A COMPARISON ON STATISTICAL TECHNIQUES BETWEEN SINGLE AND COMBINED FORECASTS : A CASE STUDY ON TRAFFIC ACCIDENT DATA IN BANGKOK

JUNTANEE VONGSPANICH

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE (BIOSTATISTICS)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2000

ISBN 974-664-477-7

Thesis
entitled

A COMPARISON ON STATISTICAL TECHNIQUES BETWEEN SINGLE AND COMBINED FORECASTS : A CASE STUDY ON TRAFFIC ACCIDENT DATA IN BANGKOK

Juntanee Vongspanich
....
Miss Juntanee Vongspanich
Candidate

Thavatchai Vorapongsathorn
....
Associate Prof. Thavatchai Vorapongsathorn, Ph.D.
Major-Advisor

Chukiat Viwatwongkasem
....
Associate Prof. Chukiat Viwatwongkasem, M.Sc
Co-Advisor

....
Prof. Liangchai Limlomwongse, Ph.D.
Dean
Faculty of Graduate Studies

....
Assistant Prof. Junya Pattaraarchachai, Sc.D.
Chairman
Master of Science Programme in Biostatistics
Faculty of Public Health

Thesis
entitled

A COMPARISON ON STATISTICAL TECHNIQUES BETWEEN SINGLE AND COMBINED FORECASTS : A CASE STUDY ON TRAFFIC ACCIDENT DATA IN BANGKOK

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Biostatistics)
on
July 20, 2000

Juntanee Vongspanich
Miss Juntanee Vongspanich
Candidate

Thavatchai Vorapongsathorn
Associate Prof. Thavatchai Vorapongsathorn, Ph.D.
Chairman

Chukiat Viwatwongkasem
Associate Prof. Chukiat Viwatwongkasem, M.Sc.
Member

Assistant Prof. Dusit Sujirarat, M.Sc.
Member

Prof. Liangchai Limlomwongse, Ph.D.
Dean
Faculty of Graduate Studies
Mahidol University

Kanda Vathanophas
Associate Prof. Kanda Vathanophas, M.D., M.Sc.in Hygiene.(P.H. Microbiology)
Dean
Faculty of Public Health
Mahidol University

ACKNOWLEDGEMENT

There were many persons involved in this study. I would like to express my sincere gratitude and deep appreciation to my advisor, Associate Professor Dr. Thavatchai Vorapongsathorn for his advices, helpfulness, comments and understanding throughout the process of writing this study.

My grateful deep appreciation have been given to my co-advisor, Associate Professor Chukiat Viwatwongkasem and Assistant Professor Dusit Sujirarat, for their kindness and generous suggestions.

Sincere appreciation is also expressed to everybody who help this thesis to be completed.

In particular, I wish to pay tribute to my parents and my cousins for their encouragement and support my studies.

Juntanee Vongspanich

3836005 PHBS/M : MAJOR : BIOSTATISTICS; M.Sc. (BIOSTATISTICS)
KEY WORDS : FORECASTING / ARIMA / HOLT-WINTERS / LEAST-SQUARES METHOD / COMBINED FORECAST

JUNTANEE VONGSPANICH: A COMPARISON ON STATISTICAL TECHNIQUES BETWEEN SINGLE AND COMBINED FORECASTS : A CASE STUDY ON TRAFFIC ACCIDENT DATA IN BANGKOK. THESIS ADVISOR: THAVACHAI VORAPONGSATHORN, Ph.D., CHUKIAT VIWATWONGKASEM M.Sc. 114 p. ISBN 974-664-477-7

Forecasting is used to predict future values of a variable based on known past values of that variable or other related variables. There are many factors to consider when choosing the most appropriate forecasting procedure for a given set of conditions. The purpose of this study was to compare three single and combined forecasting methods using traffic accident time series data in Bangkok. The single techniques considered were the Least Squares procedure for fitting an exponential curve, Holt-Winters and the Box-Jenkins methods. Three combined forecasts techniques were calculated using pairs of single forecasting procedures: Least Squares + Holt-Winters, Least Squares + Box-Jenkins and Holt-Winters + Box-Jenkins. Comparisons were made using data from different times in the past (5, 12, 36 and 60 months) to forecast future periods (3, 6 and 12 months). Methods were combined using Bates, Granger and Newbold procedures to take weighted averages. Then, the statistical indicators, the mean square error (MSE), the average mean square error (AMSE) and the mean absolute percentage errors (MAPE) were calculated to evaluate the accuracy of each method.

The findings of this study revealed that on average the combined forecast techniques yielded higher accuracy in forecasting the number of traffic accidents in Bangkok than the single forecast approaches. For the single forecast methods, the Least Squares method was appropriate for forecasting short distances in the future (3 months) using the past data from 5 and 12 months. The Box-Jenkins procedure was efficient for forecasting 3 months in the future using data from 60 months in the past. The Holt and Winters method was appropriate for forecasting farther into the future (6 and 12 months) using data from 12 and 60 months in the past. For the combined forecasting procedures, combining Holt-Winters and the Box-Jenkins methods gave low errors in forecasting for both short and long periods using data from 60 months in the past. This study recommends using this efficient combined method to reliably predict the number of traffic accidents. These data should be very useful in policy making and administrative planning in order to prevent and reduce the number of traffic accidents in Bangkok.

