
4	8.76	5.40	8.60	28.38	8.73	4.24	8.83	1.73	8.54	1.55	8.78	1.80
4.5	7.17	4.47	6.95	23.46	7.14	3.56	7.18	1.63	6.97	1.47	7.11	1.68

Continued on next page

Continued

(d) Case 4

NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	301.08	300.00	301.08	300.00	301.08	300.00	301.08	300.00	301.08	300.00	301.08	300.00
0.5	109.11	47.18	107.87	41.11	109.10	37.67	109.84	38.92	107.71	45.27	109.30	50.40
1	59.17	18.65	58.03	15.87	59.17	14.30	59.61	14.74	58.01	17.53	59.00	20.05
1.5	37.43	9.84	36.66	8.42	37.32	7.57	37.49	7.71	36.71	9.08	37.41	10.47
2	25.81	6.08	25.20	5.30	25.75	4.79	25.72	4.80	25.13	5.50	25.76	6.33
2.5	18.75	4.16	18.20	3.72	18.65	3.39	18.72	3.34	18.34	3.70	18.75	4.23
3	14.18	3.07	13.81	2.83	14.09	2.60	14.24	2.52	13.78	2.70	14.06	3.04
3.5	11.05	2.40	10.73	2.27	10.95	2.12	11.04	2.02	10.68	2.10	10.92	2.33
4	8.81	1.97	8.57	1.91	8.72	1.80	8.80	1.70	8.57	1.72	8.73	1.88
4.5	7.14	1.68	6.92	1.66	7.11	1.58	7.17	1.49	7.01	1.48	7.13	1.59

It is interesting to note that the residual-based method is not favorable in Case 1 when both eigenvalues are positive. This case would appear to be reminiscent of the univariate AR(1) model. For the univariate model the eigenvalue can be interpreted as the autocorrelation. For the univariate case Wardell et al. (1992, 1994) showed that with positive autocorrelation the special cause chart (residual-based chart) performs poorly relative to a traditional Shewhart chart on the observed data. However, it would be rash to conclude that residual-based Hotelling's T^2 always will perform poorly when both eigenvalues are positive. Indeed this inference is wrong. To show this we provide two counter examples. In both case we fix $\phi_1 = 1$ and $\phi_2 = -0.16$. The eigenvalues are the same as in Case 1. However we chose two different realizations of Φ , shown as Case 5 and 6 in Table 2-2.

Table 2-4 shows ARL comparisons for Case 5 and Case 6. We see that in both cases the residual-based Hotelling's T^2 performs better for most mean shift scenarios. For example for Case 5, in some directions such as $\pi / 6$ the residual-based method has a larger ARL when there a small shift. However, as the size of the shift increases, the ARL decreases quickly and even becomes smaller than the observed data based method. This situation echoes what Wardell et al. (1992, 1994) found for univariate case, namely that the residual based method is more sensitive to large shifts.

Table 2-4. ARL's comparison for Case 5 and 6. NC= Noncentrality. "Z" stands for observed data based Hotelling's T^2 chart, "A" stands for residuals based Hotelling's T^2 chart.

(a) Case 5

NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	300.45	300.00	300.45	300.00	300.45	300.00	300.45	300.00	300.45	300.00	300.45	300.00
0.5	113.60	102.99	113.98	223.97	99.88	104.01	114.19	81.01	113.66	73.86	100.12	73.74
1	62.49	50.12	62.54	159.72	52.70	53.29	62.27	38.30	61.64	33.92	52.43	33.63
1.5	39.79	27.62	39.51	108.03	32.57	32.01	39.93	22.38	39.42	19.57	32.62	19.20
2	27.71	16.19	27.48	69.51	22.20	21.03	27.79	14.75	27.71	12.85	22.31	12.43
2.5	20.44	9.88	20.37	42.86	16.18	14.65	20.54	10.52	20.37	9.19	16.22	8.74
3	15.88	6.25	15.71	25.54	12.34	10.64	15.80	7.95	15.71	6.99	12.33	6.52
3.5	12.62	4.11	12.49	14.88	9.79	7.99	12.57	6.28	12.43	5.56	9.76	5.09
4	10.27	2.84	10.19	8.60	7.92	6.17	10.25	5.13	10.19	4.59	7.96	4.11
4.5	8.55	2.09	8.42	5.05	6.58	4.87	8.53	4.31	8.41	3.90	6.63	3.43

Continued on next page

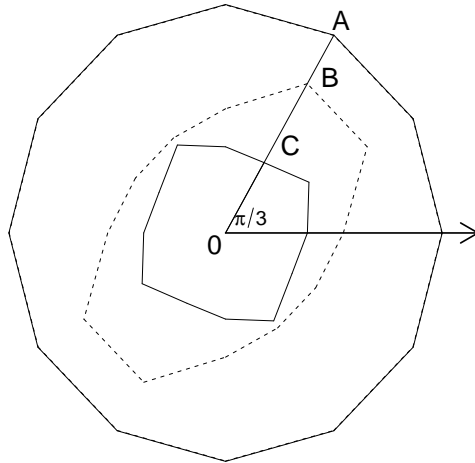
Continued

(b) Case 6

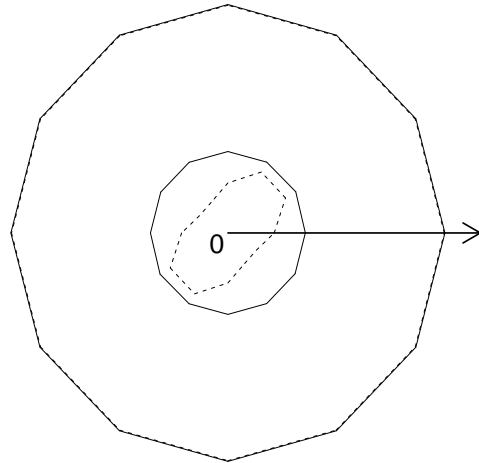
NC	0		$\frac{\pi}{6}$		$\frac{\pi}{3}$		$\frac{\pi}{2}$		$\frac{2\pi}{3}$		$\frac{5\pi}{6}$	
	Z	A	Z	A	Z	A	Z	A	Z	A	Z	A
0	300.07	300.00	300.07	300.00	300.07	300.00	300.07	300.00	300.07	300.00	300.07	300.00
0.5	103.48	69.88	150.38	97.70	98.27	235.86	103.05	70.37	149.71	59.71	97.97	62.05
1	54.89	31.48	93.27	48.35	51.20	162.30	54.59	30.52	93.30	25.67	50.98	27.07
1.5	34.26	17.98	64.89	28.14	31.67	98.77	34.46	16.29	64.54	14.38	31.72	15.32
2	23.60	11.73	47.92	17.90	21.64	54.39	23.63	9.72	48.03	9.32	21.62	10.02
2.5	17.29	8.36	37.27	12.06	15.71	27.75	17.36	6.25	37.14	6.65	15.71	7.20
3	13.18	6.34	29.41	8.47	12.01	13.49	13.26	4.26	29.70	5.07	11.98	5.54
3.5	10.49	5.04	24.19	6.15	9.50	6.49	10.49	3.07	24.13	4.07	9.52	4.48
4	8.53	4.16	19.94	4.61	7.71	3.30	8.58	2.33	19.78	3.39	7.71	3.77
4.5	7.09	3.53	16.92	3.55	6.45	1.92	7.08	1.86	16.81	2.92	6.43	3.26

To visualize the ARL comparisons, we have summarized part of the data from Table 2-3 and Table 2-4 in Figure 2-10. It is shown in polar coordinates how the ARL changes as a function of the shift size and direction. The solid circles represent the observed data based method and the dashed circles the residual-based method. Only cases with non-centrality parameters 0 and 0.5 are shown. Take Figure 2-10(a) on Case 1 for example. When the non-centrality parameter is 0 (no mean shift), the ARL's of both methods are 300 no matter of the direction of the shift. Thus the solid and dashed circles overlap and the radius is 300, shown as the outer circle. Point A on this circle, $(300, \pi/3)$, represents that ARL equals 300 when a shift takes place at angle $\pi/3$ with non-centrality 0 (essentially no mean shift). When the non-centrality parameter increases to 0.5, implying a mean shift, both circles shrink. When the shift direction equals $\pi/3$, the observed data based method shrinks to C (on the solid circle) while the residual-based method shrinks to B (on the dashed circle). For Case 1, the solid circle shrinks more than the dashed circle. This means the raw data based method generally has smaller ARL's regardless of the shift direction. However, in Case 2 and 4, the situation is the opposite. Case 3, 5 and 6 provide more interesting scenarios. The dashed circles do not shrink evenly in all directions. In fact, in some directions the radial distance of the residual-based method is much shorter than the observed data based method. In other directions, the ARL's for the residual-based method do not shrink as much as the observed data based method. Another interesting feature of the observed data based method is that it appears to be more robust to the change in the shift direction, indicated by the fact that the solid line shrinks more evenly in different directions. For the residual-based method, it is not so. Here changes in the direction of the shift influences the ARL much more.

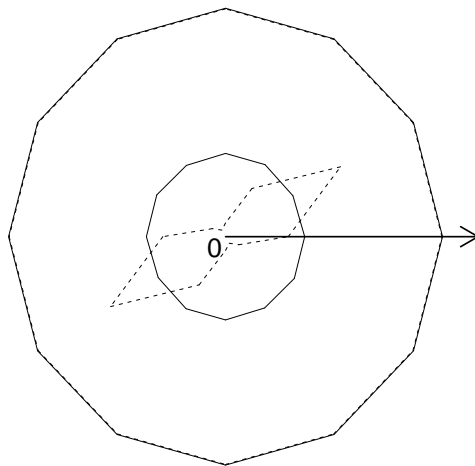
Note also that the polar diagrams in Figure 2-10 are made by connecting together the 12 points $\pi/6$ apart. If more points were used the curves would be smoother.



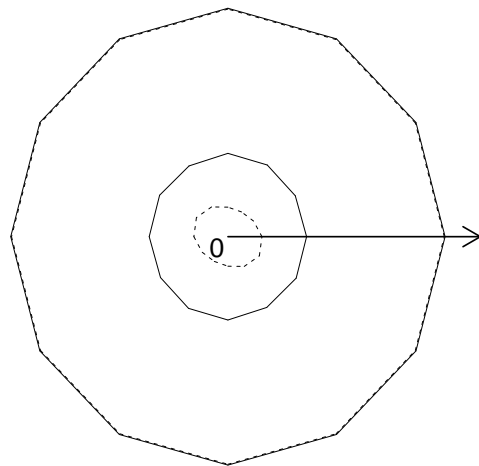
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Continued on next page

Continued

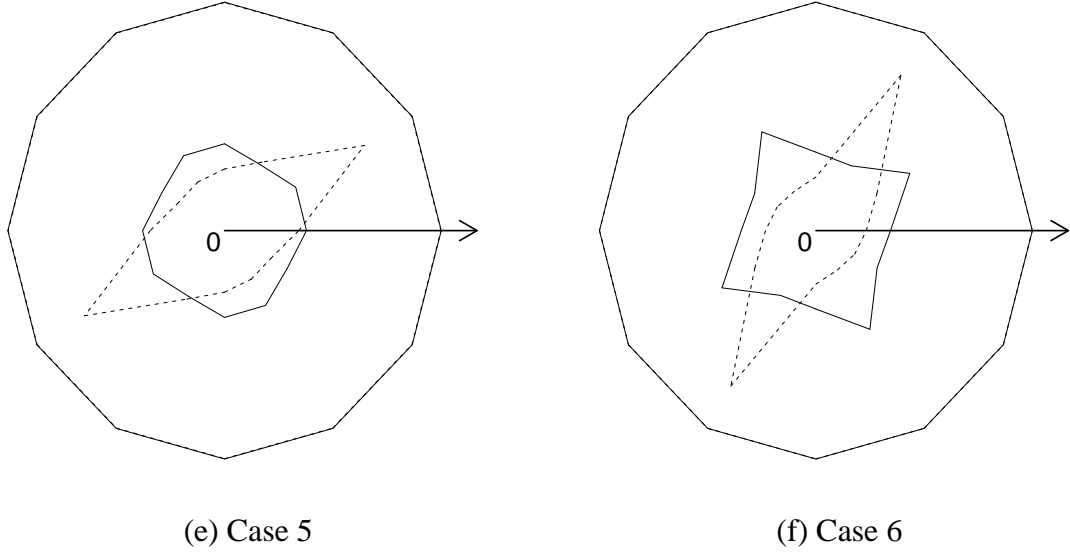


Figure 2-10. ARL's Comparison by graphs.

The difference in appearance between Case 1, Case 5 and Case 6 is noteworthy. They all have the same eigenvalues, but the performance of the residual-based method differs significantly. By fixing $\phi_1 = 1$ and $\phi_2 = -0.16$ the parameters $\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}$ lie on a two-dimension surface in \mathfrak{R}^4 . To make the problem more tractable, we assume each of the parameters remains within $[-2, 2]$. This is reasonable assumption since parameters outside this region would be unusual in practice. And, since $\phi_{22} = 1 - \phi_{11} \in [-2, 2]$, it is inferred that $\phi_{11} \in [-1, 2]$. Because $\phi_{11}\phi_{22} = \phi_{11}(1 - \phi_{11})$ is symmetric around $\phi_{11} = 0.5$, we may further confine our discussion to $\phi_{11} \in [0.5, 2]$. Now suppose we fix ϕ_{11} . Through $\phi_1 = \phi_{11} + \phi_{22} = 1$ and $\phi_2 = \phi_{12}\phi_{21} - \phi_{11}\phi_{22} = -0.16$, ϕ_{22} is also fixed and so are $\phi_{11}\phi_{22}$ and $\phi_{12}\phi_{21}$. Thus for each fixed value of ϕ_{11} , the feasible region is a pair of hyperbola in $\phi_{12} - \phi_{21}$ coordinates. As shown in Figure 2-11, the two pairs in quadrant I and III represents $\phi_{11} = 0.5$ and $\phi_{11} = 0.7$ respectively (moving inward), the pair overlaid on the axes

represents $\phi_{11} = 0.8$ and the pairs in quadrant II and IV represent $\phi_{11} = 0.9$, $\phi_{11} = 1.1$, $\phi_{11} = 1.3$, $\phi_{11} = 1.5$, $\phi_{11} = 1.7$, $\phi_{11} = 1.9$, $\phi_{11} = 2$ respectively (moving outward). The two pairs of accentuated hyperbola represent the boundaries of possible values of ϕ_{12} and ϕ_{21} . Note they are smaller than $[-2, 2] \times [-2, 2]$ because of the two restrictions. Each point in the $\phi_{12} - \phi_{21}$ coordinates is mapped into a identical Φ matrix (with ϕ_{12} and ϕ_{21} fixed, ϕ_{11} and ϕ_{22} can be determined through $\phi_1 = 1$ and $\phi_2 = -0.16$).

To explore the effect of how these parameters affect the performance of the residual-based method, we use solid lines to represent regions where the residual based method is competitive; by “competitive” we mean performing at least as well as the raw data based method in some kind of shifts. We use dashed line for the converse. It is interesting to note that for fixed ϕ_{11} and ϕ_{22} (in one pair of hyperbola), if either ϕ_{12} or ϕ_{21} is large (larger than 0.6~0.7 generally), the residual-based method performs better. If either ϕ_{11} and ϕ_{22} are larger than 0.9, no matter what the values of ϕ_{12} or ϕ_{21} are, the residual based method performs better. This explains the difference between Case 1, Case 5 and Case 6.

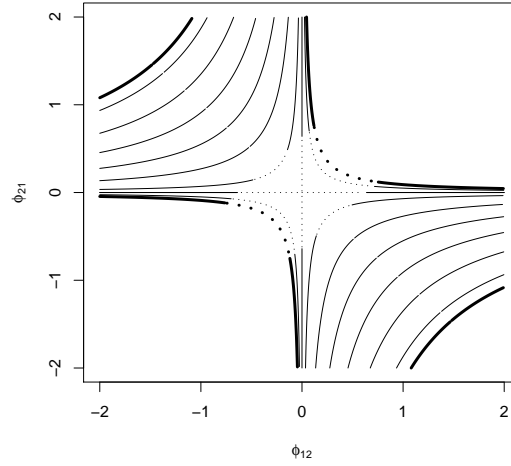


Figure 2-11. ϕ_{21} , ϕ_{22} Space when eigenvalues are 0.2 and 0.8 respectively.

2.9 Conclusion

We have proposed a method for monitoring multivariate processes based on two charts, a common cause plot based on predicted values from a fitted VARIMA model and a special cause chart that is a Hotelling's T^2 chart based on the residuals from that model. We provided a detailed discussion of a two-dimensional VAR(1) model, and showed the performance of the proposed method relative to a traditional T^2 chart applied directly to the unfiltered data. In this case, no uniformly superior performance can be claimed; in many cases the performance depends on the time series parameters and the shift. It should be anticipated that when the time series model gets more complicated, and there are more parameters, it will be more complicated. However, our discussion of the behavior for a VAR(1) model provides enough evidence to indicate potential benefits of the new method.

We like to point out that the ARL performance is not the only reason we think the residuals based method should be used. The fundamental issue is that rather than

proposing an ad hoc algorithm, we ought to identify and estimate a time series model before we attempt to control it. Without knowing the autocorrelation structure and fitting an appropriate time series model to a process, we do not fully understand the process. That said, we do not suggest replacing any of the traditional control charts. The combination of both will help achieve better control of processes. The advantage of the model based method is that it takes into account the auto- and cross- correlation.

We agree with Box and Paniagua-Quiñones (2007) that it makes more sense to model industrial process as weakly non-stationary process in which case the traditional Hotelling's T^2 would be inappropriate. Since the residuals are approximately random if the model fits well, all of the assumptions of traditional multivariate process control charts are met. Moreover, the residual chart can be used to detect changes in the structure of the time series while observed data based methods are not necessarily sensitive to structural changes.

From the perspective of applications, the problem with the proposed method is the difficulty of fitting VARIMA models. However, for a cross-correlated and autocorrelated multivariate process, we find it often worthwhile to understand the process better by fitting a model. Even if we only implement Hotelling's T^2 chart based on raw data, we still need to know the model to control the ARL0, as discussed in Section 2.5. Besides the difficulty of fitting VARIMA models there is also the problem that the number of parameters increase rapidly with an increase of the time series dimensions. A process with more than ten dimensions is almost intractable. Thus we suggest considering various dimension reducing approaches, as will be discussed in Chapter 4 of this dissertation.

CHAPTER 3

A CLASS OF MARKOV CHAIN MODELS FOR AVERAGE RUN LENGTH COMPUTATION FOR AUTOCORRELATED PROCESSES

The ARL of conventional control charts is typically computed assuming temporal independence. However, this assumption is frequently violated in practical applications. Alternative ARL computations have often been conducted via time consuming and not necessarily very accurate simulations. In this chapter, we develop a class of Markov chain models for evaluating the run length performance of traditional control charts for autocorrelated processes. We show extensions from the simplest univariate AR(1) model to the general univariate ARMA(p, q) model, and the multivariate VARMA(p, q) time series case. The results of the proposed method are compared with simulation results. The Markov chain methods for processes discussed in this chapter are implemented in MATLAB and are available online as supplemental materials.

3.1 Introduction

As introduced in Section 2.1.4, the ARL has become the standard metric for the evaluation of the performance of control charts. In the literature it is often assumed that the observations are temporally independent. This typically implies that the control chart statistic being charted (monitored) also is independent. If independent, the run length has a geometric distribution with mean equal to p^{-1} , where p is the probability of the statistic falling outside the control limits. However, in many important applications, the data and hence the statistic monitored is autocorrelated, which implies that the traditional ARL computations may be seriously misleading.

Autocorrelation of the monitored statistic occur sometimes even when the data is assumed independent. For example the Cumulative Sum (CUSUM) and the Exponentially Weighted Moving Average (EWMA) are autocorrelated even when the underlying process being monitored is independent. To compute the ARL for such situations, Brook and Evans (1972) introduced a Markov chain approximation and used that to analyze the performance of the univariate CUSUM control chart. Lucas and Saccucci (1990) applied the same technique to provide design recommendations to the univariate EWMA control chart. Runger and Prabhu (1996) extended this idea to a multivariate EWMA control chart.

To deal with an autocorrelated process, ad-hoc control charts as well as model based monitoring algorithms have been developed. For example, Berthouex, Hunter and Pallesen (1978) suggested filtering the process data through an autoregressive integrated moving average (ARIMA) time series model and applying regular control charts to the residuals. Alwan and Roberts (1988) extended this idea and proposed a special-cause chart based on residuals from fitting a univariate time series model as well as a common cause chart based on the one-step prediction. Jiang (2001) used the Markov chain approximation to compute the ARL for his ARMA chart, a special case of which includes the special-cause chart (Alwan and Roberts, 1988) as showed in Jiang et al. (2000) for univariate processes.

That said, quality practitioners frequently, advertently or inadvertently, apply conventional control charts such as Shewhart control charts and Hotelling's T^2 control charts when processes are autocorrelated. Moreover, for robustness studies it is important to compute the ARL performance under alternative assumptions. The purpose of this

chapter is to provide a class of Markov chain approximations for ARL computations when a given time series model of the process is assumed known. We hope that a more accurate ARL computation will provide guidance for setting more appropriate control limits for conventional control charts. We acknowledge that conventional control charts sometimes are applied to monitor non-stationary processes, with no constant first and second moments. However, we confine our discussion to stationary time series processes. The novelties are that we work on conventional control charts, that we extend the Markov chain approach from univariate cases to multivariate cases, and that we use polar coordinates to divide the in-control state space in the multivariate case.

This chapter is organized as follows. In Section 3.2 we review the basic idea of using a Markov chain approximation to calculate the ARL illustrated with an example of a univariate AR(1) model. In Section 3.3 we present an extension to the general ARMA(p, q) model. In Section 3.4 we extend the Markov chain approximation to the multivariate time series case and illustrate it with an example of a first order Vector Autoregressive VAR(1) model. Computational accuracy, timing and strategies are discussed in Section 3.5. Finally we present a conclusion and discussion in Section 3.6.

3.2 The Univariate AR(1) process

Consider an AR(1) model $z_t = \phi z_{t-1} + a_t$ where a_t is white noise with $E\{a_t\} = 0$, $V\{a_t\} = 1$, and $z_t = z_t - \mu_t$, $\mu_t = E(z_t)$. We further assume that the parameters of the AR(1) model are known. When the mean is constant, we can assume without loss of generality that $\mu_t = 0$ and $z_t = z_t$. We declare observations outside the interval $C = [-c, c]$ to be out-of-control. The Markov chain method for computing the ARL is

based on the Markov property of the AR(1) model. As illustrated in Figure 3-1, the real line is divided into $s+1$ subintervals, with s subintervals within the interval C . One straightforward way of dividing C is $S_i = [-c + 2c(i-1)/s, -c + 2ci/s)$, $i = 1, 2, \dots, s$ and then to define the $(s+1)$ -th “interval” as $S_{s+1} = (-\infty, -c) \cup [c, \infty)$, out-of-control zone.

We then define a Markov chain X_t with states $\{1, 2, \dots, s+1\}$ and let

$$X_t = \sum_{i=1}^{s+1} i \cdot I(z_t \in S_i), \text{ where } I(\cdot) \text{ is indicator function defined such that } X_t \text{ is in state } i$$

if $z_t \in S_i$. The ARL computation then proceeds as follows: Starting from any in-control state, how many steps does it take to arrive at the out-of-control state for the first time.

Symbolically we denote this time by $\tau_1 = \min \{n \geq 1, \text{ s.t. } X_n = s+1 \mid X_0 \neq s+1\}$.

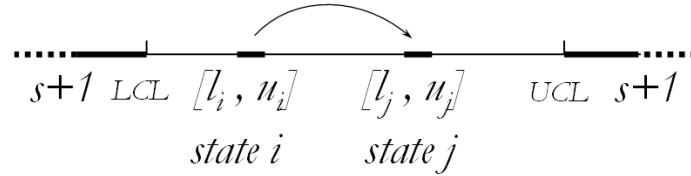


Figure 3-1. Markov Chain State Space for AR(1) model.

Brook and Evans (1972) proved that ARL is given by

$$\tau_1 = \boldsymbol{\phi}'(\mathbf{I} - \mathbf{R})^{-1} \mathbf{1}, \quad (3.1)$$

where $\boldsymbol{\phi}$ is a $s \times 1$ dimensional vector where i -th element $\phi(i)$ is the probability

$P(X_0 = i \mid X_0 \neq s+1)$, $i = 1, 2, \dots, s$. Further \mathbf{R} is a $s \times s$ matrix with (i, j) -th element

$$r_{ij} = P(X_t = j \mid X_{t-1} = i), \quad (3.2)$$

and $\mathbf{1}$ is a $s \times 1$ unit vector.

Brook and Evans (1972) used factorial moments to prove (3.1). However, we provide an alternative proof in APPENDIX C.

Consider a step mean shift for the AR(1) model at time T such that

$$z_t - \mu_t = \phi(z_{t-1} - \mu_{t-1}) + a_t, \quad (3.3)$$

where,

$$\begin{aligned} a_t &\sim N(0,1) \\ \mu_t &= \begin{cases} 0 & t < T \\ \delta & t \geq T. \end{cases} \end{aligned} \quad (3.4)$$

Combining (3.4) and (3.3) we get

$$z_t = \begin{cases} \phi z_{t-1} + a_t & t < T \\ \delta + \phi z_{t-1} + a_t & t = T \\ (1-\phi)\delta + \phi z_{t-1} + a_t & t > T. \end{cases} \quad (3.5)$$

Now suppose the whole process starts from an in-control state at time $t = T - 1$ and that a step-shift in the mean occurs at time T . The probability that the shift is detected at time T is

$$\begin{aligned} p_1 &= P(X_T = s+1 | X_{T-1} \neq s+1) = P(z_T \in S_{s+1} | z_{T-1} \notin S_{s+1}) \\ &= \frac{P(z_T \in S_{s+1}, z_{T-1} \notin S_{s+1})}{P(z_{T-1} \notin S_{s+1})} = 1 - \frac{\int_{-c}^c \int_{-c}^c f_z(z_1) f_a(z_2 - \phi z_1 - \delta) dz_2 dz_1}{\int_{-c}^c f_z(z) dz}, \end{aligned} \quad (3.6)$$

where $f_z(\cdot)$ is the probability density function (pdf.) of z_t and $f_a(\cdot)$ is the pdf. of a_t .

If the shift has not been detected at time T , which means $X_T \neq s+1$, the whole Markov chain starts with initial probabilities,

$$\begin{aligned} \phi(i) &= P(X_T = i | X_T \neq s+1, X_{T-1} \neq s+1) \\ &= \frac{P(X_T = i, X_T \neq s+1, X_{T-1} \neq s+1)}{P(X_T \neq s+1, X_{T-1} \neq s+1)} \\ &= \frac{P(X_T = i, X_T \neq s+1 | X_{T-1} \neq s+1)}{P(X_T \neq s+1 | X_{T-1} \neq s+1)} \\ &= \frac{P(X_T = i | X_{T-1} \neq s+1)}{P(X_T \neq s+1 | X_{T-1} \neq s+1)}. \end{aligned} \quad (3.7)$$

The nominator of $\phi(i)$ is given by

$$P(X_T = i | X_{T-1} \neq s+1) = \frac{\int_{-c}^c \int_{l_i}^{u_i} f_z(z_1) f_a(z_2 - \phi z_1 - \delta) dz_2 dz_1}{\int_{-c}^c f_z(z) dz},$$

and the denominator is

$$P(X_T \neq s+1 | X_{T-1} \neq s+1) = \sum_{i=1}^s P(X_T \neq i | X_{T-1} \neq s+1).$$

Moreover the transition probability after time T is

$$\begin{aligned} r_{ij} &= P(X_t = j | X_{t-1} = i) = \frac{P(z_t \in S_j, z_{t-1} \in S_i)}{P(z_{t-1} \in S_i)} \\ &= \frac{\int_{l_i}^{u_i} \int_{l_j}^{u_j} f_z(z_1) f_a(z_2 - \phi z_1 - (1-\phi)\delta) dz_2 dz_1}{\int_{l_i}^{u_i} f_z(z) dz}. \end{aligned} \quad (3.8)$$

Using (3.1), we can compute the ARL as

$$\begin{aligned} ARL &= p_1 \cdot 1 + (1-p_1)(1+\tau_1) \\ &= p_1 \cdot 1 + (1-p_1)(1 + \boldsymbol{\phi}'(\mathbf{I} - \mathbf{R})^{-1} \mathbf{1}), \end{aligned} \quad (3.9)$$

where p_1 , $\boldsymbol{\phi}$ and \mathbf{R} are computed through (3.6), (3.7) and (3.8) respectively.

