TOPICS IN UNIVARIATE TIME SERIES ANALYSIS

WITH BUSINESS APPLICATIONS

A Dissertation Presented

by

DAVIT KHACHATRYAN

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

DOCTOR OF PHILOSOPHY

September 2010

Isenberg School of Management

TOPICS IN UNIVARIATE TIME SERIES ANALYSIS

WITH BUSINESS APPLICATIONS

A Dissertation Presented

by

DAVIT KHACHATRYAN

Approved as to style and content by:

____

Iqbal Agha, Chair

____

John Staudenmayer, Member

____

Senay Solak, Member

____

Ahmed Ghoniem, Member

____

D. Anthony Butterfield, Program Director
Isenberg School of Management

DEDICATION

Dedicated to Søren Bisgaard.

ACKNOWLEDGEMENTS

I first had the opportunity to meet Dr. Søren Bisgaard in the spring of 2006. It was an interview which resulted in my acceptance to the doctoral program in the department of Finance and Operations Management at the University of Massachusetts, and starting my doctoral work under his supervision. During each of our meetings that followed throughout the course of the next four years, I was the fortunate witness of his endless passion towards statistics. The testimony of his passion was evident not only through his research, but also from the enthusiasm with which he taught applied statistics until the very last days of his life. That fervor, that immense passion and illuminating clarity with which he communicated his knowledge to me was culminating at that very last meeting that we had in his house only couple of weeks before his death. He wanted to meet with me to discuss my future research goals and of course to make sure that everything was in order before I delivered the proposal for this dissertation.

Throughout my doctoral studies Dr. Bisgaard provided help and guidance on a myriad of occasions ranging from supervising my research to finding my summer internship projects. But above all, he has taught me applied statistics as an art of scientific inquiry, placing empiricism and experimentation at the cornerstone of the discipline. I would like to express my deepest appreciation to Dr. Bisgaard for his exceptionally devoted supervision and for all his invaluable teachings without which neither this work nor most of my personal and professional development during the last four years would have been possible. It would be hard to separate Sue Ellen Bisgaard when thanking Dr. Bisgaard, and I am very appreciative of her caring and warmth, as well as for her believing in me. This dissertation is dedicated to Dr. Søren Bisgaard.

I would like to express my sincere gratitude to Dr. Iqbal Agha for his unwavering support during the last four years and for his dedication to serve as the chair of my dissertation committee. Without his constant support I would not have made it—especially through the difficult times that followed the passing of Dr. Søren Bisgaard.

I would like to thank Dr. John Staudenmayer for serving on my dissertation committee and for being my first teacher of mathematical statistics at the University of Massachusetts. I am indebted to Dr. Staudenmayer's wonderful teaching, which emphasized the mathematical formality but nevertheless did not overshadow the empirical beauty of statistics. It is due to his teachings that I not only started to comprehend the elegance of statistics, but was also fired up to dig deeper into the discipline.

I am very thankful to Dr. Senay Solak and Dr. Ahmed Ghoniem who, in spite of my short notice, eagerly agreed to be part of my dissertation committee. In addition to serving on my committee, they have shown support and eagerness to help in spite of their hectic schedules, their own research and teaching responsibilities. I am very appreciative of their support.

I would also like to express my gratitude to Dr. John Buonaccorsi who always has kept his doors open to me during the last six years and has shown great help and support on numerous occasions. His erudition of statistics will always be a topic that I will look up to as he will always be remembered as the "walking encyclopedia of statistics," the way I often characterize him to others.

Special thanks to Dr. Anna Nagurney, Dr. Anthony Butterfield, Dr. Marta Calas and Dr. Ben Branch for their teaching, encouragement and guidance.

Finally, I could not have climbed this mountain without the love and support of my parents and family. No part of this dissertation would exist without their help. The extent of their importance is incomprehensible.