3836005 PHBS/M: สาขาวิชา: ชีวสถิติ; วทม. (ชีวสถิติ)

จันทนี วงศ์พานิช : การเปรียบเทียบเทคนิคทางสถิติโดยวิธีพยากรณ์เดี่ยวและวิธีพยากรณ์รวม : กรณีศึกษาข้อมูลอุบัติเหตุจราจรในเขตกรุงเทพมหานคร (A COMPARISON ON STATISTICAL TECHNIQUES BETWEEN SINGLE AND COMBINED FORECASTS : A CASE STUDY ON TRAFFIC ACCIDENT DATA IN BANGKOK) คณะกรรมการควบคุมวิทยานิพนธ์ : ธวัชชัย วรพงศธร, Ph.D, ชูเกียรติ วิวัฒน์วงศ์เกษม, M.Sc. 114 หน้า ISBN 974-664-477-7

การพยากรณ์เป็นการคาดการณ์ค่าของตัวแปรในอนาคตโดยใช้ค่าของตัวแปรนั้นหรือค่าของตัวแปรที่เกี่ยวข้องในอดีตมาช่วยในการพยากรณ์ การเลือกวิธีพยากรณ์ที่เหมาะสมจะต้องพิจารณาปัจจัยหลายปัจจัยประกอบกัน การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีพยากรณ์เดี่ยวกับวิธีพยากรณ์รวม โดยใช้ข้อมูลอนุกรมเวลาของจำนวนคดีอุบัติเหตุจราจรทางบกที่เกิดขึ้นในกรุงเทพมหานคร วิธีพยากรณ์เดี่ยวมี 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุดที่มีแนวโน้มลักษณะเอ็กซ์โพเนนเชียล วิธีโฮลท์-วินเทอร์ และวิธีบ๊อกซ์-เจนกินส์ สำหรับวิธีพยากรณ์รวมได้ใช้วิธีเฉลี่ยถ่วงน้ำหนักของเบทส์ แกรงเจอร์ และนิวโบลด์ เป็นเทคนิคในการรวมวิธีพยากรณ์เดี่ยว 3 คู่เข้าด้วยกัน วิธีพยากรณ์รวมได้แก่ วิธีกำลังสองน้อยที่สุดรวมกับวิธีโฮลท์-วินเทอร์ วิธีกำลังสองน้อยที่สุดรวมกับวิธีบ๊อกซ์-เจนกินส์ และวิธีโฮลท์-วินเทอร์รวมกับวิธีบ๊อกซ์-เจนกินส์ การเปรียบเทียบกระทำภายใต้สถานการณ์ของการใช้ข้อมูลอุบัติเหตุจราจรที่เกิดขึ้นในอดีตในช่วง 5, 12, 36 และ 60 เดือน ทำการพยากรณ์ข้อมูลอุบัติเหตุจราจรที่เกิดขึ้นล่วงหน้าในช่วงเวลา 3, 6 และ 12 เดือน แล้วคำนวณค่าความคลาดเคลื่อนจากการพยากรณ์เพื่อประเมินความแม่นตรงของวิธีพยากรณ์แต่ละวิธี

ผลการศึกษาพบว่า โดยเฉลี่ยแล้ววิธีพยากรณ์รวมสามารถพยากรณ์ข้อมูลอุบัติเหตุจราจรที่เกิดขึ้นล่วงหน้าในกรุงเทพมหานคร ได้แม่นตรงกว่าวิธีพยากรณ์เดี่ยว สำหรับวิธีพยากรณ์เดี่ยววิธีกำลังสองน้อยที่สุดเหมาะสมสำหรับพยากรณ์ข้อมูลที่จะเกิดในอนาคตในช่วงสั้นๆ 3 เดือน โดยใช้ข้อมูลที่เกิดขึ้นในอดีต 5 และ 12 เดือน เช่นเดียวกับวิธีบ๊อกซ์-เจนกินส์ เหมาะสมสำหรับพยากรณ์ข้อมูลที่เกิดขึ้นในอนาคตในช่วงสั้นๆ 3 เดือน โดยใช้ข้อมูลที่เกิดขึ้นในอดีต 60 เดือน สำหรับวิธีโฮลท์-วินเทอร์เหมาะสมสำหรับพยากรณ์ข้อมูลที่จะเกิดในอนาคตในช่วง 6 ถึง 12 เดือน โดยใช้ข้อมูลในอดีต 12 และ 60 เดือน วิธีพยากรณ์รวมที่เหมาะสมได้แก่วิธีโฮลท์-วินเทอร์รวมกับวิธีบ๊อกซ์-เจนกินส์ ซึ่งสามารถพยากรณ์ข้อมูลที่จะเกิดในอนาคตได้ทั้งในระยะสั้นและระยะยาว โดยใช้ข้อมูลในอดีต 60 เดือน ผลพยากรณ์ของวิธีนี้มีความคลาดเคลื่อนต่ำ การศึกษานี้จึงเสนอแนะให้ใช้วิธีพยากรณ์รวมนี้ในการทำนายข้อมูลอุบัติเหตุจราจร ซึ่งคาดว่าข้อมูลนี้จะเป็นประโยชน์ในการกำหนดนโยบายและการวางแผนการบริหารงานเพื่อป้องกันและลดอุบัติเหตุจราจรในกรุงเทพมหานครให้ลดน้อยลงได้

CONTENTS

LIST OF TABLES

Appendix A

LIST OF FIGURES

Figures **Page**

Appendix D.

CHAPTER I

INTRODUCTION

A. Background and Problem

Forecasting is the prediction of future values of a variable based on known past values of that variable or other related variables. Forecasts also may be based on expert judgements, which in turn are based on historical data and experience (1).

Forecasting is an important aid in effective and efficient planning and deciding about the development of every organization. There were many forecasting methods. However, quantitative forecasting techniques vary considerably, having been developed by diverse disciplines for different purposes. Taesombut (2) has summarized the important factors to be considered when choosing the following forecasting procedures: Simple Exponential Smoothing methods, Holt and Winter's Seasonal Smoothing methods, Decomposition methods, the Box – Jenkins methods, Regression methods, Econometric models and Multivariate Autoregressive Moving Average (ARMA) models. These important factors were the purpose of the forecast, nature of data, length of forecast, size of data and costs of the forecast (Table 1).

These procedures have been called the single forecasts or traditional approaches. They have been extensively applied to forecast the data in various areas, such as economics (3), agriculture, industrial products, financial and banking, international trading, taxation and tour (4), epidemiology and public health (5) (6).