To validate this Markov chain method we computed results for selected parameters of three different AR(1) models for both the Markov chain method and the simulation method as showed in Table 3-1. The standard deviation of each simulated ARL was estimated on the entire sample of simulation outputs and is recorded after the simulated ARL entries (\pm s.d.). Undoubtedly, the more the simulation rounds, the smaller the standard deviation of the simulated ARL. The simulations were each run 90,000 times to give a reasonable standard deviation; so were the simulations in the remaining part of this chapter. For the Markov chain method, the number of subintervals, s , was chosen to

be 10. Both methods have been implemented in MATLAB and executed at a DELL personal workstation (Intel Pentium Dual CPU E2180@2GHz, 2GHz, 1.99G of RAM). Computing time has been recorded and provided in parenthesis following the ARL entries. Notably, for the same underlying time series model, computing time for the Markov chain method is independent of the size of the shift and the parameters while the simulation time is nearly proportional to the ARL, i.e., the larger the ARL, the longer the simulation time. As it can be seen, the results from the Markov chain method match the simulation results very well while the Markov chain method took significantly less time than the simulation method.

Table 3-1: Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on AR(1) models.

	$\phi = 0$		$\phi = 0.5$		$\phi = -0.6$	
	MC	Simulated	MC	Simulated	MC	<i>Simulated</i>
$\delta = 0$	300.00 (0.78s)	299.51 ± 1.00 (315.09s)	322.50 (0.83s)	323.52 ± 1.08 (323.52s)	341.81 (0.81s)	342.67 ± 1.14 (360.88s)
$\delta = 0.5\sigma_z$	129.24 (0.78s)	128.76 ± 0.43 (136.53s)	147.82 (0.81s)	148.15 ± 0.49 (156.45s)	138.57 (0.81s)	139.07 ± 0.46 (147.22s)
$\delta = \sigma_z$	37.70 (0.75s)	37.87 ± 0.12 (40.92s)	47.06 (0.78s)	47.36 ± 0.16 (51.23s)	40.08 (0.81s)	40.16 ± 0.13 (43.67s)

3.3 Extension to the Univariate ARMA(p, q) process

The Markov chain method can easily be applied to the AR(1) model because the process z_t is Markovian. For more general ARMA(p, q) models, we need to use a transformation before the Markov chain method can be applied; see e.g. Box, Jenkins and Reinsel (1995). We provide a few illustrations.

3.3.1 Extension to the Univariate AR(2) Process

Consider an AR(2) model $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$, where a_t is white noise with $E\{a_t\} = 0$, $V\{a_t\} = 1$, and $z_t = z_t - \mu_t$, $\mu_t = E(z_t)$. Again for a process with constant mean, we may without loss of generality assume that $\mu_t = 0$ and $z_t = z_t$. This univariate AR(2) model can be presented in a multivariate AR(1) (Markovian state space) model as

$$\begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} a_t. \quad (3.10)$$

The new statistic $\mathbf{w}_t = [z_t \quad z_{t-1}]'$ is Markovian. Thus the methodology from Section 3.2 immediately applies. The state space for \mathbf{w}_t is illustrated in Figure 3-2. Although z_t and z_{t-1} are correlated, the control mechanism is imposed on observations at each single time point. Thus the in-control state space is a square which is subdivided into an $m \times m$ grid, each sub-square representing one state and all the points outside the in-control square the $(m \times m + 1)$ -th state. Notice that the transition matrix is $(m^2 + 1) \times (m^2 + 1)$ but very sparse. The transition probability from state $[j_1 \quad i]$ to $[k \quad j_2]$ is only non-zero when $j_1 = j_2$ since “ z_t ” at this time point has to become “ z_{t-1} ” at the next time point.

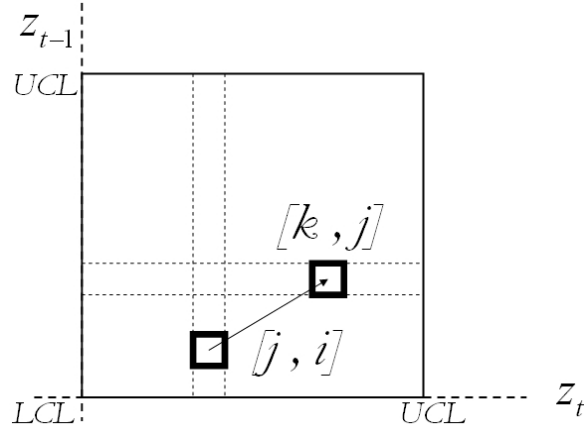


Figure 3-2. Markov Chain State Space for AR(2) model.

Again consider a step shift model

$$z_t - \mu_t = \phi_1 (z_{t-1} - \mu_{t-1}) + \phi_2 (z_{t-2} - \mu_{t-2}) + a_t, \quad (3.11)$$

with a_t and μ_t follow (3.4) as well. A four-stage model for z_t follows immediately:

$$z_t = \begin{cases} \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t & t < T \\ \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t & t = T \\ (1 - \phi_1) \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t & t = T + 1 \\ (1 - \phi_1 - \phi_2) \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t & t > T + 1. \end{cases} \quad (3.12)$$

Assume p_1 is the probability that the shift is detected right away ($t = T$), p_2 is the probability that the shift is not detected until the second step ($t = T + 1$), and z_t follows a Markov chain model for $t > T + 1$. The ARL is then

$$ARL = p_1 \cdot 1 + (1 - p_1) p_2 \cdot 2 + (1 - p_1)(1 - p_2)(2 + \tau_1). \quad (3.13)$$

Comparisons between the Markov chain method and simulations for the AR(2) process is provided in Table 3-2. In the Markov chain method, m was chosen to be 10 and thus s was 100. Again, the results from the Markov chain method closely match those from the simulations but took significantly less time.

Table 3-2. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on AR(2) models.

	$\phi_1 = 0$ $\phi_2 = 0$		$\phi_1 = 0.5$ $\phi_2 = 0.2$		$\phi_1 = -0.5$ $\phi_2 = -0.2$	
	MC	Simulated	MC	Simulated	MC	Simulated
$\delta = 0$	300.71 (9.03s)	301.09 ± 1.00 (368.82s)	369.88 (8.61s)	368.15 ± 1.22 (449.95s)	311.93 (8.67s)	312.47 ± 1.04 (381.72s)
$\delta = 0.5\sigma_z$	129.36 (8.41s)	129.71 ± 0.43 (160.00s)	181.29 (8.67s)	181.40 ± 0.60 (222.95s)	129.86 (8.59s)	130.28 ± 0.44 (160.22s)
$\delta = \sigma_z$	37.71 (8.51s)	37.55 ± 0.12 (47.98s)	62.51 (8.59s)	62.40 ± 0.21 (78.08s)	36.78 (8.62s)	36.99 ± 0.12 (47.18s)

3.3.2 Extension to the Univariate ARMA(1, 1) Process

For the ARMA(1, 1) model $z_t = \phi z_{t-1} + a_t - \theta a_{t-1}$ with the assumptions as above,

the state space vector $\mathbf{w}_t = [z_t \quad a_t]'$ can be transformed to be Markovian process as

follows

$$\begin{bmatrix} z_t \\ a_t \end{bmatrix} = \begin{bmatrix} \phi & -\theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} a_t. \quad (3.14)$$

The state space for \mathbf{w}_t is illustrated in Figure 3-3. Theoretically control limits are only imposed on z_t , not on a_t , implying that the in-control state space will be an infinite area. However, as a pragmatic device we imposed a large artificial limit $[-6, 6]$ on a_t so that the in-control state space becomes a finite rectangle. It has to be noted that by doing this the computed ARL will be shorter than it really is because an abnormal a_t will be

considered to be out-control without upsetting z_t . However the bias is small with control limits on a_t chosen to be large.

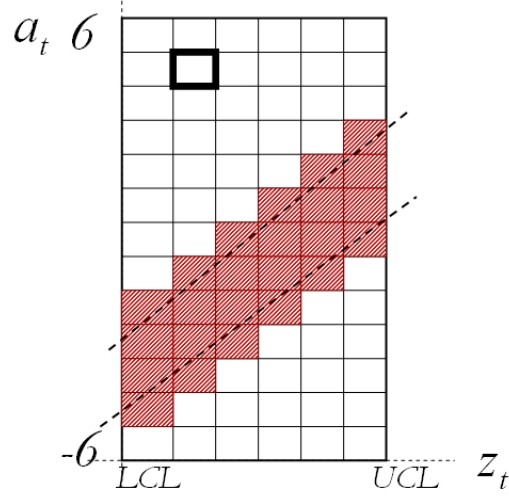


Figure 3-3. Markov Chain State Space for ARMA(1,1) model.

The rectangle is divided into an $m \times n$ grid with each sub-square representing one state and all the points outside the in-control rectangle representing the $(m \times n + 1)$ -th state. The transition matrix is $(mn + 1) \times (mn + 1)$ but is again very sparse. In APPENDIX D we show that once the initial state is fixed, two parallel lines with slopes of 1 will be determined and that the ending states can only be between the two lines or by crossed through the lines.

For illustration we consider a step shift ARMA(1, 1) model. Thus a three-stage model for z_t is

$$z_t = \begin{cases} \phi z_{t-1} + a_t - \theta a_{t-1} & t < T \\ \delta + \phi z_{t-1} + a_t - \theta a_{t-1} & t = T \\ (1 - \phi)\delta + \phi z_{t-1} + a_t \theta a_{t-1} & t > T. \end{cases} \quad (3.15)$$

Similarly, the ARL is given by

$$ARL = p_1 \cdot 1 + (1 - p_1)(1 + \tau_1), \quad (3.16)$$

with p_1 again being the probability that the shift is detected at time T and τ_1 determined by Markov chain model after time $T+1$.

Comparison between the Markov chain method and the simulations for the ARMA(1,1) is showed in *Table 3-3*. For the Markov chain method, both m and n were chosen to be 15 and thus s was 225. Again, the results from the Markov chain method closely match those from the simulations except for a slight discrepancy in case 2 ($\phi = 0.8, \theta = 0.5$). For ARMA(1,1) model, the two methods are comparable in computing time. The realization of the Markov chain method on ARMA(1,1) process requires a set of control flow statements that drastically increase the computing time compared with cases for AR(1) and AR(2) models. MATLAB is especially inefficient with control flow statements; had it been implemented with C/C++ or other programming languages, the increased computing time might not be as significant.

Table 3-3. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on ARMA(1,1) models.

	$\phi = 0.5$ $\theta = 0.8$		$\phi = 0.8$ $\theta = 0.5$		$\phi = -0.4$ $\theta = 0.2$	
	MC	Simulated	MC	Simulated	MC	Simulated
$\delta = 0$	303.85 (238.65s)	301.49 ± 1.01 (361.39s)	316.63 (242.43s)	322.13 ± 1.07 (385.47s)	326.60 (201.39s)	326.80 ± 1.09 (389.91s)
$\delta = 0.5\sigma_z$	128.47 (238.00s)	128.12 ± 0.42 (154.23s)	146.00 (247.62s)	153.21 ± 0.51 (183.69s)	133.56 (201.92s)	133.07 ± 0.44 (159.72s)
$\delta = \sigma_z$	36.62 (228.26s)	36.05 ± 0.12 (45.28s)	47.00 (244.73)	51.14 ± 0.17 (62.69s)	38.25 (192.14s)	38.30 ± 0.13 (48.14s)

3.3.3 Extension to the Univariate ARMA(p, q) Process

For the ARMA(p, q) model $z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$, a similar extension can easily be made by constructing a Markovian state space vector. A general model for ARMA(p, q) model can be written in state space format as,

$$\begin{bmatrix} z_t \\ z_{t-1} \\ \mathbf{M} \\ z_{t-p+1} \\ a_t \\ \mathbf{M} \\ \mathbf{M} \\ a_{t-q+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \mathbf{L} & \phi_p & -\theta_1 & \mathbf{L} & \mathbf{L} & -\theta_q \\ 1 & 0 & \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{M} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{M} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{M} & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{M} \\ \mathbf{M} & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{M} \\ 0 & \mathbf{L} & \mathbf{L} & \mathbf{L} & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ z_{t-2} \\ \mathbf{M} \\ z_{t-p} \\ a_{t-1} \\ \mathbf{M} \\ \mathbf{M} \\ a_{t-q} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \mathbf{M} \\ 0 \\ 1 \\ 0 \\ \mathbf{M} \\ 0 \end{bmatrix} a_t \quad (3.17)$$

For any ARMA(p, q) model, (3.17) provides a $p + q$ dimensional state space vector model. An alternative classical state space model, see e.g. Reinsel (1997), for the ARMA(p, q) model is given by

$$\mathbf{Z}_t = \begin{bmatrix} 0 & 1 & 0 & \mathbf{L} & 0 \\ 0 & 0 & 1 & \mathbf{L} & 0 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ 0 & 0 & 0 & \mathbf{L} & 1 \\ \phi_r & \phi_{r-1} & \phi_{r-2} & \mathbf{L} & \phi_1 \end{bmatrix} \mathbf{Z}_{t-1} + \begin{bmatrix} 1 \\ \psi_1 \\ \mathbf{M} \\ \psi_{r-1} \end{bmatrix} \varepsilon_t \quad (3.18)$$

where $\mathbf{Z}'_t = [\hat{z}_t(0), \hat{z}_t(1), \mathbf{L}, \hat{z}_t(r-1)]$, $\hat{z}_t(j)$ is the expected value of z_{t+j}

conditional on all the observations up to time point t , ϕ_i 's are the autoregressive

coefficients from the ARMA(p, q) model ($\phi_i = 0$ if $i > p$) and ψ_j 's are the coefficients

for the ARMA model converted to infinite MA form, $z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$. Note that the

dimension r of \mathbf{Z}_t , which is computed as $\max(p, q+1)$, is smaller than that of model

(3.17). Note further that the first element of \mathbf{Z}_t , $\hat{z}_t(0)$, is essentially z_t , on which we will impose the control limit for the computation.

We acknowledge that as either p or q increases, the complexity of the computations increase geometrically. However, for most practical applications applying the principle of parsimonious model building, p and q rarely exceed 2.

3.4 Extension to the Multivariate ARMA(p, q) process

In many industrial applications the processes to be monitored are multivariate. An extension from univariate time series model to multivariate time series model is therefore desirable. We demonstrate this extension with the example of the VAR(1) model of dimension k . To facilitate easy visualization, we will choose $k = 2$.

Consider a VAR(1) model $\mathbf{z}_t = \Phi \mathbf{z}_{t-1} + \mathbf{a}_t$ where $E(\mathbf{a}_t) = \mathbf{0}$, $Var(\mathbf{a}_t) = \mathbf{I}$ and again without loss of generality, $\mathbf{z}_t = \mathbf{z}_t - \boldsymbol{\mu}_t$ where $\boldsymbol{\mu}_t = \mathbf{0}$. The state space for \mathbf{z}_t is shown in Figure 3-4. The in-control area is an ellipse. Using polar coordinates it can be divided into an $m \times n$ mesh of sectors each representing one state and the large sector outside the in-control ellipse representing the $(m \times n + 1)$ -th state. The transition matrix is a $(mn + 1) \times (mn + 1)$ matrix. By subdividing the in-control zone after converting to polar coordinates rather than using a Cartesian grid we achieve better accuracy. The polar coordinate transformation can easily be generalized to higher dimensions.

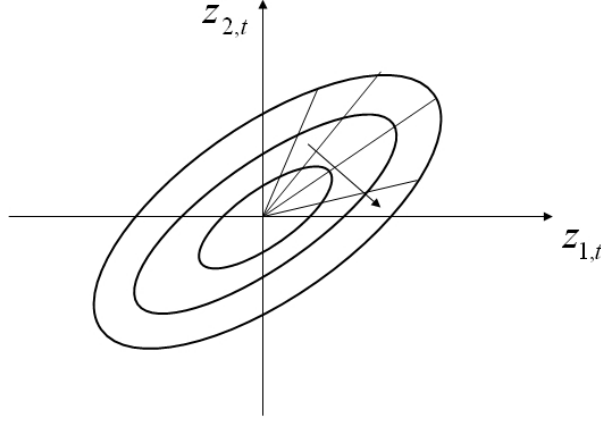


Figure 3-4. Markov Chain State Space of VAR(1) model.

In the multivariate case we use a step shift model similar to that is used for the univariate case. The result is again a three-stage model for \mathbf{z}_t given as

$$\mathbf{z}_t = \begin{cases} \Phi \mathbf{z}_{t-1} + \mathbf{a}_t & t < T \\ \delta + \Phi \mathbf{z}_{t-1} + \mathbf{a}_t & t = T \\ (\mathbf{I} - \Phi) \delta + \Phi \mathbf{z}_{t-1} + \mathbf{a}_t & t > T . \end{cases} \quad (3.19)$$

Equation (3.19) looks similar to (3.5) and works almost the same way. The difference is that computing the elements of transition matrix is more tedious:

$$\begin{aligned} r_{ij} &= P(X_t = j | X_{t-1} = i) = \frac{P(z_t \in S_j, z_{t-1} \in S_i)}{P(z_{t-1} \in S_i)} \\ &= \frac{\int_{r_{li}}^{r_{ui}} \int_{\gamma_{li}}^{\gamma_{ui}} \int_{r_{lj}}^{r_{uj}} \int_{\gamma_{lj}}^{\gamma_{uj}} f_a(g(r_2, \gamma_2) - (\mathbf{I} - \Phi)\delta - \Phi g(r_1, \gamma_1)) |J| d\gamma_2 dr_2 d\gamma_1 dr_1}{\int_{r_{li}}^{r_{ui}} \int_{\gamma_{li}}^{\gamma_{ui}} f_{z,\delta}(g(r, \gamma)) |J|^2 d\gamma dr} , \end{aligned} \quad (3.20)$$

where the $g(\cdot, \cdot)$ is the transformation from polar to Cartesian coordinates and J is the corresponding Jacobian matrix.

Comparisons between the Markov chain method and simulations for the VAR(1) are shown in Table 3-4. Note that unlike the univariate case, the step shift δ is a two dimensional vector. Thus we need two polar parameters, α the angle and λ the length,

to describe it. For the Markov chain method, m and n were chosen to be 10 and 4 respectively, and thus s was 40. Again, the results from the Markov chain method match the results from the simulations very well. However, both methods took significantly more computing time than their counterparts in univariate cases.

Table 3-4. Comparison of computed ARL between the Markov chain method (MC) and the simulation method (Simulated) on VAR(1) models.

	$\Phi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$		$\Phi = \begin{bmatrix} 0.2 & 0.4 \\ 0.1 & 0.1 \end{bmatrix}$		$\Phi = \begin{bmatrix} -0.4 & 0.3 \\ 0.5 & -0.1 \end{bmatrix}$	
	MC	Simulated	MC	Simulated	MC	Simulated
$\lambda = 0$	299.93	299.98	305.24	302.92	322.66	325.97
$\alpha = 0$	(668.35s)	± 1.00 (999.69s)	(690.81s)	± 1.01 (1053.70s)	(680.50s)	± 1.09 (1119.50s)
$\lambda = 0.5$	108.45	108.71	112.59	112.21	114.22	116.07
$\alpha = 0$	(665.01s)	± 0.36 (360.20s)	(697.92s)	± 0.37 (406.50s)	(672.64s)	± 0.39 (399.67s)
$\lambda = 1$	57.21	57.17	60.23	60.12	60.23	60.49
$\alpha = 0$	(675.75s)	± 0.19 (190.55)	(691.02s)	± 0.20 (217.94s)	(682.42s)	± 0.21 (213.59s)

The extension to a VARMA(p, q) process is similar to the one going from the univariate AR(1) to the univariate ARMA(p, q) process. However, the computation becomes more complicated, especially when the vector dimension is larger than two.

3.5 Computation Efficiency

The Markov chain approximation that we have proposed raises a number of computational issues that we will briefly discuss in this section.

3.5.1 How Fine Should the Mesh be?

To compute the ARLs, we need to determine the number s of subintervals into which the in-control region is divided. ARLs have been computed against a set of s values and are illustrated in Figure 3-5 through Figure 3-8. For comparison we also simulated the ARLs. However, note that a simulated ARL is a pseudo random number. Hence we provide for the simulations the ARL plus the standard deviation ($\overline{ARL} \pm s.d.$) to judge the convergence of the Markov chain method.

For the AR(1) model, s was chosen to be 10 for the computation in Table 3-1. However, for the AR(1) case 1 ($\phi = 0$), Figure 3-5 shows that $s = 2$ provides result accurate enough; for AR(1) case 2 ($\phi = 0.5$), Figure 3-6 shows that $s = 4$ works well enough. Thus we infer that the higher the autocorrelation, the larger s has to be to achieve the same level of accuracy. In fact, when the autocorrelation is 0, there is no need for the Markov chain method; instead the ARL can be computed via the geometric distribution.

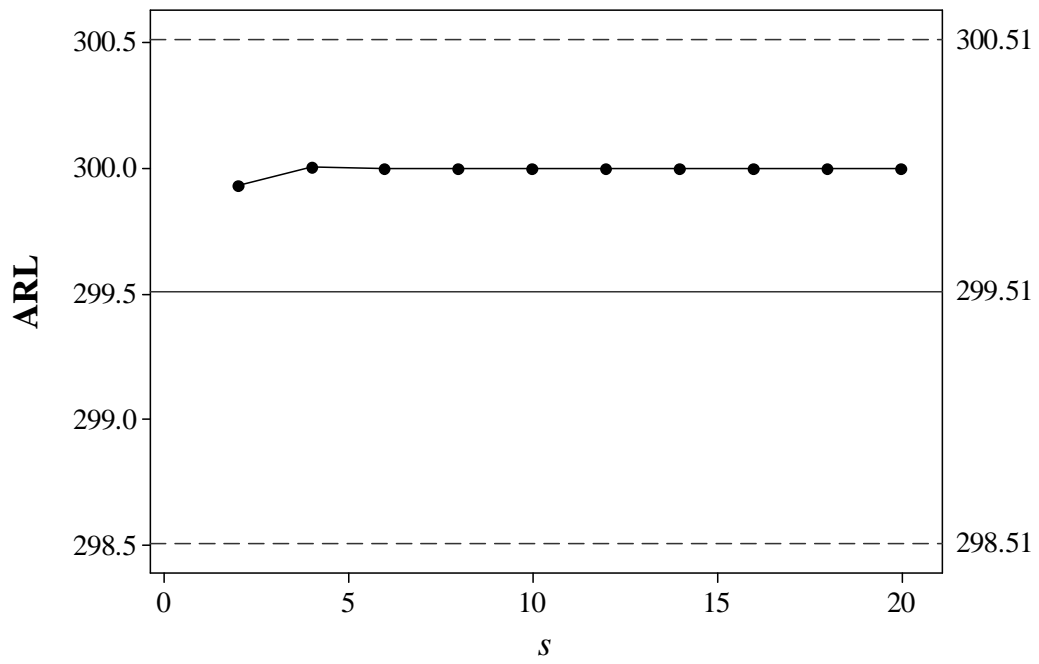


Figure 3-5. ARL v.s. s for AR(1) model, $\phi = 0$.

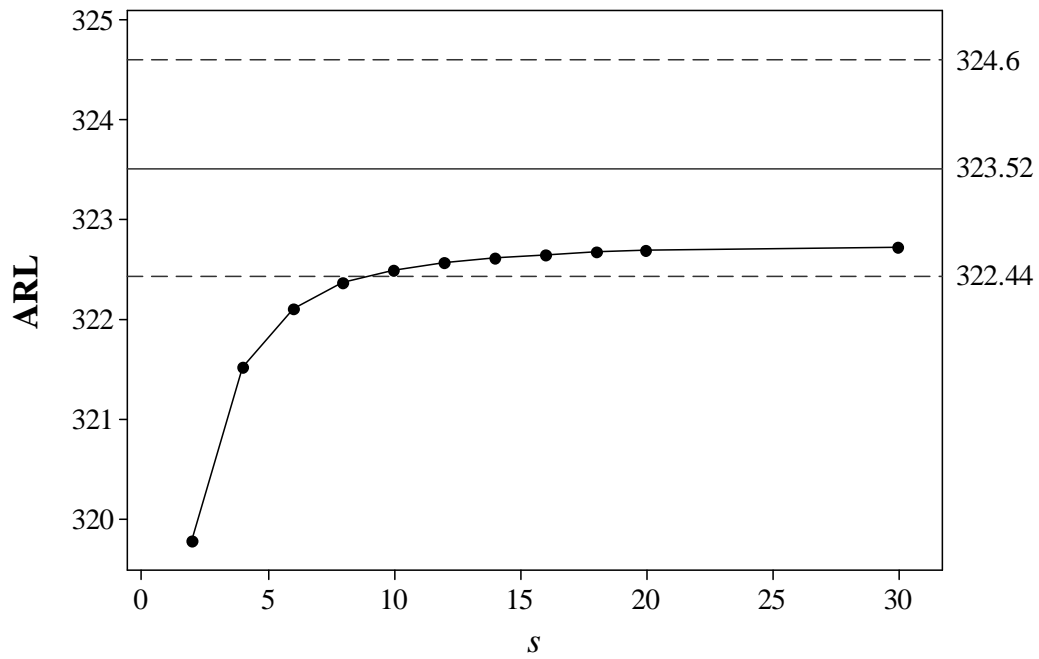


Figure 3-6. ARL v.s. s for AR(1) model, $\phi = 0.5$.

For the AR(2) model, s was chosen to be 100 for the computation in Table 3-2.

However, Figure 3-7 shows for case 2 ($\phi_1 = 0.5, \phi_2 = 0.2$), $s = 36$ or even $s = 16$ provides sufficient accuracy.

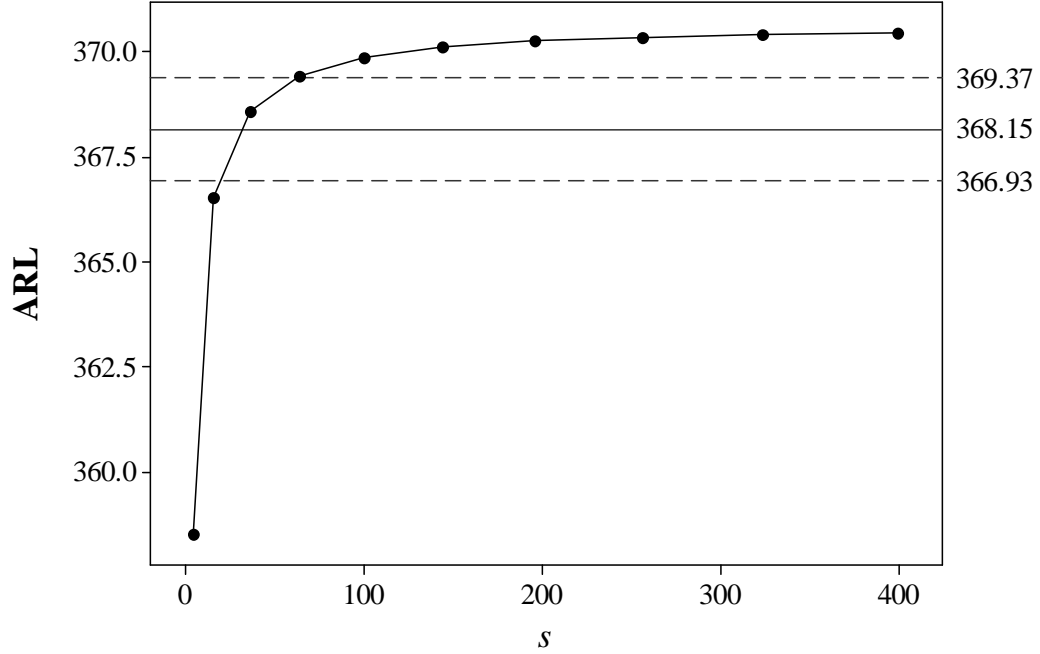


Figure 3-7. ARL v.s. s for AR(2) model, $\phi_1 = 0.5, \phi_2 = 0.2$.

