APPLICATION OF INTERVENTION ANALYSIS TO TIME SERIES DATA ON ECONOMIC AND SOCIAL PROBLEMS (WITH LOGNORMAL NOISE)

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of
Dokuz Eylül University
In Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Statistics

by
Esin FİRUZAN

January, 2004
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We certify that we have read the thesis, entitled **"APPLICATION OF INTERVENTION ANALYSIS TO TIME SERIES DATA ON ECONOMIC AND SOCIAL PROBLEMS (WITH LOGNORMAL NOISE)"** completed by Esin FİRUZAN under supervision of Yrd. Doç.Dr. C.Cengiz ÇELİKOĞLU and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Yrd. Doç.Dr. C. Cengiz ÇELİKOĞLU

Supervisor

Prof. Dr. Serdar KURT

Jury Member

Prof. Dr. Şenay ÜÇDOĞRUK

Jury Member

Prof. Dr. Ahmet ÖZMEN

Jury Member

Prof. Dr. Levent ŞENYAY

Jury Member

Approved by the
Graduate School of Natural and Applied Sciences

Prof.Dr. Cahit HELVACI
Director

ACKNOWLEDGMENT

I would like to express my appreciation and gratitude to my advisor Yard. Doç.Dr. C.Cengiz ÇELİKOĞLU for his encouragement, advice and support during my study. I greatly appreciate the useful dicussions and contributions of Prof. Dr. Ahmet ÖZMEN. His valuable suggestions, constant input of creative ideas and enthusiastic guidance provided me with the great privilege of learning new attitudes towards research life. I also want to thank my thesis committee Prof. Dr. Serdar KURT and Prof. Dr. Şenay ÜÇDOĞRUK for their generosity and helping me throughout the course of this work.

I also wish to express my special thanks to Umay KOÇER, Aylin ALIN and Tufan ALIN for their special help. I am also grateful to the providers of data set used in this thesis and especially to Korkut ÖZKAN that kindly provided the traffic accident data analyzed in Chapter 4.

I would especially like to dedicate this work to my husband, Yard. Doç. Dr. Ali Rıza FİRUZAN for his support and understanding that I have always received and for his love, constant support and dedication that are sources of inspiration and strength for being a better person everyday.

Esin FİRUZAN

ABSTRACT

Traffic accidents have been big problem during the last years because of the fundamental changes in social economic life. In this study, the effect of radar application in Izmir was examined by using two approaches. The first approach consisted of adapting traffic accidents to the lognormal distribution. This approach depends on assumption that accidents occurred on the routes are correlated each other. After the adaptation to the lognormal distribution, the estimators of the parameters of lognormal distribution were estimated by using moment estimation and modified moment estimation methods. The validity of the underlying assumption is supported by goodness of fit tests.

The second approach, involved time series model of traffic accidents before intervention (the radar application) as a Box-Jenkins univariate time series model. As expected, this model proved to be inadequate in explaining the postintervention period, and the model was modified to incorporate the expected form of the intervention effects. Intervention effects were measured quantitatively on time series of traffic accidents.

Results showed that the radar application produced significant reduction on accidents occurred on the routes.

Keywords: Time Series, Intervention Analysis, Lognormal Distribution, Traffic Accident Series

ÖZET

Bu çalışma, trafik kazaları verilerinin lognormal dağılıma adapte edilmesi ve trafik kazalarında radar uygulamasının, kazaların oluşumu üzerindeki etkisini araştırma aşamalarından oluşmaktadır. Radar uygulaması, trafik kazaları serisine müdahale olarak ele alınmış ve müdahalenin trafik kazaları serisi üzerindeki niceliksel etkisi araştırılmaya çalışılmıştır.

Zaman serisinde müdahale analizi, gürültü değişkeni için stokastik bir model, müdahale için dinamik bir model olmak üzere iki aşamada gerçekleştirilmiştir. Daha sonra en uygun ARIMA modeli belirlenerek model parametreleri tahminlemeye çalışılmıştır. Bunun için Moment Tahminleme ve Uyarlanmış Moment Tahminleme Yöntemleri kullanılmıştır. Beyaz gürültüden farklı olmadığından emin olunan model artıkları teşhis edilerek müdahalenin seriye etkileri ölçülmeye çalışılmıştır.

Elde edilen sonuçlar, radar uygulamasının yol üzerindeki kazaların meydana gelişinde azaltıcı etki yarattığı yönünde olmuştur.

Anahtar Kelimeler : Zaman Serisi, Müdahale Analizi, Lognormal Dağılım, Trafik Kaza Serisi

CONTENTS

Page

Chapter One
INTRODUCTION

Chapter Two
FUNDAMENTAL CONCEPTS

Chapter Three
INTERVENTION ANALYSIS IN TIME SERIES

Chapter Four
APPLICATION

Chapter Five
CONCLUSIONS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER ONE
INTRODUCTION

1.1. Introduction

Much of statistical methodology is concerned with models in which the observations are assumed to vary independently. In many applications, dependence between the observations is regarded as a nuisance. Many sets of data in business, economics, engineering and natural sciences appear as a time series where observations are dependent and where the nature of this dependence is of interest in itself. The body of techniques available for the analysis of such series of dependent observations is called time series analysis. It is necessary to set up a hypothetical probability model to represent the data. An important part of the analysis of a time series is the selection of a suitable probability model (or class of models) for the data. An overview of the time series models contains a non-technical description of five classes of models such have been developed to deal with a wide range of practical situations.

First model is univariate models which predict future values of the variable of interest solely on the basis of the historical pattern of that variable, assuming that the historical pattern will continue. Another model is transfer function models which are Box-Jenkins causal models. These models predict future values of a time series on the basis of past values of the time series and on the basis of values of one or more other time series related to the time series to be predicted. Time series are frequently affected by certain external events or circumstances such as holidays, strikes, advertising promotions and environmental regulations. Methods for estimating transfer function models based on deterministic perturbations of the input, such as step, pulse, and sinusoidal changes, have not always been successful. This is because, for perturbations of a magnitude that are relevant and tolerable, the response of the system may be masked by uncontrollable disturbances referred to collectively as noise. Intervention models are used for such as a strike, a holiday or a change in definition of a variable. Intervention analysis is developed to obtain a quantitative measure of the impact of the intervention event on the time series of interest. Other models are Multivariate Stochastic Models and Multivariate Transfer Function Models. Multivariate stochastic models can represent several dependent series with mutual interactions. Multivariate transfer function models can be used to relate several mutually interacting dependent variables to several independent variables. In this study, intervention models will be focused on.

Many applications have been solved using intervention analysis. Box and Tiao (1975) used intervention models to study and quantify the impact of air pollution controls on smog-producing oxidant levels in the Los Angeles area and of economic controls on the consumer price index in the United States. Then Deutsch and Alt (1977), Hay and McClearly (1979) applied intervention analysis in the area of gun control law. Wichern and Jones (1977) analyzed advertising and its impact on market share using intervention time series.

Cauley and Im (1988) have examined the effects of increasing security measures on the number of terrorist incidents. Singer and McDowall (1988) considered intervention in the area of juvenile offender law, Martinez-Schnell and Zaidi (1989) utilized the intervention analysis model to investigate deaths caused by motor vehicles. And also Fomby and Hayes (1990) have examined the war on poverty has been examined to see if it has had an impact on income distribution. Deadman and Pyle (1993) used intervention analysis to study the influence of the abolition of capital punishment on the homicide rate in Britain. In addition, Foreman (1993) aided public policy makers who wished to know if airline industry deregulation has had an effect on air safety.

Recent increases in the number of traffic accidents and deaths in these accidents in Turkey make it necessary to conduct many studies on identification of the causes of these accidents and methods to prevent these accidents.

Government must put into practice some obligations because of the increment in the number of deaths and injuries due to the traffic accidents. Government forces the drivers to obey the rules applying not only fines but also various controls. Due to the improvements in technology, new applications developed in traffic control to make it easier.

The aim of this study is to obtain quantitative measurement for radar control which is considered as an intervention on traffic accident series. The impact of radar application is inspected by selecting a specified route where the radar application has just started and has intense traffic. The intervention analysis is performed on real data which are adapted lognormal distribution. It is the first study which analyses nonnormal distribution problems. Besides, through intervention analysis, to take precautions against any unusual values in the time series of traffic accidents that might have resulted as a consequence of the intervention event.

This study contains five chapters. The general information about the overall research is in the first chapter. The second chapter contains fundamental concepts used in intervention analysis. Intervention concepts in time series, application of intervention analysis and interpretation the impact of the intervention are mentioned in chapter three. The fourth chapter aims are to find the effect of impact of radar control on traffic accident series. The fifth chapter contains the results of the study and some suggestions based on these results.

CHAPTER TWO
FUNDAMENTAL CONCEPTS

2.1. Stochastic Processes

The statistical approach to forecasting is based on the construction of a model. A model that explains a mechanism, which is regarded as being capable of having produced the observations in question, is almost invariably stochastic. Stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws. Mathematically, a stochastic process may be defined as a collection of random variables which are ordered in time and defined at time points that are either continuous or discrete. The time series to be analyzed may then be thought of as one particular realization, produced by the underlying probability mechanism, of the system under study. In other words, in analyzing a time series it is regarded as a realization of a stochastic process.

$Z(\omega, t)$ are time indexed random variables where ω belongs to a sample space and t belongs to an index set. For a given ω, $Z(\omega, t)$, as a function of t is called a realization. The population that consists of all possible realizations is called the ensemble in stochastic process. $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\}$ is a finite set of random variables $\{Z(\omega,t): t = 0, \pm 1, \pm 2, \ldots\}$ is a stochastic process. The n-dimensional distribution function is defined by

$$F(z_{t_1},\ldots,z_{t_n}) = p\{\omega : z(\omega,t_1) \le z_{t_1},\ldots,z(\omega,t_n) \le z_{t_n}\} \tag{2.1}$$

If its one dimensional distribution function is time invariant a process has to be the first order stationary in distribution for any integers t_1, k and t_1+k; if $F(z_{t_1},...,z_{t_n}) = F(z_{t_1+k},...,z_{t_n+k})$ for any integers ($t_1, t_2, ..., t_n$) and k lags, process has to be the n^{th} order stationary in distribution. If $F(z_{t_1},...,z_{t_n}) = F(z_{t_1+k},...,z_{t_n+k})$ is true for any n, a process is strongly and completely stationary. Stationary stochastic processes are based on the assumption that the process is in a particular state of statistical equilibrium.

2.2. The Autocovariance and the Autocorrelation Functions

"Time series may be stationary or nonstationary. Stationary series characterized by a kind of statistical equilibrium around a constant mean level as well as a constant dispersion around that mean level" (Yaffee,&McGee 2000, p.5). Autocovariance function plays very important role for describing a stochastic process. When a stochastic process is stationary, its time domain properties can be summarized by autocovariance function.

The autocovariance function γ_k, which is the covariance of Z_t with Z_{t+k} can be described by following equation.

$$\gamma_k = Cov\,(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \tag{2.2}$$

"The autocovariance remains the same regardless of the point of temporal reference. Under these circumstances, the autocovariance depends only on the number of time periods between the two points of temporal references" (Yaffee,& McGee, 2000, pp.5-6).

The autocovariances are standardized by dividing them by the variance of the process. The autocorrelation function (ACF) is defined by equation (2.3).

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \tag{2.3}$$

For a stationary process the autocovariance function γ_k *and the autocorrelation function* ρ_k *have the following properties:*

1. $\gamma_0 = Var(Z_t)$; $\rho_0 = 1$
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ *and* $\rho_k = \rho_{-k}$, for all k, i.e., γ_k and ρ_k *are even functions and hence symmetric about the time origin, k=0. This follows from the fact that the time difference between* Z_t *and* Z_{t+k} *and* Z_t *and* Z_{t-k} *are the same. Therefore, the autocorrelation function is often plotted only for the nonnegative lags (Wei, 1990, p.10).*

2.3. The Partial Autocorrelation Functions

The partial autocorrelation (PACF) between Z_t and Z_{t+k} which is denoted by P_k is defined by equation (2.4).

$$P_k = \frac{Cov\left[\left(Z_t - \hat{Z}_t\right)\left(Z_{t+k} - \hat{Z}_{t+k}\right)\right]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}} \qquad (2.4)$$

The important point is that its behaviour is the opposite of the behaviour exhibited by the autocorrelation function. "If a series is stationary, the magnitude of the autocorrelation attenuates fairly rapidly, whereas if the series is nonstationary, the autocorrelation diminishes gradually over time" (Yaffee,& McGee, 2000, p.6).

"For a pure AR(p) process, the theoretical partial autocorrelations are zero at lags beyond p, while for an MA process they die away gradually. If the observations are generated by an AR(p) process, the sample partial autocorrelations beyond lag p are normally distributed with mean zero and variance" (Harvey, 1993, p.75).

2.4 Stationary Stochastic Process

Stationary processes are defined as a random process where all of its statistical properties (such as mean, variance, covariance and higher moments) do not vary with time or purified from the periodic waves.

2.4.1 The Gaussian Process

If the joint probability distribution of observations associated with any set of times is a multivariate normal distribution, the stochastic process is called a normal or Gaussian process.

"A stochastic process $X(t)$, $t \geq 0$ is called a Gaussian, or a normal, process if $X(t_1), \ldots, X(t_n)$ has a multivariate normal distribution for all $t_1, \ldots, t_n$" (Ross, 1993, p.476).

If $\{X(t), t \geq 0\}$, then as each of $X(t_1), X(t_2), \ldots, X(t_n)$ can be expressed as a linear combination of the independent normal random variables $X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$. This process is called as Gaussian process.

Since the multivariate normal distribution is uniquely characterized by its moments of first and second order, strictly stationary and weakly stationary are equivalent for a Gaussian process. Many areas in statistics, most time series results are established for Gaussian processes.

Noise

The term "noise" was first used in communication engineering as a result of the undesired acoustic effects accompanying spontaneous electric fluctuations in receivers. Noise means something that interferes with the undesired signal. In many statistical methods such as regression analysis and categorical data analysis, the noise is the randomly fluctuating part and the signal is the unknown deterministic parameter.

Noise is treated as a stochastic process of irregular fluctuations. In applications, the noise process is often assumed to be stationary and ergodic. That is, its statistical properties can be described completely by just one sample over a long period. "The analysis of autocorrelation functions is useful in fitting autoregressive-moving average (ARMA) models to the noise process" (Canada, 1985, p.252).

The Gaussian Noise

"In many cases, noise is best described as a Gaussian process. The assumed normality can be justified by the central limit theorem if the noise is composed of many small independent (or weakly dependent) random effects" (Canada, 1985, p.253).

The main advantage of using the Gaussian assumption is that, the best linear estimators of the signal is optimal under the criterion of mean squared error. It means that there is no need to consider nonlinear theory in signal estimation in Gaussian systems. In signal detection, the likelihood ratio statistic is an optimal test statistic, and many results have been established on absolute continuity and the Radon- Nikodym derivative between the two measures induced by pure noise and signal plus noise, when both measures are Gaussian (Canada, 1985, p.253).

Even though optimal statistics can be obtained under the Gaussian condition, robust statistics are desired so that decisions based on statistics are less sensitive to the Gaussian assumption. "Various kinds of Non-gaussian noise take place in different situations. Some are generated from Gaussian noise through nonlinear devices." (Canada, 1985, p.254) Sometimes the noise is simply measurement error; in engineering applications the noise may be interference some kind of process.

2.4.2 White Noise Process

Once examining the characteristics of nonlinear models, distinguishing between independent and uncorrelated random variables can be very important. White noise is defined as a sequence of independent and identically distributed (*i.i.d.*) random variables with constant mean and variance. It is sometimes possible to make non-trivial predictions from the series which have the white noise property. When successive values are merely uncorrelated it is called strict white noise. "If successive values follow a normal (Gaussian) distribution, then zero correlation implies independence so that Gaussian uncorrelated white noise is strict white noise. However, with nonlinear models, distributions are generally nonnormal and zero correlation need not imply independence" (Chatfield, 1999, pp.197-198).

White noise is a stationary stochastic process with constant spectral density. The term white is borrowed from optics, where white light has been used to signify uniform energy distribution among the colors. Actually the analogy is not correct since in optics the uniform energy distribution of white light is based on wavelength rather than frequency. "Discrete white noise is simply an uncorrelated wide-sense stationary time series with zero mean. ARMA processes are derived by discrete-time white noise. Continuous-time white noise $X(t)$ statisfies formally $E(X(t)) = 0$ and $E[X(t)X(t+s)] = \sigma^2 \delta(t-s)$, where $\delta(.)$ is the Dirac delta function" (Canada, 1985, p.254).

Since an entirely flat spectral density distribution implies infinite power, white noise does not exist in practice. Nevertheless, the use of a white noise model is justified in many aspects. Many real data sets, such as aircraft flight test data radar return data, and passive conar detection data, involve an additive random noise that has large bandwidth compared to that of the signal. In some other problems, noise may be best described as a linear transformation of white noise More importantly, it is often much easier analytically and computationally to deal with white noise (Canada, 1985, p.254).

A stationary white noise process $\{a_t\}$has;

$$\gamma_k = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad \text{the autocovariance function}$$

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad \text{the autocorrelation function and}$$

$$\rho_{kk} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \quad \text{the partial autocorrelation function}$$

The ACF and PACF of a white noise process are shown in Figure 2.1 (Wei, 1990, p.16).