In recently, there are many statisticians (3) (7) (8) have suggested the combination of forecast, an alternative to the traditional approach, to aggregate information from different forecasting methods. This eliminates the problem of having to select a single method and rely exclusively on its forecasts.

Table 1. Comparing Quantitative Forecasting Methods by Several Factors.

Factors	Forecasting Methods						
	SES	HWS	D	BJ	REG	ECO	MARMA
1.Movement of Data							
Horizontal	✓	-	✓	✓	-	-	✓
Trend	-	✓	✓	✓	✓	✓	✓
Seasonal	-	✓	✓	✓	✓	✓	✓
Cyclical	-	-	✓	-	✓	✓	-
2.The length of times							
Immediate Term	✓	✓	✓	✓	-	-	-
Short Term	✓	✓	✓	✓	✓	✓	✓
Medium Term	-	-	-	-	✓	✓	✓
Long Term	-	-	-	-	✓	✓	✓
3.Size of Data (Trend)							
(s- The length of the seasonal)	10	15 2(s)	30 6(s)	30 6(s)	30 6(s)	small 100	60 8(s)
4.Costs	1	2	3	5	4	7	6

SES	Simple Exponential Smoothing Methods
HWS	Holt and Winter's Seasonal Smoothing Methods
D	Decomposition Methods
BJ	The Box – Jenkins Methods
REG	Regression Methods
ECO	Econometric Models
MARMA	Multivariate Autoregressive Moving Average (ARMA) Models

The combination of forecasts has been tried before on a somewhat limited basis with considerable success. Methods for combining forecasts in terms of weighted averages were discussed in Newbold and Granger (3). In Newbold and Granger (3), the authors concluded that:

> It does appear…that Box – Jenkins forecasts can frequently be improved upon by combination with either Holt – Winters or stepwise autoregressive forecasts, and results indicate that in any particular forecasting situation combining is well worth trying, as it requires very little effort. Further improvement is frequently obtained by considering a combination of all three types of forecast (p.143).

From various studies of forecasting (3)(8), the results revealed that the combined forecast methods give minimum mean square error (MSE) and mean absolute percentage error (MAPE) better than the single forecast methods. In overall, the combined forecasts were better than the single methods (7)(8). However, this conclusion was come from the researches, which were mostly based on the simulation of data or even the actual data from main parts of economics and industry. So, it needs to confirm this result by applying the single and combined forecasts to the actual data in different fields, especially in health and traffic accidents.

In case of Thailand, there are few evidences of studies dealing with an application of appropriate forecasting techniques in preventing and reducing the traffic accident problems. In particular in Bangkok, the number of traffic accidents has been increasing as a consequence of increase number of automobiles. From annual reports of police information system center in 1991-1995, morbidity and mortality from road traffic accidents in Bangkok were double increase, eventhough there were many underreports. For example, there were 10,701 injured persons in 1991 and increased to 21,697 cases in 1995 (9).

The traffic accidents in Bangkok create not only health and psychological problems but also socioeconomic losses in society (9)(10). The economic loss in Bangkok was estimated as 154 million Baht in 1991 and increased to 497 million Baht in 1995 (9). Each year, traffic accidents caused high number of manpower loss who were otherwise formed and important component for the development of the country. In view of health loss, the injured persons need longer time for treatment more than other disease and for this reason the cost of treatment is expensive. In overall, the injured and disable persons have become a big burden both to their families and the community.

At present, the traffic accident problem in Bangkok has become more serious and need to be solved or reduced continuously. One way in helping to reduce the problem is to get the traffic information in future for traffic planning. This information can be predicted by using the appropriate forecasting technique. The reliable traffic accident information will be surely useful to many persons and organizations involved for policy making and administrative planning, such as policeman, traffic authorities, hospitals and insurance companies.

Since there are many forecasting methods, this study proposes to determine the appropriate forecast approach to predict reliable traffic accident in different future periods.

B. Objectives

This study has the main objectives as follows:

1. To compare the accuracy between single and combined forecast methods using the mean absolute percentage error (MAPE) as the accuracy indicator.
2. To determine appropriate sizes of the past traffic accident data used for each single forecast method.
3. To find an appropriate forecast technique for predicting the traffic accident data in Bangkok.

C. Scope of the Study

This study used the traffic accident data in Bangkok from police information system center, the royal Thai police department. The data were monthly time series during January, 1992 to December, 1996. The data were forecast forward and compared with actual data during January, 1997 to December, 1997.

D. Definitions of Variables

Accident

Accident means an event independent of the will of man, caused by a quickly action extraneous manifesting itself by injury of the body or mind (10).

Accuracy

Accuracy refers to the correctness of the forecast as measured against actual events. Accuracy can be measured using such dimensions as mean squared error (MSE); mean absolute percentage error (MAPE); or mean percentage error or bias (MPE) (1).

ARIMA

ARIMA is abbreviated from autoregressive (AR) –integrated (I) –moving average (MA).

Autoregression (AR) is a form of regression, but instead of the dependent variable (the item to be forecast) being related to independent variables, it is related to past values of itself at varying time lags. Thus an autoregressive model would express the forecast as a function of previous values of that time series.

Integrated (I): This is often an element of time series models where one or more of the differences of the time series are included in the model. The term comes from the fact that the original series may be recreated from a differenced series by a process of "integration" (involving a summation in the typical discrete environment).

Moving Average (MA): There are two distinct meanings to this term. First, for a time series, the moving average of order K means the average (mean) value of the last *K* observations. Second, in Box-Jenkins modeling the MA in ARIMA stands for "moving average" and means that the value of the time series at time *t* is influenced by a current error term and (possibly) weighted error terms in the past (1).

Autoregressive/Moving Average (ARMA) Scheme

It is a type of time-series forecasting model, which can be autoregressive (AR) in form, moving average (MA) in form, or a combination of the two (ARMA). In an ARMA model, the series to be forecast is expressed as a function of both previous values of the series (autoregressive terms) and previous error values from forecasting (the moving average terms) (1).