For the AR(1,1) model, s was chosen to be 225 for computation of Case 2 ($\phi = 0.8, \theta = 0.5$) in Table 3-3. Figure 3-8 does show slower convergence compared with Figure 3-5 through Figure 3-7. However, $s = 81$ or $s = 144$ provides sufficient accuracy.

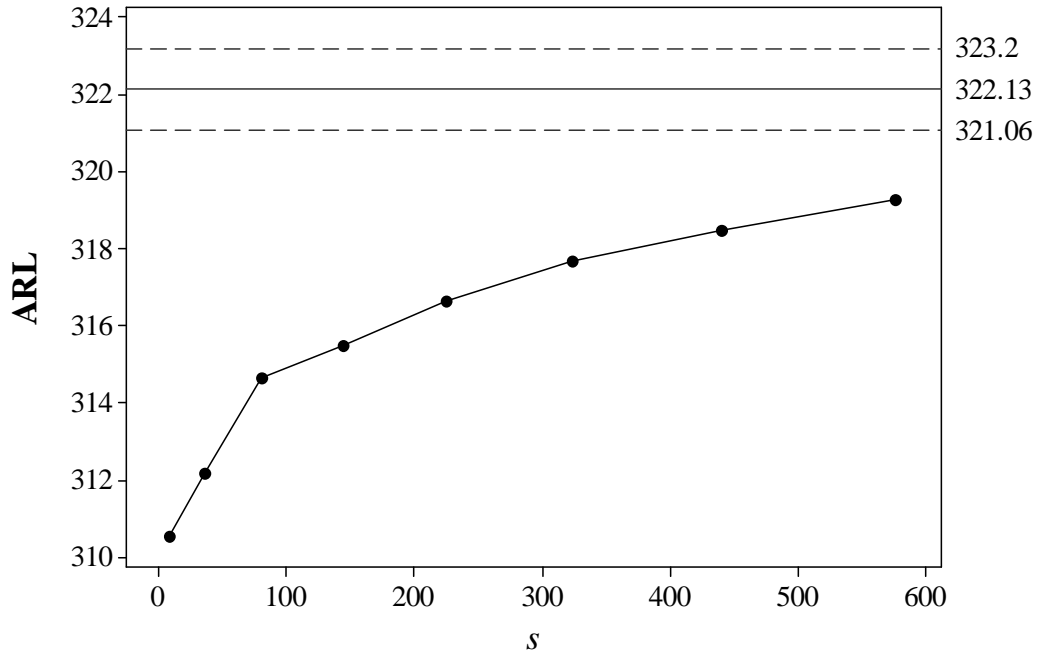


Figure 3-8. ARL v.s. s for ARMA(1,1) model, $\phi = 0.8$, $\theta = 0.5$.

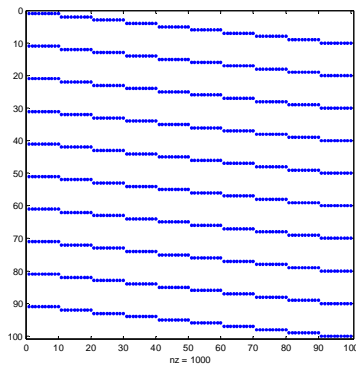
As a rule of thumb, in univariate cases, we have found that for the one-dimension in-control region $s = 10$, and for the two-dimension in-control region, $s = 100$ provides sufficient accuracy. For the two-dimension AR(1) case, we have found that $s = 40$ works well.

One way to improve the computation accuracy of the Markov chain method is through assuming the functional relation $ARL(s) = ARL(\infty) + A/s + B/s^2$ between the ARL and s as suggested by Brook and Evans (1972); see also Jiang (2001). By computing pairs of ARL and s , OLS regression provides a reasonable estimate of $ARL(\infty)$. For the four cases shown in Figure 3-5 through Figure 3-8, the $ARL(\infty)$ estimates are 300.0, 322.9, 370.6, 319.8 respectively. This regression model appears to improve the accuracy. However, the least square estimates of the regression coefficients A

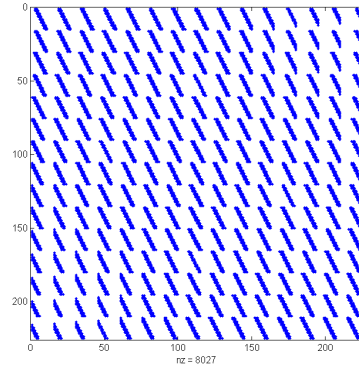
and B vary from case to case. Thus there does not appear to be a universal set of coefficients that apply to every condition to save computation cost.

3.5.2 Matrix Computation

As was pointed out above, the transition matrices for AR(2) and ARMA(1,1) cases are sparse. Figure 3-9 illustrates the transition matrices graphically for these two cases. In the AR(2) case, with $s = 100$, the non-zero entries are 1000 out of a 10000 total entries or 10%. In the ARMA(1,1) case, with $s = 225$, the non-zero entries account for 8027 out of 50625 entries or about 16%. In MATLAB, the function “sparse” can be invoked to reduce memory storage and accelerate computing time.



(a) AR(2), $s = 100$



(b) AR(1,1), $s = 225$

Figure 3-9. Sparse transition matrices.

In general, matrix computation may be subject to large computation errors, especially for large dimension sparse matrices. However, a 100×100 matrix for the AR(2) process, or a 225×225 matrix for the ARMA(1,1) process, is by today's standards not a large matrix. To measure the approximation error of a large sparse matrix \mathbf{A} , the computational inverse for $(\mathbf{I} - \mathbf{R})$ with \mathbf{R} defined in (3.2), we use the Frobenius norm of the matrix $(\mathbf{I} - \mathbf{A} \cdot (\mathbf{I} - \mathbf{R}))$. If \mathbf{A} is close to $(\mathbf{I} - \mathbf{R})^{-1}$, $\mathbf{A} \cdot (\mathbf{I} - \mathbf{R})$ is close to \mathbf{I} and the

Frobenius norm of the matrix $(\mathbf{I} - \mathbf{A} \cdot (\mathbf{I} - \mathbf{R}))$ will be close to 0. For the 100×100 matrix for AR(2) computation, the Frobenius norm is $3.4133e^{-014}$, which indicates the approximation error is very small.

Moreover the matrix computation is not where the major computing time was spent. In the AR(2) case with $s = 100$, for example, the integral computation using the lattice method (Sloan and Joe, 1994) to calculate the non-zero elements of the transition matrix took 9.92s while the matrix computation only took 0.02s.

To reduce the integral computing time, one alternative is to use midpoint multiplication. However, in that case, a finer mesh is required to achieve similar accuracy and a much larger transition matrix will be needed. In spite of the computing cost saved for each entry in the transition matrix, the computation efficiency will be offset by the larger number of entries to be computed, and a larger, sparser transition matrix which will result not only costly computing, but also raises accuracy problem because of the size.

3.6 Conclusion and Discussion

In this chapter we have proposed a class of Markov chain models for the approximate ARL calculations for control algorithms where the monitored processes exhibit autocorrelation. The suggested Markov chain approximation is versatile and we show that it is easily generalized to multivariate time series process. It also performs well compared to simulations.

One potential issue is deciding how fine a mesh to divide the in-control space to make the computation as accurate as needed. Clearly the further it is divided the better the results. However, a linear increase in the number of mesh points and hence states

result in a quadratic increase in the dimension of the transition matrix and thus the computational effort. Fortunately, if the numerical computations of the multiple integral based on lattice method is used for the transition probabilities as demonstrated in this chapter, the accuracy is fairly high even if the in-control space is divided into only 10 grid points for the univariate AR(1) model and a 10×4 grid for the VAR(1) model. Further the CPU time was only seconds for the univariate AR(1) models and about 10 minutes for the two-dimensional VAR(1) model, which is considerably shorter than what would be needed for comparable accuracy using simulation methods.

CHAPTER 4

COMMON FACTORS AND LINEAR RELATIONSHIPS OF MULTIVARIATE TIME SERIES

Due to the ever-growing data made available by modern technologies, the effort to understand business processes is often challenged by co-existing high-dimensionality and autocorrelation. High-dimensionality obscures the true latent factors. In traditional multivariate literature, Principal Components Analysis is the standard tool for dimension reduction. For autocorrelated processes, however, PCA fails to take into account the time structure information, i.e., the autocorrelation information. Thus it is arguable that PCA is still the best choice. In this chapter I propose an enhanced dimension reduction method which by design takes into account both cross-correlation and autocorrelation information. I demonstrate through both a case study and a simulation study that the proposed method is better than PCA in dimension reduction and discovering latent factors for multivariate autocorrelated processes.

4.1 Motivation

Figure 4-1 shows a plot of the Furnace data introduced in Section 1.1.1. One interesting pattern is that all the dimensions of the data seem to move up and down at the same time. A similar pattern can be found in the Hog data displayed in Figure 1-3. In fact, it is common to observe such “co-movement” in multivariate time series processes. Consider for example the stock market. Stock prices move up and down as random walks. However, there seems to be an invisible hand that keeps most of the stocks moving in the same direction, the direction indicated by the Dow Jones Industrial Average index (DJIA) or the Standard and Poor’s 500 index (S&P500).

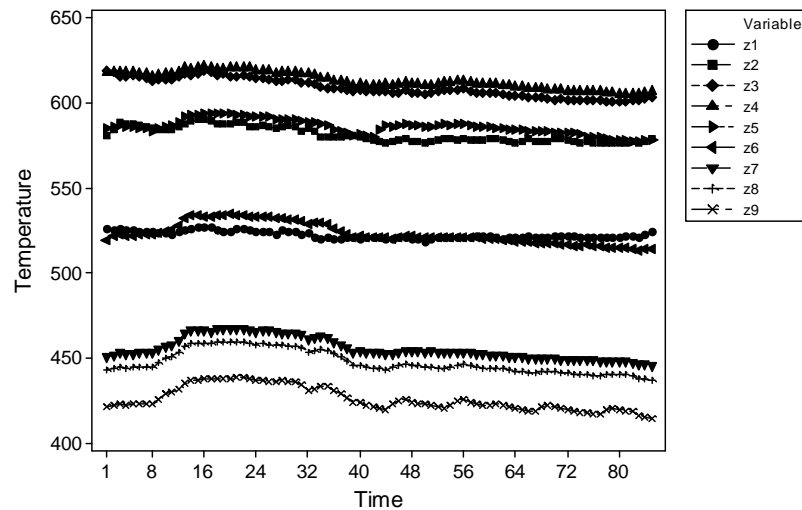


Figure 4-1. A traditional time series plot of temperatures measured at nine different locations of a large industrial furnace.

This phenomenon leads us to believe that the large dimension of multiple time series often might be driven by a smaller number of common sources, often referred to as common factors or latent variables. In the furnace example, a common factor could be an overall trend computed by taking the average of the temperatures across the furnace. In the Hog data example, it is conjectured that the GDP, a proxy for the overall economy, could be a common factor. Similarly, in the stock market, the market factor, suggested by Sharpe (1963, 1964) for the Capital Asset Pricing Model (CAPM), may serve as a common factor.

To discover the common factors for a multivariate time series process is an important issue from many perspectives. The immediate benefit of dimension reduction is that it frees us from the so-called “curse of dimensionality.” It is often a formidable task to fit a high-dimensional time series model with a large number of parameters. The problem becomes more accessible once we can reduce the problem to focusing on only a few common factors. Furthermore, finding a small number of common factors helps us

interpret the variability of the data. A reasonable short list of common factors may not only explain most of the in-sample variability, but also a large portion of the out-of-sample variability. A further reason is most relevant in the quality control context. Rather than overwhelming our eyes with every single observation, we may only need to monitor a few common factors. Last but not least, if a linear combination of the observed variables does not contain any common factors, it implies that there might be a stationary relationship among these variables. This relationship often provides insights into the underlying structure of the data and hence allows us to interpret the data. In the econometrics literature, the Error-Correction Model (ECM, see Engle and Granger, 1987) is build upon relationships of this kind.

4.2 Literature Review

Principal Components Analysis, Factor Analysis, and Canonical Correlation Analysis (CCA) have been used as tools to discover common factors in the general multivariate setting. These techniques involve calculations based on the contemporary variance-covariance matrix. Thus PCA in particular does not utilize any time-related information. For multivariate time series, endeavors have been made to devise better methods. Badcock et al (2004) devised a time-structure based but not a model based method. Meanwhile other researchers start with assuming a certain model and base their methods on that model. Approaches taken by them include the Canonical Correlation Analysis by Box and Tiao (1977), the maximum likelihood estimation method for cointegrated model by Johansen (1988), the Reduced Rank model by Velu, Reinsel and Wichern (1986) and the Dynamic Factor model by Peña and Poncela (2004). In this

section, important literature is reviewed with an emphasis on the method by Box and Tiao (1977).

4.2.1 The PCA Method

PCA applies to either the variance-covariance matrix Σ_z or the correlation matrix ρ_z of the observed data \mathbf{z}_t . I will focus on the covariance matrix based PCA only. The method is based on finding a set of projection vectors \mathbf{e}_i which are mutually orthonormal and sequentially maximize the variance of the transformed variables $\mathbf{e}_i' \mathbf{z}_t$ (Johnson and Wichern, 2002):

$$\max_{\mathbf{e}_i \perp \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{i-1}} \frac{\mathbf{e}_i' \Sigma_z \mathbf{e}_i}{\mathbf{e}_i' \mathbf{e}_i} \quad (4.1)$$

Unfortunately PCA is developed on the contemporary covariance matrix only. Thus this method does not capture any structure in the data that manifests itself in the autocorrelation and temporal cross-correlation structure. However, auto- and lagged cross- correlation are key features of time series processes. Thus we question the usefulness of the variability criteria used in PCA to decompose the data. Specifically we question the ability of PCA to retrieve the common factors in time series data.

4.2.2 The Box-Tiao Method

Box and Tiao (1977) introduced a new projection method based on decomposing the data according to “predictability” instead of “variability.” Their method, which we will call the Box-Tiao method, is essentially a Canonical Correlation Analysis between the vector of current observations and a set of past or lagged observations.

The derivation of the method is based on assuming a typical vector autoregressive time series model

$$\mathbf{z}_t = \hat{\mathbf{z}}_{t-1}(1) + \boldsymbol{\varepsilon}_t, \quad (4.2)$$

where $\hat{\mathbf{z}}_{t-1}(1)$ is a linear combination of past observations, representing the best prediction of \mathbf{z}_t made at time $t-1$. Accordingly for the variance, we have

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Sigma}_{\hat{z}} + \boldsymbol{\Sigma}_{\varepsilon}. \quad (4.3)$$

where $\boldsymbol{\Sigma}_z = \text{Var}(\mathbf{z}_t)$, $\boldsymbol{\Sigma}_{\hat{z}} = \text{Var}(\hat{\mathbf{z}}_{t-1}(1))$ and $\boldsymbol{\Sigma}_{\varepsilon} = \text{Var}(\boldsymbol{\varepsilon}_t)$.

By performing a linear transformation \mathbf{m}_i on both sides of (4.2), we get

$$\mathbf{m}_i' \mathbf{z}_t = \mathbf{m}_i' \hat{\mathbf{z}}_{t-1}(1) + \mathbf{m}_i' \boldsymbol{\varepsilon}_t. \quad (4.4)$$

The Box-Tiao method then amounts to finding a set of transformation vectors \mathbf{m}_i , $i = 1, K, k$, which sequentially maximize the predictability of the transformed variables $\mathbf{m}_i' \mathbf{z}_t$. The predictability is defined as how much variability of the transformed variable is explained by that of its best linear estimate at time $t-1$. Defined in this way is it similar to the R^2 criterion used in regression. The problem is thus formulated as

$$\max \frac{\mathbf{m}_i' \boldsymbol{\Sigma}_{\hat{z}} \mathbf{m}_i}{\mathbf{m}_i' \boldsymbol{\Sigma}_z \mathbf{m}_i}. \quad (4.5)$$

If we set $\mathbf{a}_i' = \mathbf{m}_i' \boldsymbol{\Sigma}_z^{1/2}$, (4.5) becomes,

$$\max \frac{\mathbf{m}_i' \boldsymbol{\Sigma}_z^{1/2} \cdot \boldsymbol{\Sigma}_z^{-1/2} \boldsymbol{\Sigma}_{\hat{z}} \boldsymbol{\Sigma}_z^{-1/2} \cdot \boldsymbol{\Sigma}_z^{1/2} \mathbf{m}_i}{\mathbf{m}_i' \boldsymbol{\Sigma}_z^{1/2} \cdot \boldsymbol{\Sigma}_z^{1/2} \mathbf{m}_i} = \max \frac{\mathbf{a}_i' \cdot \boldsymbol{\Sigma}_z^{-1/2} \boldsymbol{\Sigma}_{\hat{z}} \boldsymbol{\Sigma}_z^{-1/2} \cdot \mathbf{a}_i}{\mathbf{a}_i' \mathbf{a}_i}. \quad (4.6)$$

It is now clear that the sequence of maxima in (4.6) is the set of eigenvalues λ_i 's of $\Sigma_z^{-1/2} \Sigma_{\hat{z}} \Sigma_z^{-1/2}$, ranked from the highest to the lowest, with the \mathbf{a}_i 's being the corresponding eigenvectors. More explicitly substituting back $\mathbf{a}_i' = \mathbf{m}_i' \Sigma_z^{1/2}$ we get

$$\Sigma_z^{-1/2} \Sigma_{\hat{z}} \Sigma_z^{-1/2} \cdot \mathbf{a}_i = \lambda_i \mathbf{a}_i \Rightarrow \Sigma_z^{-1} \Sigma_{\hat{z}} \cdot \mathbf{m}_i = \lambda_i \mathbf{m}_i. \quad (4.7)$$

Thus it is immediately seen that the vector \mathbf{m}_i is an eigenvector of $\Sigma_z^{-1} \Sigma_{\hat{z}}$; the corresponding eigenvalue λ_i is the predictability of the i -th transformed variable.

The Box-Tiao components with predictability close to zero or close to one have interesting implications. Those close to zero represent transformed time series that basically are unpredictable random shocks. Thus these linear combinations of the original time series are stable relationships among the component variables. They are related to the concept of cointegration which we will pursue in more details in Section 4.3.2. The Box-Tiao components with eigenvalues close to one represent nearly non-stationary time series.

The original work by Box and Tiao (1977) is based on assuming a stationary vector time series. For non-stationary multivariate time series, we may therefore question the viability of the Box-Tiao method because it depends on the covariance matrices which are not defined in this situation. Velu et al. (1987) justified the Box-Tiao method in this case by analyzing asymptotic behaviors of the covariance matrices.

The Box-Tiao method provides an alternative to the PCA method and is more appealing in many ways. The foremost reason perhaps resides in our intuition that common factors differentiate themselves from each other in terms of their dynamics

rather than their variability, and it is the dynamics, not the variability, that does not change over time.

4.2.3 PCA on $\hat{\Sigma}_a$

So far I have focused on the variance and autocorrelation of $\{\mathbf{z}_t\}$. Another interesting perspective is to look at the random shocks \mathbf{a}_t , which drive the time series in multivariate ARIMA processes through $\Phi(B)\mathbf{z}_t = \Theta(B)\mathbf{a}_t$.

Tiao, Tsay and Wang (1993) illustrated the usefulness of several different linear transformation methods by studying 1-month, 3-month and 6-month Taiwan's interest-rate. They first fitted a VAR(4) model and then applied PCA to the residual covariance matrix $\hat{\Sigma}_a$. For this data set they observed that $\hat{\Sigma}_a$ was nearly singular. It turned out that by doing so, they were able to achieve parsimony parameterization and reveal interesting and meaningful features of the interest-rate. They also showed that this approach produced transformations very similar to the Box-Tiao transformation.

4.2.4 The Badcock Method

Badcock et al. (2004) developed a method for retrieving “temporally structured” components. Specifically they were seeking components that exhibit distinct differences in the persistence of autocorrelation. Their method was based on earlier work by Switzer (1985) used in image reconstruction.

Above we defined $\Gamma_l = \text{Cov}(\mathbf{z}_t, \mathbf{z}_{t-l})$. After applying a linear transformation \mathbf{f}_l to \mathbf{z}_t , the autocovariance of the transformed variable $\mathbf{f}_l' \mathbf{z}_t$ is $\mathbf{f}_l' \Gamma_l \mathbf{f}_l$. The Badcock method

then amounts to finding a set of transformation vectors \mathbf{f}_i , $i = 1, \dots, k$, which sequentially maximize the autocorrelation of the transformed variables $\mathbf{f}_i' \mathbf{z}_t$:

$$\max \frac{\mathbf{f}_i' \Gamma \mathbf{f}_i}{\mathbf{f}_i' \Sigma \mathbf{f}_i} . \quad (4.8)$$

This idea has some similarities to the Box-Tiao method but is essentially different. I will discuss this method in more detail in Section 4.3.3.

4.2.5 Dynamic Factor Model

In a dynamic factor model it is assumed that the observed vector of k time series are generated by $r < k$ common factors \mathbf{y}_t , plus a measurement error:

$$\begin{aligned} \mathbf{z}_t &= \mathbf{P} \mathbf{y}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\Phi}_y(B) \mathbf{y}_t &= \boldsymbol{\Theta}_y(B) \mathbf{a}_t, \end{aligned} \quad (4.9)$$

where $\boldsymbol{\varepsilon}_t$ is a white-noise sequence with full-rank covariance matrix Σ_ε . The matrix \mathbf{P} is a $k \times r$ loading matrix and \mathbf{y}_t is assumed to follow a r -dimensional VARMA(p, q) process. In what follows \mathbf{y}_t represents the common factors. If we further assume that $\boldsymbol{\Phi}_y(B)$ and $\boldsymbol{\Theta}_y(B)$ are diagonal then the model is called an uncoupled-factors model. Otherwise it is called a coupled-factors model.

A key problem in dynamic factor modeling is to identify the number r of latent factors. In the stationary case Peña and Box (1987) developed a procedure to identify this number by looking at the eigenvalues of lagged covariance matrices. Peña et al. (2004) generalized this method to the non-stationary case by analyzing the asymptotic behavior of re-defined covariance matrices.

4.3 Reflections on the Literature

4.3.1 The Box-Tiao Method and Canonical Correlation Analysis

Canonical correlation analysis originally developed by Hotelling (1936) seeks to identify and quantify the associations between two sets of variables. It focuses on finding linear combinations of the variables in one set and linear combinations of another set of variables that sequentially have the highest correlation. The idea is to first determine the pair of linear combinations having the largest correlation. Next, the pair of linear combinations having the largest correlation among all pairs uncorrelated with the initially selected pair, etc. The pairs of linear combinations are called the canonical variables, and their correlations are called canonical correlations. For technical details, see Johnson and Wichern (2002).

I now briefly show that for a VAR(p) process, the Box-Tiao method is the canonical correlation analysis between one set \mathbf{z}_t and another set $\{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots, \mathbf{z}_{t-p}\}$. The predictability of the i -th Box-Tiao component corresponds to the square of the canonical correlation of the i -th pair of canonical variables.

The VAR(p) model is readily expressed as,

$$\mathbf{z}_t = \Phi_1 \mathbf{z}_{t-1} + L + \Phi_p \mathbf{z}_{t-p} + \mathbf{a}_t. \quad (4.10)$$

The Yule-Walker equations (see Reinsel 1997) for this model implies that

$$\begin{bmatrix} \Gamma_1' \\ \Gamma_2' \\ M \\ \Gamma_p' \end{bmatrix} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & L & \Gamma_{p-1} \\ \Gamma_{-1} & \Gamma_0 & L & \Gamma_{p-2} \\ M & M & O & M \\ \Gamma_{1-p} & \Gamma_{2-p} & L & \Gamma_0 \end{bmatrix} \begin{bmatrix} \Phi_1' \\ \Phi_2' \\ M \\ \Phi_p' \end{bmatrix}. \quad (4.11)$$

The Canonical Correlation analysis between the set $\{\mathbf{z}_t\}$ (set 1) and the set $\{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \mathbf{K}, \mathbf{z}_{t-p}\}$ (set 2) amounts to finding the eigenvectors of $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$, where $\mathbf{\Sigma}_{ii}$ is the variance-covariance matrix of set i , and $\mathbf{\Sigma}_{ij}$ is the variance between set i and set j . Specifically with the Yule-Walker equations (4.11) substituted in, $\mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1} \mathbf{\Sigma}_{21}$ can be expressed as,

$$\begin{aligned} & \mathbf{\Sigma}_z^{-1} \begin{bmatrix} \mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \mathbf{K}, \mathbf{\Gamma}_p \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1 & \mathbf{L} & \mathbf{\Gamma}_{p-1} \\ \mathbf{\Gamma}_{-1} & \mathbf{\Gamma}_0 & \mathbf{L} & \mathbf{\Gamma}_{p-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{\Gamma}_{1-p} & \mathbf{\Gamma}_{2-p} & \mathbf{L} & \mathbf{\Gamma}_0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Gamma}'_1 \\ \mathbf{\Gamma}'_2 \\ \mathbf{M} \\ \mathbf{\Gamma}'_p \end{bmatrix} \\ &= \mathbf{\Sigma}_z^{-1} \begin{bmatrix} \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{K}, \mathbf{\Phi}_p \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1 & \mathbf{L} & \mathbf{\Gamma}_{p-1} \\ \mathbf{\Gamma}_{-1} & \mathbf{\Gamma}_0 & \mathbf{L} & \mathbf{\Gamma}_{p-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{\Gamma}_{1-p} & \mathbf{\Gamma}_{2-p} & \mathbf{L} & \mathbf{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}'_1 \\ \mathbf{\Phi}'_2 \\ \mathbf{M} \\ \mathbf{\Phi}'_p \end{bmatrix}. \end{aligned} \quad (4.12)$$

As demonstrated in section 4.2.2, the Box-Tiao transformations are the eigenvectors of $\mathbf{\Sigma}_z^{-1} \mathbf{\Sigma}_{\hat{z}}$, which is the same as (4.12) by recognizing that

$$\begin{aligned} & \mathbf{\Sigma}_{\hat{z}} = \text{Cov}(\mathbf{\Phi}_1 \mathbf{z}_{t-1} + \mathbf{L} + \mathbf{\Phi}_p \mathbf{z}_{t-p}) \\ &= \begin{bmatrix} \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{K}, \mathbf{\Phi}_p \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1 & \mathbf{L} & \mathbf{\Gamma}_{p-1} \\ \mathbf{\Gamma}_{-1} & \mathbf{\Gamma}_0 & \mathbf{L} & \mathbf{\Gamma}_{p-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{\Gamma}_{1-p} & \mathbf{\Gamma}_{2-p} & \mathbf{L} & \mathbf{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}'_1 \\ \mathbf{\Phi}'_2 \\ \mathbf{M} \\ \mathbf{\Phi}'_p \end{bmatrix}. \end{aligned} \quad (4.13)$$

Thus we see the relationship between Box-Tiao transformation and the canonical correlation analysis. The first Box-Tiao component is the linear transformation of the set $\{\mathbf{z}_t\}$ which has the largest correlation with a corresponding linear transformation on the second set $\{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \mathbf{K}, \mathbf{z}_{t-p}\}$.