ABSTRACT

TOPICS IN UNIVARIATE TIME SERIES ANALYSIS WITH BUSINESS APPLICATIONS

SEPTEMBER 2010

DAVIT KHACHATRYAN

B.Sc., YEREVAN STATE UNIVERSITY

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Søren Bisgaard

Recent technological advances in sensor and computer technology allow the observation of business and industrial processes at fairly high frequencies. For example, data used for monitoring critical parameters of industrial furnaces, conveyor belts or chemical processes may be sampled every minute or second. A high sampling rate is also possible in business related processes such as mail order distribution, fast food restaurant operations, and electronic commerce. Data obtained from frequently monitored business processes are likely to be autocorrelated time series that may or may not be stationary. If left alone, processes will typically not be stable, and hence they will usually not posses a fixed mean, thus exhibiting homogeneous non-stationarity. For monitoring, control, and forecasting purposes of such potentially non-stationary processes it is often important to

develop an understanding of the dynamic properties of processes. However, it is sometimes difficult if not impossible to conduct deliberate experiments on full scale industrial plants or business processes to gain the necessary insight of their dynamic properties. Fortunately, intentional or inadvertent process changes that occur in the course of normal operation sometimes offer an opportunity to identify and estimate aspects of the dynamic behavior.

To determine if a time series is stationary, the standard exploratory data analytic approach is to check that the sample autocorrelation function (ACF) fades out relatively quickly. An alternative, and at times a sounder approach is to use the variogram – a data exploratory tool widely used in spatial (geo) statistics for the investigation of spatial correlation of data. The first objective of this dissertation is to derive the basic properties of the variogram and to provide the literature on confidence intervals for the variogram. We then show how to use the multivariate Delta method to derive asymptotic confidence intervals for the variogram that are both practical and computationally appealing.

The second objective of this dissertation is to review the theory of dynamic process modeling based on time series intervention analysis and to show how this theory can be used for an assessment of the dynamic properties of business and industrial processes. This is accompanied by a detailed example of the study of a large scale ceramic plant that was exposed to an intentional but unplanned structural change (a quasi experiment).

The third objective of this dissertation concerns the analysis of multiple interventions. Multiple interventions occur either as a result of multiple changes made to the same process or because of a single change having non-homogeneous effects on the

time series. For evaluating the effects of undertaken structural changes, it is important to assess and compare the effects, such as gains or losses, of multiple interventions. A statistical hypothesis test for comparing the effects among multiple interventions on process dynamics is developed. Further, we investigate the statistical power of the suggested test and elucidate the results with examples.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Discrete time series represent the evolution of a stochastic process over the time. The research on statistical time series dates back to Yule (1927), who used harmonic analysis and difference equations for analyzing periodicities in time series. The application of the method proposed by Yule concerned the forecasting of periodicity in business cycles based on the number of sunspots (Wolfer's Sunspot Numbers). Box and Jenkins (Box and Jenkins (1968)) were among the first to pull together a coherent framework for the statistical analysis of discrete time series and apply their methodology to a wide range of business and industrial applications including control, monitoring and forecasting of industrial processes. The illustration of the pernicious effects of misspecification of the noise model in classical regression analysis was yet another cornerstone in the development of time series analysis. See Box and Newbold (1971) for an elegant exhibition of the need for time series and transfer function modeling when dealing with time dependent phenomena. Comprehensive introduction to time series modeling can be found in Box, *et al.* (1994) and Brockwell and Davis (1987) among many others.

In this chapter we provide a concise introduction to the theory and applications of univariate time series modeling. In Section 1.1 we briefly outline the theoretical fundamentals of univariate time series modeling which will be used in the dissertation. After a general discussion of the business applications of times series in Section 1.2 we provide four examples (sections 1.2.1-1.2.4) of time series which we will utilize

throughout the dissertation for elucidation of our theoretical results. Finally, we provide the plan of the dissertation in Section 1.3.

1.1 Time Series Modeling and ARIMA Models

Suppose a time series z_t $(t = 1, 2, ..., n)$ is sampled from a stochastic process Z_t and that we are interested in analyzing the process dynamics. By definition, Z_t is an autoregressive integrated moving average process ARIMA(p, d, q) of orders *p, d*, and *q* if it can be represented in the following form

$$\varphi(B)Z_t = C + \theta(B)a_t$$

where $\varphi(B) = \phi(B)(1-B)^d$ is the generalized autoregressive operator, $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ is the autoregressive operator, $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$ is the moving average operator, *B* is the backshift operator, *C* is a constant and *d* denotes the order of differencing. Moreover, errors a_t are assumed to be white noise, distributed normally and independently from one another with mean zero and constant variance; equivalently $a_t \overset{iid}{\sim} N(0, \sigma^2)$. Furthermore, Z_t is said to be weakly stationary autoregressive moving average ARMA(*p*, *q*) process of orders p and *q*, if all the roots of the polynomial $\varphi(B)$ are outside the unit circle (in absolute value), in other words if $\varphi(B) = \phi(B)$ and $|B| > 1$ for the roots of $\phi(B)$. Similarly, Z_t is called invertible if the roots of the moving average polynomial $\theta(B)$ are all greater than one, in absolute value. It can be shown (Wold (1954)) that every weakly stationary process Z_t can be represented as