Box-Jenkins Method

Box-Jenkins Method is a single method of forecasting, which was developed by Box and Jenkins. The procedure of this technique consists of three major steps: model identification, parameter estimation and residual diagnostic checking.

Combined Forecasts

Combined forecasts means the forecast methods which used the combination of single forecasting methods in terms of weighted averages, simple averages and Bayesian methods. This study use combined forecasts via Weighted Averages by Bates, Granger and Newbold (BGN 's method) (11).

Differencing

Differencing is a method of converting nonstationary time-series data to stationary form. When a time series is nonstationary, it can often be made stationary by taking first differences of the series, creating a new time series of successive differences, (X_t - X_{t-1}). If first differences do not convert the series to stationary form, then first differences of first differences can be created. This is called second-order differencing, (X_t - X_{t-2}) (1).

Exponential Smoothing, Single

It is the basic form of exponential smoothing. It uses the parameter alpha to smooth past values of the data and errors in forecast. It is most commonly used in inventory control systems where many items are to be forecast and low cost is a primary concern (1).

Forecasting

Forecasting is the prediction of values of a variable based on known past values of that variable or other related variables. Forecasts also may be based on expert judgments, which in turn are based on historical data and experience (1).

Holt-Winters' Exponential Smoothing Method

Holt-Winters' Exponential Smoothing Method is a technique of forecasting. Winters extended Holt's exponential smoothing method by including an extra equation that is used to adjust the forecast to reflect seasonality. This form of exponential smoothing can thus account for data series that include both trend and seasonal elements. It uses three smoothing parameters controlling the level, trend and seasonality (12).

Least Squares Method

This approach to estimating the parameter values is an equation minimizes the squares of the deviations that result from fitting that particular model (1).

Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error is the mean or average of the sum of all of the percentage errors for a given data set taken without regard to sign. That is, their absolute values are summed and the average computed. It is one measure of accuracy commonly used in quantitative methods of forecasting (1).

Mean Squared Error (MSE)

The mean squared error is a measure of accuracy computed by squaring the individual error for each item in a data set and then finding the average or mean value of the sum of those squares. The mean squared error gives greater weight to large errors than to small errors because the errors are squared before being summed (1).

Multivariate ARMA model

Multivariate ARMA model is a forecast model, which allows several time series. Each of the series is forecast using a function of its own past, the past of each of the other series, and past errors. This is in contrast to dynamic regression models where there is only one forecast variable and it is assumed that the explanatory variables do not depend on the past of forecast variable (1).

Nonstationary

Nonstationary is a form of time series data, which do not have a constant mean and/or a constant variance. A visual inspection of the plotted time series can help to determine if either or both of these conditions exist, and the set of autocorrelations for the time series can be used to confirm the presence of nonstationarity or not (1).

Stationary

Stationary is a form of time series data, which have a constant mean and a constant variance. More formally, a series is stationary if its statistical properties are independent of the particular time period during which it is observed (1).

Time Series

Time series refers to an ordered sequence of values of a variable observed at equally spaced time interval (1).

Traffic Accident

Traffic accident is an accident involving at least one moving vehicle. This study used the traffic accident data in Bangkok from police information system center, the royal Thai police department. The data were monthly time series during January, 1992 to December, 1996 (10).

White Noise

White noise refers to there is no pattern whatsoever in the data series. Thus, it is analogous to a series that is completely random (1).

CHAPTER II

LITERATURE REVIEW

Review of the literature in this chapter was divided into 6 main parts as follows:

A. Time Series

B. Autocorrelation and Partial Autocorrelation Coefficients

C. Box – Jenkins (ARIMA) Time Series

D. Combined Forecast

E. Comparison of Forecast and Actuality

F. Traffic Accident

A. Time Series

In this topic, two important properties of time series were reviewed: types of variation and stationary time series.

1. Types of Variation (12)

Traditional methods of time series analysis are mainly concerned with decomposing a series into a trend, a seasonal variation, and other "irregular" fluctuation. This approach is not always the best approach but is still often useful. These different sources of variation are described in more detail.

1.1. *Seasonal Effect*. Many time series, such as sales figures and temperature reading, exhibit a variation, are annual in period. These types of variation are easy to understand, and we see that they can be measured or removed from the data to give de-seasonalized data. For a quarterly seasonal pattern, the data might be similar to Figure1.

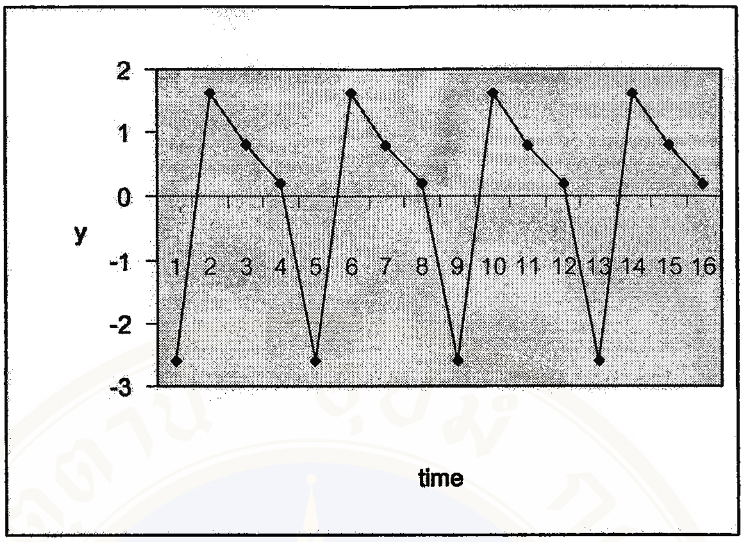


Figure 1. Seasonal Data Pattern

1.2. *Other Cyclic Change.* Apart from seasonal effects, some time series exhibit variation at a fixed period due to some other physical cause. An example is daily variation in temperature. In addition, some time series exhibit oscillations which do not have a fixed period but which are predictable to some extent. For example, economic data may be affected by business cycle with a period varying between about 5 and 7 years. This type of patterns is shown in Figure 2.