Note that the predictability of the i -th Box-Tiao component corresponds to the i -th eigenvalue λ_i of (4.12), while the correlation of the the i -th pair of canonical variables corresponds to the square-root of the i -th eigenvalue λ_i of (4.12). Note also that eigenvalues essentially are the R^2 's in regression analysis.

4.3.2 The Box-Tiao Method, Cointegration and Co-movement

When a univariate time series have d unit-roots, it is said to be integrated of order d , denoted as $I(d)$. For non-stationary multivariate time series, it is possible for there to be a linear combination of integrated variables that is stationary. These variables are said to be cointegrated. More specifically, the components of the vector $\mathbf{z}_t = [z_{1t}, z_{2t}, \dots, z_{kt}]'$ are said to be cointegrated of order d, b , denoted by $\mathbf{z}_t \sim CI(d, b)$ if all components of \mathbf{z}_t are integrated of order d and there exists a vector $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_k]'$ such that the linear combination $\boldsymbol{\beta}'\mathbf{z}_t$ is integrated of order $d - b$ where $b > 0$. See Engle and Granger (1987) for more technical details.

If cointegration does exist, being able to recognize it may help us avoid over-differencing the data and prevent the fitted model from being non-invertible. In the econometric literature, cointegration is an important issue because it implies a long-run economic equilibrium.

Developed a decade earlier than the cointegration econometrics literature, the Box-Tiao method is based on the same idea. Consider an integrated multivariate time series process. If there exists a Box-Tiao component with predictability near zero, then the related Box-Tiao component is a purely random noise series which implies a stable

relationship among the integrated component variables. Such a relation is generally called co-movement. In the special case where the original vector time series is non-stationary then it is called a cointegrated relationship. Note if a Box-Tiao component of non-stationary vector time series has predictability between zero and one (i.e., the Box-Tiao component follow a general ARMA model), it also indicates co-movement but not cointegration. On the other hand, a Box-Tiao component with predictability near one, is a latent variable that when mixed into all the component variables makes every component variable look non-stationary. Table 4-1 summarizes how predictability λ and unit roots determine whether the case is cointegration or co-movement.

Table 4-1. Cointegration and Co-movement.

	Whether there are Unit Roots	
	Yes	No
$\lambda = 0$	Cointegration Or Co-movement	Co-movement
$0 < \lambda < 1$	Cointegration	Neither

We conclude that the Box-Tiao method can be used to identify co-movement as well as cointegration. Consider for example a vector time series where there is Box-Tiao component with predictability equal to 0.6 as well as a component with eigenvalue close to zero. Although the time series may not be integrated, the Box-Tiao component with predictability 0.6 contributes to every component variables in the process and influences them to co-move to some extent.

4.3.3 The Box-Tiao Method and the Badcock Method

The Box-Tiao method and the Badcock method have certain similarities. They are both based on an assumption that common factors differentiate themselves from each

other in terms of time structure. The former uses predictability as the criterion while the latter uses autocorrelation. In many cases, the two methods produce similar transformations. For example, if the time series follows a VAR(1) model, the Box-Tiao transformation vectors are eigenvectors of $\Sigma_z^{-1}\Phi\Sigma_z\Phi'$ while the Badcock transformation vectors are of the eigenvectors of $\Sigma_z^{-1}(\Phi\Sigma_z + \Sigma_z'\Phi')$.

Based on my research I find that the Box-Tiao method is more appealing. The predictability criterion used by the Box-Tiao method is better because it takes into account the whole history of \mathbf{z}_t while the Badcock method only considers the autocorrelation at one particular lag. Badcock et al. (2004) suggests to pick the specific lag in a more or less ad hoc and arbitrary fashion. Secondly, Badcock et al. (2004) suggested using $(\Gamma_l + \Gamma_{-l})/2$ to approximate Γ_l without accessing the difference and its impact. From this perspective, it appears to be an ad hoc method not based on the particular time series model.

Despite these misgivings, the Badcock method is easier to use because it does not necessitate the fitting of a model. However, this “advantage” may not be worth it as we give up the benefits that come from knowing the underlying model.

4.3.4 A Simple Augmented Model

To gain better understanding of how the PCA and Box-Tiao methods work, I propose a simple augmented model to study the performance of the two methods. The model is a dynamic factor model with time-structured noise. It is a special case of the Dynamic Factor Model (4.9).

In real world applications the noise is not necessarily “white.” For example, in the furnace example, the ambient temperature could affect the temperature readings from the thermocouples and it itself could be autocorrelated. Although this added noise has as strong time structures as common factors, we assume that it has a small loading in the factor model. When retrieving common factors, we are not particularly interested in this added noise.

I will assume that the dimension of the observations \mathbf{z}_t is k , that there are $r < k$ dimensional *common factors* vector \mathbf{y}_t and s dimensional time structured *noise factors* vector \mathbf{n}_t . Further I assume that $k \geq r + s$ and the model is

$$\begin{aligned}\mathbf{z}_t &= \mathbf{P}\mathbf{y}_t + \mathbf{Q}\mathbf{n}_t + \boldsymbol{\varepsilon}_t = [\mathbf{P}\mathbf{Q}] \begin{bmatrix} \mathbf{y}_t \\ \mathbf{n}_t \end{bmatrix} + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\Phi}_y(B)\mathbf{y}_t &= \boldsymbol{\Theta}_y(B)\mathbf{a}_t \\ \boldsymbol{\Phi}_n(B)\mathbf{n}_t &= \boldsymbol{\Theta}_n(B)\mathbf{b}_t,\end{aligned}\tag{4.14}$$

where $\boldsymbol{\varepsilon}_t$, \mathbf{a}_t and \mathbf{b}_t are white noises uncorrelated with each other. Further $\boldsymbol{\Phi}_y(B)$, $\boldsymbol{\Theta}_y(B)$, $\boldsymbol{\Phi}_n(B)$, $\boldsymbol{\Theta}_n(B)$, $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ are diagonal matrices so that all time-structured components and the noise are uncoupled. It is also assumed that \mathbf{n}_t has a smaller loading than \mathbf{y}_t , i.e., $\max_{j=1 \sim s} (q_{ij}) \ll \max_{j=1 \sim r} (p_{ij})$, $\forall i$, where p_{ij} is the ij -th element of matrix \mathbf{P} and q_{ij} is the ij -th element of matrix \mathbf{Q} . $k \times (r + s)$ matrix $[\mathbf{P}\mathbf{Q}]$ is of full rank $r + s$. Without loss of generality, it is further assumed that $\boldsymbol{\Sigma}_y = \mathbf{I}$ and $\boldsymbol{\Sigma}_n = \mathbf{I}$. The goal of dimension reduction is to retrieve \mathbf{y}_t .

4.3.5 A Simulation Study

To gain insights to how PCA and the BT method works I now provide a discussion of a simulation study based on (4.14). I choose r to be 2 and s to be 1. Three independent time series were simulated. They were designed to differ in both variability and autocorrelation structure. The model was

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \text{ where } \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad (4.15)$$

$$n_{1t} = 0.95n_{1t-1} + b_{1t} \text{ where } b_{1t} \sim N(0, 0.05).$$

The first series y_{1t} is white noise, but it has large variability. The second series y_{2t} and the noise series n_{1t} are similar in time structure, both positively autocorrelated, yet are different in variability. The variance of y_{2t} , denoted by $\sigma_{y_2}^2$, is 11.11 and the variance of n_{1t} , denoted by $\sigma_{n_1}^2$, is 0.51. Figure 4-2 shows the time series plot of y_1 , y_2 and n_1 against the same scale. Obviously n_1 is dominated by y_{1t} and y_{2t} in variability while exhibiting a clear time series structure.

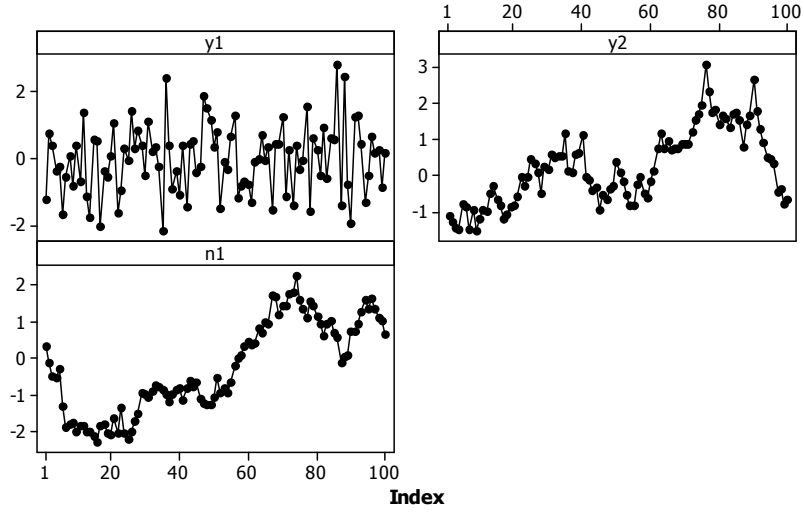


Figure 4-2. Time Series plot of y_1 , y_2 and n_1 .

Next I create nine time series by adding up linear combinations of the underlying common and noise factors $\mathbf{x}_t = [\mathbf{y}'_t \mathbf{M}_t]'$ and a random noise vector $\boldsymbol{\varepsilon}_t$

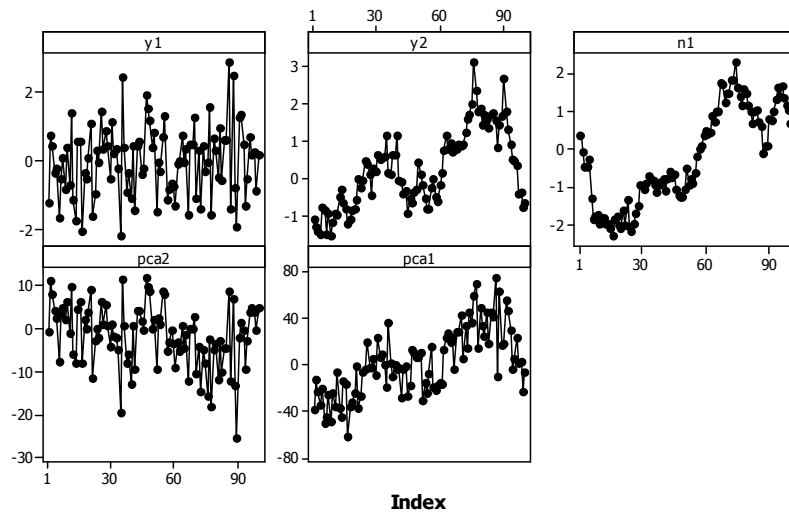
$$\begin{bmatrix} z_{1t} \\ z_{2t} \\ \mathbf{M} \\ z_{9t} \end{bmatrix} = \begin{bmatrix} 2 & 9 & 1 \\ 6 & 3 & 2 \\ 4 & 1 & 1 \\ 9 & 7 & 2 \\ 4 & 4 & 1 \\ 6 & 9 & 3 \\ 2 & 5 & 1 \\ 8 & 3 & 2 \\ 5 & 7 & 2 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ n_{1t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \mathbf{M} \\ \varepsilon_{9t} \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \mathbf{M} \\ \varepsilon_{9t} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{I}_9) \quad (4.16)$$

Note that if not for $\boldsymbol{\varepsilon}_t$ the variance-covariance matrix of \mathbf{z}_t would be singular.

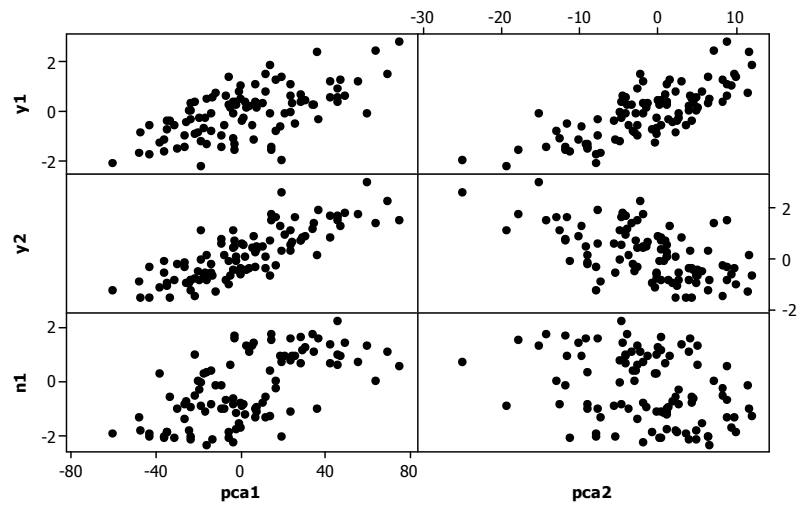
This added-in noise disguises the true dimensionality. Also note the linear transforming matrix is chosen such that the weight on the noise factor n_1 will not make it dominate the

common factors. This design assures n_1 does not contribute much variability to the observed data set \mathbf{z}_t .

The PCA method suggests the first two PCs explain 99.7% of variability. However, when we compare them with y_1 and y_2 , as shown in Figure 4-3, the results is not satisfying. The PCs are not able to track y_1 and y_2 accurately. Both are confounded mix of underlying components.



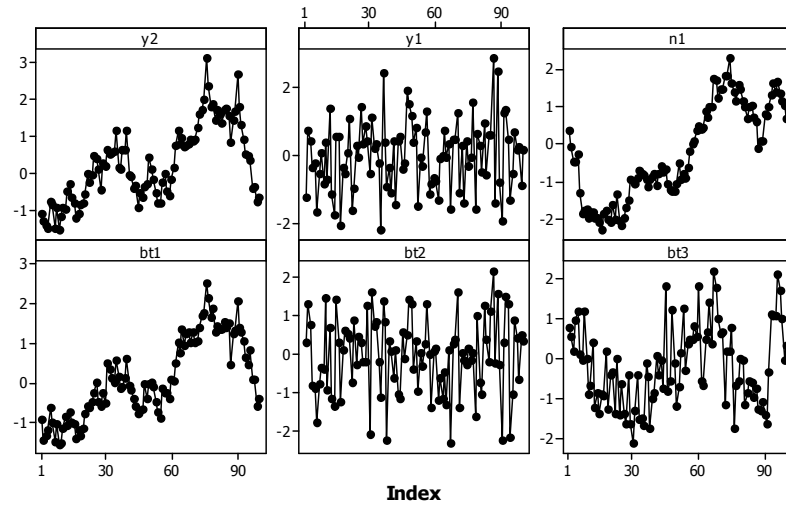
(a)



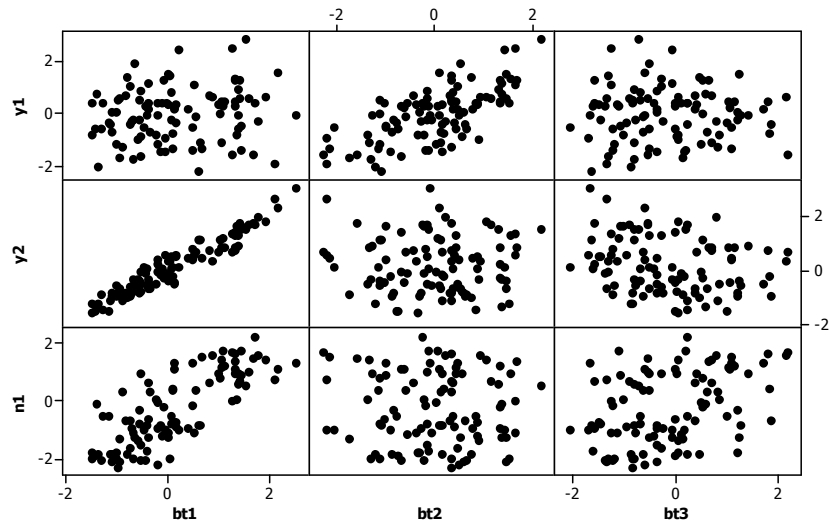
(b)

Figure 4-3. (a) The Time Series plot of y_1 , y_2 , n_1 and the first two PCs; (b) The Matrix plot of y_1 , y_2 and n_1 against two PCs.

The Box-Tiao Method provides interesting insights. Shown in Figure 4-4, the first Box-Tiao component tracks y_2 well while y_1 is not tracked out at all.



(a)



(b)

Figure 4-4. (a) The Time Series plot of y_1 , y_2 , n_1 and the first three Box-Tiao components; (b) The Matrix plot of y_1 , y_2 and n_1 against the first three Box-Tiao components.

This simulation result is not surprising. By design the Box-Tiao transformation seeks the components with predictability ranking from the highest to the lowest. The second common factor y_2 has a clear time structure and a considerable variability. The noise factor n_1 has a clear time structure as well but because of its low variability, it

easily gets contaminated. The first common factor y_1 does not have predictability at all because it is white noise; it is thus hard to be singled out from other white noise.

No conclusion is attempted to draw from one simulation study, but it is conjectured when either method is applied alone, problem happens. They are each designed for a special function; either function working alone will not produce the true underlying components.

4.3.6 Other Application of the Box-Tiao Method

Much of the past research built upon the Box-Tiao method focused on unit roots and cointegration. Bossaerts (1988) suggested that the Box-Tiao components could be directly tested for the presence of unit roots using the methods and critical values provided by Dickey and Fuller (1979). Bewley et al. (1994) found that the Box-Tiao estimator of the cointegrating parameter (the Box-Tiao eigenvectors) performs well compared to Johansen's (1988) maximum likelihood estimator. Bewley and Yang (1995) proposed cointegration tests based on the Box-Tiao method and found that compared to Johansen's (1988) and Engle and Yoo's (1987) tests, the Box-Tiao based tests are more robust to correlation between the disturbances in the cointegrating relationships and those generating the common trends. In a more recent paper by Anderson et al. (2006), it was shown that compared with Johansen's test (1988), the Box-Tiao based test shows promise in dealing with small sample size, fat tails, heteroskedasticity and skewness.

4.3.7 Spurious Box-Tiao Components

Box and Tiao (1977) cautioned that their method has a potential problem if the residual covariance matrix Σ_a is singular. If Σ_a is of rank k_1 , it is possible to transform the VAR(p) time series model of \mathbf{z}_t to

$$\begin{bmatrix} \mathbf{z}_{1t} \\ \mathbf{z}_{2t} \end{bmatrix} = \sum_{l=1}^p \begin{bmatrix} \Phi_{11}^{(l)} & \Phi_{12}^{(l)} \\ \Phi_{21}^{(l)} & \Phi_{22}^{(l)} \end{bmatrix} \mathbf{B}^l \begin{bmatrix} \mathbf{z}_{1t} \\ \mathbf{z}_{2t} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{1t} \\ \mathbf{0} \end{bmatrix}, \quad (4.17)$$

where \mathbf{b}_{1t} is of dimension k_1 . In another words, the $(k - k_1)$ -dimensional vector \mathbf{z}_{2t} is completely predictable from past values $\mathbf{z}_{1(t-l)}$ and $\mathbf{z}_{2(t-l)}$, $l = 1, 2, \dots, p$. Thus it is possible that eigenvalues close to one could come from either integrated components or a singular Σ_a . So far we have no method for distinguishing between these two cases.

Another singularity problem occurs when there is lagged linear redundancy in \mathbf{z}_t (e.g., $z_{i,t} = z_{j,t-1}$). In this case, one of the canonical correlations would be exactly equal to 1. However, it does not mean that we have an integrated component at all. Rather it comes from the lagged relationship and we want to know how to treat this case.

If we rewrite the VARMA model as $\mathbf{z}_t = \Phi^{-1}(B)\Theta(B)\mathbf{a}_t$, then it is clear that that singularity of either of the matrices $\Phi(B)$ and $\Theta(B)$ will cause singularity of the observed data. As will be demonstrated with the furnace data in the next section, these spurious components make it tricky to use the Box-Tiao method to tackle dimension reduction problem alone.

4.4 A Combined method of both the PCA and the Box-Tiao Methods

The analysis in Section 4.3.5 suggests neither the PCA method nor the Box-Tiao method on its own can retrieve common factors exactly. In Section 4.4.1 we further demonstrate the problem of applying the Box-Tiao method individually through both the furnace data and a simulated example. In Section 4.4.2 we introduce a two-step procedure of sequentially applying the PCA method and the Box-Tiao method. In Section 4.4.3 we prove the effectiveness of the combined method based on some approximation. In Section 4.4.4 we further demonstrate its effectiveness with a revisit to the simulated example discussed in Section 4.3.5 and in Section 4.4.5 we revisit the furnace example.

4.4.1 The Problem of Using the Box-Tiao method Individually

Recall the furnace example we introduced in Section 1.1.1. Figure 1-2 shows temperature measurements at nine locations of the ceramic furnace. By eyeballing each individual series in Figure 1-2 we find none of them resemble an I.I.D process. They all exhibit certain dynamic and even some non-stationarity. More interestingly, z_{6t} through z_{9t} all look alike; they reach peaks and troughs around the same time. It seems plausible that the real dimension is much smaller than nine, say, two or three.

If we rely on the Box-Tiao method to pick the important factors, however, we soon run into troubles. Assuming an AR(1) model, the Box-Tiao analysis generates a set of components with predictability shown in Figure 4-5.

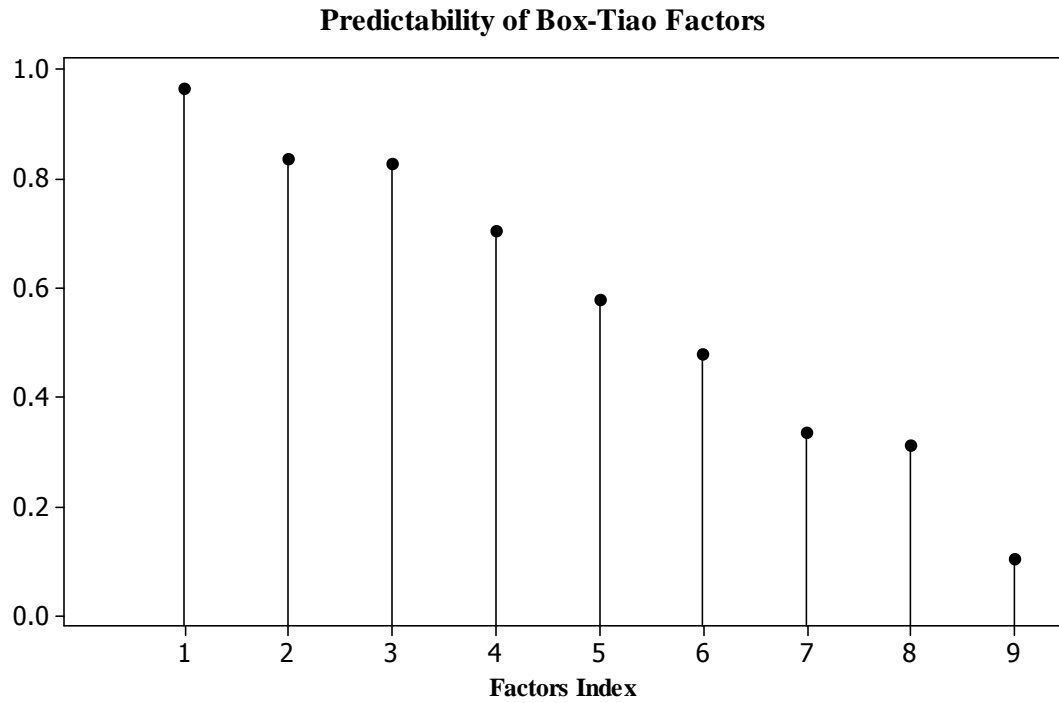


Figure 4-5. Predictability of Box-Tiao Factors of the Furnace data.

Among the nine Box-Tiao components, five have predictability higher than 0.5 and eight of them are obviously autocorrelated. However, does it mean five or eight of these components are important underlying factors? A plot of the correlation of these Box-Tiao factors (labeled BT1~BT9) with each observed variable $z_1 \sim z_9$, as shown in Figure 4-6, tells an interesting story. The fourth Box-Tiao component BT4, for example, despite of having a predictability as high as 0.71, has low correlation with all observed variables (the highest correlation is with z_3 and it is 0.45); or stated in another way, it has limited contribution to the overall variability. Same argument applies to the second, fifth, seventh, eighth and ninth components.

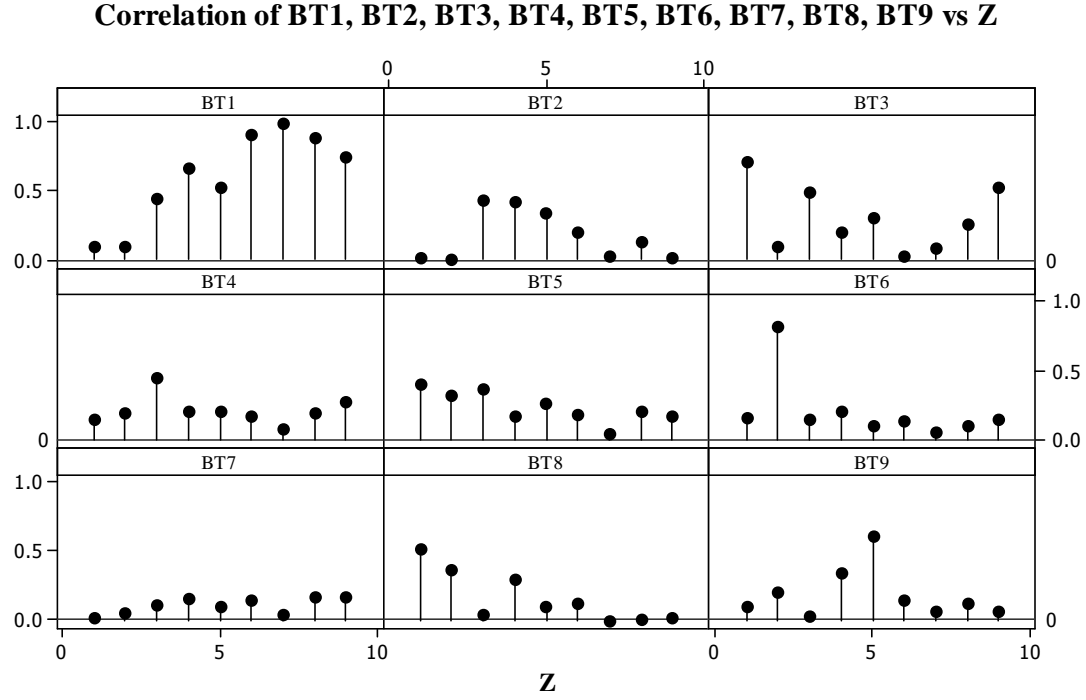


Figure 4-6. Correlation of each Box-Tiao factor with each dimension of observed data.

Together with our prior knowledge from Section 4.3.7 on how possible suspicious components could emerge, we see the Box-Tiao method alone may not lead to meaningful dimension reduction. The component with high predictability could result from a highly correlated error term, lagged observation or could be a time-structured noise factor. With what follows I further demonstrate this idea with a simulation.

The Box-Tiao analysis on the Furnace data reveals that there are highly predictable components with limited contribution to variability. This confirms with the idea of the proposed augmented model (4.14) where the noise factors are separated from the common factors because of their little contribution to the overall variability. In the following I illustrate with one realization of (4.14) how noise factors could affect the Box-Tiao analysis. The observation \mathbf{z}_t and white noise $\boldsymbol{\varepsilon}_t$ are both 9×1 vectors. $\boldsymbol{\Sigma}_{\varepsilon} = \mathbf{I}_9$.

$\boldsymbol{\Sigma}_y = \mathbf{I}_3$. $\boldsymbol{\Sigma}_n = \mathbf{I}_3$. The loading matrix \mathbf{P} and \mathbf{Q} are also randomly generated and then

scaled to ensure $\max_{j=1 \sim r} (p_{ij}) \geq 1.5q_{ij}, \forall i$. \mathbf{y}_t and \mathbf{n}_t are both simulated from a VAR(1)

model with first-order autoregressive coefficient matrix Φ_y and Φ_n respectively:

$$\Phi_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} \quad \Phi_n = \begin{bmatrix} 0.75 & 0 & 0 \\ 0 & 0.85 & 0 \\ 0 & 0 & 0.95 \end{bmatrix}$$

Only one common factor is autocorrelated. However, one run of this simulation shows at least three retrieved Box-Tiao components are predictable with predictability equal to 0.93, 0.47 and 0.26. The reason behind this is that the highly autocorrelated noise factors contaminate the analysis.