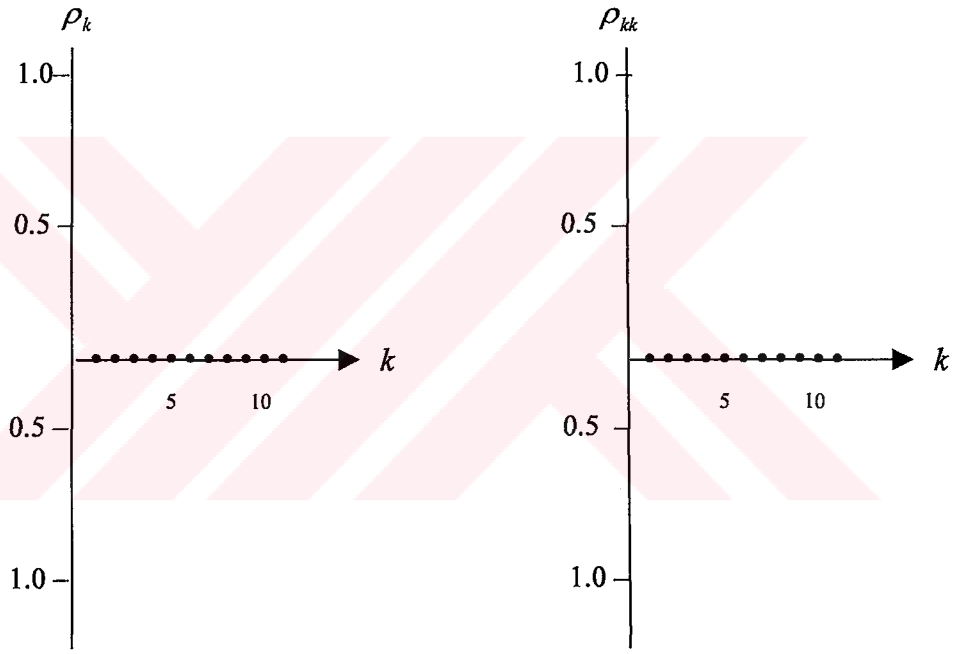


Figure 2.1. ACF and PACF of a white noise process

2.4.3 Autoregressive (AR) Process

When the value of a series at a current time period is a function of its immediately previous value plus some error, then underlying generating mechanism is called an autoregressive process. In this process, the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a shock a_t. Let $Z_t, Z_{t-1}, Z_{t-2}, \ldots$ denote the values of a process at equally spaced times

t, t-1, t-2,... This process is denoted by AR(p). Then, the nature of this relationship may be expressed as follows:

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + a_t \tag{2.5}$$

or

$$\phi_p(B) Z_t = a_t \tag{2.6}$$

where $\phi_p(B)=(1-\phi_1 B-\ldots-\phi_p B^p)$.

Since $\sum_{j=1}^{\infty} |\pi_j| = \sum_{j=1}^{p} |\phi_j| < \infty$, the process is always invertible. To be stationary, the roots of $\phi_p(B)$=0 must lie outside of the unit circle.

2.4.4 Moving Average (MA) Process

Moving average processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time. The moving average process is denoted as MA(q). It is given by

$$Z_t = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \tag{2.7}$$

or

$$Z_t = \theta(B) a_t \tag{2.8}$$

where $\theta_q(B)=(1-\theta_1 B-\ldots-\theta_q B^q)$.

Because $1+\theta_1^2 + \ldots + \theta_q^2 < \infty$, a finite moving average process is always stationary. This moving average process is invertible if the roots of $\theta(B)$ = 0 lie outside of the unit circle.

2.4.5 Autoregressive Moving Average (ARMA) Process

In general, a large number of parameters reduce efficiency in estimation. Thus, in model building, it may be necessary to include both autoregressive and moving average terms in a model. This leads to the following useful mixed autoregressive moving average (ARMA) process:

$$\phi_p(B)Z_t = \theta_q(B)a_t \qquad (2.9)$$

If the roots of $\theta_q(B)=0$ lie outside the unit circle, the process will be invertible. If the roots of $\phi_p(B)=0$ lie outside the unit circle, the process is stationary. Also, it is assumed that $\theta_q(B)=0$ and $\phi_p(B)=0$ share no common roots. This process is denoted by ARMA(p,q).

2.5 Nonstationary Stochastic Process

Many applied time series, particularly from economic and business areas, are nonstationary and in particular do not vary about a fixed mean. Although the general level about which fluctuations are occurring may be different at different times, the broad behaviour of the series, when differences in level are allowed for, may be similar. It is illustrated the construction of a very useful class of homogenous nonstationary time series models- the autoregressive integrated moving average (ARIMA) models.

2.5.1 Random Walk and Drift

The random walk plays a central role in all principal model-building procedures. At each point in time, the series moves randomly away from its current position. The stochastic processes were stationary. Although stationarity is a fundamental concept for the analysis of the time series, stationary models are clearly not appropriate for modeling a series. The simplest non-stationary process is the random walk

$$y_t = \phi y_{t-1} + \varepsilon_t \qquad (2.10)$$

which can be obtained by setting $\phi = 1$. It means the current observation is equal to the previous observation plus a random disturbance term. There is no parameter to estimate since ϕ is equal to one. The first difference of this model is stationary.

"Even if the series is a random walk, there is nothing to prevent fitting an AR(1) model by regressing y_t on y_{t-1}. The consequences of doing so are that the resulting estimator is biased downwards." (Harvey, 1990, p.29)

Nonstationarity is described by random walk, drift, trend, or changing variance. If each realization of the stochastic process appears to be random fluctuations, the series of movements will be a random walk. If the series exhibits sparse movements around a level before the end of the time horizon under consideration, it exhibits random walk plus drift. Drift is random variation around a nonzero mean.

2.5.2 The General Nonseasonal Autoregressive Integrated Moving Average Process

ARIMA models are associated primarily with Box and Jenkins (1976). They developed a model selection methodology based on identification, estimation and diagnostic checking. If the model fails the diagnostic checks entire cycle will be repeated until a satisfactory model is obtained. This methodology is essentially applied on modeling stationary processes.

The attraction of the ARMA(p,q) model is providing a restrictive representation of a stationary stochastic process. It may be extended to encompass a much wider class of nonstationary models by differencing. If the difference operator is applied d times before an ARMA(p,q) representation is appropriate, the variable is said to follow an autoregressive integrated moving average process of order (p,d,q). This is abbreviated as ARIMA(p,d,q).

Predictions for an ARIMA model are made by depicting it as a nonstationary ARMA model as in Equation (2.11) and then proceeding exactly as in the stationary case. The general stationary ARMA(p,q) process,

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)\, a_t \tag{2.11}$$

where the stationary AR operator $\phi_p(B)=(1-\phi_1 B-\ldots-\phi_p B^p)$ and the invertible MA operator $\theta_p(B)=(1-\theta_1 B-\ldots-\theta_q B^q)$ share no common factor. The parameter $-\theta_0$ plays very different roles for $d = 0$ and $d > 0$. If $d = 0$ original process is stationary and has an ARIMA$(p,0,q)$. If a series requires first, differencing to convert it stationary, then it is distributed ARIMA$(p,1,q)$. The series is then investigated for autoregressive or moving average components.

2.5.3 The General Seasonal Autoregressive Integrated Moving Average Models

Traditional methods presented in the previous section are based on assumptions that the seasonal component is deterministic and independent of other nonseasonal components. However, many time series are not so well behaved. More likely, the seasonal component may be stochastic and correlated with nonseasonal components. Seasonal model as well as regular ARIMA models has parameters that must meet the bound of stationarity and invertibility. The seasonal autoregressive models ARIMA $(p,d,q)(P,D,Q)_s$ need to be stationary for analysis. For stationarity to exist, both the regular and the seasonal autoregressive parameters need to lie within the bounds of stationarity. That is, $-1 < \varphi_p, \Phi_s < +1$.

The bounds of invertibility similarly must hold for multiplicative seasonal moving average models. Hence, the series

$$Z_t = (1-\theta_1 B)(1-\Theta_{12} B^{12})e_t \tag{2.12}$$

would have to possess regular and seasonal parameters that lie within the same bounds of invertibility ($|\theta_1|, |\theta_s| < +1$) for the mixed seasonal moving average model to be stationary. If the moving average parameters were confined to this range, the product of these factors would also be confined to these bounds.

CHAPTER THREE
INTERVENTION ANALYSIS IN TIME SERIES

3.1 Intervention in Time Series

Policy changes or sudden decisions in public and private sectors are result in effect some changes in certain response variables occurring in the form of time series. Such intrusions to a time series are usually referred to as interventions. Examples of specific events of such intervention effects are given in the real life as follows:

- the effect on inflation of the creation of the Canadian Anti-inflation Board in October 1975,
- the impact on the number of the traffic fatalities of introducing the 55 miles-per-hour speed limit in the United States in 1974,
- the influence on sales of a change in advertising strategy,
- the impact of the Arab oil embargo,
- the impact of air pollution control,
- the impact of the politics on economy.

Interventions can affect the response variable in several ways. They can not only change the level of a series abruptly or after a short delay but also deflect a series going downward, causing it to drift up, or effect some other form of change. "When the intervention occurred, it must be determined whether there is evidence to suggest that a corresponding change has occurred in the time series and if so, in determining the nature and magnitude of this change." (Abraham& Ledolter, 1986, p.355)

3.2 Description of Intervention Variables

An intervention variable is active only for limited period over the whole length of data record being analyzed, and is used to describe either exceptional or abnormal events or variables with infrequent movements. "An intervention variable may only be used as independent variable." (McLead, 1983, p.10-5)

In the setting of intervention analysis, it is assumed that an intervention event has occurred at a known point in time T of a time series. It concerns to determine whether there is any evidence of a change or effect, of an expected kind, on the time series Y_t under study associated with the event.

It is considered that transfer function models are used for estimating the magnitude of the effects of intervention, modeling the nature of intervention and possible abnormal behaviour in associated time series. Box and Tiao (1975) have provided a strategy for modeling the effect of interventions. They consider transfer function-noise models of the form as following;

$$Y_t = v(B)\xi_t + N_t = \frac{\omega(B)B^b}{\delta(B)}\xi_t + N_t \qquad (3.1)$$

where the term $v(B)\,\xi_t$ represents the effects of the intervention event in terms of the deterministic input series ξ_t, and N_t is the noise series which represents the background observed series Y_t without the intervention effects in equation (3.1). "It is assumed that N_t follows an ARIMA(p,d,q) model, $\varphi(B)N_t = \theta(B)a_t$, with $\varphi(B) = \phi(B)(1-B)^d$. Multiplicative seasonal autoregressive integrated moving-average (ARIMA) time series models can also be included for N_t." (Box etal., 1994, p.463)

"The intervention variables ξ_t in equation (3.1) are usually taken as indicator variables indicating the intervention at t=T" (Canada, 1985, p.209). There are two common types of deterministic input variables ξ_t that have been found useful to represent the impact of intervention events on a time series. Both of these are indicator variables taking only the values 0 and 1 to denote the nonoccurrence and occurrence of intervention respectively.

One type represents an intervention occurring in time T that remains in effect thereafter. That is, the intervention is a step function,

$$S_t^{(T)} = \begin{cases} 0 & t < T, \\ 1 & t \geq T. \end{cases} \qquad (3.2)$$

$S_t^{(T)}$ is usually referred to as a step input and this denotes the nonoccurrence and occurrence of interventions. It is typically used to represent the effects of an intervention that are expected to remain permanently after time T to some extent.

The other one type represents an intervention taking place at only one time period. Thus, it is a pulse function,

$$P_t^{(T)} = \begin{cases} 1 & t = T, \\ 0 & t \neq T. \end{cases} \qquad (3.3)$$

The pulse input $P_t^{(T)}$ takes the value 1 at the time of intervention and zeros elsewhere. It represents the effects of an intervention that are temporary or transient and will die out after time T. (Box etal, 1994, p.463)

The pulse function can be produced by differencing the step function $S_t^{(T)}$ in Equation (3.4). That is,

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)} = (1-B)S_t^{(T)} \qquad (3.4)$$

"Therefore, an intervention model can be represented equally well with the step function or with the pulse function. The use of a specific form is usually based on the convenience of interpretation" (Wei, 1990, p.185).

These indicator input variables are used in many situations where the effects of the intervention cannot be represented as the response to a qualitative variable, because such a qualitative variable does not exist or it is impractical or impossible to obtain measurements on such a variable.

"Model 3.1 is really a special case of the transfer function model with the input variables are indicator variables. It can also be seen as a distributed lag model discussed extensively in the econometrics literature with the exogenous variables taken as indicator variables" (Canada, 1985, p.209). The transfer function *v(B)* is usually estimated from data in the transfer function modeling, whereas in the intervention analysis it is postulated on the basis of the change expected. "Because of the deterministic nature of the indicator input series ξ_t in equation (3.1), identification of the structure of the intervention model operator *v(B)* cannot be based on the technique of prewhitening" (Box etal., 1994, p.463). Instead, it is desirable to postulate the form of the intervention model through consideration of the mechanisms that might cause the change or effect and the implied form of the change that would be expected. "In addition, the identification may be aided by direct inspection of the data to suggest the form of effect due to the known event, and supplementary evidence may sometimes be available from examination of the residuals from a model fitted before the intervention term is introduced"(Box etal., 1994, p.463).

Several different response patterns are possible through different choices of the transfer function. Table 3.1 and Table 3.2 show the responses for various simple transfer functions with both step and pulse indicators as input.

Table 3.1. Forms of Intervention Responses to Step Input

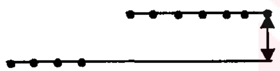
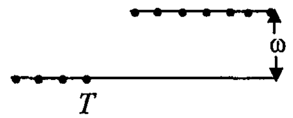
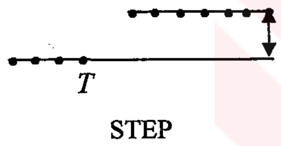
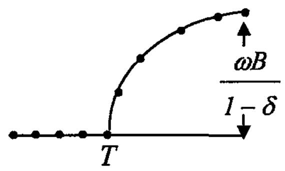
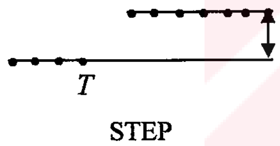
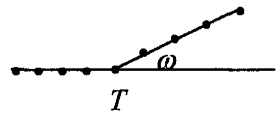
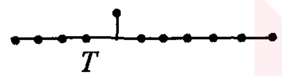
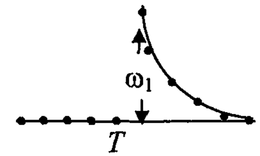
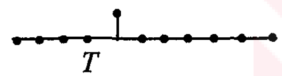
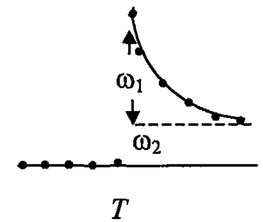
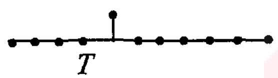
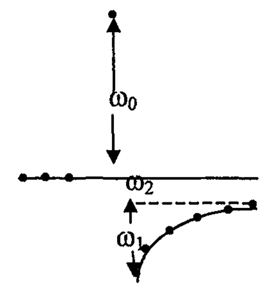
Types of Response	Input	Response Function	Output
A permanent step change in level of unknown magnitude ω after time T		$Y_t = \omega B S_t^{(T)} \quad \omega > 0$	
A gradual change with rate δ that eventually approaches the long-run change in level equal to $\frac{\omega}{(1-\delta)}$		$Y_t = \frac{\omega B}{1-\delta B} S_t^{(T)} \quad \omega, \delta > 0$	
Linearly changing without limit		$Y_t = \frac{\omega B}{1-B} S_t^{(T)} \quad \omega > 0$	

Table 3.2. Forms of Intervention Responses to Pulse Input

Types of Response	Input	Response Function	Output
Sudden change after time *T* of unknown ω_1, followed by a gradual decay of rate δ back to the original preintervention level with no permanent effect	 PULSE	$Y_t = \frac{\omega_1 B}{1-\delta B} P_t^{(T)}$	
Abrupt start and gradual decay to a permanent level	 PULSE	$Y_t = \left\{ \frac{\omega_1 B}{1-\delta B} + \frac{\omega_2 B}{1-B} \right\} P_t^{(T)}$	
More complex forms of response can be obtained by various linear combinations of the simpler forms	 PULSE	$Y_t = \left\{ \omega_0 + \frac{\omega_1 B}{1-\delta B} + \frac{\omega_2 B}{1-B} \right\} P_t^{(T)}$	

Different combinations of step and pulse inputs can produce various responses. For example, the response function may be as following in Figure 3.1.

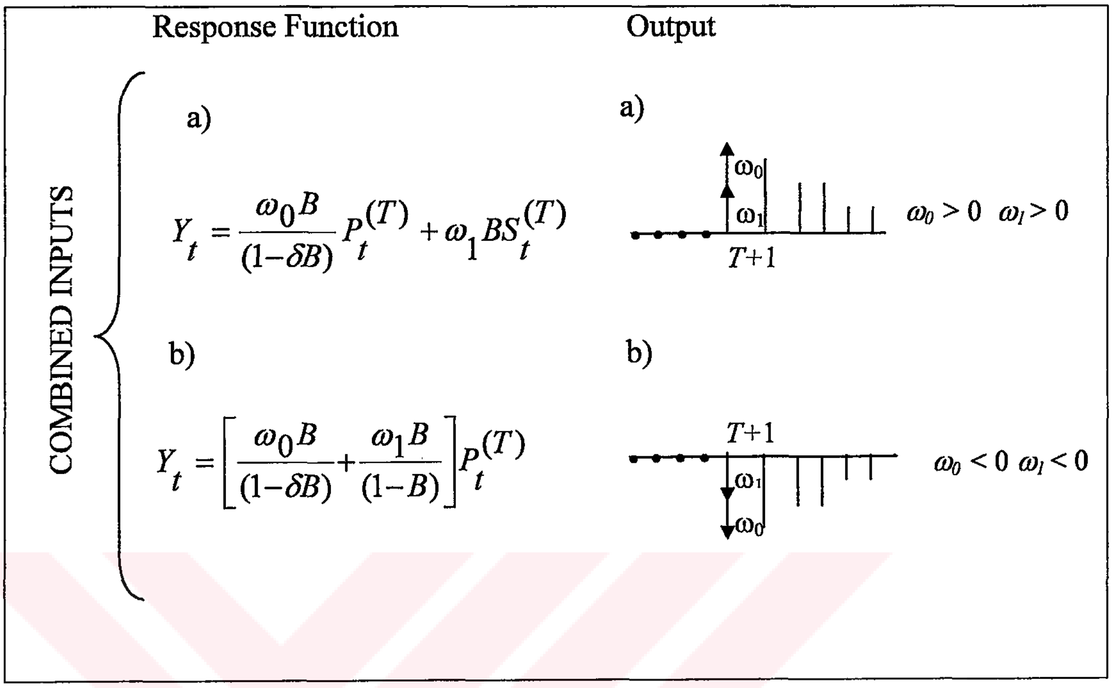


Figure 3.1.Response Functions

The model that is shown in Figure 3.1.a or Figure 3.1.b is useful to represent the phenomenon in which an intervention response process that tapers off gradually but leaves a permanent residue effect in the system. The impact of an intervention such as advertising on sales can be represented as shown in Figure 3.1 (a), and the effect of a price or a tax increase on imports may be represented in Figure 3.1. (b).