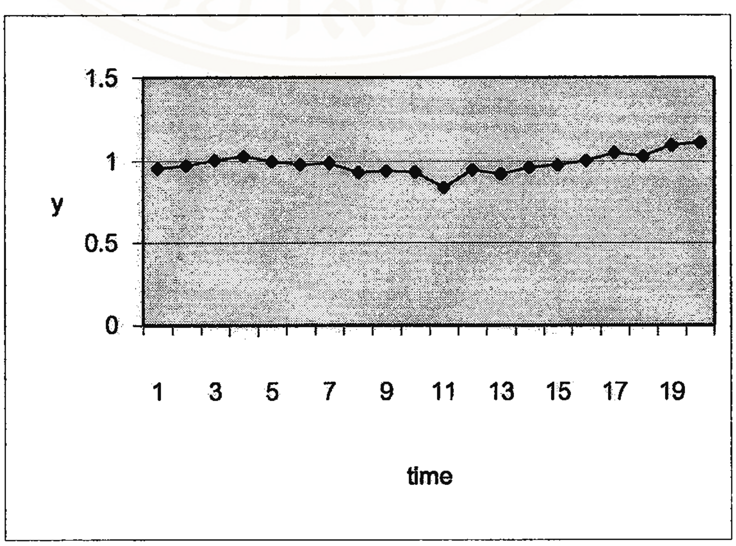


Figure 2. Cyclical Data Pattern

1.3. *Trend.* This may be loosely defined as "long term change in the mean". A difficulty with this definition is deciding what is meant by 'long term'. For example, climatic variables sometimes exhibit cyclic variation over a very long time – period such as 50 years. If one just had 20 years data, this long – term oscillation would appear to be a trend, but if several hundred years' data were available, the long – term oscillation would be visible. Nevertheless in the short term it may still be more meaningful to think of such a long – term oscillation as a trend. Thus, in speaking of a "trend", we must take into account the number of observations available and make a subjective assessment of what is 'long term'. This type of pattern is shown in Figure 3.

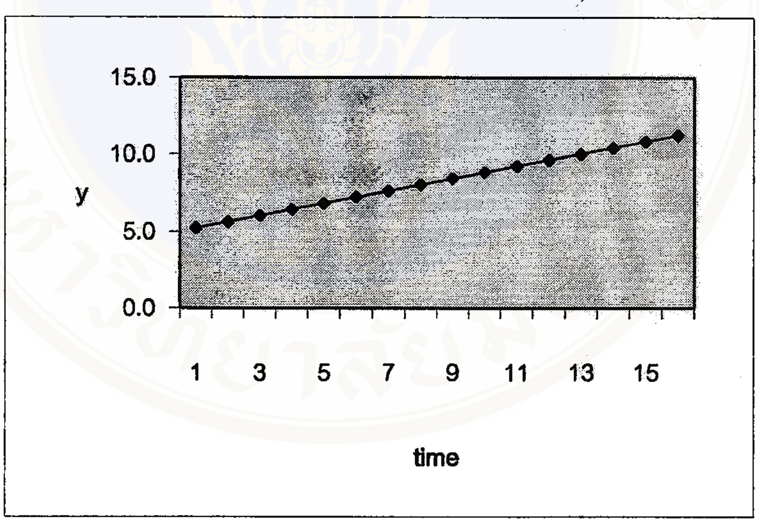


Figure 3. Trend Data Pattern

1.4. *Other Irregular Fluctuations.* After trend and cyclic variations have been removed from a set of data, examine various techniques for analyzing series of this type to see if some of the apparently irregular variation may be explained in terms of probability models, such as moving average or autoregressive models. This type of patterns is shown in Figure 4.

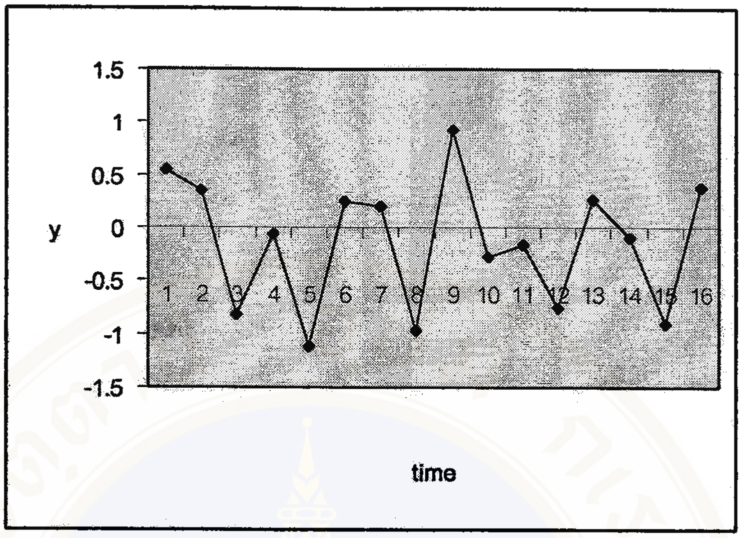


Figure 4. Irregular Data Pattern

2. Stationary Time Series

Broadly speaking, a time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance, and if strictly periodic variations have been removed.

Most of the probability theory of time series is concerned with stationary time series, and for this reason, time – series analysis often requires one to turn a non – stationary series into a stationary one so as to use this theory. For example, one may remove the trend and seasonal variation from a set of data and then try to model the variation in the residuals by means of a stationary stochastic process.