4.4.2 A Two-step Combined Method

The PCA and Box-Tiao methods use contemporary and temporary covariance information respectively and use different optimization criteria. However, both methods consist of performing a linear transformation of the data. Hence if combined, the new transformation will also be a linear transformation. It is not so easy to see what properties a combined transformation will have. However, I conjecture a combination of both may have interesting properties. The combined method consists of a two-step procedure:

Step 1: Apply the PCA method, keeping the first r components according to the scree plot.

Step 2: Apply the Box-Tiao method on the remaining r components. The transformed components will be the final results.

The idea behind this procedure is that in the first step, the noise, regardless of its time structure, will be peeled off because dimensions with small variability will be

discarded. Thus after the first step, what is left over is hopefully a linearly transformed \mathbf{y}_t . After applying the Box-Tiao method which will order linearly transformed variable according to predictability, exactly the \mathbf{y}_t will appear as the final results.

Note that applying the two methods in a reverse order does not work. If we apply the Box-Tiao method first, the variability of the transformed time series will be distorted and it is no longer possible to order the time series according to variability as is hoped to be done with PCA.

4.4.3 Demonstration of the Combined Method

In this section I demonstrate through deduction how the combined method combines the merits of both the PCA and the Box-Tiao methods.

Proposition 1: PCA method cannot always retrieve the underlying components \mathbf{y}_t from \mathbf{z}_t in the model (4.14).

Proof. I prove the *Proposition 1* with an approximation that $\boldsymbol{\varepsilon}_t$ is 0. The same approximation is also used in *Proposition 2 and 3*.

Since

$$\mathbf{z}_t = [\mathbf{P}\mathbf{M}\mathbf{Q}] \begin{bmatrix} \mathbf{y}_t \\ \mathbf{n}_t \end{bmatrix},$$

to be able to retrieve the first component, there has to be a transformation vector \mathbf{e}_i' , so that $\mathbf{e}_i' [\mathbf{P}\mathbf{M}\mathbf{Q}] = [1 \ 0 \ \dots \ 0]$, which means $\mathbf{e}_i \perp \mathbf{p}_j, \forall j \neq 1$. Similar pattern follows for other transformation vectors.

Consider a simple case where $m = 2$, $r = 2$, $s = 0$ and $\cos\langle \mathbf{p}_1, \mathbf{p}_2 \rangle \neq 0$. Further assume that $\boldsymbol{\varepsilon}_t$ is very small and $\boldsymbol{\Sigma}_\varepsilon$ is close to $\mathbf{0}$. In this two-dimension case, to retrieve the first component, as implied above, $\mathbf{e}_1 \perp \mathbf{p}_2$ and through $\cos\langle \mathbf{p}_1, \mathbf{p}_2 \rangle \neq 0$ we infer $\cos\langle \mathbf{e}_1, \mathbf{p}_1 \rangle \neq \pm 1$, which means \mathbf{e}_1 is independent of \mathbf{p}_1 . To retrieve the second component, $\mathbf{e}_2 \perp \mathbf{p}_1$ and $\mathbf{e}_2 \perp \mathbf{e}_1$, which is impossible in two dimension space since \mathbf{e}_1 is independent of \mathbf{p}_1 . This counterexample indicates that PCA is not capable to retrieve true factors exactly.

■

In the meanwhile the Box-Tiao method alone cannot discriminate the noise \mathbf{n}_t from the more important common factors \mathbf{y}_t . A heuristic argument will be given through three steps.

Proposition 2: for any VARMA(p,q) process $\boldsymbol{\Phi}(L)\mathbf{x}_t = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t$, if all $\boldsymbol{\Phi}_j$, $j = 1, \dots, p$, $\boldsymbol{\Theta}_j$, $j = 1, \dots, q$ and $\boldsymbol{\Sigma}_\varepsilon$ are diagonal, then the Box-Tiao transformation of \mathbf{x}_t is \mathbf{x}_t .

Proof: If $\boldsymbol{\Phi}_j$, $\boldsymbol{\Theta}_j$ and $\boldsymbol{\Sigma}_\varepsilon$ are all diagonal, then $\boldsymbol{\Gamma}_0(\mathbf{x}) = \text{Cov}(\mathbf{x}_t)$ is diagonal.

For the equivalent expression $\mathbf{x}_t = \hat{\mathbf{x}}_{t-1}(1) + \boldsymbol{\varepsilon}_t$, define $\boldsymbol{\Gamma}_0(\hat{\mathbf{x}}) = \text{Cov}(\hat{\mathbf{x}}_{t-1}(1))$, thus we have

$\boldsymbol{\Gamma}_0(\mathbf{x}) = \boldsymbol{\Gamma}_0(\hat{\mathbf{x}}) + \boldsymbol{\Sigma}_\varepsilon$. Since $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Gamma}_0(\mathbf{x})$ are diagonal, then $\boldsymbol{\Gamma}_0(\hat{\mathbf{x}})$ is also diagonal.

Now the Box-Tiao transformation vectors are the eigenvectors of $\mathbf{Q} = \boldsymbol{\Gamma}_0^{-1}(\mathbf{x})\boldsymbol{\Gamma}_0(\hat{\mathbf{x}})$.

Thus since $\boldsymbol{\Gamma}_0(\mathbf{x})$ and $\boldsymbol{\Gamma}_0(\hat{\mathbf{x}})$ are both diagonal, \mathbf{Q} is also diagonal. It then follows that

the eigenvectors of \mathbf{Q} are the columns of \mathbf{I} . Thus the Box-Tiao transformation of \mathbf{x}_t is

$$\mathbf{I}\mathbf{x}_t = \mathbf{x}_t.$$

■

Proposition 3: the Box-Tiao transformation is invariant to a full rank linear transformation. Specifically if the Box-Tiao transformation of \mathbf{x}_t gives \mathbf{y}_t , the Box-Tiao transformation on $\mathbf{z}_t = \mathbf{T}\mathbf{x}_t$ is also \mathbf{y}_t , for any full rank square matrix \mathbf{T} .

Proof: Suppose $\mathbf{x}_t = \hat{\mathbf{x}}_{t-1}(1) + \boldsymbol{\varepsilon}_t$, $\Gamma_0(\mathbf{x}) = \text{Cov}(\mathbf{x}_t)$ and $\Gamma_0(\hat{\mathbf{x}}) = \text{Cov}(\hat{\mathbf{x}}_{t-1}(1))$. We then have $\mathbf{y}_t = \mathbf{T}\mathbf{x}_t = \mathbf{T}\hat{\mathbf{x}}_{t-1}(1) + \mathbf{T}\boldsymbol{\varepsilon}_t$, $\Gamma_0(\mathbf{y}) = \text{Cov}(\mathbf{y}_t) = \mathbf{T}\Gamma_0(\mathbf{x})\mathbf{T}'$ and $\Gamma_0(\hat{\mathbf{y}}) = \text{Cov}(\hat{\mathbf{y}}_{t-1}(1)) = \text{Cov}(\mathbf{T}\hat{\mathbf{x}}_{t-1}(1)) = \mathbf{T}\Gamma_0(\hat{\mathbf{x}})\mathbf{T}'$. Box-Tiao transformation vectors are eigenvectors of $\mathbf{Q} = \Gamma_0^{-1}(\mathbf{x})\Gamma_0(\hat{\mathbf{x}})$ for \mathbf{x}_t and eigenvectors of

$$\tilde{\mathbf{Q}} = \Gamma_0^{-1}(\mathbf{y})\Gamma_0(\hat{\mathbf{y}}) = \mathbf{T}'^{-1}\mathbf{Q}\mathbf{T}' \text{ for } \mathbf{y}_t.$$

If $\mathbf{Q}\mathbf{m} = \lambda\mathbf{m}$, then $\mathbf{T}'^{-1}\mathbf{Q}\mathbf{T}'\mathbf{T}'^{-1}\mathbf{m} = \mathbf{T}'^{-1}\lambda\mathbf{m}$. Now let $\tilde{\mathbf{Q}} = \mathbf{T}'^{-1}\mathbf{Q}\mathbf{T}'$ then

$$\tilde{\mathbf{Q}}(\mathbf{T}'^{-1}\mathbf{m}) = \lambda(\mathbf{T}'^{-1}\mathbf{m}). \text{ Further denote by } \tilde{\mathbf{m}} = \mathbf{T}'^{-1}\mathbf{m}, \text{ then } \tilde{\mathbf{m}} \text{ is an eigenvector of } \tilde{\mathbf{Q}}.$$

The Box-Tiao transformation of $\mathbf{y}_t = \mathbf{T}\mathbf{x}_t$ associated with eigenvector $\tilde{\mathbf{m}}$ is

$\tilde{\mathbf{m}}'\mathbf{y}_t = \tilde{\mathbf{m}}'\mathbf{T}^{-1}\mathbf{T}\mathbf{x}_t = \tilde{\mathbf{m}}'\mathbf{x}_t$. Thus we see that this is equivalent to the Box-Tiao component of \mathbf{x}_t associated with the eigenvector \mathbf{m} of \mathbf{Q} .

■

Finally based on *Proposition 2 and 3* I demonstrate by way of a heuristic argument that applying the Box-Tiao method alone retrieves both \mathbf{y}_t and \mathbf{n}_t in the

simplified model (4.14). I acknowledge that my argument is not a proof because it is based on assuming that $\boldsymbol{\varepsilon}_t$ is 0.

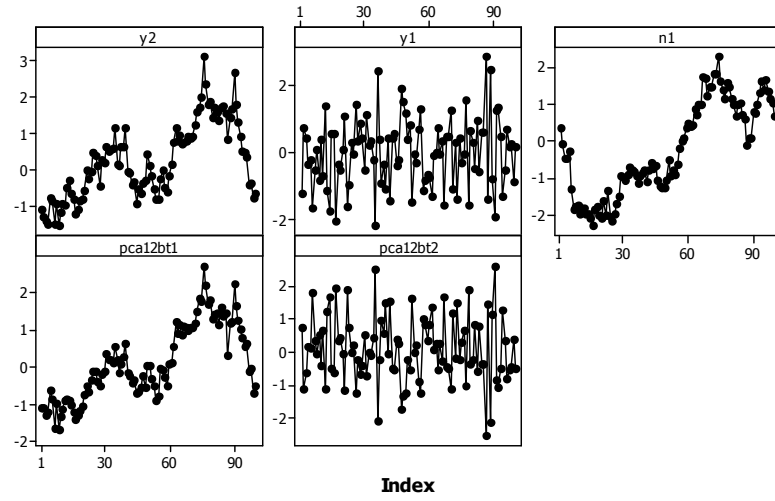
When $m = r + s$, as discussed above, the Box-Tiao transformation of $[\mathbf{y}'_t, \mathbf{n}'_t]'$ is an identity transformation. Thus because of the invariance of the Box-Tiao transformation, the linear transformation $\mathbf{z}_t = [\mathbf{P}\mathbf{M}\mathbf{Q}][\mathbf{y}'_t, \mathbf{n}'_t]'$ is $[\mathbf{y}'_t, \mathbf{n}'_t]'$.

When $m > r + s$, the first $r + s$ components will still be $[\mathbf{y}'_t, \mathbf{n}'_t]'$ while the remaining components will be 0.

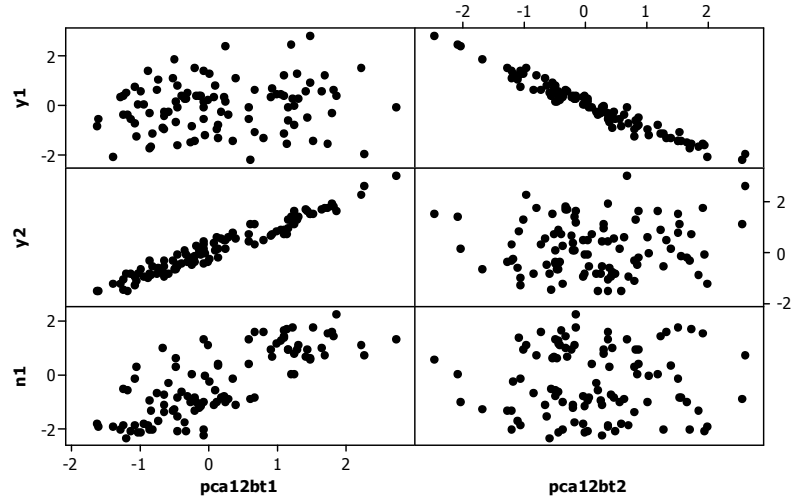
My conclusion based on above reasoning is that the Box-Tiao transformation retrieves both \mathbf{y}_t and \mathbf{n}_t . However, it cannot further distinguish \mathbf{y}_t from \mathbf{n}_t .

4.4.4 Simulated Study Revisit

As a demonstration of this two-step procedure, I try it on the simulated data from Section 4.3.5. After the first step I kept the first two PCs and then transform those again with the Box-Tiao method. Figure 4-7 shows how closely the two final retrieved components match the true underlying components y_1 and y_2 . Component “pca12bt1” is the first Box-Tiao component on the first two PCs; component “pca12bt2” is the second Box-Tiao component on the first two PCs.



(a)



(b)

Figure 4-7. (a) The Time Series plot of y_1 , y_2 , n_1 and the two PCA-BT components; (b) The Matrix plot of y_1 , y_2 and n_1 against the two PCA-BT components.

To see whether we can generalize the result above and justify the two-step procedure for the combined method, a more sophisticated simulation was designed to compare the three methods and see how well they retrieve the common factors.

The simulation is based on (4.14) and constructed as follows. The observation \mathbf{z}_t and white noise $\boldsymbol{\varepsilon}_t$ are both 9×1 vectors. The variance-covariance matrix of $\boldsymbol{\varepsilon}_t$ is

$\Sigma_{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_9)$. The underlying components \mathbf{y}_t and time-structured noises \mathbf{n}_t are both 3×1 vectors and their variance are $\Sigma_y \sim N(\mathbf{0}, \mathbf{I}_3)$ and $\Sigma_n \sim N(\mathbf{0}, \mathbf{I}_3)$ respectively. The random variability contributed by \mathbf{n}_t and ε_t are at similarly low level, and much smaller than the variability contributed by \mathbf{y}_t . The loading matrix \mathbf{P} and \mathbf{Q} are also randomly generated and then scaled to ensure $\max_{j=1 \sim r} (p_{ij}) \geq 3q_{ij}, \forall i$. \mathbf{y}_t and \mathbf{n}_t are both simulated from a VAR(1) model with first-order autoregressive coefficient matrix Φ_y and Φ_n respectively:

$$\Phi_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} \quad \Phi_n = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix}$$

For each run, 100 observations of \mathbf{z}_t were generated. Three methods were applied and for each of them, retrieved components were matched to the true factors and correlation between each matched pair was calculated. Simulation was run 1000 times and the averaged correlations were computed and shown in table 1.

Table 4-2. Correlation between each factor and its matching component.

Correlation Between	PCA	Box-Tiao	Combined Method
1 st factor and its matching component	0.7369354	0.2559151	0.960456
2 nd factor and its matching component	0.7366181	0.7395445	0.9461788
3 rd factor and its matching component	0.764179	0.966231	0.9831987

From Table 4-2 we see the combined method retrieves three components which matches the true underlying components very well, with correlations close to one. The Box-Tiao method is capable of matching the third factor y_{3t} because it has a obvious time structure (ϕ_1 as high as 0.9). However, it does not retrieve the first factor y_{1t} at all.

This is because the first factor does not have a time structure and the Box-Tiao method probably confounds the time-structured noise with the not-time-structured underlying component. PCA works fine generally for all these factors but none of them has been matched exactly. The correlations are all below 0.8.

Based on the simulations the combined method appears to be promising and it provides intuition about how it works.

4.4.5 Furnace Data Revisited

Apply the two step scheme on the Furnance data: keep the first four PCs and then conduct the Box-Tiao transformation. Figure 4-8 shows time series plots and autocorrelation function plots of the four components.

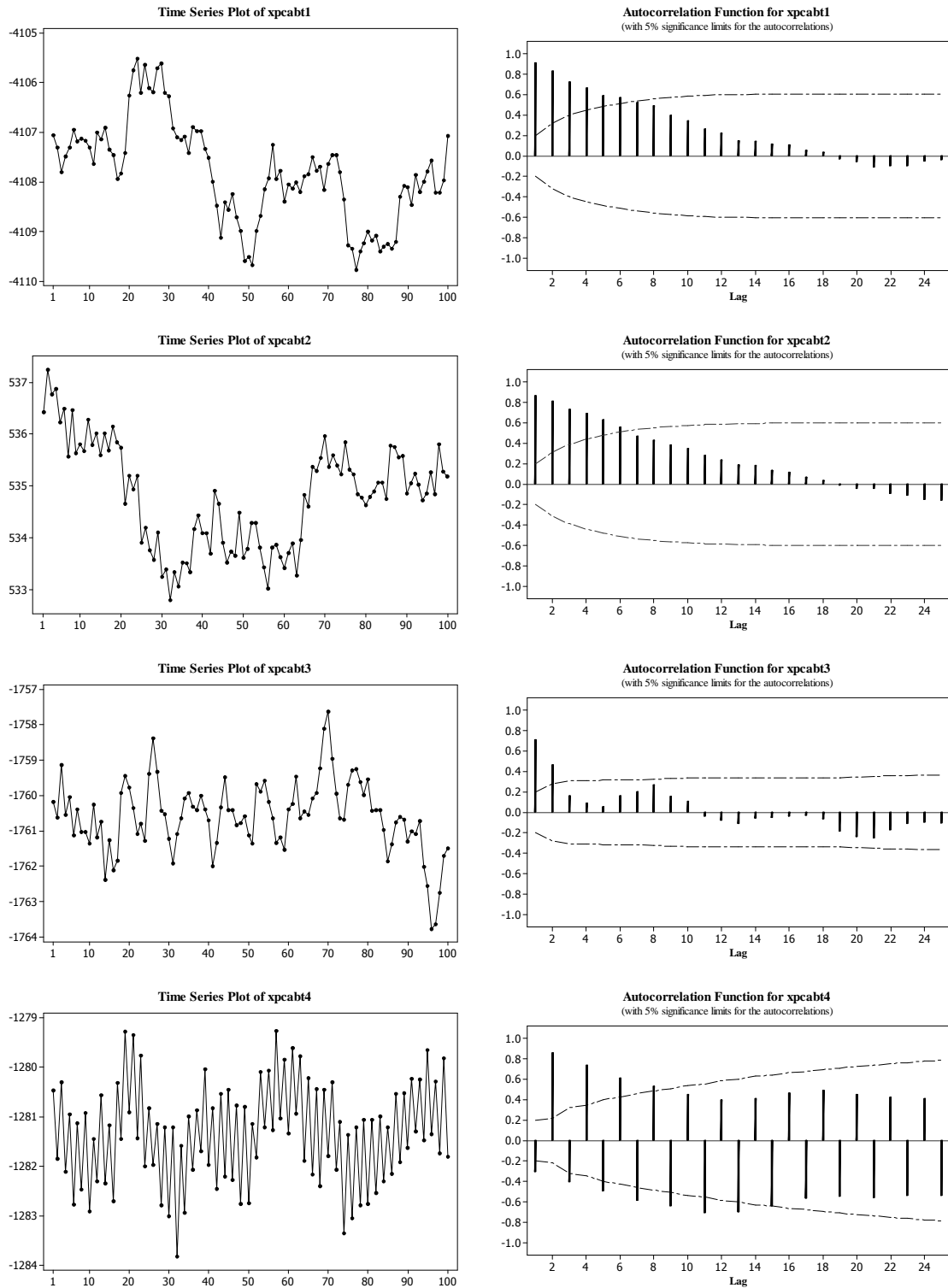


Figure 4-8. The Four PCA-BT Factors on Furnace data: Time Series plots and Autocorrelation Function.

The last component has an interesting autocorrelation pattern. It seems that autocorrelations at odd lags and those at even lags have totally separate tracks. It is as if the last component is a channel for two different variables (possibly correlated) X and Y, which are alternating to show through this channel. The two variables are autocorrelated and cross-correlated. It coincides with the fact that there are two cooling devices switching back and forth every 20 minutes to cool the furnace. Since this is hourly data, two consecutive data records happen to be influenced by different cooling devices.

Further more, all four Box-Tiao components are confirmed to be important common factors by their autocorrelation with the observed variables, as shown in Figure 4-9.

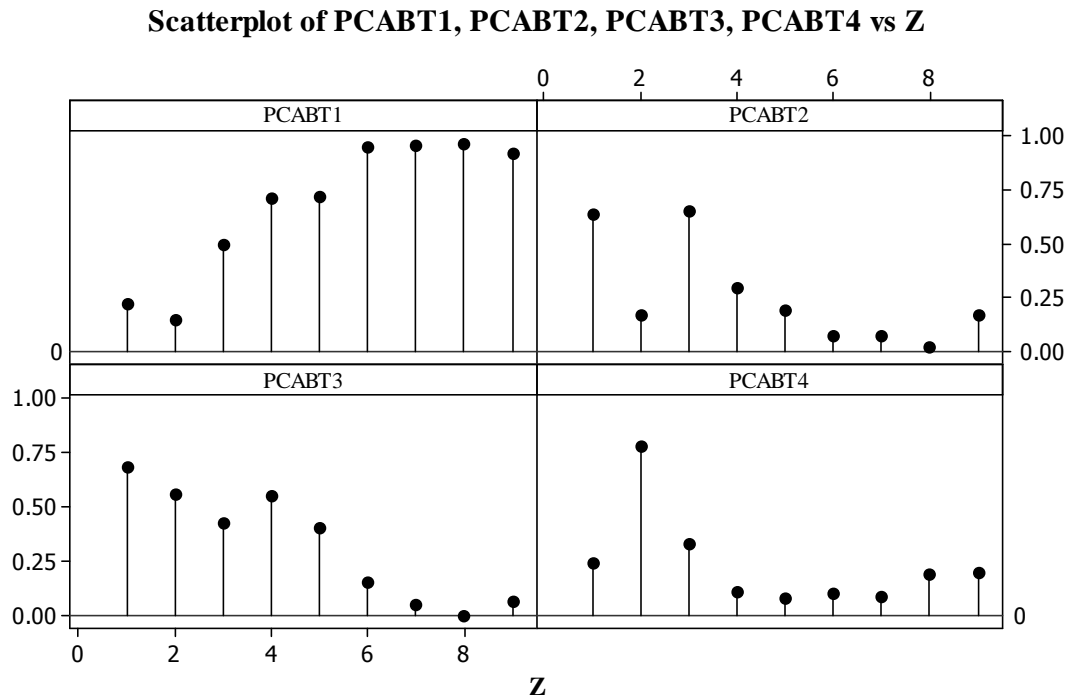


Figure 4-9. Correlation between the four BT-PCA factors and each dimension of the observed data.

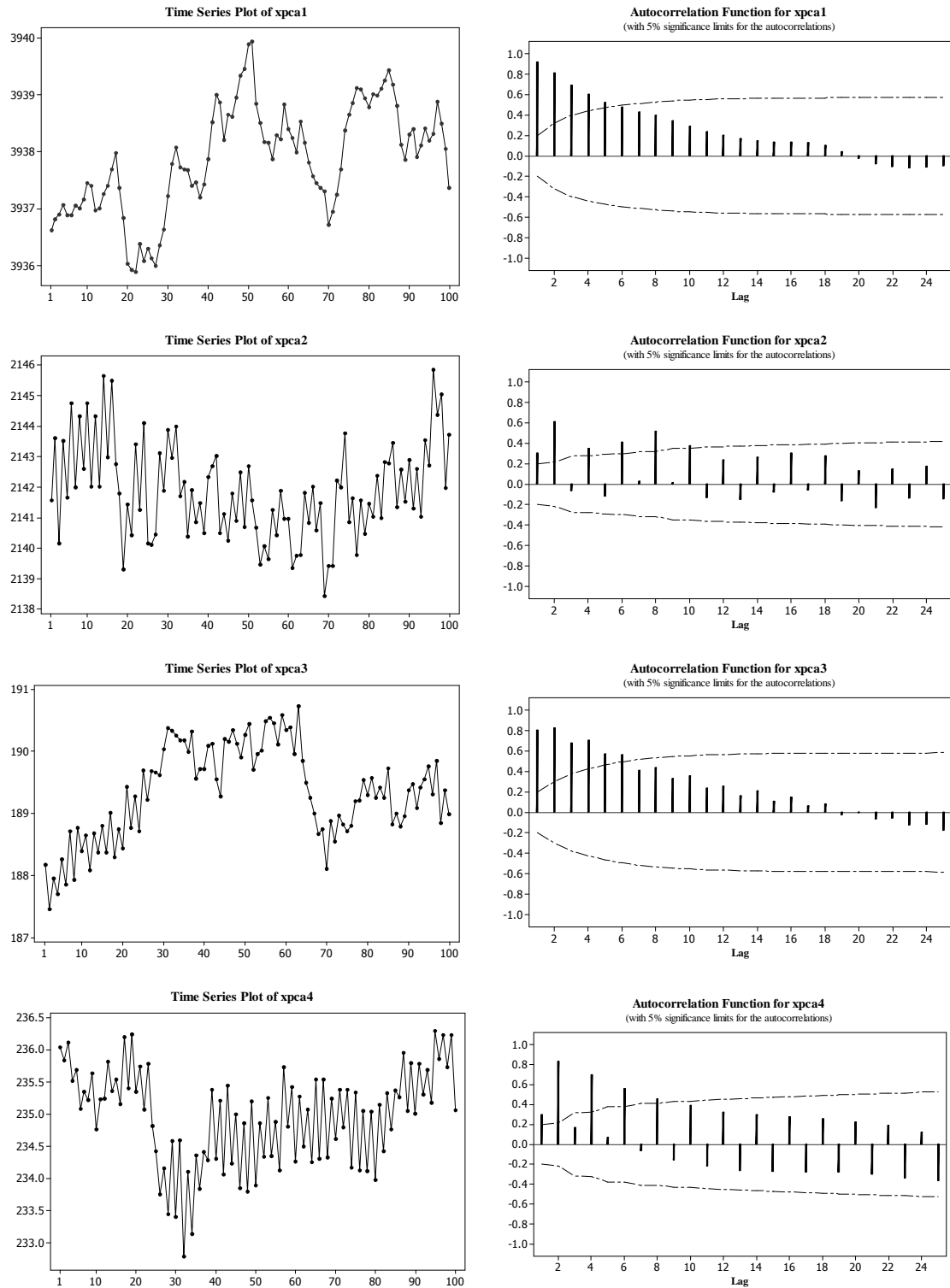


Figure 4-10. The Four PCAs on Furnace data: Time Series plots and Autocorrelation Fuction.

The two-step scheme produces a satisfactory analysis. If, however, we stop at the PCA step and keep the four PCs, what factors will we get? Figure 4-10 provides time series plots and autocorrelation function plots of the first four PCs for comparisons.

We find out the alternating autocorrelation effect appear in all last three PCs, which indicates that they are all contaminated by the PCABT4 – the switching cooling factor, to some extent. This further demonstrates how PCA treats data against the variability criteria and may be incompetent to differentiate factors according to their dynamics.

4.5 Business Applications

Dimension reduction has wide applications to many business fields. In this section, I focus discussion on Quality Management and Finance.