$$Z_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B) a_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ (with $\psi_0 = 1$) and μ is the mean of the process. Further, for every weakly stationary process there exists an autocorrelation function (ACF) which describes the structure of correlation present in the series. The autocorrelation function, by definition is

$$\rho_h = \frac{Cov(Z_t, Z_{t-h})}{\sigma_z^2} \qquad h = 0, 1, 2, \ldots$$

where Cov denotes the covariance and σ_z^2 is the variance of the process Z_t.

Usually there is a seasonal component present in the time series structure. Modeling and analysis of seasonal time series can be viewed as a generalization of the non-seasonal approach (see, e.g. Box, *et al*. (1994)). The seasonal time series Z_t is general multiplicative seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$ process if

$$\varphi(B)\Phi'(B^s)Z_t = C + \theta(B)\Theta(B^s)a_t \tag{1.1}$$

where $\Phi'(B^s) = \Phi(B^s)(1-B^s)^D$, D denotes the order of seasonal differencing, S denotes the seasonality, $\Phi(B^s) = 1 - \sum_{i=1}^{P} \Phi_i B^{is}$, $\Theta(B^s) = 1 - \sum_{i=1}^{Q} \Theta_i B^{is}$ and C is a constant.

1.2 Business Applications of Time Series

The business applications of time series modeling are numerous. The proliferation of information technology has profoundly altered business operations, enabling business practitioners to collect data at high frequencies. In industrial control engineering

applications, due to the ease of sampling and recent advances in sensor and computer technology, the automated sensors (such as air flow sensors, photoeyes and thermocouples) can currently record enormous amounts of data sampled fairly frequently. Similarly, in business related applications the data often come in the form of transactional data and the data is often sampled at point of sales (POS) devices such cash registers and retail scanners (Liu, *et al*. (2001)).

The time ordered data arising from the above discussed applications are both cross and auto correlated thus rendering the time series methodology an indispensable part of the empirical investigation. In statistical process control the use of time series methods are fundamental for control, monitoring and forecasting the process behavior. It is worthwhile mentioning that the classical Schewhart control charts (Schewhart (1931)) have been designed under the assumption of independence of the sampled data. Similarly, most of the earlier work in statistical process control rested on the classical conjecture of independence. The proliferation of automatic sensors which enabled the data collection at relatively higher frequencies thus challenged the assumption of independence due to the autocorrelation present in the sampled data. Time series analysis thus provides a more precise and accurate approach for modeling and analysis of data arising from statistical process control applications. See Alwan and Roberts (1988), Stone and Taylor (1995) and Box and Luceno (1997) for a general discussion of time series modeling in statistical process control.

Data mining on collected time series enables businesses to improve their operations management, since the useful patterns found in data often provide for better labor scheduling, inventory management and product preparation, which in turn can

significantly affect the bottom line of the particular business entity. Time series methods have proven to be useful also in understanding products – life cycles and popularity trends of products which can further improve the operations of business entities. For instance Liu, *et al.* (2001) illustrate the use of time series data mining to improve the operation of a fast foot restaurant franchise. See applications of time series forecasting in analyzing natural gas consumption (Liu and Lin (1991)), as well as the modeling of the impact of temperature on natural electricity consumption (Al-Zayer and Al-Ibrahim (1996)), and forecasting of a call center intraday arrivals at a U.K. based bank (Taylor (2008)).