B. Autocorrelation and Partial Autocorrelation Coefficients (1,12)

An important guide to the properties of time series is provided by a series of quantities called sample autocorrelation coefficients, which measure the correlation between observations at different distances apart. These coefficients often provide insight into the probability model, which generated the data. Given N pairs of observations on two variable x and y, the correlation coefficient is given by

$$r = \frac{\Sigma\,(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\Sigma(x_i - \bar{x})^2\,\Sigma(y_i - \bar{y})^2]}} \qquad (1)$$

A similar idea can apply to time series to see if successive observations are correlated. Given N observations $x_1,\ldots, x_N$, on a discrete time series, form (N –1) pairs of observations, namely $(x_1, x_2), (x_2, x_3),\ldots,(x_{N-1}, x_N)$. Regarding the first observation in each pair as one variable, and the second observation as a second variable, the correlation coefficient between x_t and x_{t+1} is given by

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\left[\sum_{t=1}^{N-1}(x_t - \bar{x}_{(1)})^2\ \sum_{t=1}^{N-1}(x_{t+1} - \bar{x}_{(2)})^2\right]}} \qquad (2)$$

by analogy with equation (1), where

$\bar{x}_{(1)} = \sum_{t=1}^{N-1} x_t/(N-1)$ is the mean of the first (N – 1) observations and

$\bar{x}_{(2)} = \sum_{t=2}^{N-1} x_t/(N-1)$ is the mean of the last (N –1) observations.

As the coefficient given by equation (2) measures correlation between successive observations, it is called an autocorrelation coefficient or serial correlation coefficient.

For N reasonably large, r_1 is approximately given by

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})/(N-1)}{\sum_{t=1}^{N}(x_t - \bar{x})^2/N} \qquad (3)$$

where $\bar{x} = \sum_{t=1}^{N} x_t/N$ = Overall mean

Some authors also drop the factor N/ (N-1), which is closed to one for large N to give

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2} \tag{4}$$

In a similar way find the correlation between observations a distance k apart, which is given by

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2} \tag{5}$$

This is called the autocorrelation coefficient at lag k. The autocorrelation coefficients are usually calculated by computing the series of autocovariance coefficients $\{C_k\}$, which is defined by analogy with the usual covariance formula as

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) = \text{autocovariance coefficient at lag k.} \tag{6}$$

$$r_k = C_k / C_0 \tag{7}$$

for k = 1, 2,…,m, where $m \leq N$. There is usually little point in calculating r_k for values of k greater than about N/4. Note that some authors suggest

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

rather than equation (6), but there is little difference for large N.

A partial correlation coefficient is the measure of the relationship between two variables when the effect of other variables has been removed or held constant. Similarly, the partial autocorrelation coefficient (p_{kk}) is the measure of the relationship between the stationary time-series variables Z_t and Z_{t+k} when the effect of the intervening variables $Z_{t+1}, Z_{t+2}, \ldots, Z_{t+k-1}$ has been removed. This adjustment is made to see if the correlation between Z_t and Z_{t+k} is due to the intervening variables or if indeed three is something else causing the relationship. The behavior of the partial correlation coefficients (pacf's) for the stationary time series, along with the corresponding acf, is used to identify a tentative ARIMA model.

The sample partial autocorrelation coefficient can be computed using the following formula (where j takes on values from 1 to k-1):

$$r_{kk} = \frac{r_k - \Sigma\,(r_{k-1,j})\,(r_{k-j})}{1 - \Sigma\,(r_{k-1,j})\,(r_j)}$$

where r_k = the autocorrelation coefficient for k lags apart;

r_{kj} = the partial autocorrelation coefficient for k lags apart when the effect of j intervening lags has been removed; calculated by

$r_{kj} = r_{k-1,j} - (r_{kk})\,(r_{k-1,k-j})$; note that, by definition, $r_{11} = r_1$.

Because of the recursive nature of the formula, any attempt to compute the partial correlation coefficient (pacf) by hand is quite lengthy and time-consuming. So, the partial correlation coefficients (pacf) are mainly computed by a computer.

C. Box –Jenkins (ARIMA) Time Series

Helfenstein (5) reported that Box – Jenkins models may be useful for forecasting. Forecasts of epidemiological time series are needed for many reasons. For example, public health organizations may wish to know what frequencies of diseases have to be expected in the future in order better to plan the distribution of resources. Forecasting can also be used as a complementary method to intervention analysis. A forecast obtained from data before the intervention may be compared with actual data obtained after the intervention.

Studies on individual subjects may be interesting and relevant in basic medical research and in clinical applications. In basic research they have great potential for the investigation of biological mechanisms. In clinical research they may allow physicians' assessment of individual treatment effects. Decisions on treatment strategy may be based on knowledge of the stochastic processes representing the observed time series, thus allowing full use to be made of the recorded data. Even though monitoring techniques have improved in recent years, natural limitations of the approach may arise, of course, when it is necessary.

Milner (6) studied forecasting models for first, return and total attendance at accident and emergency (A&E) departments and yearly forecasts were developed ten years ago for all the health districts in the Trent region in England. The one-yearly forecasts had been checked against the 1986 actual figures and found accurate for first attendance but less accurate for return attendance. The forecasts for 1993 and 1994 were much further from the actual figures than the 1986 forecasts, with an increasing bias towards overestimation, particularly for reattendance. Whether a first attender is reviewed at a further visit may depend on local medical policy, which itself may vary with personnel changes. The one-off original ARIMA forecasts for new attendance for 1994 were no better than the district projections made in 1984, but they were better than the Trent Regional Health Authority guidelines. The ten-year strategic plan for Trent Regional Health Authority overestimated the increase in the number of first attendance at A&E departments in the Trent region. The forecasting methodology on which it was based could be improved by incorporating the ARIMA method into planning at the health district level. New forecasts or updated ones need to be calculated yearly.