4.5.1 Statistical Process Control

Dimension reduction methods, especially the PCA method has been suggested for SPC; see Jackson (1991), Runger and Alt (1996). For a long time it was reasoned that attention should be focused on principal components (PCs) with highest eigenvalues (major PCs). However, it turns out that the PCs with the lowest eigenvalues (PCA residuals) are also important. Runger and Alt (1996) suggested applying multivariate control charts on the PCA residuals. His argument is that the major PCs reflect system dynamics and common causes while the PCA residuals tend to be white noise and channels whereby assignable causes may enter the system. Thus focusing on PCA residuals better positions us in detecting assignable causes. This idea appears to be analogous to the residual based Hotellings T^2 chart I proposed in Chapter 2. The

difference is that time series model residuals are replaced by PCA residuals. Runger's methodology is innovative yet vulnerable in terms of the assumption that the PCA residuals will be uncorrelated over time. This is not necessarily true. One counter example was provided by Badcock (2004). A promising alternative is to use Box-Tiao components with close to zero eigenvalues instead of PC components. I will refer to this approach as Box-Tiao residuals. These components are close to white noise and most possibly associated with assignable causes. Monitoring mechanism built on these components will be succinct summary of the data yet provide an efficient tool to look into assignable causes.

4.5.2 Applications in Asset Pricing

In this section I will only outline that dimension reduction also plays an important role in Finance. Thus the method discussed herein in my thesis will apply quite widely. However, it is not my goal to provide a lengthy discussion of problems in Finance.

An important area in Finance is Asset Pricing. In Asset Pricing, one key issue is how to determine the market price for risk and appropriately measure risk for a single asset assuming market equilibrium. One economic model used to solve this problem was developed almost simultaneously by Sharpe (1963, 1964) and Treynor (1961), while Mossin (1966), Lintner (1965, 1969) and Black (1972) developed it further. This model is normally referred to as the Capital Asset Pricing Model (CAPM). In recognizing some negative empirical results on testing CAPM and the strong assumptions CAPM requires, a multifactor model was developed by Merton (1973), stating that any factors that influence the growth of consumption should also price individual assets. Later the arbitrage pricing theory (APT), formulated by Ross (1976), assuming that random returns

are linear functions of a number of economy-wide exogenous random factors and an idiosyncratic random variable, offers a testable multifactor alternative to CAPM.

Although APT is more flexible than CAPM because CAPM is a special case of APT, there is no guidance with respect to the number and the nature of the underlying pricing factors. This not only creates a problem on testing APT, but also becomes a problem itself. To address this, research was motivated to extract the true factors, testing APT or a combination of both.

Typically, researchers relied on two approaches to identify the underlying factors. One approach is to use a set of pre-specified macroeconomic variables/factors. Chen, Roll and Ross (1980) showed that exposures to widely discussed macroeconomic variables such as innovations in inflation and the term structure generate risk premium. A more successful model was proposed by Fama and French (1992, 1993) and they, based on previous literature, proposed using size and book-to-market equity as the other underlying factors besides the overall market factor.

The other approach is to use statistical factor methods. Chamberlain and Rothschild (1976), Lehmann and Modest (1988) used the PCA method to extract underlying components. Connor and Korajczyk (1987) extended Chamberlain and Rothschild's work and proposed an asymptotic principal components technique. Roll and Ross (1980) used maximum likelihood factor analysis and tested APT based on their estimated factor loadings.

It would thus be interesting to replace the PCA method in the literature with the Box-Tiao method, or even the combined method. The PCA method, though widely used, does not necessarily provide satisfying results. It is not guaranteed that the Box-Tiao

method would work better, but it provides an interesting alternative to explore. To compare, I apply both the PCA and the Box-Tiao methods to the return data on portfolios formed on each size decile (stocks are ranked according to size – market capitalization, and then divided into ten groups; portfolios are formed for each group). And then look at the correlation between these empirical factors and theoretical proposed pricing factors – the Market factor ($R_m - R_f$), the size factor (SMB) and the book-to-equity factor (HML). Figure 4-11 provides a graphical illustration of such correlations. PCA1~PCA3 represent the first three PCs. PCA12BT1 represents the Box-Tiao transformed first component on first two PCs. PCA12BT2 is named in the same fashion. It is obvious that only PCA1 has a reasonably high correlation with one of the theoretical factors – the Market factor while the other two PCs do not. The two-step resulted factors are able to match both the Market factor and the size factor; and the correlation is high (0.971 and 0.747 respectively). It is obvious that Box-Tiao transformation refines the factors and improves the retrieving results.

Correlation between Rm-Rf, SMB, HML and pca1, pca2, pca3, pca12bt1, pca12bt2

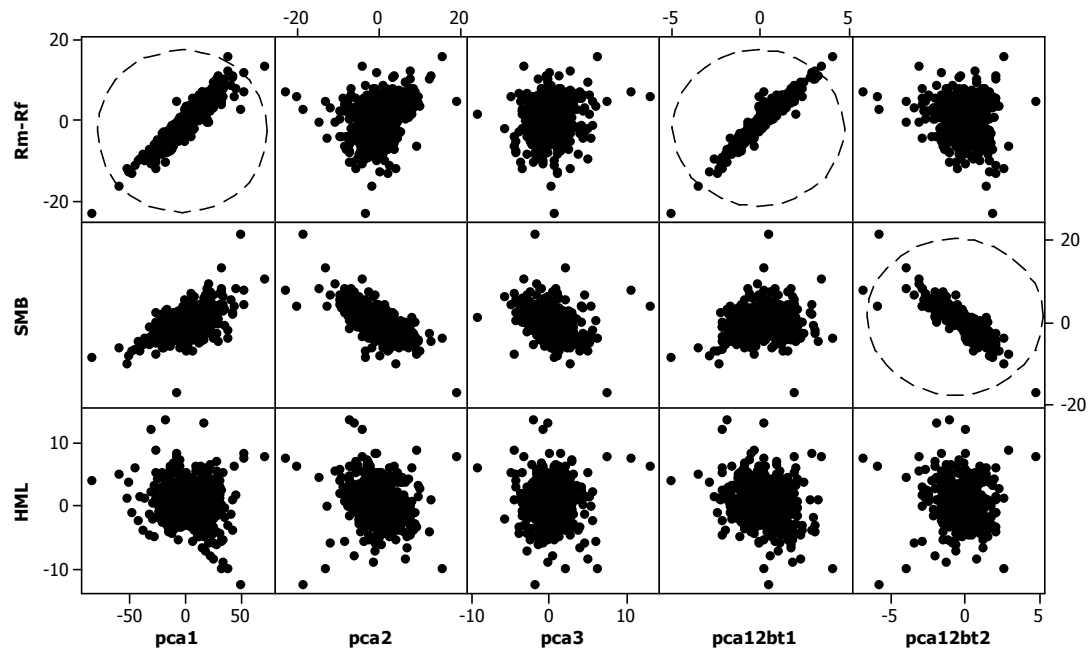


Figure 4-11. How well do results from statistical factor retrieving methods match with theoretical proposed pricing factors.

CHAPTER 5

VISUALIZING PRINCIPAL COMPONENTS ANALYSIS FOR MULTIVARIATE PROCESS DATA

In this chapter, we suggest a simple method for visualizing the results of Principal Components Analysis (PCA) intended to complement existing graphical methods for multivariate time series data applicable for process analysis and control. The idea is to visualize multivariate data as a surface that in turn can be decomposed with PCA. The surface plots developed in this chapter are intended for statistical process analysis but may also help visualize economic data and in particular co-integration.

5.1 Motivation

Data obtained from industrial processes are increasingly multivariate in nature. Data visualization is critical to the understanding of complex data structures from such processes and an important aid in their statistical analysis. The literature on data visualization is highly developed for data structures with a single response. Cleveland (1985, 1993) provide comprehensive overviews of methods and approaches while Tufte (1983) and Wainer (2004) provide less technical, but lucid surveys of statistical graphics and its history. Of course, anyone concerned about data analysis can benefit from reading Tukey's (1962) philosophical exposition of fundamental principles.

Graphical methods for multivariate data structures are less developed. The reasons are obvious; multivariate data structures are more complex and constitute a richer class of problems while we remain essentially confined to two-dimensional representations. A reasonably comprehensive overview of the current state of the art can be gained by consulting Gabriel (1985), Gnanadesikan (1997), Jolliffe (2002), Rencher (2002), Johnson and Wichern (2002), Jackson (2003) and the references herein.

When it comes to visualization of multiple time series structures, the literature is sparse. The standard references, Reinsel (1997), Lütkepohl (1993), Brockwell and Davis (1991) and Peña, Tiao and Tsay (2001), hardly consider the topic except in terms of standard univariate time series plots in separate panels or in overlaid format. An exception is Cressie (1993) whose emphasis is on spatial structures rather than on time series.

The purpose of this research is to describe a graphical data analytic method that may help elucidate certain structural properties of multiple time series data, both temporary and contemporary. As indicated, the traditional approach is to either overlay the time series on a single plotting frame or plot the series on separate panels. Those approaches provide insight into the temporary nature of time series, and in some cases, cross correlations and other contemporary relationships. We do not suggest replacing these useful approaches. Rather, the objective is to show that complementary insight sometimes can be gained by also viewing multiple time series as a surface. Further, we show how the surface can be decomposed in meaningful ways using PCA. One advantage of the described method is that it can handle high-dimensional data vectors.

The application of our method is primarily geared toward statistical process analysis and control where multiple time series are common. However, we have also found that the method may be useful for multivariate time series data structures from other domains.

The remainder of this chapter is organized as follows. In Section 5.2, we explain the idea of using surface plots. In Section 5.3, we introduce a decomposition of the original data set using principal components (PCs). In Section 5.4, we provide two

examples using the PCA techniques, one from *statistical process control* and one from economics, both cases with obvious time series structure. We provide additional discussion on approximating the surfaces and visualizing cointegration in Section 5.5 and Section 5.6 respectively. Finally in Section 5.7, we provide a conclusion.

5.2 Graphical Representation of Multiple Time Series

Driven by the desire to visualize events and patterns over time, multiple time series have, since the 10th or 11th century and certainly since the time of Playfair (1786), typically been represented graphically as overlaid time series plots or as separate plots carefully aligned to allow comparison; see Tufte (1983, p. 28). Most often, the plots have a common abscissa (time) scale, but the ordinate may either have the same or different scales for each time series. Figure 4-1 shows such a traditional plot of a multiple time series from engineering process quality control. As was introduced in Chapter 4, this data set contains simultaneous observations of the temperature at nine different locations of a large industrial furnace. The temperatures are measured with thermocouples located at equal distance along the length of the furnace. For technical reasons, the temperatures are supposed to exhibit an arch-shaped profile, with the highest temperature (hot spot) at location 4 (z4). It is important to the operation of the furnace to maintain this temperature profile over time. This type of problem is common in industrial process monitoring and control; see Kourti and MacGregor (1996).

Although Figure 4-1 conveys information about a common trend or comovement along the time axis of the nine time series, it does not clearly provide an insight to the contemporary relationship (the profile) among variables. In Figure 5-1, we have represented the same data as a three-dimensional surface plot. The three axes are (i) a

dimension index, in this case, counting equidistant physical locations from one to nine; *(ii)* time; and *(iii)* the data scale, in this case, temperature. From Figure 5-1, we see the patterns of the nine time series evolve in time, as well as the interrelationship among the variables. Figure 5-2 shows a contour plot of the same data. None of these representations is “better” than any other. Rather, the three representations complement each other and provide additional insight to the structure of the time series. Indeed, depending on the orientation of the surface in Figure 5-1, some details may be hidden. However, an additional advantage of plotting multivariate data as a surface is realized with modern

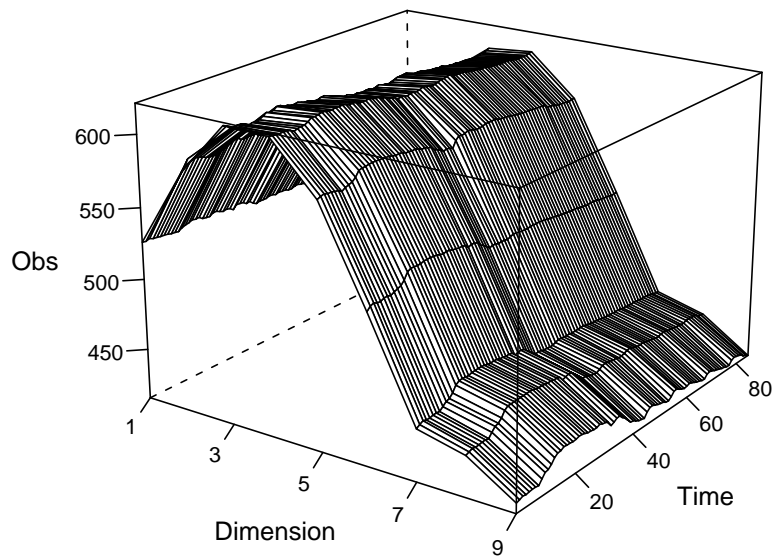


Figure 5-1. Surface plot of the industrial furnace temperature profile.

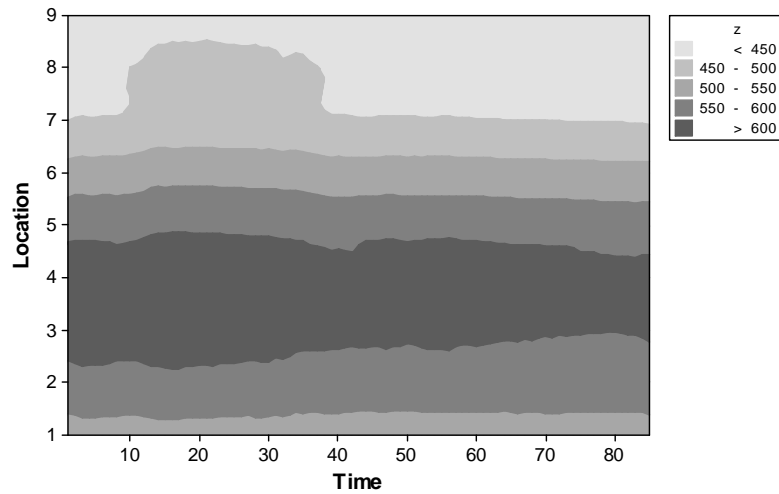


Figure 5-2. Contour plot of the furnace temperature profile.

5.3 Principal Components Decomposition

The idea behind PCA is to transform multivariate data by introducing a new set of rotated orthogonal linear coordinates so that the sample variances of the transformed data are in decreasing order of magnitude. A comprehensive introduction to PCA is provided by Johnson and Wichern (2002). The application of PCA to time series data is described by Tsay (2005), Tiao et al. (1993), and Peña et al. (2001). As pointed out by Anderson (1963), when PCA is applied to time series data, one should be cautious especially for small sample sizes because the sample variance may be seriously biased, or worse, if the time series is nonstationary, not be particularly meaningful. In this context, we consider PCA entirely as an exploratory, data-analytic, and descriptive tool.

Most applications of PCA involve an interpretation of the PCs as projections of original data onto a new set of rotated coordinate axes. In this research, we project the PCs back to the original coordinate axes. This allows us to decompose the original data according to the PCs. With the aid of the surface plot, we then visualize what the PCs

represent in the original scales. In the following discussion, we have adopted the notations used by Johnson and Wichern (2002); we refer to this text for definitions and proofs.

Suppose we have p time series each of n observations. Let x_{ik} denote the i th observation of the k th time series. The $n \times p$ data set can then be summarized as the matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & \mathbf{K} & x_{1p} \\ \mathbf{M} & & \mathbf{M} \\ x_{i1} & \mathbf{O} & x_{ip} \\ \mathbf{M} & & \mathbf{M} \\ x_{n1} & \mathbf{A} & x_{np} \end{pmatrix}.$$

Further let \mathbf{S} be the estimated dispersion matrix of \mathbf{X} . We assume \mathbf{S} is non-singular. The i th principal axis is defined as the i th orthonormal eigenvector $\hat{\mathbf{e}}_i$ of \mathbf{S} (we use the “hat” circumflex on \mathbf{e}_i because $\hat{\mathbf{e}}_i$ is the sample eigenvector of the sample dispersion matrix \mathbf{S} , whereas \mathbf{e}_i is the eigenvector of the “true” dispersion matrix $\mathbf{\Sigma}$; the same convention is used for other quantities below) and $\hat{\mathbf{P}} = (\hat{\mathbf{e}}_1, \mathbf{K}, \hat{\mathbf{e}}_p)$. Further

$\mathbf{S} = \hat{\mathbf{P}} \hat{\mathbf{\Lambda}} \hat{\mathbf{P}}' = \sum_{i=1}^p \hat{\lambda}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$ where $\hat{\lambda}_i$ are the eigenvalues of \mathbf{S} . The i th principal component is

the projection of \mathbf{X} on the i th principal axis, $\hat{\mathbf{y}}_i = \mathbf{X} \hat{\mathbf{e}}_i$. Now let $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}_1, \mathbf{K}, \hat{\mathbf{y}}_p)$ then

$\hat{\mathbf{Y}} = \mathbf{X} \hat{\mathbf{P}}$. Since $\hat{\mathbf{P}}' = \hat{\mathbf{P}}^{-1}$, the inverse transformation is given by

$$\mathbf{X} = \hat{\mathbf{Y}} \hat{\mathbf{P}}^{-1} = \hat{\mathbf{Y}} \hat{\mathbf{P}}' = \begin{bmatrix} \hat{\mathbf{y}}_1, & \mathbf{L}, & \hat{\mathbf{y}}_p \end{bmatrix} \begin{bmatrix} \hat{\mathbf{e}}_1' \\ \mathbf{M} \\ \hat{\mathbf{e}}_p' \end{bmatrix} = \hat{\mathbf{y}}_1 \hat{\mathbf{e}}_1' + \mathbf{L} + \hat{\mathbf{y}}_p \hat{\mathbf{e}}_p' = \mathbf{C}_1 + \mathbf{L} + \mathbf{C}_p,$$

where each of the $n \times p$ matrices $\mathbf{C}_i = \hat{\mathbf{y}}_i \hat{\mathbf{e}}_i'$, $i = 1, K, p$ are of rank 1. Those p rank-one matrices can be displayed as surface plots either individually or combined in any way we please. Typically we may be interested in plotting $\mathbf{X}_k = \sum_{i=1}^k \mathbf{C}_i$, $k = 1, K, p$ for an increasing sequence of k 's and hence essentially reconstruct or approximate the original data set with a lower rank approximation leaving out only residual variability. As we demonstrate in the following sections, this method of decomposition, reconstruction and approximation of the data matrix \mathbf{X} facilitates meaningful ways of looking at the original data that provide visual images of PCs. A similar decomposition can be achieved using singular value decomposition, see e.g. Seber (2004).

Above we explained the decomposition in terms of the eigenvector, eigenvalue decomposition of the dispersion matrix. By replacing the dispersion matrix by the estimated correlation matrix \mathbf{R} , we can accomplish a similar decomposition. However, because PCA is not scale invariant, those will typically not be the same. In some cases, either one or the other decomposition will provide a more meaningful visualization.

We hasten to note that the above decomposition is based on the spectral decomposition of a symmetric matrix, a purely algebraic result. Nowhere does it involve distributional or temporal independence assumptions of the data. Thus, we consider it a data analytic tool useful for data exploration and visualization but not for inference.

5.4 Examples

In this section, we use two examples to illustrate the idea of using surface plots and PCA decomposition. The first example is the nine-dimensional furnace temperature data introduced above and the second is a five dimensional data set from economics.

5.4.1 Example 1: The Furnace Temperature Data

As an illustration, we first apply the decomposition method to the furnace temperature data. Table 5-1 provides a summary of a standard principal components analysis based on the covariance matrix. Figure 5-3 shows the eigenvector elements (loadings) versus the (vector) dimension indices.

Table 5-1. Principal Components Analysis Based on the Covariance Matrix of the Furnace-Temperature Data

Eigenvalue	251.22	18.24	4.19	2.13	0.83	0.40	0.10	0.04	0.01
Proportion	0.906	0.066	0.015	0.008	0.003	0.001	0.000	0.000	0.000
Cumulative	0.906	0.972	0.987	0.995	0.998	0.999	1.000	1.000	1.000
Eigenvectors	$\hat{\mathbf{e}}_1$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$	$\hat{\mathbf{e}}_6$	$\hat{\mathbf{e}}_7$	$\hat{\mathbf{e}}_8$	$\hat{\mathbf{e}}_9$
1	0.088	-0.275	-0.279	-0.424	-0.157	0.763	-0.219	-0.073	0.011
2	0.244	-0.405	-0.411	-0.521	0.165	-0.526	0.163	0.069	-0.019
3	0.309	-0.603	0.123	0.466	-0.087	0.177	0.516	-0.049	-0.007
4	0.294	-0.364	0.188	0.226	-0.069	-0.186	-0.777	0.222	-0.040
5	0.258	0.032	0.790	-0.518	-0.132	0.000	0.145	0.039	0.004
6	0.397	0.116	0.041	0.048	0.504	0.029	-0.144	-0.695	0.257
7	0.403	0.270	-0.066	0.050	0.413	0.199	0.067	0.324	-0.662
8	0.422	0.292	-0.157	0.062	-0.023	0.094	0.091	0.499	0.664
9	0.432	0.302	-0.221	0.051	-0.702	-0.163	0.006	-0.315	-0.229

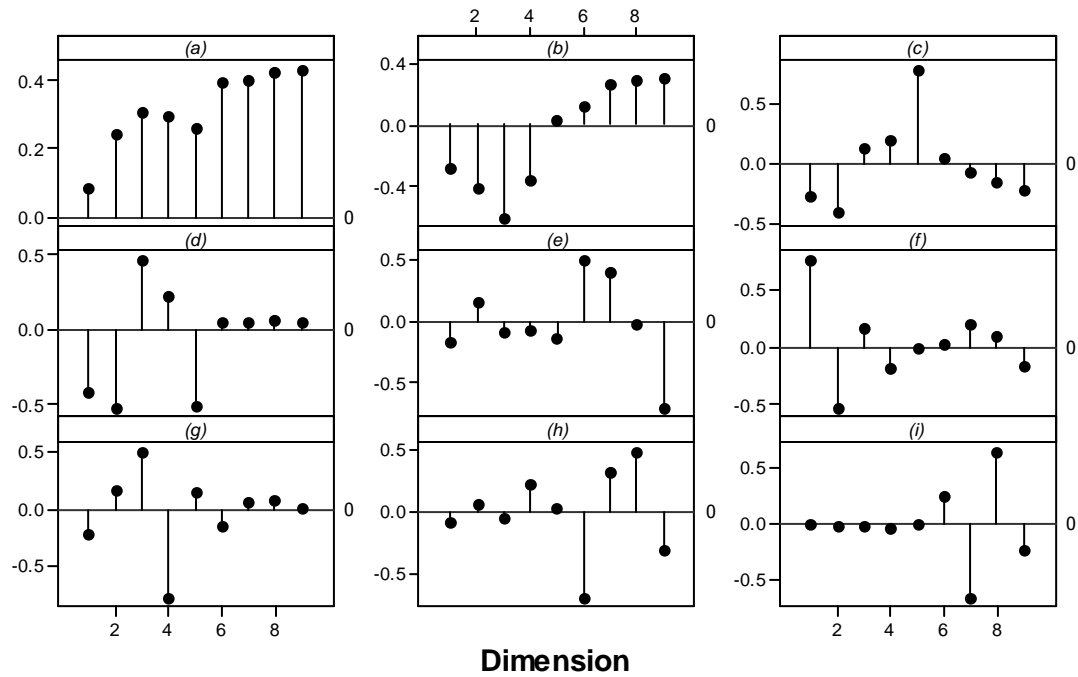
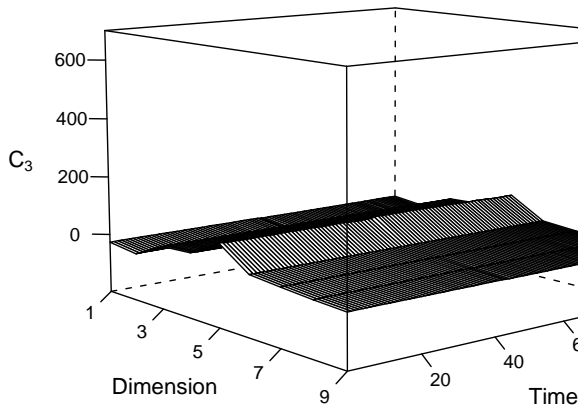
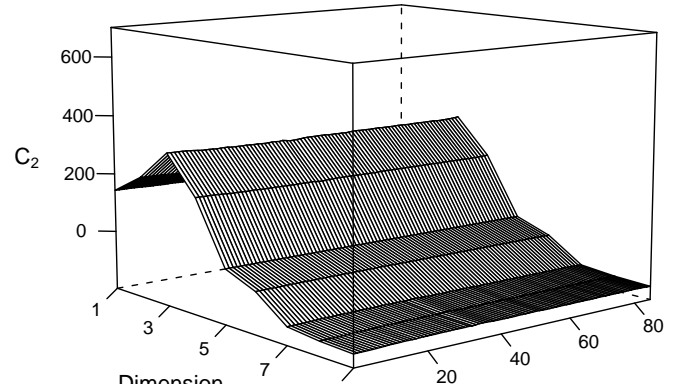
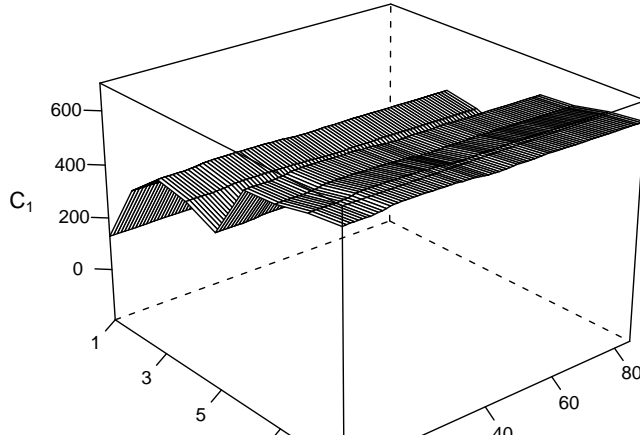


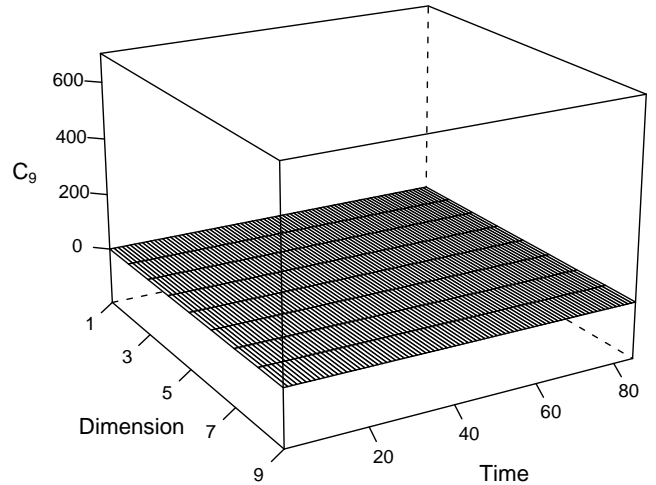
Figure 5-3. Plot of the Eigenvector Elements Versus Their Dimension Index Based on the Covariance Matrix of the Furnace Data with (a) for \hat{e}_1 , (b) for \hat{e}_2 and so forth.

From Table 5-1, we note that approximately 97% of the variability in the furnace-temperature data is accounted for by the first two principal components. Further, we see from Figure 5-3 that the first set of loadings (first eigenvector elements) indicate that the first PC is an average across all temperatures. From Figure 5-3, we also see that the second PC is a contrast between the front (locations 1, 2, 3, 4) and the back end (locations 5, 6, 7, 8, 9) of the furnace. Both of these interpretations make sense from an engineering point of view. Figure 5-4 shows principal component surfaces for the first, second, third and ninth components. Here we can take advantage of the surface plots to see how the data profile gets decomposed. From Figure 5-4(a) and Figure 5-4(b), we see C_1 and C_2 contribute substantially to the profile, while C_9 , shown in Figure 5-4(d), does not

contribute much. In Figure 5-5, we have experimented with yet another representation of the nine-dimensional time series, this time using a three-dimensional scatter plot.