More generally, a response may be represented as a rational function $\frac{\omega(B)B^b}{\delta(B)}$ *where* $\omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$ *and* $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r$ *are polynomials in B, b is the time delay for the intervention effect, and the weights* ω_j*'s in the polynomial* $\omega(B)$ *often represent the expected initial effects of the intervention. The polynomial* $\delta(B)$*, on the other hand, measures the behaviour of the permanent effect of the intervention. The roots of* $\delta(B) = 0$ *are assumed to be on or outside the unit circle. The unit root represents an impact that increases linearly, and* *the root outside the unit circle represents a phenomenon that has a gradual response.* (Wei, 1990, p.187)

The following additional points concerning the intervention models are worthy of note. "The function Y_t represents the additional effect of the intervention event over the noise or "background" series N_t. Hence when possible, the model $N_t = [\theta(B)/\varphi(B)]a_t$ for the noise is identified based on the usual procedures applied to the time series observations available before the date of the intervention, that is, Y_t, $t < T$" (Box etal., 1994, p.465). Also, it is assumed in Model (3.1) that only the level of the series is affected by the intervention and, in particular, that the form and the parameters of the time series model for N_t are the same before and after the intervention. "One should also recognize that there can be considerable differences in the accuracy with which the intervention model parameters can be estimated depending on whether the noise N_t is stationary or nonstationary, as well as on whether permanent or transitory effects are postulated" (Box etal., 1994, p.465).

"For multiple intervention inputs, the general class of models is as follows:

$$Y_t = \sum_{j=1}^{k} \frac{\omega_j(B)B^{b_j}}{\delta_j(B)}\xi_{jt} + \frac{\theta(B)}{\psi(B)}a_t \tag{3.5}$$

where ξ_{jt}, $j = 1, 2, \ldots, k$ are intervention variables. These intervention variables can be either step or pulse functions" (Wei, 1990, p.187). Moreover, the parameter estimates and their standard errors for the intervention model are obtained by the least squares method of estimation for transfer function-noise.

More generally, they can be proper indicator variables. The form $\omega_j(B)B^{b_j} / \delta_j(B)$ *for the jth intervention is postulated based on the expected form of the response given knowledge of the intervention. The main purpose of the models is to measure the effect of the interventions. Thus, with respect to the intervention variables* ξ_{jt}, *the time series free of intervention is called the noise series and denoted* *by N_t, and its model is hence known as the noise model. The noise model $[\theta(B)/\psi(B)]a_t$ is usually identified using the univariate model identification procedure based on the time series Z_t before the date of intervention. If diagnostic checking of the model reveals no model inadequacy, then it can make appropriate inferences about the intervention. Otherwise, appropriate modifications must be made to model, and estimation and diagnostic checking repeated.* (Wei, 1990, p.187)

3.3 Procedure for Building an Intervention Model

In some situations, it is known that exceptional external events have affected the variables being forecast. Such exceptional external events involve some temporary inducement to change other variables and sometimes it is difficult to quantify these effects. "The effect of an exceptional event, such as a strike, or holiday, may be to produce one or more large residuals in the univariate model. Such large residuals may be a distorting influence on the structure of the model tentatively entertained at the identification stage, the values of estimated parameters and the magnitude of the residual variance"(McLead, 1983, p.9-8). For a particular time (month, year, etc.), these external activities took place are characterized by a dummy variable in the form of an impulse of unit height and in the same particular time states which there is no external event are characterized by zeros. Some other series can be related to the external series by a transfer function model. This transfer function model can be used to explore various hypotheses concerning the dynamic (or lagged) relationship between two variables. Such dynamic models, involving the dummy variables as independent variables, are called intervention models. Practical applications of intervention analysis have included the following:

- The effect of different kinds of promotional activity on sales,
- The effect of changes in politics or legislation (represented by a step function consisting of zeros before the policy change and ones after the change) on business and time series in economy.

An intervention analysis, performed by the user's request, may be employed to avoid extreme values that would influence the next steps of the analysis:

specification, estimation, test for adequacy, and forecasting (Mélard,&Pasteels 2000, p. 501). In the intervention analysis, the input series will be in the form of a simple pulse or step indicator function to indicate the presence or absence of the event. Initially, it will be assumed that the timing of the intervention event is known. "The method is then generalized to study the impact of the events when the timing of interventions is unknown and hence leads to the general time series outlier analysis" (Wei, 1990, p.188).

Traditionally a Student's t-test is used for estimating and testing for a change in the mean levels before and after intervention. "Such a test may not be adequate when the data occur in the form of a time series. This is because (a) in these cases the successive observations are usually serially correlated and often nonstationary, and (b) the form of the change may not be a step as required by the t-test but it could be a gradual increase (decrease), a ramp increase (decrease), or any other form of the change"(Box&Tiao, 1975, p.70).

When the intervention variable is being used to describe an exceptional or abnormal event, the user has to formulate the variable from his/her knowledge of the event. The formulation must produce an intervention variable, which is consistent with the dependent series to which it is to be related.

The length of an intervention variable is conditioned by the dependent series to which it is to be related. Because, in general, it is not possible to back forecast an intervention variable without knowledge of the mechanism used for generation, then it is necessary to supply back forecasts for any intervention variables for which the generating mechanism are not known. In all cases, 21 back forecasts must be provided.

Thus, an intervention variable must always have 21 more observations than its corresponding dependent variable. It is assumed that the first 21 observations are backforecasts. When an intervention variable is used for forecasting, both backwards *and forward forecasts must be provided. The number of backward and forward forecasts required is*

•backward forecasts = 21

•forward forecasts = forecast lead time (LT)

Thus, an intervention variable used for forecasting must always have 21 + LT more observations than its corresponding dependent variable. It is assumed that the first 21 observations are the backforecasts, and the last LT observations are the forward forecasts. (McLead, 1983, p.10-5)

Building the intervention model is the iterative process illustrated as Figure 3.2. The data requirements and the situations in which continuous and intervention variables can be used are summarized in Table 3.3. (McLead, 1983, 10-7)

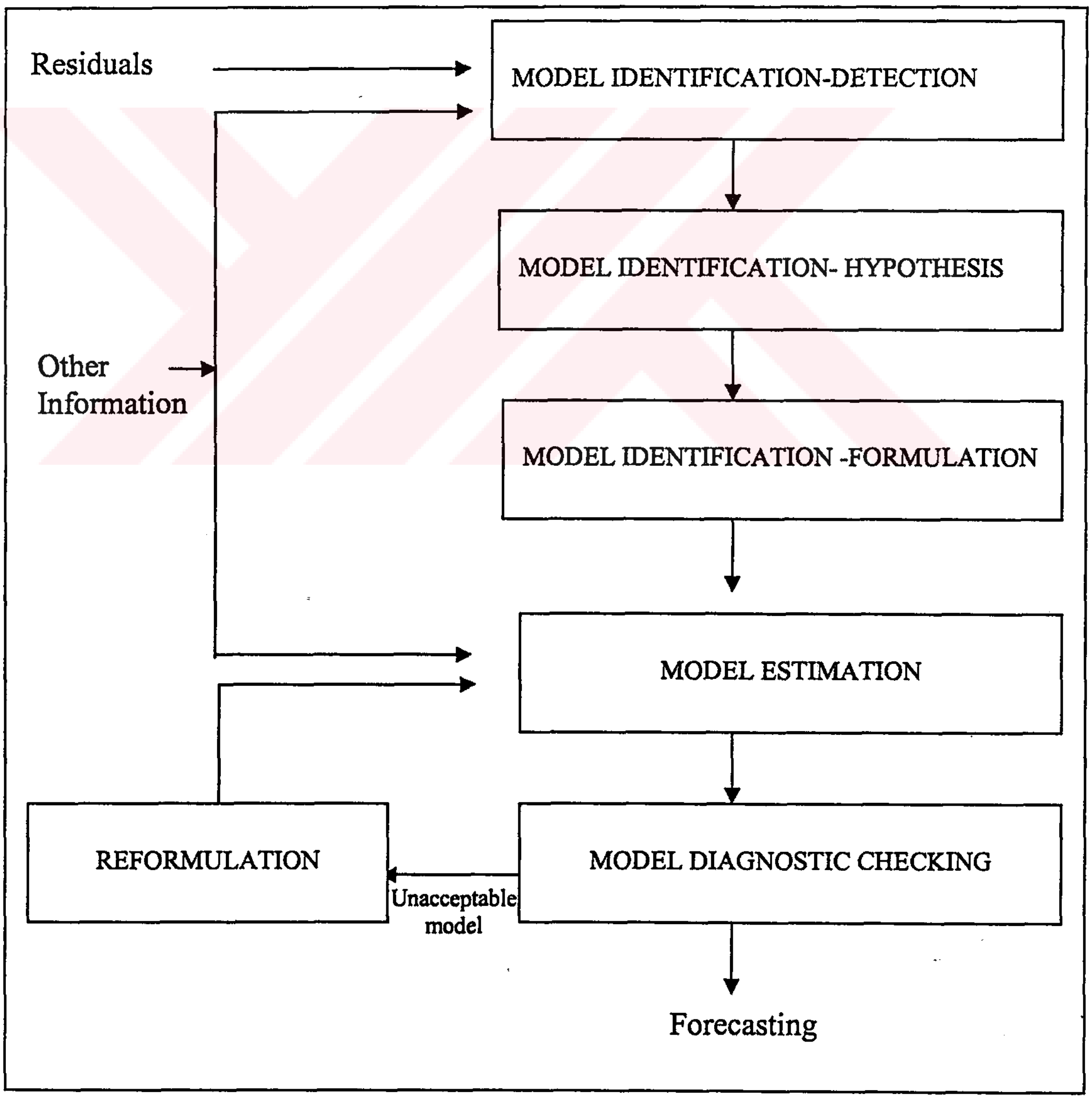
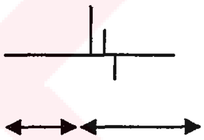
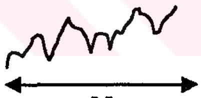
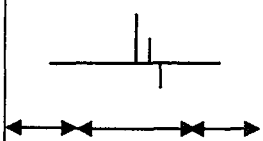


Figure 3.2. Stages in the Iterative Approach to Model Building

Table 3.3. Data Requirements for Transfer Function/Intervention–Noise Models

Purpose	Variable Description	Variable Type Allowed	Series Length	Number of Data
IDENTIFICATION	Dependent (Y_t) and Independent (X_t)	Continuous		N
ESTIMATION AND CHECKING	Dependent (Y_t)	Continuous		N
	Independent (X_t)	Continuous		N
	Independent (ξ_t)	Intervention		$N+21$
FORECASTING	Dependent (Y_t)	Continuous	 	N $N + LT$
	Independent (ξ_t)	Intervention		$N+LT+21$

N: Number of data points in dependent variable, *LT*: Forecast lead-time (McLead, 1983, p.13-8)

Iterative Process of the Intervention Analysis

An overall strategy for performing an intervention analysis is outlined in the following steps.

1. *Using the preintervention data, build a time series model for Z_t adopting the iterative strategy of specification, estimation and diagnostic checks.*
2. *Given knowledge of the interventions, frame models for change that describe what is expected to occur.*
3. *Estimate the parameters of the joint model, perform diagnostic checks, modify the structure if necessary, and arrive at a final model.*
4. *Make appropriate inferences about parameters* (Canada, 1985, p.211).

Assumptions of the Intervention Analysis

It should be emphasized that the analysis described assumes that

1. *the time series model and its parameters before and after the intervention are the same, and*
2. *there are no other events or interventions coinciding with the particular one being considered.*

It is important to keep these assumptions in mind when drawing conclusions about any intervention (Canada, 1985, p.212).

3.3.1 Model Identification

Model building for a process containing an intervention cannot proceed by pre-whitening since the input variable is predetermined and the shock occurs once. The investigator must consider whether the effect is likely to be permanent or temporary, whether the onset will be subject to a delay and whether the impact will be sudden or gradual. As in other cases, the choice is not irrevocable, but a good initial specification will help to ensure convergence of the estimation procedure.

Using ξ_t to denote P_t or S_t as appropriate and *v(B)* to denote its impulse function, the model becomes

$$Y_t = v(B)\xi_t + \frac{\theta(B)}{\phi(B)}\varepsilon_t \qquad (3.6)$$

The next step in model identification is the usual one of selecting the ARIMA scheme and the available tools.

A problem arises in that a major jump caused by ξ_t would distort estimates of the autocorrelations. One of four possible approaches may be used, depending on the particular series:

a) *use only observations prior to the intervention or, less commonly, only some time after the intervention;*
b) *use observations both before and after but exclude all pairs using observations in the interval $[t_0, t_0+k]$, in which the effects of the intervention are deemed to have worked themselves out;*
c) *estimate the residuals ignoring any error structure, and then plot the sample autocorrelation functions;*

$$e_t = y_t - \hat{v}(B)\xi_t \qquad (3.7)$$

d) *use the whole series, ignoring the effects of the intervention.*

Each method has its attractions, depending upon the behaviour of the phenomenon under consideration; use of (c) is advised whenever the effects of the intervention are large and fairly sudden, whereas (b) is probably best for rather short series (Kendall&Ord, 1992, p.227).

3.3.1.1 Model Identification-Intervention Detection

In most cases, an exceptional event is detected from the presence of one or more different from normal or usual residuals, which may occur after fitting a univarite stochastic or transfer function-noise model.

When data set is fitting any univariate or transfer function-noise model, it is important to examine the residual series values. There are two points to examine the residual series. The latter is whether the residuals would be expected on the assumption of a normal distribution and whether any of these residuals are associated with a known exceptional event. The former is direction of behaviours of the residual series. "Especially it is looked forward that whether any groups or runs of residuals, which appears to behave in an untypical way, associated with a known exceptional event. After the examination of the residuals against a quantified test, the residuals are normally tested against twice their standard deviation." (McLead, 1983, p. 13-5)

There is a class of problem requiring the use of intervention analysis, which does not involve the detection step. Typical of this type of problems are situations in which price and advertising effects are being investigated and in which the value of the series moves only occasionally. This type of problem is one in which a priori knowledge would suggest a relationship should exist, but where it is by no means certain that the effect of movements in the intervention variable is going to be reflected in abnormally large residuals, or groups or runs of large residuals.

3.3.1.2 Model Identification- Hypothesis

The objective of developing an intervention model is obtain a quantitative measure of the effect of an exceptional event. A necessary intermediate step in obtaining this quantitative measurement is a qualitative description of the intervention.

The qualitative description draws together all the known facts about the exceptional events and its effects, and is aimed in particular at describing the nature of the exceptional events and the possible range of effects that have been induced in the series being analysed.

The hypothesis is based almost exclusively on the model builders' knowledge of the situation and only relies on the time series data being analysed at the point at which the question is asked; "Does the hypothesis make sense in the light of the data?" It is not unusual at this point for a hypothesis to be rejected because it is so obviously at variance with the behaviour of the data." (McLead, 1983, p.13-5)

For the exceptional event, there are two crucial points. One important point is that whether the event occurs at one point in time or permanent. For example, a strike normally only occurs at one point in time and is permanent. As another example, a change in the law only occurs at one point in time but is permanent. Another important point is that if the event occurs at more than one point in time, it is crucial whether all the individual instances can be considered simultaneously or they have to be considered separately. And also whether the event can be quantified in any way or not is another important point.

Investigator also has to be determining the effect of the exceptional event. He has to decide that the effect of the exceptional event is transitory or permanent. In addition, if the effect is transitory over, it is important that what period the effect lasts. For transitory or permanent effects, it must be determined that what the form and direction of the effect is.

3.3.1.3 Model Identification – Formulation

Formulation involves the translation of a qualitative description of the intervention into a form in which it can be quantified. The formulation involves three distinct steps:

1. Quantification of intervention variable.
2. Designing an intervention mechanism which, when it operates on the intervention variable, will produce the desired effect on the series being analysed.
3. Incorporating into the intervention model the features necessary to describe the behaviour of the major part of the time series which is unaffected by the intervention variable.

a. Quantification of Intervention Variable

The quantification of an intervention variable divides itself into two parts as formulating a structure and deciding upon a precise numerical representation. An intervention variable may have any of a wide range of structures. In practice two structures pulse and step are commonly used.

The way in which a precise numerical representation can be assigned to the intervention variable will depend upon which of two classes the variables fall into:

CLASS 1: which contains variables, which cannot have a meaningful scale of measurement assigned to them e.g. strikes, changes in a law, etc.

CLASS 2 : which contains variables which are measured on a meaningful scale e.g. price, promotional activity.

In the case of class 2 variables some scale of measurement is usually available and it is a relatively simple matter to decide on a precise scale of measurement given that the structure will already have been formulated.

b. Design of Intervention Mechanism

To be able to design an intervention mechanism, first it is necessary to understand the response characteristics of the autoregregressive and moving average operators in a transfer function.

The design of the intervention mechanism involves the selection of a transfer function, which contains the appropriate autoregressive and moving average parameters, to ensure that the formulated intervention variable will produce the desired effect in the series being analysed. In this context the nonseasonal differencing operator may be used as a special case of an autoregressive operator with a δ value of 1.

c. Noise Structure

The formulation of an intervention (variable and mechanism) is designed only to describe the anomalous behaviour of a time series over a limited period of time. To complete the description of the time series it is necessary to formulate a structure for the noise N_t. This structure is obtained by univariate stochastic or transfer function-noise model that has already been developed to describe the time series.

"The univariate stochastic or transfer function-noise model that is used to determine the structure is normally the model, which was used to originally detect the presence of the exceptional event". (McLead, 1983, p.13-18)

It is not always the case that the model used to originally detect the presence of an exceptional event is used to formulate the noise structure. The reason for this is that the exceptional event may be distorting the series to such an extent that the true model structure is hidden, and a model developed on the whole of the series may be inadequate representation. If it is believed that this situation exists, and sufficient data is available, a univariate or transfer function-noise model would be developed on the longest section of the series which was unaffected by the exceptional event this model would then be used to formulate the noise structure.

In general, if the noise is at the same structure, it is assumed that the univariate stochastic model, developed to describe the time series or the noise part of a transfer function-noise model developed to describe the time series. In both situations these models may have been developed over a length of the time series, which may or may not have included the effect of the exceptional event.

There are two fundamental advantages of formulating the noise structure in this way:

1. *The univariate or transfer function-noise model from which the noise is derived represents the best understanding that will be available of the time*

series and thus provides the most adequate means available to identify the noise structure.

2. *In the event of the intervention mechanism, failing to describe the effect of the exceptional event, or having parameters which are not significant, then, the intervention-noise model degenerates the univariate stochastic (or transfer function-noise) model.*(McLead, 1983, p.13-18)

In the case in which a transfer function-noise model is being used by a means of detection and formulating the noise structure, then the fact that the model includes the effect of specific independent variables must also be taken into consideration.