The ARIMA forecasting methodology is designed to operate within a general framework of updating and control. Thus long-term forecasts must be updated yearly as new information is received. However, the long-term forecasts depend crucially on the assessment of the underlying trend. The ARIMA forecasting approach with updating ensures that any evidence indicating a changing trend is quickly detected. Such changes may necessitate a review of the forecasting model, leading to different ARIMA parameters and hence vastly different forecasts.

Makridakis and Hibon (13) studied the various aspects of Box – Jenkins methodology applied to ARIMA models. The major conclusion has been that way that the data are made stationary in its mean is the most important factor determining post - sample forecasting accuracies. Most importantly, when the trend in the data is identified and extrapolated using the same procedure as in other methods that have been found to be more accurate in empirical studies than ARMA models perform consistently better than the models selected through the Box-Jenkins methodology. In addition, it was concluded that using seasonally adjusted data improves post-sample accuracies in a small but consistent manner, and that log and power transformation also contributed to small improvements in post-sample accuracies which become more pronounced for long forecasting horizons. Finally, it was concluded that AR(1), AR(2) or ARMA(1,1) models produced as accurate post-sample prediction as those found by applying the automatic version of the Box – Jenkins methodology. The authors (12) suggested that it was neither necessary, as far as post-sample accuracy was concerned, to study the autocorrelations and partial autocorrelations to determine the most appropriate ARMA models, nor to make sure that the residuals of such a model were necessarily random.

D. Combined forecast

Bates and Granger (7) studied two separate sets of forecasts of airline passenger data, which had been combined to form a composite set of forecasts. The main conclusion was that the composite set of forecasts could yield lower mean – square error than either of the original forecasts. Past errors of each of the original forecasts were used to determine the weights to attach to these two original forecasts in forming the combined forecasts, and different methods of deriving these weights were examined. A number of methods of combining two sets of forecasts were presented, and it was to be noted that, providing the sets of forecasts each contain some independent "information", the combined forecasts could yield improvements.

One unexpected conclusion was that, though the methods suggested for combining forecasts allow the weights to change, this could often lead to better forecasts than those could that would have resulted from the application of a constant weight determined after noting all the individual forecast errors.

Finally, though the comments in this study had related solely to combining two forecasts, there was every reason to combine more than two forecasts.

Deutsch (11) studied the combination of forecasts using changing weights derived from switching regression models or from smooth transition regression models. The regimes associated with the switches might not be known to the forecaster and thus needed to be estimated. Several approaches to this problem were considered. In two empirical examples, these time–varying combining procedures produced smaller, in some cases substantially smaller, out – of –sample squared forecast errors than those obtained using the simple linear combining model. Preliminary results from these time – varying combination methods seemed very encouraging. The time – varying methods considered in this study were very easy to implement and could result in a substantial reduction in sum of squared forecast errors. The results from the empirical examples demonstrated that the combining schemes proposed in the study could actually generate forecasts superior to those of sophisticated forecasting models, which would suggest that there were the ways in which these sophisticated forecasting models could be improved.

Newbold and Granger (3) studied a number of procedures for forecasting a time series from its own current and past values. Forecasting performances of three methods, Box-Jenkins, Holt – Winters and stepwise autoregression were compared over a large sample of economic time series. The possibility of combining individual forecasts in the production of an overall forecast was explored and present empirical results, which indicated that such a procedure could frequently be profitable. In order to discuss in any meaningful way of particular forecasting procedures, it was imperative to analyze their performances on actual data. They had examined the behavior of various univariate time series forecasting methods when applied to a wide collection of economic time series. Box – Jenkins forecasts required a good deal more time and considerably more skill to compute than do competitors. However, they had found that there was a corresponding pay-off, in the sense that the Box-Jenkins forecasts do seem to be better than those derived from two fully automatic procedures, the Holt- Winters method and stepwise autoregression for sizeable majority of the time series in our sample. The tendency of the Box-Jenkins method to produce superior forecasts was particularly marked over the short run (and, indeed, for seasonal series continues in the longer run). Moreover, the gains (in terms of average squared forecast error) from use of the Box-Jenkins approach could be, and often were, quite substantial. These substantial gains did, in fact, persist when forecasting several steps ahead for many series. It did appear, however, that Box-Jenkins forecasts could frequently be improved upon by combination with either Holt-Winters or stepwise autoregression forecasts. Further improvement was frequently obtained by considering a combination of all three types of forecast.

In Winkler's study (8) dealing with combined forecasts by weighted average, five procedures for estimating weights were investigated, and two approaches appeared to be superior to the others. These two procedures provided forecasts that were more accurate overall than forecasts from individual methods. Furthermore, they were superior to forecasts found from a simple unweighted average of the same methods. The results concerning the relative merits of the different weighting schemes were consistent with previous results of Newbold and Granger (3) in the sense that were considered in many more series, many more forecasting methods, and several time horizons instead of just one-step-ahead forecasts. However an argument for combining forecasts was that by aggregating information from different forecasting methods we might be able to generate forecasts that were more accurate than those from the individual methods. In Winkler's study (8), the combined forecasts were indeed more accurate under most conditions, with large time horizons providing some exceptions. In overall basis, the combined forecasts were better than the individual methods.

Clemen (14) tested a variety of different methods in the context of combining forecasts of GNP from four major econometric models. The methods included one in which forecasting errors were jointly normally distributed and several variants of this model as well as some simpler procedures and a Baysian approach with a prior distribution based on exchange ability of forecasters. The results indicated that a simple average, the normal model with an independence assumption, and the Baysian model performed better than the other approaches.

E. Comparison of Forecast and Actuality

Ngamaugpak (4) studied the efficiency of seven forecasting methods namely, Double Exponential Smoothing, Holt's two parameters, Holt's two parameters (DLS), Holt's three parameters, Adaptive Filtering, Box-Jenkins and Winter Smoothing technique. Also, the combined forecast methods were compared. The method proposed by Bates and Granger (7) and the Switching Regression model in which signs, sizes and correlation of forecasting error were considered to be index of the model.