(c)



(d)

Figure 5-4. Principal component surfaces for the Furnace Temperature Data: (a) surface of the first principal component C_1 , (b) surface of the second principal component C_2 , (c) surface of the third principal component C_3 , (d) surface of the ninth principal component C_9 .

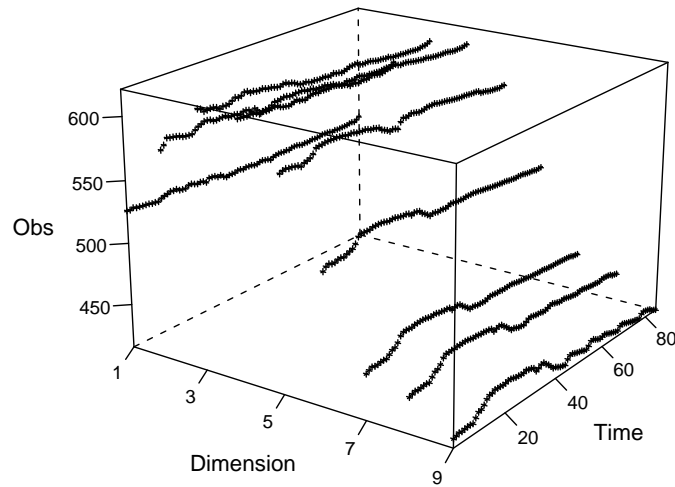


Figure 5-5. A three-dimensional scatter plot of the furnace data.

In Figure 5-2, we already represented the nine dimensional time series as a contour plot. The interesting aspect of this type of representation in the context of process analysis and control is that, if the temperature profile was constant over time, the contour lines would be parallel straight lines. In this case, we see that the process varied significantly over time, especially for locations 7, 8 and 9. This was also visible in Figure 4-1 and Figure 5-1, but was perhaps made a bit more precise in Figure 5-2.

5.4.2 Example 2: The Hog Data

In Section 1.1.2, we introduced an econometrics example of so-called “hog data.” Figure 1-3 shows the traditional graphical representation of the “hog data.” It is a five-dimensional multiple time series consisting of 82 yearly observations from 1867 to 1948 of quantities relevant to the US hog trade. The five dimensions are hog supply, hog price, corn price, corn supply, and farm wages, respectively. Please refer to Section 1.1.2 for transformations applied to the original data.

Figure 5-6 is the surface plot of the standardized hog data. Here we may question the ability of a two-dimensional plot to convey the information in a three-dimensional surface, but again, with modern dynamic-graphics software, we can always rotate the surface to see hidden details. Although the surface plot of the hog data does not have as nice a profile as the temperature data, we include this example because once we apply the decomposition on the data, the individual component surfaces provide us with intuition about the data structure and cointegration, a topic we pursue in the Section 5.6.

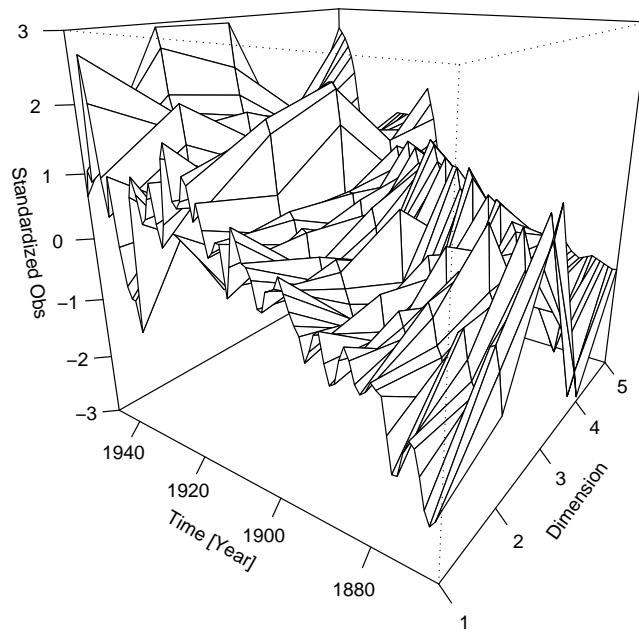


Figure 5-6. Surface plot of the Hog Data.

Because of the incommensurate scales of the five time series, we decided to base the PCA on the correlation matrix. However, it could as well be carried out using the dispersion matrix. Table 5-2 provides a summary of a standard PCA based on the correlation matrix. Figure 5-7 shows the eigenvector elements (loadings) versus the (vector) dimension indices.

Table 5-2. Principal Components Analysis based on the Correlation Matrix of the Hog Data

Eigenvalue	3.5870	1.0691	0.2534	0.0630	0.0275
Proportion	0.717	0.214	0.051	0.013	0.006
Cumulative	0.717	0.931	0.982	0.994	1.000
Eigenvectors	$\hat{\mathbf{e}}_1$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
1	0.446	-0.370	-0.730	-0.089	-0.351
2	0.491	0.227	0.513	-0.081	-0.661
3	0.386	0.631	-0.264	0.575	0.228
4	0.380	-0.635	0.358	0.494	0.283
5	0.516	0.100	0.076	-0.641	0.554

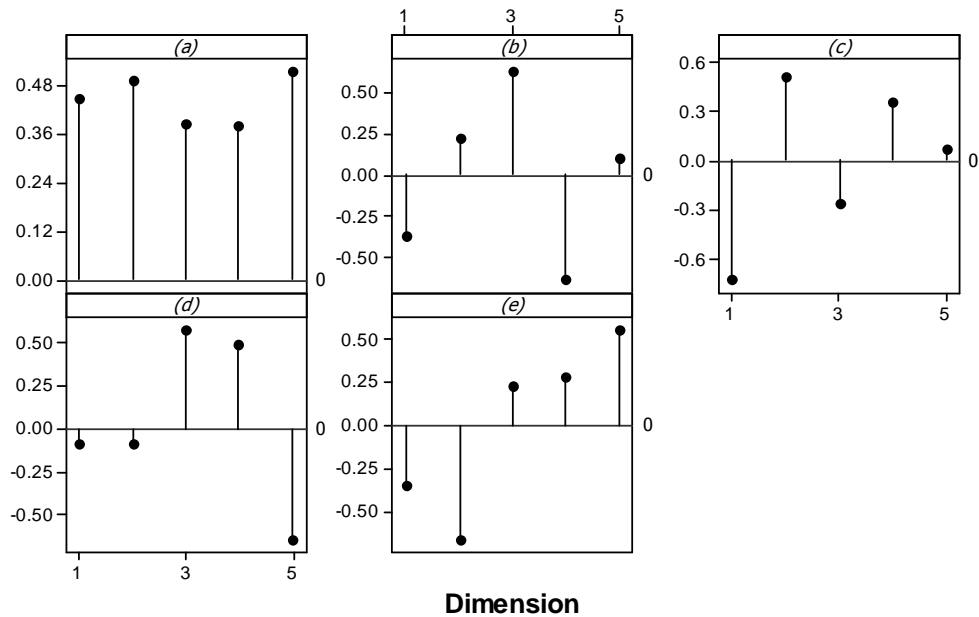


Figure 5-7. Plot of the eigenvector elements versus their dimension index based on the correlation matrix of the Hog data with (a) for $\hat{\mathbf{e}}_1$, (b) for $\hat{\mathbf{e}}_2$ and so forth.

Next, we decompose the data in Figure 5-6 into five separate PC surfaces as shown in Figure 5-8. We see from Table 5-2 that \mathbf{C}_1 , the first PC surface, representing 71.7% of the total variation and shown in Figure 5-8(a), can be interpreted as a general trend visible in all five time series. We also notice the large increase around the years of

World War I, followed by a significant dip across all time series corresponding to the effect of the 1930's economic depression. Further, we see a rapid surge in general economic activity at the advent of World War II.

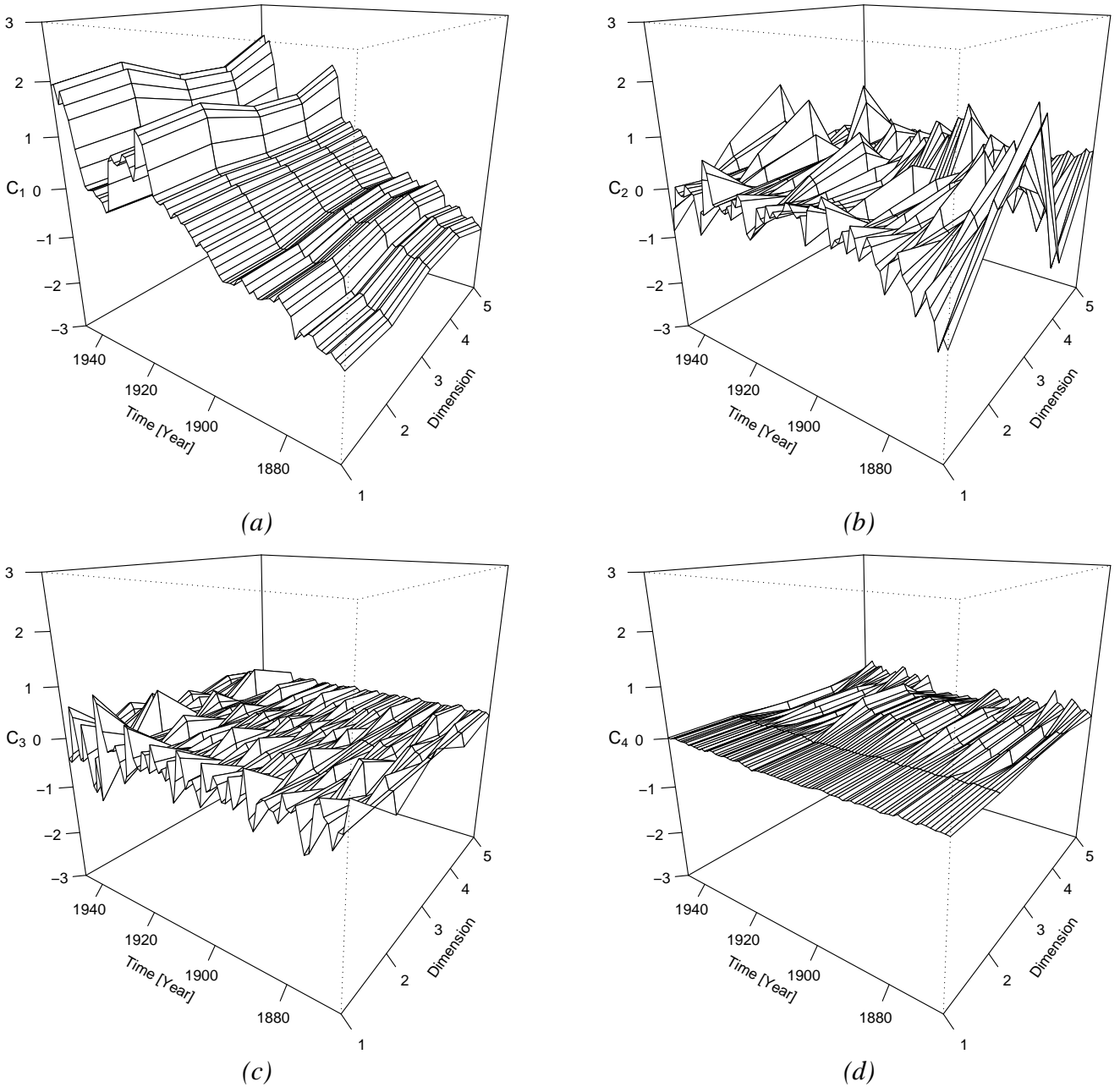


Figure 5-8. Principal component surfaces for the Hog Data: (a) surface of the first principal component C_1 , (b) surface of the second principal component C_2 , (c) surface of the third principal component C_3 , (d) surface of the fourth principal component C_4 .

From Figure 5-7, we see that all elements of the first eigenvector (loadings) are more or less the same and hence interpretable as an average across the time series. Thus, we conclude that the first PC is a contemporaneous average that, over time, exhibits a common trend. In other words, the first PC appears to be a general index of (agricultural) economic activity. We believe that the visual image of this underlying trend as a surface explains this well.

The second PC surface C_2 shown in Figure 5-8(b), accounting for 21.4% of the variability, seems to represent a noisy but stationary contrast between time series 2 and 3 versus 1 and 4, with less weight on time series 5. This pattern is further indicated by inspection of the loadings of the second eigenvector shown in Figure 5-7. The third PC surface, shown in Figure 5-8(c), represents 5.1% of the total variability. The primary feature of this surface is that it is almost stationary in time, but shows a contrast between time series 2 and 4 versus 1 and 3. The fourth PC surface C_4 shown in Figure 5-8(d), representing 1.3%, is essentially a non-informative random variation and may not be worth serious interpretation. So is C_5 , and, hence, it is not depicted here.

5.5 Further Discussion: Adding Up the Surfaces

We now proceed to plot the data by adding up each of the rank-one surfaces. In other words, we suggest constructing rank $j = 1, K, p-1$ approximations of the original data set. This is indicated in Figure 5-9 for the hog data showing \mathbf{X}_k , $k = 2, 3$. Note how the surfaces approach the original surface as k increases. The surface plots show how principal component surfaces add up to the original surface. By definition \mathbf{X}_1 is the same

as \mathbf{C}_1 in Figure 5-8(a). However, \mathbf{X}_2 , showed in Figure 5-9(a), is already a close approximation to the original surface given in Figure 5-6. Adding the 3rd PC surfaces, as showed in Figure 5-9(b), mostly adds details or noises and does not materially change the general structure of the data. Indeed it might be easier to see the general structure of the data from the surface plot of a lower rank approximation, in this case \mathbf{X}_2 , because most of the noise is eliminated.

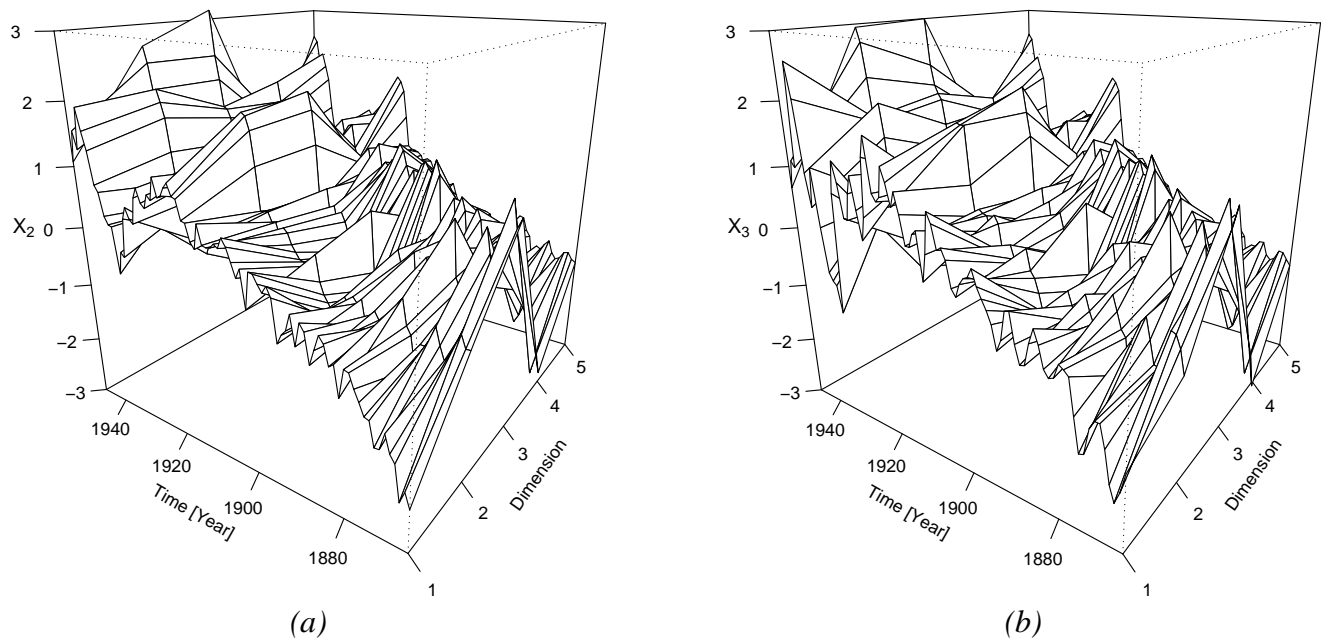


Figure 5-9. Adding up the rank one principal component surfaces for the Hog data: (a) surface of the added-up first two principal components \mathbf{X}_2 , (b) surface of the added-up first three principal components \mathbf{X}_3 .

We have also added up the rank 1 surfaces for the furnace temperature data. Those are shown in Figure 5-10, from \mathbf{X}_2 in (a) to \mathbf{X}_4 in (b), to \mathbf{X}_6 in (c), then to \mathbf{X} in (d). It is interesting to note that the first two PC's explains 97.2% of the variability but the surface plot reconstructed from only the first two PC's (Figure 5-10a) is quite different from the raw data surface (Figure 5-10d) in profile. The reason is the PC decomposition in this

case was based on the covariance; the first PC surface does account for the largest variance component, but its location has been shifted and the profile reveals the shift.

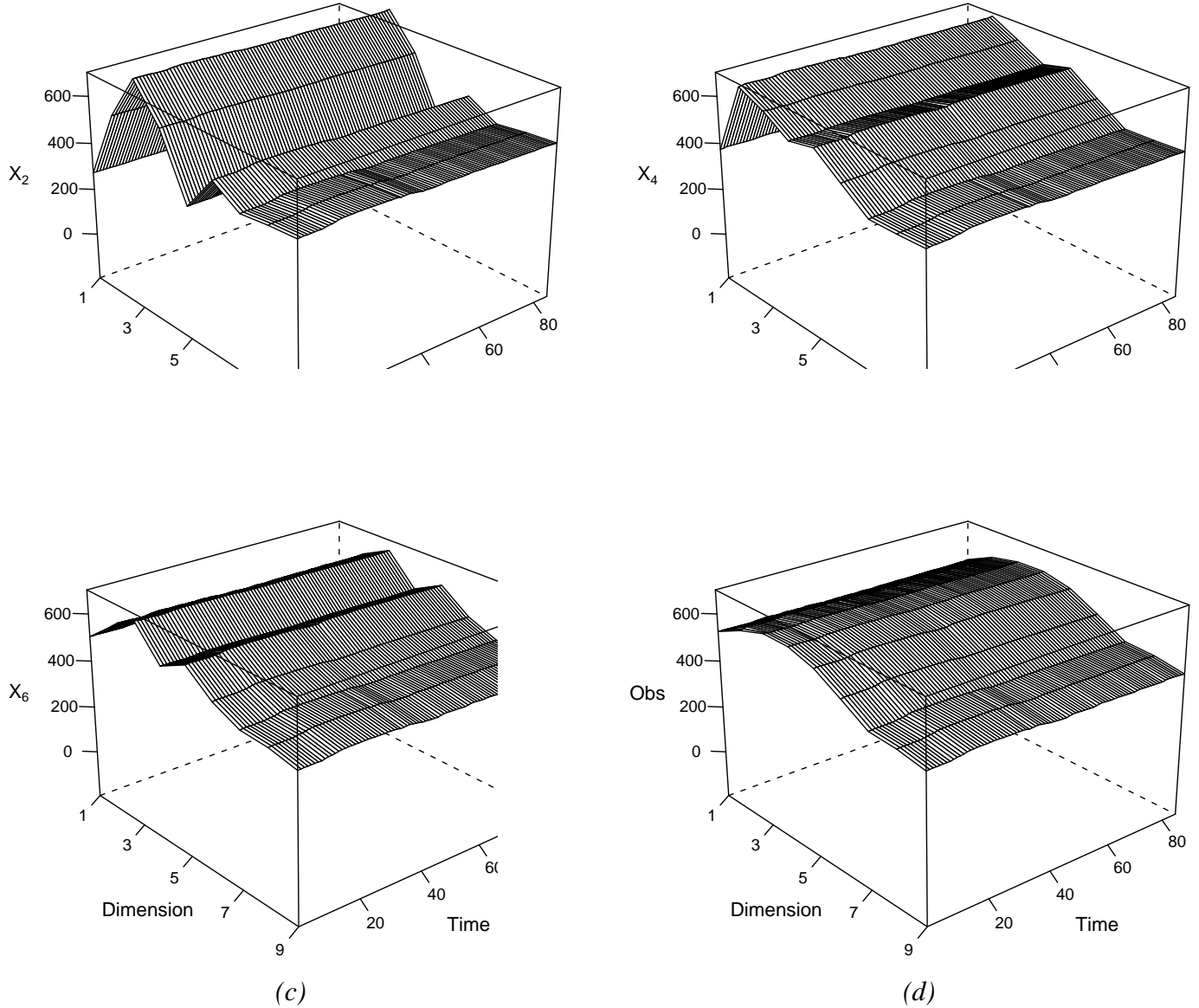


Figure 5-10. Adding up the rank one principal component surfaces for the Furnace data: (a) surface of the added-up first two principal components X_2 , (b) surface of the added-up first four principal components X_4 , (c) surface of the added-up first six principal components X_6 , (d) surface of the added-up all principal components X .

The above suggests two things: first, using PCA based on the covariance matrix may show that the PCs explain most of the variance, but that does not imply that the data reconstructed by the first PC will approximate the raw data. The first PC approximation may still differ significantly in certain locations. Thus, instead of only looking at how much variance the PCs explain, we may also want to consider how well the PCs approximate the raw data in specific locations. Second, PCA reconstruction based on the covariance is complex. One way to avoid this is to use the correlation matrix, in which case the problem of location no longer is an issue. However, when working with the correlation matrix, we essentially standardize the data and lose the profile information. In the furnace example, where it was important to preserve the profile, we used the covariance matrix; while in the hog-data example, we used the correlation matrix, and the combined surface of the first two components approximated the raw-data surface quite well.

5.6 Visualization of Comovement

Cointegration refers to the phenomenon occurring in econometrics where it often is possible to find linear combinations of nonstationary time series that are themselves stationary. More generally, we can sometimes determine linear combinations of multiple time series, stationary or nonstationary, that move together. This is called comovement; see, e.g., Peña et al. (2001). Although it is not the main purpose here to discuss cointegration or comovement, it is worth noting that the surface plots offer means for visualizing comovement in a very intuitive fashion. For the hog data, we see that all the five time series, when considered individually, appear to be nonstationary. However, if we look at Figure 5-11, where the individual PCs are plotted as time series, we find the

first PC appears to be nonstationary. Indeed, it is integrated of order 1, $I(1)$. In other words, their first differences are stationary. If we now project the first PC back to the original coordinates, as illustrated in Figure 5-8(a), we get five ($p = 5$) $I(1)$ nonstationary time series. Thus, the five time series appear to be linearly dependent or driven by a single (one-dimensional) underlying trend. The general trend represented by the first PC is an overall measure of the hog-related farm economy that drives all five time series. Further, the last 4 PCs look stationary. This is also clearly visualized in Figure 5-8 and Figure 5-11. This same phenomenon may occur for industrial processes. In the furnace example, it is clear both from the physics of the process and from the plots that, when the temperature in the furnace goes up or down, all nine temperatures more or less move up or down concurrently. In other words, there might be individual differences from thermocouple to thermocouple, but in general, the temperatures exhibit comovement.

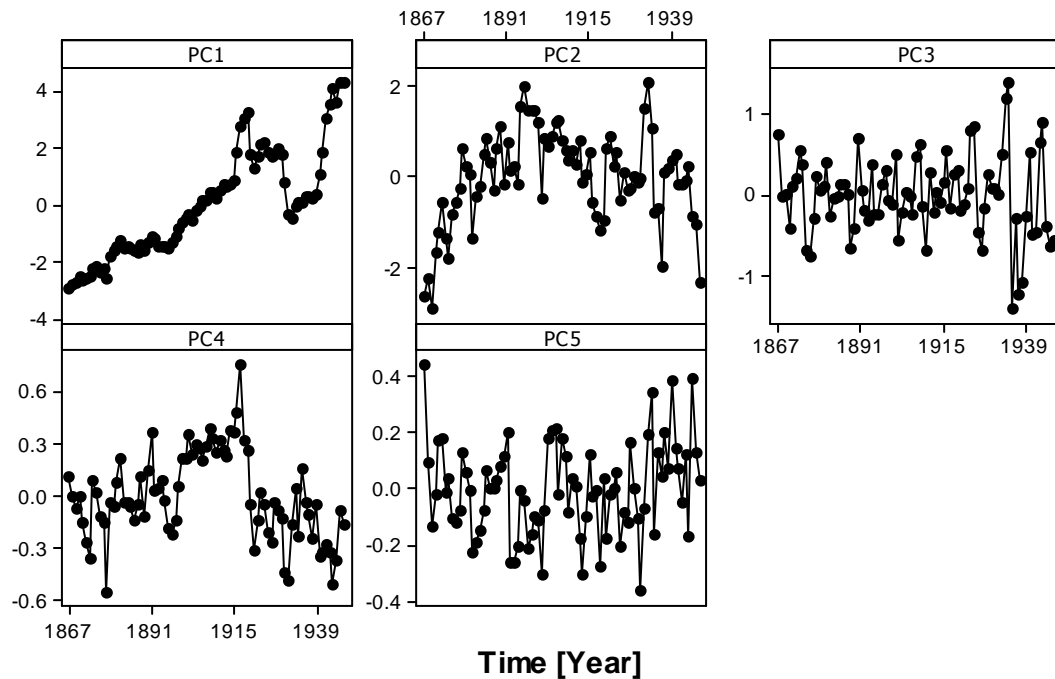


Figure 5-11. Time series plots of the five principal components, PC1, ..., PC5 for the Hog data.

5.7 Conclusion

Multivariate time series data occur in many areas of applications in engineering analysis and quality control. Methods for visualizing patterns in such data, especially when the dimensions are high, may help gain insight to the often complex structure of such data. Methods currently available for visualizing multiple process time series are relatively limited. The idea of using a surface plot makes it possible to visualize a considerable number of time series in one single plot.

The decomposition of the surfaces via spectral decomposition of the covariance or the correlation matrix appears to provide a useful way of dissecting a multivariate-process data set into meaningful components. Indeed, we think that this way of visualizing and interpreting PCs can be useful in many applications. Further, this

decomposition provides a simple way to explain the concept of co-integration and co-movement.

CHAPTER 6

CONCLUSION

Due to the ever-growing data made available by modern technologies, the effort to understand business processes is often challenged by co-existing autocorrelation and high-dimensionality. My dissertation aims to address these two issues at the same time.

Operations Management, especially Quality Management, is one example of fields in need of such a solution. Literature in designing multivariate control charts has generally assumed observations I.I.D. while that in dealing with autocorrelation it largely remains in one-dimension. In this dissertation, I fill this gap by offering a procedure to treat autocorrelation and high-dimensionality simultaneously.

The procedure starts with understanding the process, which includes plotting the data in different perspectives (Chapter 5), and retrieving meaningful common factors (Chapter 4). When the observed dimension has been reduced to a handful of representative common factors, it is suggested that a VARIMA time series model should be fitted. Based on the model, a special cause chart on residuals and a common cause chart on fitted values can then be run on Phase II data to effectively monitor the process (Chapter 2). It is acknowledged that traditional Shewhart charts are still widely used among practitioners despite the broadly existing autocorrelation. In such a situation, it is urged to use a numerical computation method (Chapter 3) to calculate accurate ARL. The calculation assuming an I.I.D. process can be misleading and result in unreasonable suggestions for control limits.

The proposed methods are illustrated with the glass furnace production case, which threads through the whole dissertation. However, the methods, especially the

dimension reduction procedure proposed in Chapter 4, has wider applications to many business and economics fields.

6.1 Future Work

The proposed common cause chart and special cause chart largely depend on the model fitted to the Phase I data. Undoubtedly, the estimation error from fitting the model will add noises to the Phase II control charts. I would like to know by how much the control charts are affected by estimation error. This research may answer the question of how long Phase I should be run to collect enough data for constructing robust control charts for Phase II.