The introduction of these independent variables may have had one of three effects on the residuals obtained from the univariate stochastic model of the dependent series.

1. *To have no effect on the abnormal residuals.*
2. *To reduce, but not completely remove, abnormal residuals.*
3. *To induce abnormal residuals.* (McLead, 1983, p. 13-18)

If situations (1) or (2) have occurred, then the correct formulation of the intervention model must take into account the effect of the specific independent variables. This is done by also introducing into the intervention model the transfer functions, which have previously been established for the specific independent variables.

If situation (3) has occurred, an intervention model should not being developed to describe the abnormal residuals associated with the dependent series. The correct course of action should be to determine what abnormal behaviour exists in the independent variables- and to correct this abnormal behaviour.

3.3.2 Model Estimation

"The procedure for estimating the parameters in an intervention model is identical to that of a transfer function containing a continuous variable. Computationally intervention variables and ordinary transfer function variables are handled in a similar way"(McLead, 1983, p.13-22).

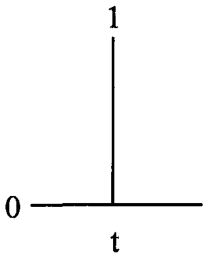
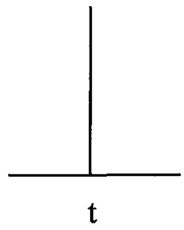
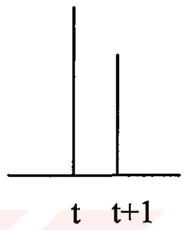
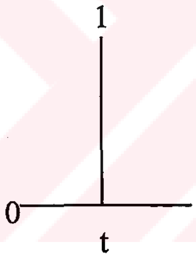
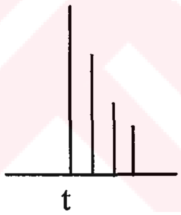
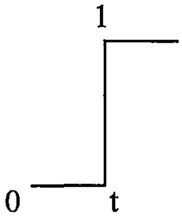
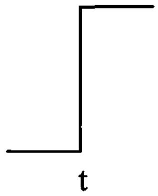
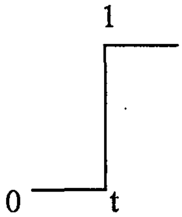
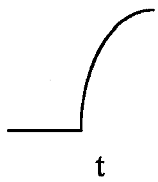
As with a transfer function-noise model the estimation process is iterative. To reduce the amount of computational effort required, and to minimize the risk of encountering a local minimum in the sum of squares surface, it is desirable to provide suitable preliminary estimates where possible.

Preliminary estimates may be required for three purposes.

1. *For the noise, preliminary estimates may be obtained from either the univariate stochastic model, or the noise part of the transfer function-noise model.*
2. *For the transfer function, in those cases in which a continuous variable (X_t) and an intervention variable (ξ_t) are introduced into the model simultaneously it is possible to obtain appropriate preliminary estimates for the transfer function for the continuous variable from the previously fitted transfer function-noise single input model.*
3. *For the intervention, the intervention mechanism presents some difficulty concerning preliminary estimates, as there is no previous model building from which preliminary estimates may be derived. It is, in fact, necessary to consider each situation in deriving preliminary estimates. Examples of some typical situations are given in Table 3.4.* (McLead, 1983, p.13-23)

In absence of any preliminary estimates, suggested values 0.1 is used.

Table 3.4. Preliminary Estimates for Intervention Mechanisms

CASE	VARIABLE	INTERVENTION	MECHANISM	STARTING VALUES
1		ω_0		Set ω_0 equal to <u>minus</u> the residual series value at time *t*.
2		$(\omega_0 - \omega_1 B)$		Set ω_0 equal to <u>minus</u> the residual series value at time *t* and ω_1 equal to the residual series value at time *t*+1.
3		$\frac{\omega_0}{(1-\delta_1 B)}$		Set ω_0 equal to <u>minus</u> the residual series value at time *t* and δ_1 equal to 0.2.
4		ω_0		As Case 1
5		$\frac{\omega_0}{(1-\delta_1 B)}$		As Case 2

(McLead, 1983, p.13-2)

Residuals Regeneration

The estimation process attempts to determine the parameter values, which minimize the sum of squares of the residuals (the a_t's). For this purpose, it is necessary to generate the residuals, using the identified model structure and the preliminary or current parameter values.

Iterative Output

The iterative nature of the estimation process enables parameter estimates to be obtained at each iteration; it is recommended that these intermediate estimates are examined in two respects in situations where convergent estimates are not obtained.

In situations in which the identified model is over parameterised (and the parameter estimates are highly correlated), the parameter values will oscillate between successive iterations and convergent parameter values may be impossible to obtain. An inspection of the iterative output will reveal this situation, which will be subsequently confirmed during model checking.

The maximum number of iterations allowed for the estimation of the model parameters may be insufficient to obtain convergent values. If convergent estimates are not obtained an examination of the iterative output will;

a) confirm that the parameter estimates are moving towards convergent values

b) indicate if the movement in the parameter estimates is still sufficiently large to warrant restarting the estimation process.

Convergence Criteria

While the objective of parameter estimation is to determine the model parameter values which minimize the sum of squares of the a_t's, the criterion used to stop the estimation process is based on the movement in the parameter values at successive iterations. The estimation algorithm stops the estimation process when all model parameter values satisfy the relationship;

$$\frac{\text{Parameter value at iteration } i \text{ - parameter value at iteration } i\text{ - 1}}{\text{Parameter value at iteration } i} < \text{Tolerance}$$

A typical value used for tolerance would be 0.00001. The only supplementary comment that needs be made is that intervention variables have to be handled in a slightly different way in respect to residual regeneration.

3.3.3 Model Diagnostic Checking

Using the formulated intervention model and preliminary estimates, the model is estimated in the normal manner, allowing sufficient iterations to achieve convergent parameter estimates.

At the end of the estimation process the following information should be obtained for model checking purposes:

- Parameter estimates and their standard errors
- Transfer function/intervention mechanism gains and their standard errors
- Estimated residual variance
- Model in factored form, with the period, frequency and damping factors of any complex operators
- Correlation matrix of the parameter estimates
- Residual series values and their standard deviation
- Histogram of the residual series values
- Table of anomalous residuals
- Auto- and partial-autocorrelation function of the residual series values and associated chi squared statistic
- Cross-correlation function between the residuals and any continuous variables, but not intervention variables.

3.3.3.1 Model Checking Strategy

In the checking process, the diagnostic checks produced at the end of the estimation must be answer to four questions:

- Does the intervention adequately describe the effect of the exceptional event?
- Which is the most appropriate intervention in those cases in which more than one mechanism has been formulated?
- Has the structure of the noise been changed?
- Has the structure of any of the transfer functions describing the effect of the continuous variables been changed?

The answer to these questions will lead to either an acceptance of the fitted model, or the provision of information that will enable all or part of the model to be reformulated. In the discussion that follows the function of each of the diagnostic checks are examined.

3.3.3.2 Parameter Estimates and Standard Errors

The parameter estimates and their standard errors are used to decide which parameters are to be retained in the model, and where the parameters are considered unnecessary to point towards a simplification of the structure in particular parts of the model. In addition, they are used to decide whether the sign and magnitude of the estimated effects make sense.

“In particular case of the intervention mechanism, it would be expected that if the mechanism had been formulated correctly, all the parameters would be of sufficient statistical significance to justify their retention in the model, and the magnitude and sign of the parameters would confirm the hypothesis that had been formulated” (McLead, 1983, p.13-25).

The evidence produced by these tests can point to number inadequacies in the model in general, and the following inadequacies in the intervention mechanism in particular.

3.3.3.3 The Gain and Its Standard Error

The gain and its standard error measure the overall or net effect of a continuous or intervention variable and are used to decide whether the overall effect makes sense and is significant.

In particular case of intervention mechanism, it would be expected that the gain could be directly related to the description of the exceptional event that had been formulated.

By correct formulation of the intervention variable and mechanism it is possible in particular cases to impose overall constraints in effect to perform a constrained estimation.

3.3.3.4 Correlation Matrix of Parameter Estimates

The correlation matrix of the parameter estimates provides the means by which it can be determined whether the parameters describing the intervention mechanism are estimated independently of the parameters describing the independent series transfer function and the noise or the parameters describing the intervention mechanism are estimated independently of each other.

"For well-structured models, it is desirable that the correlation between the parameter estimates is as small as possible, and preferably zero." (McLead, 1983, pp. 13-27,13-28)

In practice, the correlation between the estimates is seldom zero, and can be high (greater than 0.6) if part of the model is badly structured. In situations in which the correlation between parameters is high, the particular correlations provide the evidence that is necessary to decide how the model structure should be simplified.

High parameter correlation can occur either between the parameters describing the intervention mechanism in which case a simplification of the intervention mechanism is indicated and/or between the parameters describing the transfer function, the intervention and the noise –in which case either a simplification of one part of the model may be required, or complete removal of one part of the model may be necessary.

Any simplification suggested by the correlation matrix should be checked for its descriptive implications. This is particularly important if the simplification of second or higher order autoregressive operators with complex roots suggested. Such operators are describing cyclical behaviour, and this description will be lost if a second order operator is simplified to a first order operator, or may be lost if the order of a higher order operator is reduced. The case that high correlation occurs between parameters in different parts of the model is seldom.

3.3.3.5 Model in Factored Form

It is important that the models that are developed are inherently stable. The principles on which the Box-Jenkins approach is based specify very precise stability criteria.

For any model that is developed it is of prime importance that basic checks are made concerning model stability. To perform these checks, the model is presented in factored form. What is meant by this is that in those cases which an operator is of second or higher order should be factorized into real or complex factors. Each of these factors should then be examined to see if they violate the basic stability criteria.

The factors of any operator may be real and/or complex. Whether the factors are real or complex they should be tested to ensure that they lie within the acceptable region. If factors are found which lie outside the acceptable region the model will be unstable and the forecast function will be explosive.

In these circumstances, the model structure must be modified to remove the instability. In addition to providing information on stability, the factored form of the model provides which will enable a better understanding of the model to be obtained.

The second use of the factors is to suggest modifications to basic transfer function or intervention mechanism, and/or to the noise structure.

In intervention models, the factors are used in a similar way to that in which they are used in a transfer function. Two particular characteristics should be looked for in particular cancellation of autoregressive and moving average operators, and alterations of basic mechanisms.

If the structure of the intervention mechanism containing both autoregressive and moving average operators are too complex, there may be factors in the operator which cancel, and which would point towards simplification.

If the operators describing the intervention mechanism have factors close to (1-1.0B) and (1-0.0B), this can suggest alterations in the basic mechanism between the dependent and intervention variables.

3.3.3.6 Estimated Residual Variance

The estimated residual variance provides an absolute measure by determining the performance of the model by means of comparing two or more models. This test does not provide means by which the structure of a model may be simplified or elaborated.

3.3.3.7 Model Residuals and Histogram of Residuals

The last step of the model identification is examining the residuals. At the end of the step, it must be found answer to the following questions:

1. Are these residuals which are significant at a particular level, more important than that of occurred by chance?
2. Have all the symptoms of the effect of an exceptional event been removed?
3. Are the residuals approximately normally distributed?
4. Do the residuals have constant variance over time?

It is important to ask and answer all these questions. Question 1 and 2 are of particular importance in intervention analysis.

Significant Residuals

The presence of a significant residual or a group of residuals was the primary way for the need for intervention analysis was originally detected. After removing the effect of a particular exceptional event, the presence of further exceptional events may be revealed. The individual or groups of large residuals should, therefore, be viewed in the same way in which they were viewed at the detection stage.

Removal of Symptoms

If the intervention has been correctly formulated, its effect will be to reduce the size of the anomalous residuals identified at the detection stage to acceptable values not to induce further anomalous residuals immediately before or after the intervention is active. The failure in removal of the originally identified residuals or to creating new anomalous residuals is a clear indication that the intervention has not been correctly formulated.

3.3.3.8 Autocorrelation and Partialautocorrelation Function of the Residuals

The removal of the effect of the exceptional event can reveal structure in the noise, which was previously masked. The reason for calculating the autocorrelation function and partial autocorrelation function of the residuals is to check that there is no evidence that any new structure should be added to the noise, and if there is any evidence of the necessity for additional structure to decide what it should be.

The principles and actions involved here are identical to those employed in the development of a univariate stochastic model, or the noise part of a transfer function-noise model.

3.3.3.9. Cross-Correlations Between Continuous Variables and the Residuals

The removal of the effect of the exceptional event can reveal structure in a transfer function that was previously masked. The cross correlation function provides the means by which a check can be made for the presence of new structure, and if necessary the form of this structure.

The principles and detailed actions required here are identical to those normally employed in the development of a transfer function. Because an intervention variable is active over such a short period of time it is not possible to calculate a meaningful cross correlation function for either original identification purposes or checking.

In Table 3.5, a summary has been prepared of the purpose of each of the diagnostic checks, and the use of these checks in answering the four basic questions being asked about the intervention model.

Table 3.5 Using Diagnostic Checking

DIAGNOSTIC	PURPOSE	Intervention Adequacy	Appropriate Intervention	Noise Adequacy	Transfer Function Adequacy
Parameter Estimates & Standard Errors	• to decide parameter significance • if effect makes sense	YES	<u>NO</u>	YES	YES
Gains&Standard Errors	• to decide if the overall effect makes sense is significant	YES	<u>NO</u>	<u>NO</u>	YES
Estimated Residual Variance	• to provide an absolute measure of improvement a means of comparing models	YES	YES	YES	YES
Model in Factored Form	• to check model stability provide an understanding	YES	<u>NO</u>	YES	YES
Correlation Matrix of Parameter Estimates	• to determine whether parameter values are independently estimated • model simplification is possible	YES	<u>NO</u>	YES	YES

Table 3.5 (Continued)

DIAGNOSTIC	PURPOSE	Intervention Adequacy	Appropriate Intervention	Noise Adequacy	Transfer Function Adequacy
Residual Series Values & Standard Deviations Histogram **Anomalous Residuals**	• to decide whether more residuals are significant than would be expected by chance • all the symptoms needing intervention analysis have been removed • the residuals are normally distributed • the residuals have constant variance	YES	<u>NO</u>	YES	YES
ACF and PACF of the Noise	• to decide if there is evidence of defective noise • what new structure to add	<u>NO</u>	<u>NO</u>	YES	NO
C.C.F. Residuals& Continuous Variables	• to decide if there is evidence of defective transfer function structure • what new structure to add	<u>NO</u>	<u>NO</u>	<u>NO</u>	YES

3.3.4 Reformulation

As a result of model checking, evidence may be found to suggest that the model is defective in a number of ways. The reformulation is assessed while supporting the evidence of defective behaviour and a decision is made on how the structure of the model changes. This evidence can point towards the need for change in any or all of three major areas

- the intervention
- the transfer function
- the noise

It is common the evidence to suggest changes in more than one area. In the case of more than one area, it is not advisable to change more than one part of the structure at any time. Good general sets of principles are first it should be attempt to correct the largest defect. Secondly, then if correcting the largest defect fails to remove the other defects, it should be attempt to remove these defects in order of importance.

In reformulating the intervention, the diagnostic checks will point investigator towards three possibilities as illustrated Table 3.6.

Table 3.6 Possible Situations When Intervention is Reformulation

	Simplification	Elaboration	Complete Reformulation
Parameter Estimates	YES	NO	YES
Gains	NO	NO	YES
Factored Form	YES	NO	YES
Correlation Matrix	YES	NO	NO
Residual Series	NO	YES	NO

3.3.5 Forecasting

When a model containing an intervention variable is used for forecasting, it is necessary to provide external forecasts of the intervention variable. These external forecasts are required because it is not possible to develop a univariate stochastic model for the intervention variable from which meaningful forecasts can be produced.

The number of externally supplied intervention variable forecasts which must be provided must be equal to the required forecast lead time minus any delay in the intervention mechanism. Given that external forecasts of the intervention variable are available, the intervention variable is used in exactly the same way that a continuous variable with externally supplied forecasts are used.

3.4 Advantages and Disadvantages of Intervention Analysis

A time series research design may be required to detect changes in level, slope or regime of a process. Sometimes the impact of the intervention, treatment or event is not applied instantaneously. A time series research may be needed to detect a gradual, threshold, delayed or varying effect. A principal advantage of intervention analysis is that it focuses on the sequence of events, some of which may be input and others of which may be responses. Along with the covariation of input and response, this sequence of these events is necessary for the inference of causality. The modeling of the type of response reveals a sense of the response facilitates understanding of the nature of the effect, as it were.

Intervention analysis may be afflicted with problems that threaten the internal and external validity of the analysis. The researcher must be sure that the series is properly defined conceptually and operationally before data collection. Proper administration of the data collection and maintenance of the records throughout the process is necessary. The time intervals must be made small enough to capture the process to be studied. If there is trend, cycle, or seasonality inherent in the series, then the instrumentation must be calibrated to units of temporal measurement appropriate to the capture, detection, identification of these components. Without a large enough sample size for the preintervention and postintervention series, there will not be enough power to detect the differences of trend, cycle, seasonality, noise, or impact necessary for modeling and intervention analysis. The sample size plays very important role in intervention analysis. When the sample size is not enough, the impact of intervention will be biased.

CHAPTER FOUR
APPLICATION

Recent increases in the number of traffic accidents and deaths in these accidents in Turkey make it necessary to conduct many studies on identification of the causes of these accidents and methods to prevent these accidents. In the last 17 years (1984-2002), 4,586,082 traffic accidents occurred in Turkey, 106,488 were killed and 1,748,565 were injured in these accidents (WEB_1).

The recent amendments in the traffic laws have not only increased the monetary amounts of fines, but also changed the methods of punishment to persuade people to quit their habits of committing traffic crimes. Precautions taken for decreasing number of traffic accidents and deaths affect time series of accidents. The aim of this study is to obtain quantitative measure of impact of intervention events on time series of traffic accidents. Besides, another aim is, to take precautions against any unusual values in the time series of traffic accidents that might have resulted as a consequence of the intervention event through intervention analysis. This will ensure that the results of the time series analysis of the traffic accident series, such as the structure of the fitted model, estimates of model parameters, and forecasts of future values, are not seriously distorted by the influence of these unusual values.

4.1 Description of Traffic Accident Data

This thesis essentially depends on the validity of two assumptions. The first assumption is that accidents data are correlated to each other, and so the ratio of traffic accidents which occurred in the evenings to those taking place in the day time fits with lognormal model. The second assumption is that the data are most likely to be affected by certain intervention. The Box and Tiao method is used to analyze such data.