Performance of a business can be further improved by utilizing time series methods in marketing applications such as in the analysis of the promotional effects of undertaken advertising campaigns (via the internet, radio, television, etc.). The pioneering work in this area is due to Palda (1964) who pointed out the advantages of incorporating time series analysis in the investigation of the sales-advertising relationship. Subsequently, there has been a vast literature concerning the analysis of the *build-up effects* of advertising, in other words the investigation of the time it takes for an advertising campaign to become effective in increasing the sales or market share of the firm (Clarke (1976); Krishnamurti, *et al.* (1989)). In addition, another fundamental question in market research is related to the *carryover effect*, namely – the amount of time that the advertising effect persists (Clarke (1976); Bass and Clarke (1972)). Time series transfer function theory has proved to be effective in the analysis of build-up and carryover effects of advertising campaigns (Leone (1983)). An illuminating example of time series intervention modeling appears in Wichern and Jones (1977) to analyze the

effect of the American Dental Association's endorsement of the Crest toothpaste on the market shares of Crest and Colgate dentifrice. In addition to univariate models, there has been interest in multivariate time series analysis in marketing research, see for example Leone (1983) for an illustration of the applications of bivariate time series models in market research.

There has been recent growing interest in using time series models in internet traffic and computer network monitoring and management applications. In a local area network (LAN) the information flows in the form of data packets that are constantly directed from origins to destinations according to a routing mechanism. The term *network tomography* was coined by Vardi (1996) and refers to the statistical inference of the unknown intensity (usually byte counts) on the links of a computer network given the aggregate link counts. A convenient way to represent the network intensity is via a *traffic matrix* which has as its elements the volumes of traffic flowing between all possible pairs of origins and destinations in a computer network. One of the major questions in network tomography is the statistical estimation of the traffic matrix of a LAN. The accurate estimation of the traffic matrix is crucial for effective operation of the routing scheme, since the knowledge of the traffic intensity between origins and destinations is directly related to the routing decision. Time series methods are particularly useful in aforementioned situations, since the observed link counts of network are autocorrelated thus rendering the classical assumption of independence obsolete. See Vardi (1996), Medina, *et al.* (2002), and Cao *et al.* (2000) for time series modeling in network tomography.

Time series methods have also extensively been used in financial and macro economical time series (Tsay (2005)). Various econometric applications can be found in Enders (2009) and Griffiths, *et al.* (1993) among other valuable sources.

In the following four subsections we provide four examples of univariate time series which will be used in this dissertation.

1.2.1 Chemical Process Concentration Data

Figure 1.1 depicts the uncontrolled chemical process concentration readings (sampled every two hours) also known as *Series A*, that has originally been analyzed by Box, *et al*. (1994). It can be noticed from Figure 1.1 that the data are autocorrelated and possibly non-stationary. In this dissertation we will discuss the use of the variogram for investigating if this time series can be modeled as stationary.

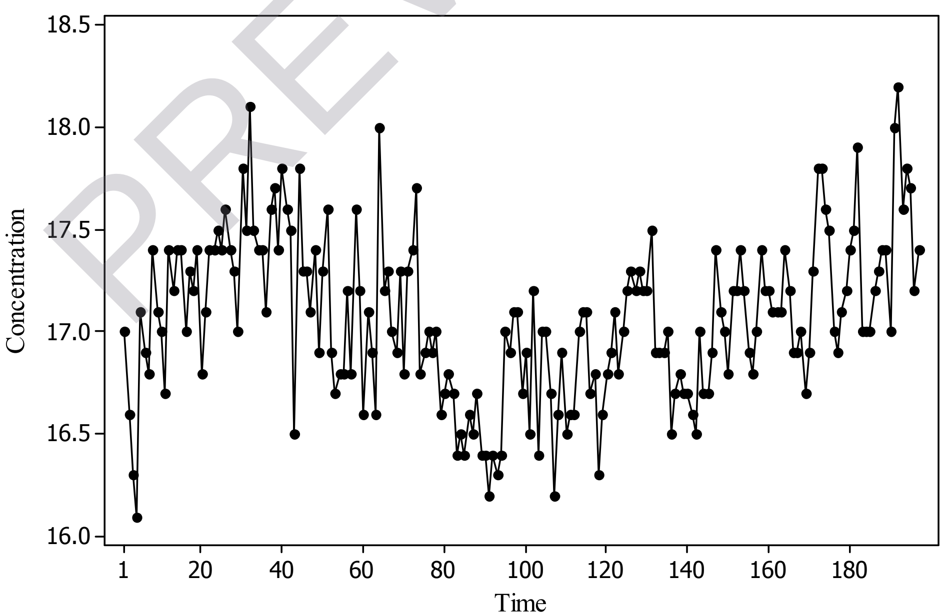

Figure 1.1. Time series plot of chemical process concentration readings (Series A).