The second branch of future work is to enhance the Box-Tiao method to deal with more versatile time series models. The Box-Tiao method is based on assuming VARIMA for the underlying process. How about ARCH, GARCH models? ARCH and GARCH models have wide applications to economics and financial data. I would like to know how the Box-Tiao method can be expanded to accommodate these models by applying the same criteria – ranking components based on predictability.

In Chapter 4, I have discussed several spurious Box-Tiao components. The two-step combined procedure can partly deal with some of them but not all, for example, the components caused by lagged linear redundancy. For such problems, a linear cross-sectional transformation falls short of a solution. It is thus of my interest to further look into such cases.

APPENDIX A

EIGENVALUES INVARIANT TO TRANSFORMATION

For a 2-dimension VAR(1) model $\tilde{\mathbf{z}}_t = \Phi \tilde{\mathbf{z}}_{t-1} + \mathbf{a}_t$, we can multiply each side by a non-singular matrix \mathbf{A} on the left,

$$\mathbf{A}\tilde{\mathbf{z}}_t = \mathbf{A}\Phi\tilde{\mathbf{z}}_{t-1} + \mathbf{A}\mathbf{a}_t, \quad (\text{A.1})$$

such that for the new innovation vector $\mathbf{a}_t^* = \mathbf{A}\mathbf{a}_t$, we have $\text{Cov}(\mathbf{a}_t^*) = \mathbf{I}$. If we transform the observation vector as $\mathbf{y}_t = \mathbf{A}\tilde{\mathbf{z}}_t$, then (A.1) can be written as,

$$\mathbf{y}_t = \mathbf{A}\tilde{\mathbf{z}}_t = \mathbf{A}\Phi\mathbf{A}^{-1}\mathbf{A}\tilde{\mathbf{z}}_{t-1} + \mathbf{A}\mathbf{a}_t = \Phi^* \mathbf{y}_{t-1} + \mathbf{a}_t^*, \quad (\text{A.2})$$

where $\Phi^* = \mathbf{A}\Phi\mathbf{A}^{-1}$.

If λ_i is an eigenvalue of Φ , i.e., $\Phi\mathbf{e}_i = \lambda_i\mathbf{e}_i$, and we let $\mathbf{f}_i = \mathbf{A}\mathbf{e}_i$ then,

$$\Phi^*\mathbf{f}_i = \mathbf{A}\Phi\mathbf{A}^{-1}\mathbf{f}_i = \mathbf{A}\Phi\mathbf{A}^{-1}\mathbf{A}\mathbf{e}_i = \mathbf{A}\Phi\mathbf{e}_i = \mathbf{A}(\lambda_i\mathbf{e}_i) = \lambda_i\mathbf{A}\mathbf{e}_i = \lambda_i\mathbf{f}_i$$

Thus λ_i is also an eigenvalue of Φ^* . Without loss of generality, we can therefore assume that $\Sigma_a = \mathbf{I}_{2 \times 2}$, because for any covariance matrix we can always by a linear transformation convert it to that form. Further, the eigenvalues, on which our discussion is mainly based, are the same for Φ and Φ^* .

APPENDIX B

ARL CALCULATION FOR RESIDUAL BASED METHOD

If we assume that the time series model is correct and all parameters have been estimated perfectly, then the residuals are independent. Under those assumptions the calculation of the ARL for the residual based method is much simpler. Let N denotes the number of steps after the step shift happens ($t = s$) but before an observation falls outside the control limits. Let T_t^2 denote the T^2 statistics at time t , and c denotes the UCL.

$$ARL = \sum_{n=0}^{\infty} n \cdot P(N = n) = \sum_{n=0}^{\infty} n \cdot \left(\prod_{t=1}^{n-1} P(T_t^2 \leq c) \right) P(T_n^2 > c)$$

If the process is under control and we have $\Pr(T_t^2 > c) = p, \forall t > s$, then

$ARL = 1/p$. However, for a step shift with a VAR(1) model that is no longer true.

Now consider the VAR(1) model $\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1(\mathbf{z}_{t-1} - \boldsymbol{\mu}) + \mathbf{a}_t$. Suppose we assume a step change at time T given by: $E(\mathbf{z}_t) = \begin{cases} \boldsymbol{\mu} & t < T \\ \boldsymbol{\mu} + \boldsymbol{\delta} & t \geq T \end{cases}$. Since we do not know when the process mean has shifted, we estimate the residuals by $\mathbf{e}_t = (\mathbf{z}_t - \boldsymbol{\mu}) - \boldsymbol{\Phi}(\mathbf{z}_{t-1} - \boldsymbol{\mu})$. Thus we have a residual mean shifted given by

$$\begin{aligned} E(\mathbf{e}_t) &= (E(\mathbf{z}_t) - \boldsymbol{\mu}) - \boldsymbol{\Phi}(E(\mathbf{z}_{t-1}) - \boldsymbol{\mu}) \\ &= \begin{cases} 0 & t < T \\ \boldsymbol{\delta} & t = T \\ \boldsymbol{\delta} - \boldsymbol{\Phi}\boldsymbol{\delta} = (\mathbf{I} - \boldsymbol{\Phi})\boldsymbol{\delta} & t \geq T + 1 \end{cases} \end{aligned}$$

Correspondingly, the non-centrality parameter induced by the shift is:

$$\lambda_{et} = t(E(\mathbf{e}_t))' \cdot \Sigma_a^{-1} \cdot E(\mathbf{e}_t) = \begin{cases} 0 & t < T \\ \boldsymbol{\delta}' \Sigma_a^{-1} \boldsymbol{\delta} & t = T \\ \boldsymbol{\delta}' (\mathbf{I} - \boldsymbol{\Phi})' \Sigma_a^{-1} (\mathbf{I} - \boldsymbol{\Phi}) \boldsymbol{\delta} & t \geq T + 1 \end{cases}$$

Because of the dynamics of the time series model, a step shift in the mean of the process causes two consecutive step shifts in the mean of the residuals. Therefore we experience two consecutive step shifts of the non-centrality parameters of the mean of the residuals. Typically, the first shift is upward and the second is downward, as illustrated in Figure B.1. These shifts thus affects $P(T_i^2 > c)$ accordingly. Now let $p_0 = P(T_T^2 > c)$ denote the probability before the shift and $p = P(T_t^2 > c)$ after the shift ($t > T$). We could get a general formula for VAR(1) ARL:

$$ARL = p_0 + \sum_{n=2}^{\infty} n(1-p_0)(1-p)^{n-2} p = \frac{1-p_0}{1-p} \left(\frac{1}{p} - p \right) + p_0$$

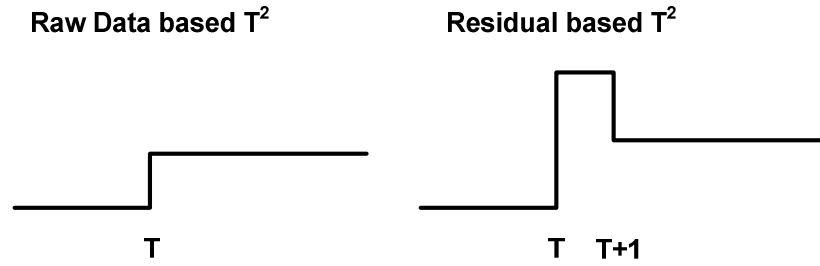


Figure B.1. Non-centrality parameters shift.

APPENDIX C

PROOF OF (3.1)

By the definition of τ_1 , i.e., starting from any in-control state, the number of steps it takes to arrive at the out-of-control state for the first time, we have,

$$\begin{aligned}\tau_1 &= \min \{n \geq 1, \text{ s.t. } X_n = s+1 \mid X_0 \neq s+1\} \\ &= \sum_{i=1}^s P(X_0 = i \mid X_0 \neq s+1) \min \{n \geq 1, \text{ s.t. } X_n = s+1 \mid X_0 = i\}.\end{aligned}\tag{C.1}$$

Now let $\tau_1(i) = \min \{n \geq 1, \text{ s.t. } X_n = s+1 \mid X_0 = i\}$, then (C.1) is

$\tau_1 = \sum_{i=1}^s \phi(i) \tau_1(i)$. Another relationship can be established by recognizing that starting at state i , the MC can arrive at state $s+1$ in the first step, or connect through state j at the first step. Assuming $q(i) = P(X_t = s+1 \mid X_{t-1} = i)$, we have

$$\tau_1(i) = q(i) + \sum_{j=1}^s r_{ij} (1 + \tau_1(j)).\tag{C.2}$$

Now define $\boldsymbol{\tau}' = [\tau_1(1), \dots, \tau_1(i), \dots, \tau_1(s)]$ and

$\mathbf{q}' = [q(1), \dots, q(i), \dots, q(s)]$. By extending (C.2) we have

$$\boldsymbol{\tau} = \mathbf{q} + \mathbf{R}(\boldsymbol{\tau} + \mathbf{1}).\tag{C.3}$$

Thus $\boldsymbol{\tau} = (\mathbf{I} - \mathbf{R})^{-1}(\mathbf{q} + \mathbf{R} \cdot \mathbf{1}) = (\mathbf{I} - \mathbf{R})^{-1} \mathbf{1}$. By substituting this back into (C.1) we

get $\tau_1 = \boldsymbol{\phi}'(\mathbf{I} - \mathbf{R})^{-1} \mathbf{1}$. This is the same as (3.1).

APPENDIX D

DETERMING THE END STATES IN FIGURE 3-3

The ARMA(1, 1) model $z_t = \phi z_{t-1} + a_t - \theta a_{t-1}$ can be written as:

$$a_t = z_t + (\theta a_{t-1} - \phi z_{t-1}).$$

This can be interpreted as a linear relationship between a_t and z_t where the slope is 1 and the intercept is $b_t = \theta a_{t-1} - \phi z_{t-1}$. If we let $\mathbf{w}_{t-1} = [z_{t-1} \quad a_{t-1}]'$ be the initial state S_0 shown as the bold rectangle in Figure 3-3, the intercept b_t will then range between $\min_b = \min_{\mathbf{w}_{t-1} \in S_0} \{\theta a_{t-1} - \phi z_{t-1}\}$ and $\max_b = \max_{\mathbf{w}_{t-1} \in S_0} \{\theta a_{t-1} - \phi z_{t-1}\}$. These two extremes can only occur at the two different vertices of the bold rectangle. Thus with the initial state fixed, the range of b_t is fixed. Therefore the final states can only be those which intersect with the lines between $a_t = z_t + \min_b$ and $a_t = z_t + \max_b$, as shown in Figure 3-3.

BIBLIOGRAPHY

- Alt, F. B. (1984), "Multivariate quality control," *Encyclopedia of Statistical Sciences*, eds. Kotz S., Johnson N. L. and Read C. R., New York: John Wiley, 110-122.
- Alwan, A. J., and Alwan, L. C. (1994), "Monitoring Autocorrelated Processes Using Multivariate Quality Control Charts," in *Proceedings of the Decision Sciences Institute Annual Meeting*, 3, 2106-2108.
- Alwan, L. C., and Roberts, H. V. (1988), "Time-Series Modeling for Statistical Process Control," *Journal of Business & Economic Statistics*, 6, 87-95.
- Anderson, T. W. (1963), "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series," *Psychometrika* 28, 1-25.
- Richard G. A., Qian, A., and Rasche, R. H., "Analysis of Panel Vector Error Correction Models Using Maximum Likelihood, the Bootstrap, and Canonical-Correlation Estimators," unpublished manuscript, August 2006.
- Apley, D. W., and Tsung, F. (2002), "The Autoregressive T^2 Chart for Monitoring Univariate Autocorrelated Processes," *Journal of Quality Technology*, 34, 80-96.
- Atienza, O. O., Tang, L. C., and Ang, B. W. (2002), "A Cusum Scheme for Autocorrelated Observations," *Journal of Quality Technology*, 34, 187-199.
- Badcock, J., Vailey, T. C., Jonathan, P., and Krzanowski, W. J. (2004), "Two Projection Methods for Use in the Analysis of Multivariate Process Data With an Illustration in Petrochemical Production," *Technometrics*. 46, 392-403.
- Bagshaw, M., and Johnson, R. A. (1975), "The Effect of Serial Correlation on the Performance of CUSUM Tests II," *Technometrics*, 17, 73-80.
- Berthouex, P. M., Hunter, W. G., and Pallesen, L. (1978), "Monitoring Sewage Treatment Plants: Some Quality Control Aspects," *Journal of Quality Control*, 10, 139-149.
- Bewley, R., Orden, D., Yang, M., and Fisher, L. A. (1994), "Comparison of Box-Tiao and Johansen canonical estimators of integrating vectors in VEC(1) models," *Journal of Econometrics*, 64, 3-27.
- Bewley, R., and Yang, M. (1995), "Tests for cointegration based on canonical correlation analysis," *Journal of the American Statistical Association*, 90, 990-996.
- Bisgaard, S., and Huang, X. (2008), "Visualizing Principal Components Analysis for Multivariate Process Data," *Journal of Quality Technology*, 40, 334-344.
- Black, F., Jensen, M. C., and Scholes, M., 1972, "The Capital Pricing Model: Some Empirical Testing," *Studies in the Theory of Capital Market*, Jensen, M. C., Ed., New York: Praeger Publisher.

- Brockwell, P. J., and Davis, R. A. (2002), *Introduction to Time Series and Forecasting* (2nd ed.), New York: Springer-Verlag.
- Bossaerts, P. (1988), "Common nonstationary components of asset prices," *Journal of Economic Dynamics and Control*, 12, 347-364.
- Brook, D., and Evan, D. A. (1972), "An Approach to the Probability Distribution of CUSUM RL," *Biometrika*, 59, 539-549.
- Box, G. E. P., and Jenkins, G. M. (1968), "Some recent advances in forecasting and control. Part 1," *Applied Statistics*, 17, 91-109.
- Box, G. E. P., Jenkins, G. M., and MacGregor, J. F. (1974), "Some Recent Advances in Forecasting and Control," *Applied Statistics*, 23, 158-179.
- Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco, CA: Holden-Day Inc.
- Box, G. E. P., and Tiao, G. C. (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika* 64, 355-365.
- Box, G. E. P. (1991), "Bounded Adjustment Charts," *Quality Engineering*, 4, 331-338.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control* (3rd ed.), New Jersey: Prentice-Hall, Inc.
- Box, G. E. P., and Luceño, A. (1997), *Statistical Control by Monitoring and Feedback Adjustment*, New York: John & Wiley & Sons, Inc.
- Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005), *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.), New York: Wiley-Interscience.
- Box, G. E. P., and Paniagua-Quiñones, C. (2007), "Two Charts: Not One," *Quality Engineering*, 19, 93-100.
- Brockwell, P. J., and Davis, R. A. (2002), *Introduction to Time Series and Forecasting*, New York: Springer.
- Campbell, J. Y., et al., 2000, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56, 1-43.
- Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281-1304.
- Chen, N., Roll, R. and Ross, S. (1986), "Economic Forces and Stock Market," *Journal of Business*, 59, 383-403.
- Cleveland, W. S. (1985), *The Elements of Graphing Data*, North Scituate, MA: Duxbury Press.

- Cleveland, W. S. (1993), *Visualizing Data*, Summit, NJ: Hobart Press.
- Connor, G., and Korajczyk, R. (1987), "Risk and Return in an Equilibrium APT," *Journal of Financial Economics*, 21, 255-289.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: John Wiley & Sons.
- Crosier, R. B. (1988), "Multivariate Generalizations of CUSUM Quality-Control Schemes," *Technometrics*, 30, 291-303.
- Deming, W. E. (1982), *Quality, Productivity and Competitive Position*, Cambridge, MA: MIT Center for Advanced Engineering Study.
- Dickey, D. A. and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.
- Durbin, J., Koopman, S. J. (2001), *Time Series Analysis by State Space Method*, Oxford University Press.
- Engle, R. F., and Granger, C. W. J. (1987), "Co-integration and Error Correction Representation, Estimation and Testing," *Econometrica*, 55, 251-276.
- Engle, R. F., and Yoo, S. (1987), "Forecasting and Testing in Cointegrated Systems," *Journal of Econometrics*, 35, 143-59.
- Fama, E. F. and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 2, 427-465.
- Fama, E. F. and French, K. R. (1993), "Common Risk Factors in the Returns on Stock and Bonds," *Journal of Financial Economics*, 33, 3-56.
- Gabriel, K. R. (1985), "Multivariate Graphics," *Encyclopedia of Statistical Sciences*, eds. D. L. Banks, C. B. Read and S. Kotz, New York: John Wiley and Sons., Vol. 6, 66-79.
- Gnanadesikan, R. (1997), *Methods for Statistical Data Analysis of Multivariate Observations*, New York: John Wiley & Sons.
- Hald, A. (1951), *Statistics with Engineering Applications*, New York: John Wiley and Sons.
- Harris, T. J. and Ross, W. H. (1991), "Statistical Process Control Procedures for Correlated Observations," *The Canadian Journal of Chemical Engineering*, 69, 48-57.

- Harvey, A. C., Todd, P. H. J. (1983), "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study," *Journal of Business & Economic Statistics*, 1, 299-307.
- Holmes, D.S. and Mergen, A.E. (1993), "Improving the Performance of The T^2 Control Chart," *Quality Engineering*, 5, 619-625.
- Hotelling, H. (1947), "Multivariate Quality Control – Illustrated by the Air Testing of Sample Bombsights," *Techniques of Statistical Analysis*, eds. C. Eisenhart, M. W. Hastay, and W. A. Wallis, New York: McGraw-Hill, 111-184.
- Huang, X., and Bisgaard, S. (2009), "A Class of Markov Chain Models for Average Run Length Computations for Autocorrelated Processes," *Technometrics*, tentatively accepted.
- .(2010), "Model-Based Multivariate Monitoring Charts for Autocorrelated Processes," submitted.
- Jackson, J. E. (1985), "Multivariate Quality Control," *Communications in Statistics – Theory and Methods*, 14, 2657-2688.
- Jackson, J. E. (1991), *A User's Guide to Principal Components*, New York: John Wiley & Sons.
- Jiang, W. (2001), "Average Run Length Computation of ARMA Charts for Stationary Process," *Commun. Statist. – Simula.*, 30, 699-716.
- Jiang, W. (2004), "Multivariate Control Charts for Monitoring Autocorrelated Processes," *Journal of Quality Technology*, 36, 367-379.
- Jiang, W., Tsui, K. L., and Woodall, W. H. (2000), "A New SPC Monitoring Method: The ARMA Chart," *Technometrics*, 42, 399-410.
- Johansen, S. (1988), "Statistical Analysis of Co-integration Vectors," *Journal of Economic Dynamic and Control*, 12, 231-254.
- Johnson, R. A., and Bagshaw, M. (1974), "The Effect of Serial Correlation on the Performance of CUSUM Tests," *Technometrics*, 16, 103-112.
- Johnson, R. A., and Langeland, T. (1991), "A Linear Combinations Test for Detecting Serial Correlation in Multivariate Samples," *Topics in Statistical Dependence*, eds. Block, H. et al., Institute of Mathematical Statistics Monograph, 299-313.
- Johnson, R. A., and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis* (5th ed.), New York: Prentice Hall.
- Jolliffe, I. T. (2002), *Principal component analysis*, New York: Springer.

- Juran, J. M. (1986), "The Quality Trilogy: A Universal Approach to Managing for Quality," *Quality Progress*, August, 19-24.
- Kourti, T., and MacGregor, J. F. (1996), "Multivariate SPC Methods for Process and Product Monitoring," *Journal of Quality Technology*, 28, 409-428.
- Kramer, H., and Schmid, W. (1997), "EWMA Charts for Multivariate Time Series," *Sequential Analysis*, 16, 131-154.
- Krieger, C. A., Champ, C. W., and Alwan, L. C. (1992), "Monitoring an Autocorrelated Process," presented at the Pittsburgh Conference on Modeling and Simulation, Pittsburgh, PA.
- Lehmann, B. N. and Modest, D. (1988), "The Empirical Foundations of the arbitrage pricing theory," *Journal of Financial Economics*, 21, 213-254.
- Lin, W. S., and Adams, B. M. (1996), "Combined Control Charts for Forecast-based Monitoring Schemes," *Journal of Quality Technology*, 28, 289-301.
- Lintner, J. (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13-37.
- Lintner, J. (1969), "The Aggregation of Investor's Diverse Judgments and Preference in Purely Competitive Security Markets," *Journal of Finance and Quantitative Analysis*, 4, 347-400.
- Loredo, E. N., Jearkpaporn, D., and Borror, C. M. (2002), "Model-based Control Chart for Autoregressive and Correlated Data," *Quality and Reliability Engineering International*, 18, 489-496.
- Lowry, C. A., Woodall, W. H., Champ, C. W., Rigdon, S. E.. (1992), "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, 34, 46-53.
- Lowry, C.A., and Montgomery, D.C. (1995), "A Review of Multivariate Control Charts," *IIE Transactions*, 27, 800-810.
- Lu, C. W., and Reynolds, M. R., Jr. (1999a), "Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes," *Journal of Quality Technology*, 31, 259-274.
- .(1999b), "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology*, 31, 166-188.
- .(2001), "Cusum Charts for Monitoring an Autocorrelated Processes," *Journal of Quality Technology*, 33, 316-334.

- Lucas, J. M., Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, 1-12.
- Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, New York: Springer-Verlag.
- Mason, R. L., and Young, J. C. (2002), *Multivariate Statistical Process Control with Industrial Applications*, Philadelphia: ASA-SIAM.
- Mastrangelo, C. M., Runger, G. C., and Montgomery, D. C. (1998), "Statistical Process Monitoring with Principal Components," *Quality and Reliability Engineering International*, 12, 203-210.
- Merton, R. C. (1973), "An intertemporal capital asset pricing model," *Econometrica*, 41, 867-887.
- Molenaar, P. C. M. (1985), "A Dynamic Factor Model for the Analysis of Multivariate Time Series," *Psychometrika*. 50, 181-202.
- Montgomery, D. C., and Mastrangelo, C. M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology*, 23, 179-193.
- Montgomery, D. C. (2008), *Introduction to Statistical Quality Control* (6th ed.), Wiley.
- Mossin, J. (1966), "Equilibrium in a Capital Asset Market," *Econometrica*, 768-783.
- Pan, X. (2005), "Notes on Shift Effects For T^2 -type Charts on Multivariate ARMA Residuals," *Computers & Industrial Engineering*, 49, 381-392.
- Peña, D., and Box, G. E. P. (1987), "Identifying a Simplifying Structure in Time Series," *Journal of the American Statistical Association*, 82, 836-843.
- Peña, D., and Poncela, P. (2004), "Forecasting with nonstationary dynamic factor models," *Journal of Econometrics*, 119, 291-321.
- Peña, D., Tiao, G. C., and Tsay, R. S. (2001), *A Course in Time Series Analysis*, New York: John Wiley & Sons.
- Pignatiello, J. J. and Runger, G. C. (1990), "Comparisons of multivariate CUSUM charts," *Journal of Quality Technology*, 22, 173-186.
- Playfair, W. (1786), *The Commercial and Political Atlas*, Cambridge: Cambridge University Press.
- Quenouille, M. H. (1957), *The Analysis of Multiple Time Series*, New York: Hafner.
- Reinsel, G. (1983), "Some Results on Multivariate Autoregressive Index Models," *Biometrika*. 70, 145-156.

- Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*, New York: Springer.
- Rencher, A. C. (2002), *Methods of Multivariate Analysis*, New York: John Wiley & Sons.
- Roll, R. and Ross, S. A. (1980), "An Empirical Investigation of the Arbitrage Pricing Theory," *The Journal of Finance*, 35, 1073-1103.
- Rosolowski, M. and Schmid, W. (2003), "EWMA Charts for Monitoring the Mean and Autocovariances of Stationary Processes," *Statistical Papers*, 47, 595-630.
- Ross, S. A. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory*, 13, 341-360.
- Runger, G. C. and Alt, F. B. (1996), "Choosing principal components for multivariate statistical process control," *Communications in Statistics: Theory and Methods*, 25, 909-922.
- Runger, G. C. and Prabhu, S. S. (1996), "A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart," *Journal of American Statistical Association*, 91, 1701-1706.
- Runger, G. C., Willemain, T. R. and Prabhu, S. S. (1995), "Average Run Lengths for Cusum Control Charts Applied to Residuals," *Communications in Statistics - Theory and Methods*, 24, 273-282.
- Seber (2004), *Multivariate Observations*, New York: John Wiley & Sons.
- Sharpe, W. F. (1963), "A Simplified Model for Portfolio Analysis," *Management Science*, 9, 277-293.
- Sharpe, W. F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425-442.
- Shewhart, W. A. (1931), *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand.
- Shu, L. J., Apley, D. W., and Tsung, F. (2002), "Autocorrelated Processes Monitoring Using Triggered Cuscore Charts," *Quality and Reliability Engineering International*, 18, 411-421.
- Sloan, I. H., and Joe, S. (1994), *Lattice Methods for Multiple Integration*, Oxford: Clarendon Press.
- Sullivan, J. H., and Woodall, W. H. (1996), "A Comparison of Multivariate Control Charts for Individual Observations," *Journal of Quality Technology*, 28, 398-408.

- Superville, C. R., and Adams, B. M. (1994), "An Evaluation of Forecast-based Quality Control Schemes," *Communications in Statistics - Simulation and Computation*, 23, 645-661.
- Switzer, P. (1985), "Min/Max Autocorrelation Factors for Multivariate Spatial Imagery," in *Computer Science and Statistics: Proceedings of the 16th Symposium on the Interface*, 13-16.
- Tiao, G. C., and Box, G. E. P. (1981), "Modeling multiple time series with applications," *Journal of the American Statistical Association*, 76, 802-816.
- Tiao, G. C., and Tsay, R. S. (1989), "Model Specification in Multivariate Time Series," *Journal of Royal Statistical Society*. 51, 157-213.
- Tiao, G. C., Tsay, R. S., and Wang, T. (1993), "Usefulness of linear transformations in multivariate time-series analysis," *Empirical Economics* 18, 567-593.
- Tracy, C. D., Young, J. C., and Mason, R. L. (1992), "Multivariate Control Charts for Individual Observations," *Journal of Quality Technology*, 24, 88-95.
- Treynor, J. (1961), "Toward a Theory of the Market Value of Risky Assets," unpublished manuscript.
- Tsay, R. S. (2005), *Analysis of Financial Time Series* (2nd ed.), New York: John Wiley and Sons.
- Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire: Graphic Press.
- Tukey, J. W. (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics* 33, 1-67.
- Vanbrackle, L. N., and Reynolds M. R. (1997), "EWMA and Cusum Control Charts in the Presence of Correlation," *Communications in Statistics - Simulation and Computation*, 26, 979-1008.
- Velu, R. P., Reinsel, G. C., and Wichern, D. W. (1986), "Reduced Rank Models for Multiple Time Series," *Biometrika*, 73, 105-118.
- Velu, R. P., Wichern, D. W. and Reinsel, G. C. (1987), "A note on non-stationarity and canonical analysis of multiple time series models," *Journal of Time Series Analysis*, 8, 479-487.
- Wainer, H. (2004), *Graphic Discovery: A Trout in the Milk and Other Visual Adventure*, Princeton: Princeton University Press.
- Wardell, D.G., Moskowitz, H. and Plante, R.D. (1992), "Control Charts in the Presence of Data Correlation," *Management Sciences*, 38, 1084-1105.

- Wardell, D.G., Moskowitz, H., and Plante, R.D. (1994), "Run-Length Distribution of Special-Cause Control Charts for Correlated Processes," *Technometrics*, 36, 3-17.
- Wierda, S. J. (1994), "Multivariate statistical process control - Recent results and directions for future research," *Statistica Neerlandica*, 48, 147-168.
- Woodall, W. H., and Ncube, M. M. (1985), "Multivariate CUSUM quality-control procedures," *Technometrics*, 27, 285-292.
- Woodall, W. H., (2000), "Controversies and Contradictions in Statistical Process Control," *Journal of Quality Technology*, 32, 341-350.