The data are obtained from the database which is gathered by Traffic Research Center, General Police Directorships and Ministry of Internal Affairs of Republic of Turkey. Statistics of traffic accidents resulted in death and injuries occurred in İzmir between January 2002 and July 2003 are evaluated. The data are arranged weekly for measuring the effect of intervention more accurate. The data arranged for 83 weeks are given in Appendix 1.

Yeşildere, Yenişehir and Karabağlar routes have been selected for examination in Izmir. These routes are connected to each other and they are effected by multiplicative factors. They are also effected by external factors (driver mistakes, characteristics of the road, driver license issue date etc.). The statistics of traffic accidents resulted in death and injuries occurred on those routes are examined. The new mobile radar control has just been started on these routes by Izmir Traffic Control Department. The new radar control has been put into practice in two different time intervals as day and evening time. The radar control which has been carried out according to day and evening time is assumed as an intervention to the series of traffic accidents.

4.2 Adaptation Traffic Accident Data to Lognormal Distribution

In past research, accident predictive models have often been developed with accident ratios as dependent variable using simple or multiple-linear regression. In this traditional approach, the dependent variable was modeled as a linear combination of highway-related interactions, under the assumption that the dependent variable follows a normal distribution. The results obtained from this approach have generally been disappointing, both in terms of the proportion of the variation in accident ratios explained by the models and the generally weak role of geometric design variables as accident predictors.

Part of the reason for the disappointing results of past research may be that multiple regression is an inappropriate approach for developing such relationships.

There are several reasons for this concern. First, accident ratios often do not follow a normal distribution. Traffic accidents are random and discrete events. Even if the traffic accident ratios are assumed to be a continuous random variable, discrete nature of accident data do not change. Second, accident frequencies are typically very small integers. It is unusual to record no accidents in the dimension being examined during the whole study. In fact, the poisson and negative binomial distributions are often more appropriate for discrete counts of events that are likely to be zero or small integer during a given time period.

Lastly, accident frequencies and accident ratios must be non-negative. But in traditional multiple-regression models there are no constraints to make predictions from negative accident frequencies or accident ratios, which will lead to meaningless results in the use of the predicted model.

Lately several studies have implemented nontraditional statistical approaches that are based on other distribution assumptions rather than normal distribution assumptions. It is also observed that many researches are conducted where nonparametric analyses are used for traffic accidents.

4.2.1 Lognormal Distribution on Traffic Accident Data

DYS_{it} denotes the corresponding traffic accidents resulted in deaths and injuries on the Yenişehir, Yeşildere and Karabağlar routes during day time and $EVGS_{it}$ denotes traffic accident sample (i =1,2,3) resulted in deaths and injuries that occurred on that road at time t in evening time for 83 weeks. The ratio is defined between geometric mean of traffic accidents in the evening time and in the day time by $R_{it} = GEVGS_{it} / GDYS_{it}$, i=1,2,3 and t=1,2,..,83. After examining the traffic accidents of evening time, it is found that the intensity of traffic accidents occurred during the evening time is greater as a result of no radar control. Knowing this fact, existence of a recursive type interdependency, between the three routes can be assumed. Since Yenişehir, Karabağlar and Yeşildere routes are connected to each other, any traffic accident occurred on any these three routes will be dependent to each other and also are affected from external factors (driver mistakes, characteristics of the road, driver license issue date etc.), the ratio R_{it} of point i at time t can be expressed as in the following recursive relationship,

$$R_{it}=L_i R_{i-1,t} \qquad i = 1,2,3; \qquad t = 0, 1, 2,...,83$$

where (L_1, L_2, L_3) is a positive random vector. Thus,

$$R_{3t}=R_0 \prod_{i=1}^{3} L_i \qquad t = 1, 2,...,83 \qquad (4.1)$$

where $R_0 = R_{0,t}$ the initial value for the ratios. In the notation, $R_{it}=R_t$ since R_{3t} is the last ratio for point i. Then ratio R_t can be rewritten as

$$R_t=R_0 \left[\prod_{i=1}^{3} L_i^{1/3}\right]^3 = R_0 \, [G_3]^3 \qquad t = 1, 2,..,83 \qquad (4.2)$$

where G_3 is the geometric mean (GM) for L_1, L_2, L_3 ($G_3 = \sqrt[3]{\prod_{i=1}^{3} L_i}$). Therefore, from the relation (4.2), it can be concluded that the ratio R_t which is the geometric mean of the ratios at the sample routes are proportional.

R_t can reflect traffic accident rate increase when it is greater than one, and decrease when it is less than one, and stability when it is equal to one. It is reasonable to assume that the events are effected by multiplicative errors rather than additive, since the error sources contribute multiplicatively to the ratio R_t. Besides, any changes in the traffic on one route will produce changes in the traffic at the other two routes in a multiplicative fashion. Depending on above mentioned assumptions and results obtained by Crow and Schimuzu (1988), R_t can be rewritten as,

$$R_t = R_0 \prod_{i=1}^{3} (1+\delta_i) \qquad t = 1, 2, \ldots, 83 \qquad (4.3)$$

where $\{\delta_i\}$ is a set of mutually independent and identically distributed random variables with $|\delta_i| < 1$. In fact relation (4.3) can be deducted from (4.1) by letting $L_i = 1+\delta_i$. Using the Taylor expansion of $\ln(1+\delta_i)$, $\ln R_t$,

$$\ln R_t = \ln R_0 + \sum_{i=1}^{3} \delta_i$$

can be approximated by using additive central limit theorem $\ln R_t$ is asymptotically normally distributed (Crow, &Schimizu, 1988, p. 5).

The random variable R is said to have a three-parameter lognormal distribution $R \sim \Lambda(\gamma, \mu, \sigma)$, if the random variable $Y = \ln(R-\gamma)$, where $R > \gamma$, is distributed normally (μ, σ^2), $\sigma > 0$. The probability density function of R is given by

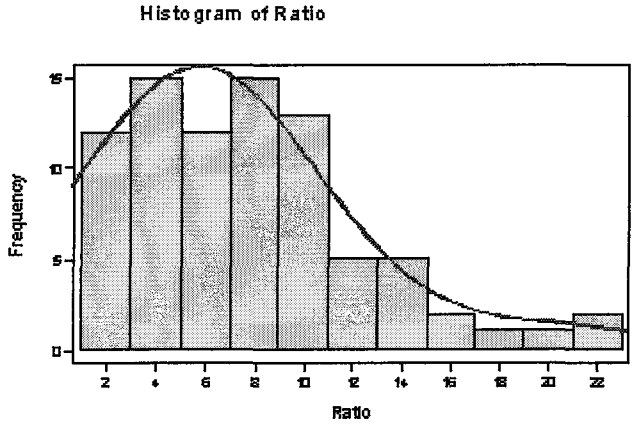
$$f(r;\gamma,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}(r-\gamma)} \exp \frac{-1}{2\sigma^2} [\ln(r-\gamma)-\mu]^2 \qquad \gamma < r < \infty \qquad \sigma > 0 \qquad (4.4)$$

The three parameter lognormal distribution can confirm the use for the geometric mean of ratios. On the other hand to overcome the strong dependency between the measurements on the road, hypothesis concerning differences between frequency of traffic accidents in days and evenings has to be tested as opposed to the traditional differences of transformation of observations. The latter does not rely on a solid statistical theory; rather it only assumes that taking the difference of the corresponding observations may eliminate the interdependency among them. On the other hand, it counts for the left-hand truncation in the observed data values. The procedure followed in this thesis can not be considered as a simple logarithmic transformation for the observed data. It is rather a distribution adjustment for the truncation part in the probability model, which is taken care of by the location; (threshold) parameter γ.

4.2.2 Lognormality Plot of Data (Q-Q Plot)

Q-Q plot is the plot of random variable's quartile versus any quartile of tested distribution. Probability distribution plot is generally used to see whether data distribution fits any specified distribution. If chosen variable fits tested distribution then the plotted points fall approximately on a straight line. Otherwise, it implies that chosen variable does not fit the tested distribution.

Based on the previous discussion on methodology and due to the truncated nature of data, it can be assumed that the data under consideration fits the lognormal distribution with three parameters (γ, μ, σ). First of all, the ratios of the measurements in the evening time to the measurements in the day time are calculated and then the geometric means for successive ratios are obtained. The geometric means can be seen in Appendix 2. In order to check the data and confirm the goodness of fit of the lognormal model, histogram of data is plotted in Figure 4.1 and the geometric means of the ratios for the traffic accidents on a lognormal probability paper plotted in Figure 4.2, which shows an adequate fit.



<u>Summary Statistics</u>

N	83
Mean	8.825
Skewness	1.079

Figure 4.1. Superimposing a Histogram with Fitted Curves

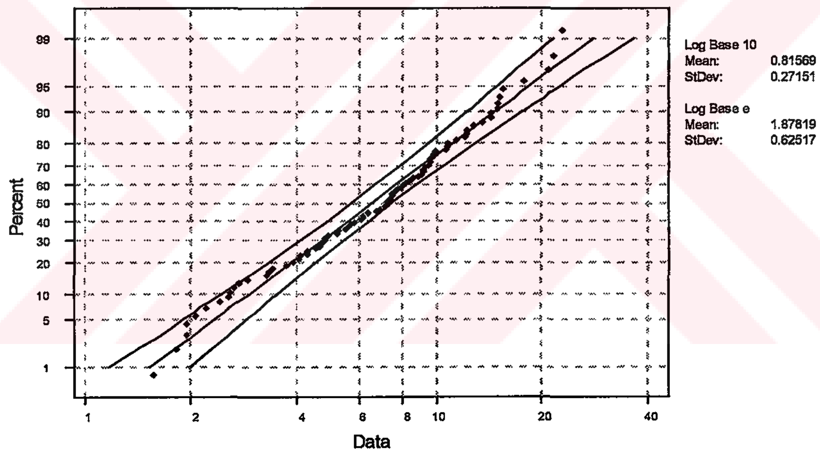
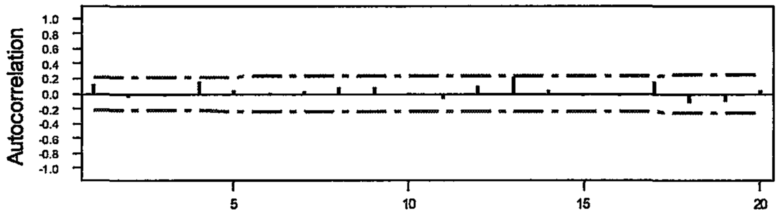


Figure 4.2. Lognormal Probability Plot for Ratios

Examination of the Sample Autocorrelation Function (SACF) of the geometric mean of the ratios plotted in Figure 4.3 shows no significant SACF at any lag, implying the independency of the geometric means of these ratios in time (Al-Khalidi, 2002, s.694).



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.13	1.16	1.40	8	0.09	0.80	5.17	15	-0.03	-0.21	12.56
2	-0.05	-0.49	1.66	9	0.08	0.68	5.75	16	-0.02	-0.19	12.62
3	-0.04	-0.33	1.79	10	0.01	0.12	5.77	17	0.17	1.35	15.57
4	0.16	1.45	4.15	11	-0.06	-0.54	6.16	18	-0.12	-0.99	17.26
5	0.04	0.35	4.30	12	0.11	0.96	7.41	19	-0.10	-0.78	18.35
6	0.01	0.06	4.30	13	0.22	1.85	12.26	20	0.04	0.30	18.52
7	0.03	0.23	4.36	14	0.05	0.38	12.49				

Figure 4.3. Autocorrelation Function for Log Adjusted Ratios

4.2.3 Point Estimation of the Three Parameters of the Lognormal Distribution

The parameters γ, μ and σ of the lognormal distribution will be estimated using the Moment Estimation (ME) and Modified Moment Estimation (MME) (Cohen, &Whitten, 1988, p.76 etal.1985). For an ordered sample of size 83, the estimating equations are

$$E(R) = \bar{r} \quad , V(R) = s^2 \text{ and } E[ln(R_1-\gamma)]=ln(r_1-\gamma) \tag{4.5}$$

where $\bar{r}$ and s^2 are the sample mean and variance (unbiased) and R_1 is the first order statistic (a random variable) in a random sample of size 83 and r_1 is the corresponding sample value.

Appropriate substitutions of the moments of lognormal into (4.5) yield,

$$\begin{aligned} \bar{r} &= \gamma + e^{\mu} e^{\sigma^2/2} \\ s^2 &= e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1) \\ r_1 &= \gamma + e^{\mu} exp(\sigma E Z_{1,n}) \end{aligned} \tag{4.6}$$

where $EZ_{1,n}$ is the expected value of the first order statistic from the standard normal distribution (0,1). Solving equations (4.6) result in:

$$\hat{\mu} = \ln\left\{s\left[e^{\hat{\sigma}^2}(e^{\hat{\sigma}^2} - 1)\right]^{-1/2}\right\} \tag{4.7}$$
$$\hat{\gamma} = \bar{r} - s(e^{\hat{\sigma}^2} - 1)^{-1/2}$$

and $\hat{\sigma}$ can be found using Appendix 3, by entering the table or the figure with $J(n,\hat{\sigma}) = s^2/(\bar{r} - r_1)^2$ to read $\hat{\sigma}$. The location (threshold) parameter, γ can be estimated by using the first equation in (4.6).

Using the procedure of this section, $J(n,\hat{\sigma})$, $\hat{\sigma}$, $\hat{\gamma}$ for the traffic accidents are obtained as follows:

$\bar{r} = 8.825$ $s^2 = 21.59$ $s = 4.6464$ $r_1 = 1.5683$ $J = 0.5515$
$\hat{\sigma} = 0.53$ (It is obtained using table and figure Appendix 3)

Using equation (4.6), $\hat{\mu} = 1.9587$, $\hat{\gamma} = 0.666$ are obtained. According to the foregoing discussion, the ratios of the geometric means have a lognormal distribution with three parameters, i.e. $R \sim \Lambda(0.666, 1.9587, 0.53)$.

The results of ME and MME are given in Table 4.1.

Table 4.1 Estimates of Lognormal Distribution for Ratio

Estimator	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{E}(R)$	$\sqrt{\hat{V}(R)}$	$\hat{\alpha}_3(R)$	$\hat{\alpha}_4(R)$
ME	0.7454	0.3362	1.900	7.825	5.9999	1.0793	5.1406
MME	0.6839	0.4813	2.092	9.780	4.6463	1.6650	8.3028

Here asymptotic standard deviations are not applicable since estimating equations failed to yield Maximum Likelihood Estimation (MLE). From the central limit theorem, $\hat{\sigma}_{\bar{x}}$=0.51017.

The differences as seen in Table 4.1 between the MME and ME are quite small Comparable differences between these estimators have been observed in most applications where MME were calculated. The superiority of the MME over ME and MLE in estimating parameters of the three-parameter lognormal distribution has been demonstrated in numerous practical applications. "ME have the disadvantage that they are not uniquely determined by their moments and that inherently large sampling errors of the third moment introduce correspondingly larger sampling errors into parameter estimates" (Cohen, &Whitten, 1988; p.76). MME are unbiased with respect to distribution mean and variance. Maximum likelihood estimators lead to inadmissible estimates and lead to questionable variance-covariance matrices. They are applicable over the entire parameter space, therefore they are more reliable. Unless α_3 is small, then the normal rather than the lognormal distribution would be a better choice as a model. As a result, since MME is more reliable than the other estimation methods, estimation of parameters of MME is used.

4.2.4 Goodness of Fit Tests of Lognormal Distribution

The ratios are then transformed to zero location by subtracting the corresponding location parameter $\hat{\gamma}$ from the measurements. After adjusting for location parameter, the natural logarithmic transformation is applied to the ratios. Depending on the results obtained in this section, the transformed random variable *ln (R-γ)* will follow a normal distribution. According to the normality plot in Figure 4.4, the transformed data are distributed as normal.

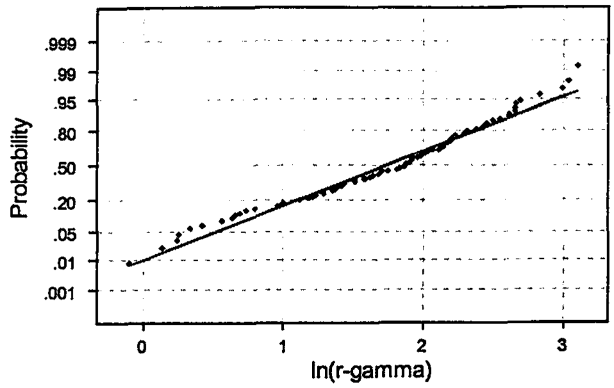


Figure 4.4 Normality Plot of Transformed Data

According to Figure 4.4, points which are on the straight line or very close to the straight line indicate that *ln(r-gamma)* are distributed normal. A straight or close to straight line indicates normality. A lot of curvature indicates non-normal data.

Furthermore, Kolmogorov-Smirnov, Anderson-Darling, and Ryan-Joiner (Similar Shapiro Wilk) Tests were applied to the data. The results are shown in Table 4.2.

Table 4.2 Goodness-of-Fit Tests for Normal Distribution

Goodness-of-Fit Tests for Normal Distribution			
Test	Statistic	DF	p-value
Kolmogorov-Smirnov(D)	0.049	Pr >D	0.059
Anderson-Darling(A-sq)	0.747	Pr>A-Sq	0.051
Ryan-Joiner(W)	0.987	Pr>W	0.097

In the entire test, p-value is greater than α (α=0.05) therefore the null hypothesis cannot be rejected, so ln(r-gamma) fits normal. Based on these results, it can be interpreted that *(r-gamma)* variable fits lognormal model.

4.3 Building Intervention Model for Traffic Accident Data

Traffic Control Department's assumption that radar control will be effective is not based on the analysis of available data but on common knowledge. Based on this assumption, the department expects traffic accidents to show some evidence of the intervention. In this section, the success of intervention is tested on available data from the standpoint of statistical significance. Interested parties, such as Traffic Control Department or Traffic Research Center can then judge that whether present legislation and its' enforcements are satisfactory.

There are inherent difficulties in statistical analysis of this kind as the data are in the form of time series in which successive observations, preintervention and postintervention are highly correlated. To see the intervention effect, one must isolate how trend, seasonality and correlated noise structure effect such a series. Box and Tiao (1975) discussed the technique of intervention analysis in two case studies, one dealing with environmental pollution and the other with economic measures.

In this thesis, following Box and Tiao (1975), a univariate time series model of the BJ type is developed by analyzing the autocorrelations and partial autocorrelations of the response series before intervention. This model is used to calculate forecasts for the postintervention period with their origin at the last observation in the preintervention period. The forecasts are compared with actual realizations to test for the effects of intervention.

If there is a significant difference, then the preintervention model is modified by certain priori specifications to incorporate the effect of intervention. The modified model is then fitted to entire series, preintervention and postintervention, and conclusions regarding the magnitude and direction of the effect of intervention are drawn.

4.3.1 Assumptions Used in the Univariate Intervention Analysis

Before attempting to develop a univariate model capturing the effects of the intervention, it is necessary to list the assumptions under which such an attempt is made. These assumptions are as follows:

1. The intensity of traffic remained unchanged during the study period,
2. The behaviours of drivers remained unchanged throughout,
3. No other major interventions occurred during the study period,
4. The noise structure of the univariate model remains unchanged preintervention and postintervention.

4.3.2 Detection of Intervention Variable –Radar Application

The data consist of weekly observations on traffic accidents which are occurred during day time and evening time on Yenişehir, Yeşildere and Karabağlar connected road for 83 weeks. The ratio of geometric mean of traffic accidents occurred during the evening time to those taking place in the day time can be seen in Figure 4.5. Meanwhile, it is assumed that conditions (weather etc.) are the same during the day time and evening time.

Since the suggestion that radar control plays important role in decreasing number of traffic accidents occurred in day time, traffic accidents series occurred in evening time is called preintervention or without intervention series. Therefore, the analysis is applied to traffic accidents occurred during evening time.

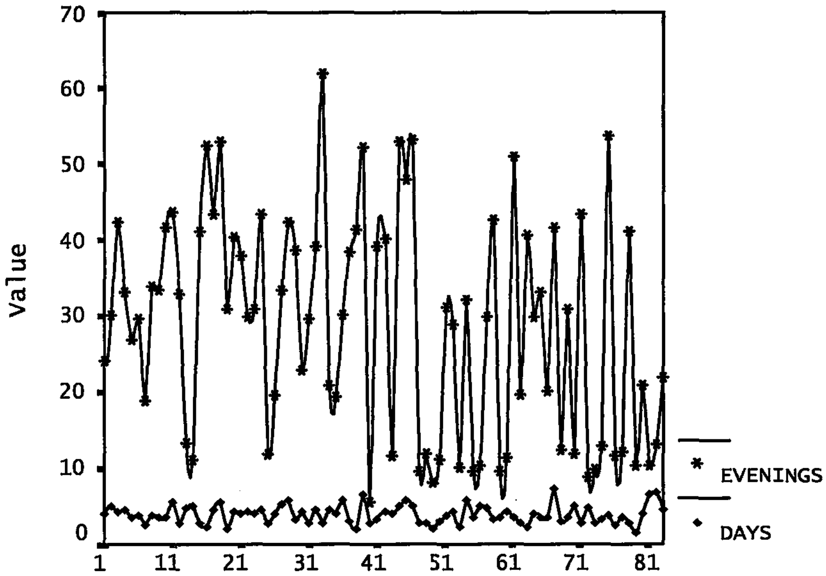
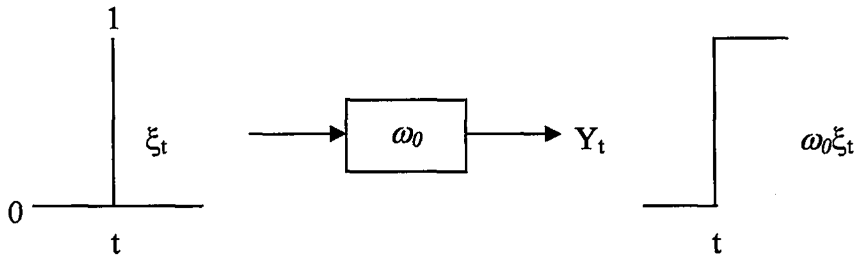


Figure 4.5 Time Series Plot of Geometric Mean Data

For the scale of measurement of intervention variable for traffic accident data can be used as follows:

IMPULSES no radar application = 0

radar application = 1



4.3.3 Developing the Preintervention Univariate Model

The graphical presentation of the traffic accidents occurred during evening time in Figure 4.6, shows a general trend during some periods, implying that the time series values are not stationary. At the same time, this figure is implying nonstationary variances. Therefore, a log transformation is necessary to achieve stationarity in variance (Al-Khalidi, 2002, p.690).

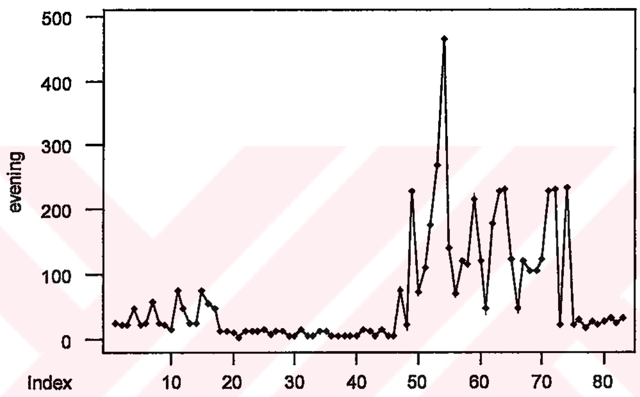
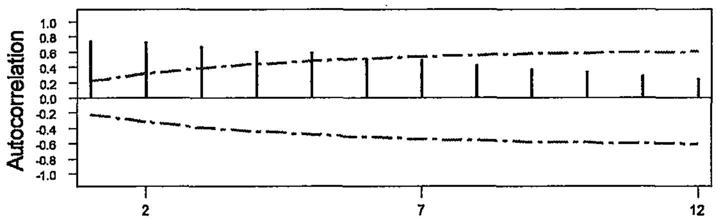


Figure 4.6 Time Series Plot for Traffic Accidents Occurred in Evenings.

In Figure 4.7 presents sample autocorrelation function of ($Z_t^* = \ln Z_t$). When examining Figure 4.7, the series shows neither quick cut-off nor quick dying down toward zero as the lag *k* increases. Thus, it means that the series values are not stationary (Al-Khalidi, 2002, p.690)



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.75	6.83	48.31	8	0.43	1.54	267.37
2	0.73	4.57	94.91	9	0.37	1.29	280.77
3	0.69	3.50	136.55	10	0.35	1.17	292.42
4	0.62	2.76	170.36	11	0.29	0.98	300.89
5	0.60	2.48	203.24	12	0.25	0.81	306.91
6	0.51	1.95	226.93				
7	0.49	1.82	249.62				

Figure 4.7. SACF for lnEvening

In order to make the series stationary in mean the first difference in Z_t^* values were taken. Figure 4.8 is belongs to time series plot of first difference of data.

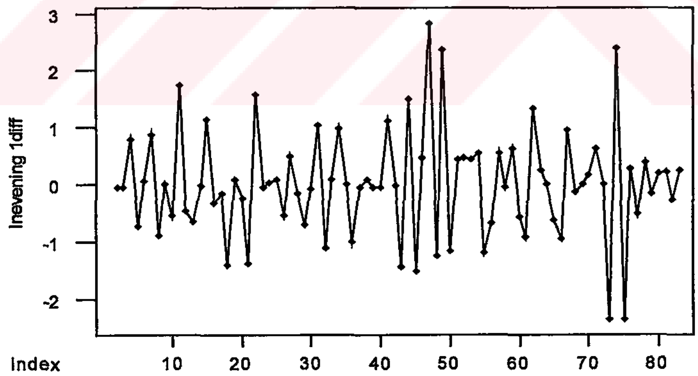
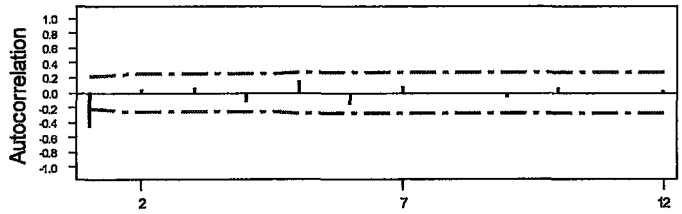


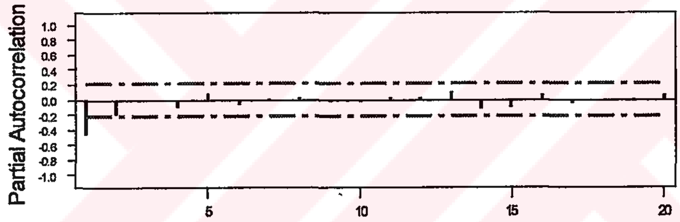
Figure 4.8. Time Series Plot for First Difference lnEvening

Figure 4.9 shows that the SACF values of W_t lies within 2σ after lag 1, i.e. the SACF shows cut off after k=1, indicating W_t values are stationary in mean (Al-Khalidi, 2002, p.690).



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.46	-4.21	18.35	8	-0.00	-0.00	25.52
2	0.05	0.40	18.59	9	-0.07	-0.49	25.95
3	0.06	0.44	18.88	10	0.06	0.42	26.28
4	-0.12	-0.91	20.17	11	-0.01	-0.10	26.30
5	0.16	1.22	22.57	12	0.02	0.15	26.34
6	-0.16	-1.15	24.81				
7	0.09	0.63	25.52				

Figure 4.9 SACF for First Difference lnEvening



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	-0.46	-4.21	8	0.02	0.18	15	-0.10	-0.94
2	-0.21	-1.89	9	-0.04	-0.33	16	0.06	0.58
3	-0.01	-0.12	10	-0.02	-0.22	17	-0.05	-0.41
4	-0.11	-1.00	11	0.03	0.23	18	-0.01	-0.06
5	0.08	0.76	12	0.03	0.26	19	0.00	0.01
6	-0.07	-0.60	13	0.11	1.01	20	0.08	0.69
7	0.01	0.05	14	-0.12	-1.10			

Figure 4.10 SPACF for First Difference lnEvening

When examining Figure 4.9 and Figure 4.10, the SACF and SPACF of the new series W_t suggests that the nonseasonal moving average model of the first order ARIMA(0, 1, 1) have best fit to the series values. The general form of ARIMA (0, 1, 1) is shown in equation 4.8.

$$W_t = (1 - \theta_1 B)\, a_t \tag{4.8}$$

The maximum likelihood estimate of θ_1 is 0.5247, which satisfies the invertibility condition ($|\theta_1| < 1$). The *t*-value 5.52 implies that θ_1 is significantly different from zero. Results, which are MINITAB Output, are shown in Appendix 4.

Finally, the goodness of fit test is carried out using Ljung-Box statistic,

$$Q' = n(n' + 2)\sum_{i=1}^{k}(n' - 1)^{-1} r_i^2(a^*) \tag{4.9}$$

where $r_i(a^*)$, is the sample autocorrelation of the residuals a^* (RSAC) at lag i, $n' = n-1$, where n is the number of observations in the original series. The calculated value of Q' (using RSAC for 24 lags) is (10.1), which is less than the tabulated value of χ^2 with 23 degrees of freedom, at significant level α = 0.05 (35.172), indicating the underlying model is accepted. MINITAB output is in Appendix 4.

4.3.4 Formulation of Intervention Model

The noise model for N_t , the best represents the traffic accidents occurred in the evening time can be written as;

$$N_t = (1 - 0.5247B)\, a_t \tag{4.10}$$

To measure the effect of intervention on the second part of the series, the analysis can be formulated using equation 4.8 and 4.10, with ξ_t as step function as follows:

$$Z_t = R(B) + (1 - 0.5247B)a_t \tag{4.11}$$

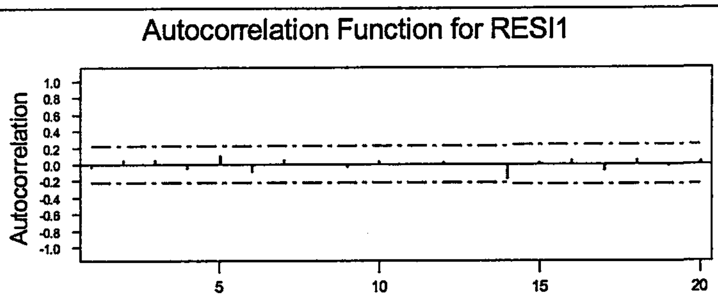
Dividing both sides of equation 4.11 by (1 – 0.5247B) and simplifying the resulted equation (using Taylor expansion with a polynomial of 4th. order),

$$K(B)Z_t = K(B)R(B)\xi_t + a_t \qquad (4.12)$$

is obtained where $K(B) = (1 + 0.5247B + 0.2753B^2 + 0.1445B^3 + 0.0.76B^4)$, $K(B)$ is a polynomial in B of order 4.

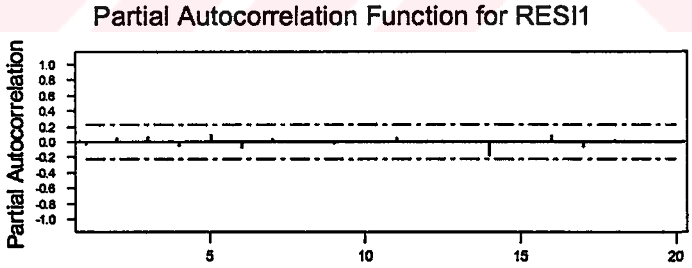
Model diagnosis entails residual analysis. If the model is properly specified and the model parameters account for all of the systematic variance, then the residuals should resemble white noise. Residual analysis is performed with the autocorrelation and partial autocorrelation function. These correlograms can be examined with reference to modified Portmanteu tests of their associated significance level. The Portmentau statistic may inflate the autocorrelation under conditions of short series. For that reason, the modified Ljung-Box statistic is used to provide better significance test. White noise residuals do not have significant p-values. These white noise p-values of the residuals should not be less than 0.05. Graphically, white noise residuals have associated spikes that do not extend beyond the confidence interval limits. The ACF and PACF plots reveal these limits as dotted lines spreading out from the midpoint of the plot. When spikes leap up beyond the limits of two standard errors on each side of the central vertical axis of no autocorrelation, then the autocorrelation or partial autocorrelation of the residuals have significant spikes with p-values less than 0.05. Indication of the significant ACF or PACF residual spikes is empirical evidence of lack of fit. The pattern of lack of fit will suggest the reparameterization of the model. Slowly attenuating autocorrelation, functions suggest further differencing.

Combination of ACF and PACF patterns indicate whether the additional terms should be moving average or autoregressive. Gradual attenuation of the ACF with a few spikes and sudden decline in PACF magnitude suggest that autoregressive parameters should be added, whereas gradual attenuation of the PACF and a few finite spikes of the ACF with sudden decline of their magnitude suggest moving average terms should be used. Once these have been properly identified and estimated, the ACF and PACF of the residuals should appear as white noise. If the parameters are all accounted for in the model, then the residuals should consist purely of white noise or unsystematic random variation. Figure 4.11 and Figure 4.12 presents the ACF and PACF of residuals.



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.05	-0.41	0.18	8	0.00	0.01	3.24	15	0.03	0.23	6.96
2	0.06	0.55	0.50	9	-0.05	-0.42	3.46	16	0.05	0.46	7.27
3	0.06	0.55	0.83	10	0.05	0.40	3.67	17	-0.08	-0.72	8.03
4	-0.06	-0.51	1.11	11	0.02	0.18	3.71	18	0.06	0.46	8.36
5	0.11	0.95	2.12	12	0.03	0.28	3.81	19	-0.02	-0.18	8.41
6	-0.10	-0.87	3.00	13	0.00	0.04	3.81	20	0.04	0.33	8.58
7	0.05	0.45	3.24	14	-0.17	-1.52	6.89				

Figure 4.11. ACF of Residuals



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	-0.05	-0.41	8	-0.00	-0.01	15	0.00	0.04
2	0.06	0.53	9	-0.03	-0.29	16	0.08	0.76
3	0.07	0.61	10	0.02	0.18	17	-0.07	-0.60
4	-0.05	-0.50	11	0.05	0.46	18	0.03	0.24
5	0.09	0.86	12	0.02	0.17	19	0.02	0.15
6	-0.09	-0.81	13	0.00	0.03	20	0.02	0.16
7	0.04	0.37	14	-0.18	-1.62			

Figure 4.12. PACF of Residuals

After examining ACF and PACF of the residuals, there is evidence that the model is adequate and residuals appear as white noise. Because no spikes leap up beyond the limits of two standard errors on each side of the central vertical axis of no autocorrelation. The result of Portmentau test is given Table 4.3.

Table 4.3. Results of Portmentau Test

Modified Box-Pierce (Ljung-Box) Chi-Square statistic				
Lag	12	24	36	48
Chi-Square	3.8(DF=11)	10.0(DF=23)	33.6(DF=35)	46.2(DF=47)

According to the result of Portmentau test, because the ARIMA model coupled with this intervention variable control for all systematic variation in the system, the residuals are white noise.

4.3.5 Estimation Intervention and Noise Parameters

Equation 4.12 can be writen as a linear regression equation:

$$Y_t = \beta X_t + a_t \tag{4.13}$$

where $Y_t = K(B)Z_t$, $X_t = K(B)\xi_t$ and $\beta = R_t(B)$

Since the observation at both time interval are taken at the same time, the same atmospheric conditions and from the same road, it is assumed that the same white noise effects a_t on the series before and after the intervention effect. Table 4.4 shows the results of regression analysis.

Table 4.4 Results of Regression Analysis

$Y_t = 58.0 - 37.2\ X_t$

Predictor	Coef	StDev	T	p-value
Constant	58.001	1.725	33.63	0.000
X_t	-37.223	1.883	-19.77	0.000
S = 15.24	R-Sq = 71.0%	R-Sq(adj) = 70.8%		

Table 4.5 Results of Analysis of Variance

Source	DF	SS	MS	F	p-value
Regression	1	90754	90754	390.90	0.000
Error	160	37146	232		
Total	161	127900			

Using the transformed time series values of data, the least squares estimates of β is -37.223. The *t*-value for β is -19.77 which is greater than the tabulated value of *t* with 160 degrees of freedom, *i.e.*, it is highly significant indicating significant effect for the intervention variable. So, radar application plays important role for decreasing number of traffic accidents. There is evidence that associated with intervention variable is a step change of approximately -37.223 units in the level of traffic accidents.

Also, the calculated *F*-value in the ANOVA for the simple linear regression model in equation 4.13 is 390.90, which is highly significant. When the regression model adapted from model of one intervention variable and noise has been examined, *R*-square shows that intervention variable accounts for the 71% of the total variability within traffic accident data.

CHAPTER FIVE
CONCLUSION

5.1 Conclusions

The use of intervention analysis and lognormal distribution analysis on economic and social areas has two main ideas. First, is to show that the data are distributed lognormal by the use of distribution theory on traffic accidents which is a major public problem nowadays. Secondly, is to apply the intervention analysis on traffic accidents.

During the study, three routes are selected where the traffic accidents occur intensely and the radar controls have just been implemented. Traffic accidents that occur on the 5 km. long Yeşildere road and also Yenişehir and Karabağlar intersections were examined.

The data between January 2002 and July 2003 are arranged weekly for measuring the effect of intervention more accurately. Since Yenişehir, Karabağlar and Yeşildere routes are connected to each other and any traffic accident occurred on any these three routes will be dependent to each other and also are affected from external factors (driver mistakes, characteristics of the road, driver license issue date etc.), existence of a recursive type interdependency between the three routes can be assumed.

Some of the data needed log-transformation, followed by taking the difference to achieve stationarity in means and variances. The sample autocorrelation function and the sample partial autocorrelation function have been used to identify the orders of the ARIMA models.

Intervention analysis is applied in two parts as preintervention and post intervention. For the first part of the series (before intervention), the maximum likelihood estimates of the parameters of the model is found and also test of significance using the t-test is carried out (after transforming the noise N_t by ARIMA model into "white" noise). For the second part of the series (after intervention), the least squares estimates for β is found and significance test using the F-test is carried out (using the same noise model as before intervention). The results of the intervention part showed it is highly significant indicating significant effect. So, radar application plays important role for decreasing the number of traffic accidents. There is evidence that intervention variable is a step change of approximately -37.223 units in the level of traffic accidents.

Moreover, the calculated F-value in the ANOVA for the simple linear regression model is highly significant. When the regression model adapted from model of one intervention variable and noise has been examined, the R-square value shows that intervention variable accounts for the 71% of the total variability within traffic accident data.

5.2 Suggestions

Suggestions based on the study can be classified in two main groups. In this study, which will be a guiding light for future studies, it is shown that the geometric mean ratios of the traffic accidents that occur on two different time intervals in Izmir fit the lognormal model. Based on this fact, fit test of other models can be examined for similar data. Besides geometric mean ratios, other transformations like multiplication of random variables, the multiplication of the powers of random variables etc. can be used to test for adapting to other distributions.

This study is very important since it is the first study in Turkey on this area. Intervention analysis can be applied to many data sets in the areas of chemistry, economy, physics and social areas as long as there is no intervention on the data. A study can be conducted on the data of not only Izmir but also Turkey as a whole. Specific routes, weather conditions and other factors that may affect the occurrence of the accident can be used as an intervention variable.

In the prediction of intervention variables, nonlinear analysis methods can also be used. Then the parameter predictions gathered from the nonlinear anaylsis can be compared with results obtained from linear analysis.

Lastly, intervention analysis can also be applied on economic time series data, which is the one of the most common time series data in applications. The only problem that researchers will face is the great reactions of economic data to Turkey's daily or even momentary changes as a result of fluctuating economy. In intervention analysis, it is assumed that there are no other interventions on the periods other then the intervention examined. The validity of this assumption is not reasonable for Turkey. For intervention analysis to give meaningful results, long term stable data is necessary. Multivariate intervention analysis can be applied to economic and many other data.

REFERENCES

Abraham, B., (1987). Application of intervention analysis to a road fatality series in Ontario. Journal of Forecasting, 6, 211-219

Abraham, B.,& Ledolter J., (1986), Statistical Methods For Forecasting, Wiley Series, USA

Akdi, Y. (2003), Zaman Serileri Analizi (Birim Kökler ve Kointegrasyon, Bıçaklar Kitabevi, Ankara,

Al-Khalidi, A.S, (2002), Measuring water sources pollution using intervention analysis in time series and lognormal model, Environmetrics, 13, pp:693-710

Bhattacharyya, M. N.,& Layton, A. P. (1979). Effectiveness of seat belt legislation on the Queensland road toll – An Australian case study in intervention analysis. Journal of the American Statistical Association, 74, 596-603

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to econo mic and environmental problems. Journal of the American Statistical Association. 70, 70-79

Box, G. E. P., Jenkins, G. M., & Reinsel G. C. (1994).Time Series Forecasting and Control. (3rd ed.) New Jersey: Prentice Hall: Englewood Cliffs.

Box, G. E. P., &Jenkins, G. M.. (1976).Time Series Analysis Forecasting and Control. (3rd ed.) San Francisco:Holden Day

Canada (1983). Encyclopedia of Statistical Sciences, 4,. John Wiley &Sons

Canada (1985). Encyclopedia of Statistical Sciences, 5,. John Wiley &Sons

Canada (1985). Encyclopedia of Statistical Sciences, 6,. John Wiley &Sons

Cauley, J., Im E. I. (1988). Intervention policy analysis of skyjacking and other terror ist incidents. American Economic Review, 78, 27-31

Chakravart A. C., Laha W.,& Roy D. (1967), Handbook of Methods of Applied Statistics Volume I, John Wiley.

Chatfield, C. (1999). The Analysis of Time Series (5th ed.). USA: Chapman & Hall.

Cohen, A. C., &Whitten B. J. (1980). Estimation in the three- parameter lognormal distribution. Journal of the American Statistical Association, 75, 399-404

Cohen, A. C.,& Whitten B. J. (1988). Parameter estimation in reliability and life span models. New York: Marcel Dekker

Cox, D. R., Hinkley, &Barndorff-Nielsen O.E. (1996). Time Series Models. Chapman & Hall.

Crow, E. L.,& Shimizu, K. (1988). Lognormal Distributions: Theory and Application New York: Marcel Dekker

Daniel, W. W. (1978). Applied Nonparametric Statistics, USA: Houghton Mifflin

Deadman, D.,& Pyle, D. C. (1993). The effect of the aboliton of capital punishment upon homicides in Great Britain: An application of intervention analysis. Journal of Applied Statistics, 20, 191-206

Fase, M. G. M., (1971). On the estimation of Lifetime Income. Journal of the American Statistical Association, 66, 686-691

Fomby, T. B., Hayes, K. J. (1990). An intervention analysis of the war on poverty. Journal of Econometrics. 43, 197-212

Fowlkes, B. E. (1979). Some methods for studying the mixture of two normal (Lognormal) distributions. Journal of the American Statistical Association, 74, 561 -575

Fox, R. T. (1995). Measuring catastrophic events on operating viability of firms: hur ricane hugo and hospitals. International Advances in Economic Research, 1, 251 - 262

Guerrero, V. M., Pena D.,& Poncela P. (2003). Measuring intervention effects on multiple time series subjected to linear restrictions: A banking example. Journal of the American Statistical Association, 98, 121-137

Haque, M. O. (1990). Preliminary evaluation of the Vitorian zero blood alcohol time An Australian case study in intervention analysis. Communications in Statistics A Theory and Methods, 19, 3881-3899

Harvey, A. C.,& Durbin, J. (1986). The effects of seat belt legislation on Britih road casualties: A case study in structural time series modelling. Journal of the Royal Statistical Society, 149, 187-227

Harvey, C. A. (1990). The Econometric Analysis of Time Series. (2nd ed.). Great Britain: Philip Alan

Harvey, C. A. (1993). Time Series Models.(2nd ed.). Harvester Wheatsheaf.

Haugh, L. D. (1976). Checking the independence of two covariance - stationary time series: A univariate residual cross – correlation approach. Journal of the American Statistical Association, 71, 378-385

Huitema, B. E., McKean J. W., & McKnight S. (1999). Autocorrelation effects on least-squares intervention analysis of short time series. Educational & Psychologi cal Measurement , 59, 767-787

Kendall, M. G., &Ord, J. K. (1992). Time Series (3rd ed.), Arnold, London.

Kotz, S. (1973). Normality versus lognormality with applications. Communication in Statistics, 1(2), 113 – 132

Kotz, S., Johnson N. L., (1970). Continuous Univariate Distributions, USA: John Wiley& Sons

Leone, R. P. (1987). Forecasting the effect of an environmental change on market performance : An intervention time –series approach. International Journal of Forecasting, 3, 463-478

McLead, G. (1983). Box-Jenkins in Practice, GSP Publication

Mélard, G., &Pasteels, J. M. (2000). Automatic ARIMA modeling including interven tions, using time series expert software. International Journal of Forecasting. 16, 497 – 508

Mills, T.C., (1990), Time Series Techniques for Economists. Cambridge University Press, New York

Murry, J. P., Stam, A.,& Lastovicka, J. L. (1993). Evaluating an anti - drinking and advertising campaign with a sample survey and time series intervention analysis Journal of the American Statistical Association, 88, 50-56

Narayan, J., & Raj S. P. (1986).Intervention analysis of a field experiment to assess the buildup effect of advertising. Journal of Marketing Research, 23, 337-345.

Narayan, J., & Considine, J. (1989). Assessing the impact of fare increases in a tran sit system by using intervention analysis. Journal of Business Research. 19, 245-54

Nelson, R. C. (1973). Applied Time Series Analysis. USA: Holden-Day

O'neill, B.,& Wells, T. W. (1972). Some recent results in lognormal parameter estimation using grouped and ungrouped data. Journal of the American Statistical Association, 67, 76-80,

Özmen, A., (1986), Zaman Serisi Analizinde Box-Jenkins Yöntemi ve Banka Mevduat Tahmininde Uygulama Denemesi, Anadolu Üniversitesi Yayınları, Eskişehir

Pole, A., West M., &Harrison J. (1999). Applied Bayesian Forecasting and Time Series Analysis. (2nd ed.). USA: Chapman & Hall.

Rao, T. S.(1993).Development in Time Series Analysis. Chapman & Hall.

Ross S. (1993). Probability Models. (5th ed.). London: Academic Press

SAS Institute Inc. (1999). Comparing Goodness-of-Fit Tests. USA

SAS Institute Inc. (1999). Goodness-of-Fit Tests for Statistical Distributions. USA

Shao, Y. E., Yue-fa, L.,& Soe-tsyr Y. (1999). Integrated application of time series multiple-interventions analysis and knowledge-based reasoning. Journal of Applied Statistics, 26, 755-767

Shen, W. (1998).Estimation of parameters of a lognormal distribution. Taiwanese Journal of Mathematics, 2, 243-250

Somers, T. M.,& Gupta, Y. P. (1994). Assessing the effect of change in advertising strategy through use of intervention methods, Journal of Information & Optimizati on Sciences, 15, 179-193

Sridharan, S., Vujic, S.,& Koopman S. J. (2003). Intervention time series analysis of crime rates. Tinbergen Institute Discussion Papers. 03-040/4, 35-54

Tanaka, K. (1996). Nonstationary and Noninvertible Theory. USA: Wiley Series

Tiao,G. C.,& Box, G. E. (1981). Modeling multiple time series with applications. Journal of the American Statistical Association, 76, 802-816

Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters extended sample autocorrelation function for stationary and nonstationary ARMA models. Journal of the American Statistical Association, 79, 84-96

Velicer, W. D.,& Redding, C. A. (1992). A time series investigation of three nicotine Regulation models. Addictive Behaviour, 17, 325-345

Wei, W. S. (1990). Time series analysis: univarite and multivariate methods. Addison-Wesley, Redwood City.

Yaffee, R.,&McGee M., (2000), Time Series Analysis and Forecasting, Academic Press, New York

WEB_1, (2003), İstanbul Police Department's web site. http://www.iem.gov.tr, 6/9/2003.

APPENDICES

Appendix 1. Traffic accidents belongs to Karabağlar, Yeşildere, Yenişehir routes

Weeks	DKarabağlar	DYenişehir	DYeşildere	EKarabağlar	EYenişehir	EYeşildere
1	3	3	7	13	13	6
2	4	5	6	11	8	12
3	3	2	13	10	12	16
4	4	2	11	8	16	14
5	2	4	6	3	5	15
6	4	2	7	14	7	12
7	3	6	1	7	15	12
8	7	2	4	15	11	13
9	6	1	7	14	13	13
10	1	10	4	17	12	8
11	7	3	8	14	11	10
12	4	6	1	11	14	14
13	8	2	7	15	12	13
14	5	5	5	8	11	15
15	1	7	3	14	9	11
16	2	6	1	12	12	13
17	3	4	8	13	12	15
18	4	6	7	14	13	12
19	1	2	4	15	11	8
20	3	4	6	18	9	7
21	7	2	5	14	12	3
22	6	3	4	5	18	11
23	7	2	5	6	14	15
24	4	2	11	14	16	11
25	1	3	8	8	12	17
26	7	2	5	5	15	12
27	3	8	6	11	17	13
28	4	9	5	12	16	10
29	1	8	4	8	14	18
30	4	2	9	10	15	14
31	3	7	1	7	10	13
32	2	8	6	5	13	12
33	3	1	8	9	11	17
34	4	8	3	11	8	14
35	1	9	7	9	15	11
36	4	6	8	14	6	12
37	3	1	9	9	5	13
38	4	2	1	13	14	8
39	7	5	8	13	8	17
40	4	6	1	11	4	4
41	1	5	7	8	6	16
42	7	2	6	12	5	17

Appendix 1. continued

Weeks	DKarabağlar	DYenişehir	DYeşildere	EKarabağlar	EYenişehir	EYeşildere
43	8	4	2	14	11	10
44	8	5	3	15	13	11
45	3	7	9	8	5	12
46	4	8	4	13	12	15
47	3	4	2	7	8	15
48	7	3	1	18	7	13
49	1	8	1	4	13	10
50	3	1	9	9	11	14
51	4	2	7	7	15	13
52	4	3	6	14	13	9
53	1	5	2	6	14	12
54	8	4	6	18	12	8
55	12	4	1	15	14	4
56	4	8	4	10	16	7
57	7	2	8	13	12	6
58	4	1	8	16	13	10
59	3	4	4	12	15	5
60	7	2	6	9	13	12
61	8	1	5	7	11	13
62	4	6	1	16	18	3
63	2	2	3	14	17	5
64	7	2	5	7	14	10
65	4	6	2	12	16	12
66	5	1	8	6	15	11
67	8	7	7	14	14	8
68	4	2	3	9	16	13
69	5	8	1	11	15	10
70	6	4	5	15	14	8
71	1	8	3	18	15	9
72	7	4	4	10	11	6
73	3	4	2	9	13	8
74	1	9	4	15	13	11
75	3	3	6	12	19	7
76	1	8	2	7	16	14
77	4	2	6	15	17	7
78	4	6	1	12	12	7
79	3	1	1	8	15	9
80	4	2	8	10	16	8
81	7	5	8	18	12	5
82	7	7	6	13	12	14
83	4	3	8	10	14	9

Appendix 2. Geometric Mean of Evenings (GME) / Geometric Mean of days (GMD)

t	GM of day	GM of evening	GMD/GME
1	3.9791	24.0500	6.0441
2	4.9324	30.1800	6.1187
3	4.2727	42.4300	9.9306
4	4.4480	33.1500	7.4529
5	3.6342	26.9000	7.4018
6	3.8259	29.5500	7.7237
7	2.6207	18.8100	7.1774
8	3.8259	33.8900	8.8581
9	3.4760	33.3300	9.5885
10	3.4200	41.7700	12.2136
11	5.5178	43.5500	7.8926
12	2.8845	32.9200	11.4127
13	4.8203	13.2700	2.7529
14	5.0000	10.9700	2.1940
15	2.7589	41.1500	14.9152
16	2.2894	52.3200	22.8529
17	4.5789	43.2800	9.4521
18	5.5178	52.9700	9.5998
19	2.0000	30.9700	15.4850
20	4.1602	40.4300	9.7184
21	4.1213	37.9600	9.2107
22	4.1602	29.9700	7.2040
23	4.1213	30.8100	7.4758
24	4.4480	43.5100	9.7820
25	2.8845	11.7700	4.0804
26	4.1213	19.6500	4.7679
27	5.2415	33.4500	6.3818
28	5.6462	42.4300	7.5148
29	3.1748	38.6300	12.1677
30	4.1602	22.8100	5.4830
31	2.7589	29.6900	10.7614
32	4.5789	39.2100	8.5633
33	2.8845	61.8900	21.4561
34	4.5789	20.7200	4.5251
35	3.9791	19.4100	4.8780
36	5.7690	30.0300	5.2054
37	3.0000	38.3700	12.7900
38	2.0000	41.3300	20.6650
39	6.5421	52.0900	7.9622
40	2.8845	5.6100	1.9449
41	3.2711	39.1600	11.9716
42	4.3795	40.0600	9.1471
43	4.0000	11.5500	2.8875
44	4.9324	52.9000	10.7249

Appendix 2. Continued

t	GM of day	GM of evening	GMD/GME
45	5.7388	47.8300	8.3345
46	5.0397	53.2800	10.5721
47	2.8845	9.4400	3.2727
48	2.7589	11.7900	4.2734
49	2.0000	8.0400	4.0200
50	3.0000	11.1500	3.7167
51	3.8259	31.0900	8.1263
52	4.1602	28.7900	6.9204
53	2.1544	10.0300	4.6555
54	5.7690	32.0100	5.5486
55	3.6342	9.4400	2.5975
56	5.0397	10.3800	2.0597
57	4.8203	29.7800	6.1781
58	3.1748	42.7600	13.4686
59	3.6342	9.6500	2.6553
60	4.3795	11.2000	2.5574
61	3.4200	51.0000	14.9125
62	2.8845	19.5200	6.7672
63	2.2894	40.6000	17.7337
64	4.1213	29.9300	7.2623
65	3.6342	33.2100	9.1381
66	3.4200	19.9600	5.8363
67	7.3186	41.6200	5.6869
68	2.8845	12.3200	4.2711
69	3.4200	30.8200	9.0118
70	4.9324	11.8900	2.4106
71	2.8845	43.4400	15.0598
72	4.8203	8.7100	1.8069
73	2.8845	9.7800	3.3905
74	3.3019	12.9000	3.9068
75	3.7798	53.6900	14.2046
76	2.5198	11.6200	4.6114
77	3.6342	12.1300	3.3377
78	2.8845	41.0300	14.2243
79	1.4422	10.2600	7.1139
80	4.0000	20.8600	5.2150
81	6.5421	10.2600	1.5683
82	6.6494	12.9700	1.9506
83	4.5789	21.8100	4.7632

Appendix 3. Table and Figures using Parameter Estimation of Lognormal Distribution

σ \ n	10	15	20	25	30	35	40	45
.01	.42615	.33575	.29061	.26272	.24341	.22906	.21786	.2088
.02	.43007	.33975	.29457	.26664	.24727	.23288	.22164	.2125
.03	.43411	.34386	.29864	.27065	.25123	.23679	.22551	.2163
.04	.43826	.34807	.30281	.27476	.25529	.24080	.22947	.2203
.05	.44253	.35239	.30708	.27897	.25945	.24491	.23354	.2243
.10	.46574	.37572	.33011	.30168	.28186	.26706	.25546	.2460
.15	.49230	.40216	.35617	.32736	.30722	.29214	.28030	.2706
.20	.52257	.43212	.38565	.35643	.33593	.32055	.30845	.2986
.25	.55700	.46603	.41901	.38933	.36845	.35275	.34038	.3303
.30	.59609	.50444	.45678	.42660	.40532	.38928	.37663	.3663
.35	.64046	.54793	.49958	.46885	.44715	.43077	.41783	.4072
.40	.69079	.59723	.54811	.51682	.49467	.47793	.46470	.4538
.45	.74792	.65317	.60322	.57132	.54871	.53161	.51807	.5070
.50	.81279	.71670	.66587	.63334	.61026	.59278	.57894	.5676
.55	.88655	.78897	.73719	.70402	.68045	.66259	.64844	.6368
.60	.97050	.87129	.81852	.78467	.76061	.74237	.72792	.7160
.65	1.06620	.96522	.91142	.87688	.85231	.83369	.81894	.8068
.70	1.17548	1.07261	1.01773	.98248	.95741	.93840	.92335	.9110
.75	1.30051	1.19562	1.13962	1.10364	1.07806	1.05868	1.04333	1.0307
.80	1.44387	1.33683	1.27966	1.24295	1.21686	1.19710	1.18145	1.1686
.85	1.60860	1.49928	1.44091	1.40345	1.37685	1.35672	1.34079	1.3277
.90	1.79834	1.68661	1.62700	1.58879	1.56168	1.54119	1.52499	1.5117
.94	1.97105	1.85730	1.79667	1.75785	1.73034	1.70955	1.69314	1.6797
.95	2.01746	1.90319	1.84230	1.80333	1.77571	1.75486	1.73840	1.7249
.96	2.06524	1.95045	1.88931	1.85018	1.82247	1.80154	1.78502	1.7715
.97	2.11446	1.99914	1.93774	1.89845	1.87064	1.84964	1.83307	1.8195
.98	2.16515	2.04929	1.98763	1.94819	1.92027	1.89921	1.88259	1.8690
.99	2.21736	2.10097	2.03904	1.99945	1.97143	1.95029	1.93362	1.9200
1.00	2.27116	2.15422	2.09202	2.05228	2.02416	2.00295	1.98623	1.9726
1.05	2.56573	2.44598	2.38244	2.34191	2.31329	2.29174	2.27477	2.2609
1.10	2.90875	2.78605	2.72112	2.67980	2.65068	2.62879	2.61158	2.5976
1.15	3.30941	3.18361	3.11725	3.07513	3.04551	3.02328	3.00584	2.9916
1.20	3.77887	3.64980	3.58196	3.53903	3.50891	3.48635	3.46868	3.4543
1.25	4.33073	4.19823	4.12886	4.08509	4.05447	4.03159	4.01369	3.9992
1.30	4.98165	4.84552	4.77456	4.72995	4.69882	4.67561	4.65750	4.6428
1.35	5.75206	5.61211	5.53950	5.49401	5.46236	5.43883	5.42050	5.4057
1.40	6.66717	6.52316	6.44882	6.40244	6.37026	6.34639	6.32785	6.3129
1.45	7.75811	7.60980	7.53366	7.48634	7.45362	7.42941	7.41064	7.3955
1.50	9.06354	8.91066	8.83263	8.78433	8.75104	8.72649	8.70750	8.6922
1.55	10.63161	10.47387	10.39383	10.34451	10.31064	10.28573	10.26650	10.2511
1.60	12.52251	12.35959	12.27743	12.22703	12.19254	12.16725	12.14778	12.1322
1.65	14.81178	14.64330	14.55889	14.50735	14.47221	14.44652	14.42680	14.4111
1.70	17.59453	17.42010	17.33329	17.28054	17.24471	17.21860	17.19860	17.1827
1.75	20.99102	20.81020	20.72082	20.66678	20.63021	20.60365	20.58335	20.5672
2.00	54.27287	54.05161	53.94619	53.88409	53.84292	53.81349	53.79132	53.7739

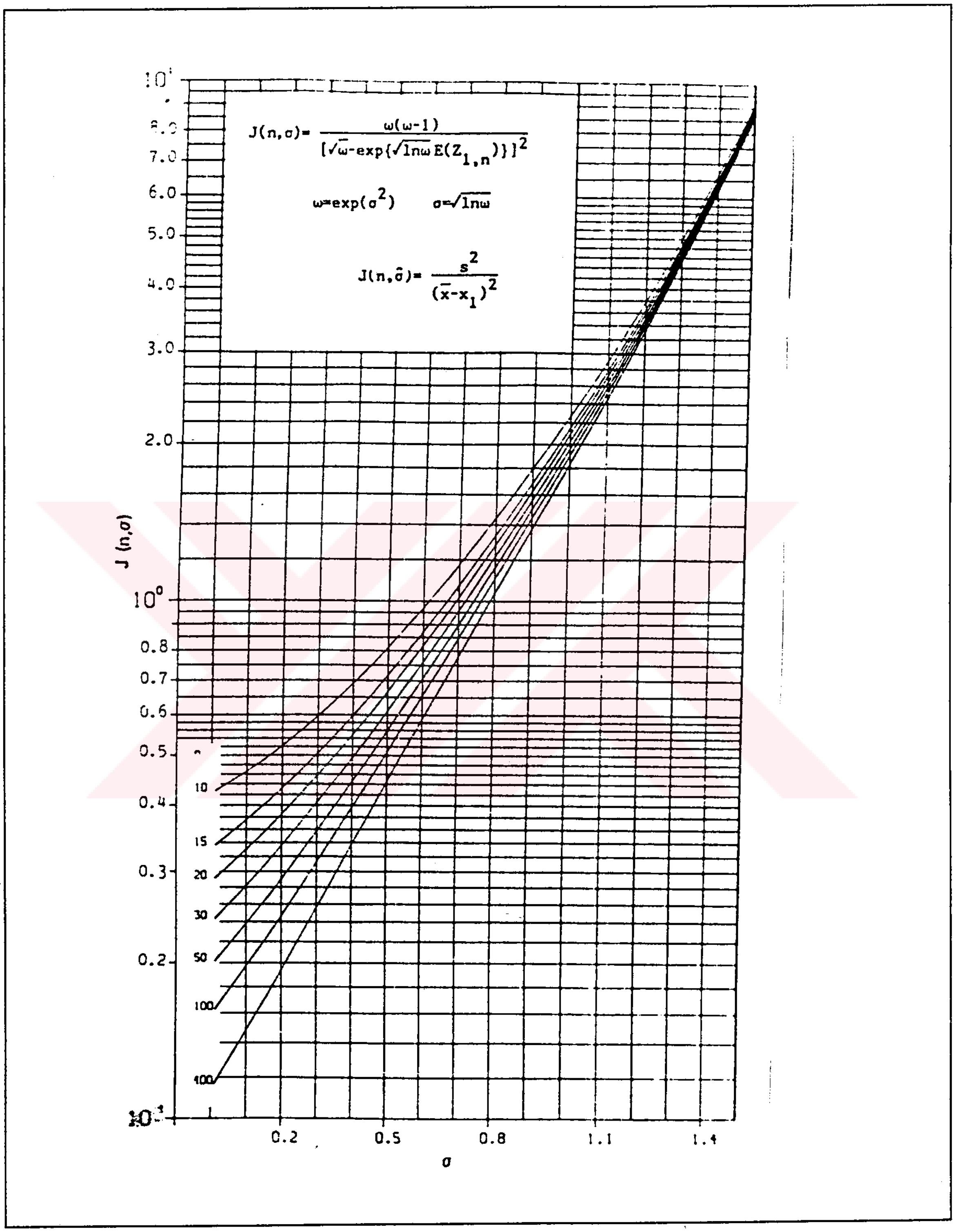
Appendix 3. Continued

δ \ n	50	55	60	65	70	75	80	85
.01	.20131	.19494	.18946	.18467	.18044	.17666	.17327	.1701
.02	.20501	.19862	.19311	.18829	.18404	.18024	.17682	.1737
.03	.20881	.20238	.19685	.19201	.18773	.18391	.18047	.1773
.04	.21270	.20625	.20068	.19581	.19151	.18767	.18421	.1810
.05	.21669	.21021	.20461	.19972	.19539	.19153	.18805	.1849
.10	.23823	.23159	.22584	.22082	.21637	.21239	.20881	.2055
.15	.26267	.25585	.24995	.24479	.24021	.23612	.23243	.2290
.20	.29039	.28340	.27735	.27204	.26733	.26312	.25932	.2558
.25	.32189	.31471	.30850	.30304	.29821	.29387	.28996	.2864
.30	.35769	.35033	.34395	.33835	.33338	.32892	.32489	.3212
.35	.39843	.39088	.38434	.37858	.37348	.36890	.36476	.3609
.40	.44483	.43709	.43038	.42448	.41924	.41454	.41029	.4064
.45	.49774	.48981	.48293	.47688	.47151	.46669	.46233	.4583
.50	.55813	.55002	.54298	.53678	.53127	.52634	.52187	.5178
.55	.62717	.61887	.61166	.60533	.59969	.59464	.59007	.5859
.60	.70618	.69770	.69034	.68386	.67810	.67294	.66827	.6640
.65	.79674	.78809	.78057	.77396	.76809	.76282	.75805	.7537
.70	.90070	.89188	.88422	.87748	.87150	.86613	.86128	.8568
.75	1.02025	1.01126	1.00346	.99661	.99052	.98506	.98012	.9756
.80	1.15796	1.14882	1.14089	1.13392	1.12773	1.12218	1.11718	1.1126
.85	1.31690	1.30761	1.29955	1.29248	1.28620	1.28058	1.27550	1.2708
.90	1.50071	1.49128	1.48311	1.47594	1.46958	1.46389	1.45875	1.4540
.94	1.66857	1.65904	1.65078	1.64354	1.63712	1.63138	1.62620	1.6214
.95	1.71375	1.70419	1.69592	1.68866	1.68223	1.67647	1.67128	1.6665
.96	1.76031	1.75073	1.74243	1.73516	1.72871	1.72294	1.71774	1.7130
.97	1.80829	1.79868	1.79036	1.78307	1.77661	1.77083	1.76562	1.7608
.98	1.85774	1.84810	1.83977	1.83246	1.82599	1.82020	1.81498	1.8102
.99	1.90870	1.89905	1.89069	1.88337	1.87688	1.87108	1.86585	1.8611
1.00	1.96124	1.95156	1.94319	1.93585	1.92935	1.92354	1.91830	1.9135
1.05	2.24945	2.23966	2.23119	2.22379	2.21723	2.21137	2.20610	2.2013
1.10	2.58595	2.57606	2.56751	2.56004	2.55343	2.54754	2.54223	2.5374
1.15	2.97992	2.96993	2.96131	2.95378	2.94713	2.94120	2.93587	2.9310
1.20	3.44248	3.43240	3.42372	3.41614	3.40946	3.40350	3.39815	3.3933
1.25	3.98722	3.97707	3.96833	3.96071	3.95399	3.94801	3.94264	3.9377
1.30	4.63077	4.62054	4.61174	4.60408	4.59734	4.59134	4.58596	4.5811
1.35	5.39352	5.38321	5.37437	5.36668	5.35991	5.35390	5.34852	5.3436
1.40	6.30062	6.29025	6.28136	6.27364	6.26685	6.26083	6.25544	6.2505
1.45	7.38317	7.37273	7.36380	7.35605	7.34924	7.34321	7.33782	7.3329
1.50	8.67978	8.66927	8.66029	8.65251	8.64569	8.63966	8.63426	8.6294
1.55	10.23852	10.22795	10.21892	10.21111	10.20428	10.19823	10.19284	10.1879
1.60	12.11953	12.10888	12.09981	12.09198	12.08512	12.07907	12.07368	12.0688
1.65	14.39827	14.38754	14.37842	14.37055	14.36367	14.35761	14.35221	14.3473
1.70	17.16976	17.15895	17.14977	17.14187	17.13497	17.12889	17.12349	17.1186
1.75	20.55419	20.54329	20.53405	20.52610	20.51918	20.51308	20.50767	20.5028
2.00	53.75998	53.74844	53.73874	53.73045	53.72329	53.71703	53.71150	53.7065

Appendix 3. Continued

σ \ n	90	95	100	125	150	200	400	1000[a]
.01	.16738	.16480	.16243	.15283	.14576	.13582	.11654	.1076
.02	.17090	.16831	.16592	.15625	.14913	.13911	.11965	.1106
.03	.17451	.17190	.16950	.15976	.15259	.14249	.12286	.1137
.04	.17822	.17559	.17317	.16337	.15614	.14596	.12616	.1169
.05	.18202	.17938	.17694	.16707	.15979	.14953	.12956	.1203
.10	.20259	.19986	.19734	.18713	.17959	.16892	.14807	.1383
.15	.22601	.22320	.22060	.21004	.20223	.19115	.16942	.1592
.20	.25270	.24980	.24712	.23621	.22812	.21664	.19403	.1834
.25	.28314	.28015	.27739	.26613	.25776	.24588	.22240	.2113
.30	.31788	.31480	.31195	.30034	.29170	.27942	.25510	.2436
.35	.35755	.35437	.35144	.33948	.33058	.31791	.29278	.2809
.40	.40287	.39962	.39660	.38430	.37515	.36209	.33620	.3239
.45	.45472	.45138	.44828	.43566	.42625	.41284	.38622	.3736
.50	.51408	.51065	.50748	.49454	.48489	.47115	.44387	.4309
.55	.58210	.57859	.57534	.56210	.55224	.53818	.51030	.4971
.60	.66013	.65654	.65323	.63971	.62964	.61529	.58689	.5735
.65	.74975	.74609	.74272	.72894	.71868	.70408	.67523	.6616
.70	.85282	.84910	.84566	.83164	.82122	.80640	.77719	.7635
.75	.97153	.96775	.96425	.95003	.93946	.92445	.89496	.8812
.80	1.10845	1.10462	1.10108	1.08667	1.07598	1.06082	1.03113	1.0173
.85	1.26667	1.26279	1.25920	1.24463	1.23384	1.21857	1.18876	1.1749
.90	1.44982	1.44590	1.44228	1.42758	1.41671	1.40135	1.37152	1.3577
.94	1.61719	1.61324	1.60960	1.59481	1.58389	1.56850	1.53870	1.5250
.95	1.66226	1.65830	1.65465	1.63985	1.62892	1.61352	1.58374	1.5701
.96	1.70870	1.70474	1.70109	1.68627	1.67533	1.65993	1.63016	1.6165
.97	1.75657	1.75261	1.74895	1.73411	1.72316	1.70776	1.67801	1.6644
.98	1.80591	1.80194	1.79828	1.78342	1.77247	1.75706	1.72734	1.7137
.99	1.85677	1.85280	1.84913	1.83426	1.82330	1.80789	1.77820	1.7646
1.00	1.90921	1.90523	1.90155	1.88667	1.87571	1.86030	1.83064	1.8171
1.05	2.19694	2.19294	2.18925	2.17431	2.16334	2.14795	2.11848	2.1051
1.10	2.53303	2.52901	2.52531	2.51034	2.49937	2.48403	2.45482	2.4417
1.15	2.92664	2.92260	2.91889	2.90392	2.89298	2.87771	2.84882	2.8359
1.20	3.38889	3.38485	3.38114	3.36618	3.35528	3.34011	3.31158	3.2989
1.25	3.93337	3.92933	3.92562	3.91070	3.89984	3.88480	3.85667	3.8443
1.30	4.57669	4.57265	4.56894	4.55407	4.54328	4.52837	4.50068	4.4886
1.35	5.33924	5.33521	5.33151	5.31669	5.30598	5.29123	5.26400	5.2522
1.40	6.24617	6.24215	6.23845	6.22371	6.21308	6.19849	6.17175	6.1602
1.45	7.32856	7.32455	7.32087	7.30620	7.29566	7.28124	7.25499	7.2438
1.50	8.62501	8.62101	8.61734	8.60276	8.59230	8.57806	8.55232	8.5414
1.55	10.18361	10.17961	10.17596	10.16145	10.15109	10.13703	10.11179	10.1012
1.60	12.06445	12.06047	12.05683	12.04241	12.03214	12.01825	11.99350	11.9832
1.65	14.34300	14.33903	14.33540	14.32105	14.31087	14.29716	14.27289	14.2628
1.70	17.11428	17.11032	17.10670	17.09243	17.08233	17.06878	17.04499	17.0352
1.75	20.49846	20.49450	20.49089	20.47669	20.46667	20.45328	20.42994	20.4204
2.00	53.70216	53.69817	53.69456	53.68049	53.67075	53.65796	53.63649	53.6281

Appendix 3. Continued



Appendix 4. ARIMA Model

ARIMA model for lnevening

Estimates at each iteration

Iteration	SSE	Parameters	
0	66.9875	0.100	0.104
1	59.8159	0.250	0.052
2	55.5837	0.400	0.019
3	54.3922	0.546	0.001
4	54.3478	0.523	0.002
5	54.3476	0.525	0.002
6	54.3476	0.525	0.002

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type Coef StDev T

Type	Coef	StDev	T
MA 1	0.5247	0.0951	5.52
Constant	0.00207	0.04341	0.05

Differencing: 1 regular difference

Number of observations: Original series 83, after differencing 82

Residuals: SS = 54.3465 (backforecasts excluded)

MS = 0.6793 DF = 80

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	3.8(DF=11)	10.1(DF=23)	33.9(DF=35)	46.2(DF=47)

Forecasts from period 83

		95 Percent Limits	
Period	Forecast	Lower	Upper
84	3.34037	1.72459	4.95616
85	3.34244	1.55346	5.13143
86	3.34451	1.39767	5.29135
87	3.34658	1.25376	5.43940
88	3.34865	1.11939	5.57790
89	3.35071	0.99290	5.70853
90	3.35278	0.87306	5.83250
91	3.35485	0.75895	5.95075
92	3.35692	0.64982	6.06402
93	3.35899	0.54507	6.17290