Apart from this objective, the efficiency of forecasting by Regression model which using factor analysis to reduce number of independent variables were studied. For forecasting comparison, the long term and short term of both nonseasonal time series were investigated by using MAPE and MSE as the evaluative criteria. In this study, the data used were monthly time series of 1985-1992. They were set consisting of wholesale Agricultural Price, Financial and Banking, International Trading, Energy

Taxation, Tour and Industrial Product Quantity. The finding of this study revealed that Box- Jenkins method was the best efficient in forecasting both long term and short term seasonal time series. Moreover, Holt's Two Parameters was also suitable for forecasting the long term nonseasonal time series. For the combined case, it revealed that Switching Regression model was more efficient in forecasting both long term and short term seasonal time series similarly. For nonseasonal time series especially wholesale Agricultural Prices Series, the efficient method was Switching Regression model using the index proposed. This index was defined by the errors of the two methods to be combined.

In conclusion, Switching Regression model was more efficient than Box-Jenkins in seasonal time series forecasting.

Luengprateep (15) compared the accuracy of forecasting time series using Holt –L model to that of Holt –D model, and also to that of ARIMA model. One hundred sets of deseasonalized data were used in the study. The measurement for comparison used were the Mean Absolute Percentage Error and Mean Square Error. For Holt – L and Holt – D model, a turbobasic computer program was used which for ARIMA model a forecast plus program was used.

The results showed that for forecasting of at most 12 lead time, ARIMA model was a more accurate than Holt-L and Holt –D models. For forecasting of medium or long range with deseasonalized data, Holt- D model was as accurate as Holt – L model. In fact, Holt- D model was more accurate especially in the case where there was a non-constant increase in the time series trend. For forecasting using Holt - D model, the model which was simple and was as accurate as any other models was the Holt-D for 2 Parameters (DLS) model for medium or long range forecast. Using the first two values of the data as the initial values S_0, T_0, The case where Holt – L model was used for medium or long range forecast, the initial values S_0, T_0, resulting from regression analysis were recommended. However, through parameter consideration in smoothing, it could be concluded that the use of initial values S_0, T_0 resulting from regression analysis and those from the first two values of the data did not affect the range of the parameters in smoothing which gave the minimum MSE and/or MAPE.

F. Traffic Accident

There were a number of studies using different forecasting techniques to investigate the trend and variation of the occurrence of road traffic accidents.

Kunaporn's study (16) aimed at applying statistical method in the analysis of the trend and seasonal variation of the occurrence of road traffic accidents in Krungthep Maha Nakorn. The numbers of accidents, persons killed, persons injured and the value of property damaged in the accidents were used to obtain the trends by using the analysis of time series non – linear trends. The study of seasonal variation was carried out by the method of ratio to moving average and the seasonal index was constructed. The data for the study were collected from the Division of Traffic Police, Police Department during the 1969-1984 period.

The trend in the number of the occurrence of road traffic accidents, the number of persons injured and the value of property damaged increased linearly and the number of persons killed increases exponentially over time. The death rates and incidence rate also increased over time but with shapes of parabolic trends. These trends however are applicable for making a relatively short period forecast only. Incidence rate of passenger cars and motorcycles was estimated by the method of moving average covering the period of one year only.

In the analysis of seasonal index, it was found that the number of accidents was highest in August and September, the rainy season. The number of persons killed was highest in January and April, the time of festival. The number of persons injured was highest in March, August and September, the time when more people migrating from the countryside while August and September were also the period within the rainy season.

Johansson (17) studied the effect of a lowered speed limit on the number of accidents in which there are fatalities, injuries and vehicle damage on Swedish motorways. Two models extending the Poisson and negative binomial count data models were used for estimation. The extended models accounted for both overdispersion and potential dependence between successive counts. The inferences of the parameters were depended on the assumed form of overdispersion. It was found that the speed limit reduction has decreased the number of accidents involving minor injuries and vehicle damage. Furthermore, the models allowing for serial correlation were shown to have the best forecasting performance.

Rock (18) examined the effect of the speed limit change from 55 to 65 mph on rural interstates and limited- access highways in Illinois at the end of April, 1987. He applied ARIMA techniques to study a monthly time series of accidents, injuries, and fatalities dating from five years before the limit increase to four years after. Two types of rural highways were investigated, those where the speed limit was raised and those where it remained at 55 mph. The impact of higher limits on mean speeds, speed variance, traffic diversion, traffic generation, speed spillover, and issues of benefits and costs were considered. The findings revealed that the higher limit led to 300 additional accidents per month in rural Illinois, with associated increases in deaths and injuries. This impact was apparent on both 65 and 55 mph roads. There was some evidence of traffic diversion from 55 to 65 mph highways plus traffic generation and speed spillover.

Martinez-Schnell, Zaidi (19) used time series models in the exploratory and confirmatory analysis of selected fatal injuries in the United States from 1972 to 1983. The researchers developed autoregresive integrated moving average (ARIMA) models for monthly, weekly, and daily series of deaths and used these models to generate hypotheses. These deaths resulted from six causes of injuries: motor vehicles, suicides, homicides, falls, drownings, and residential fires. For each cause of injury, they estimated calendar effects on the monthly death counts. They confirmed the significant effect of vehicle miles traveled on motor vehicle fatalities with a transfer function model. Finally, they applied intervention analysis to deaths due to motor vehicles.

In summary, from the literature reviewed, each single forecast and combined forecast techniques can be applied to examine the traffic accident data. For the single methods, the techniques of least square, Holt and Winters, and the Box –Jenkins were often used for time series data with trend, seasons and nonlinear. While for the combined approaches, the forecast via weighted averages by Bate, Granger and Newbold were reported in many studies. However, in choosing the appropriate forecast technique, the researcher must consider the important factors, particularly the length of time for forecasting, movement of data from curves and also sizes of past data.

CHAPTER III

MATERIALS AND METHODS

The following main topics were described in this chapter.

A. Study Design.

B. Sizes of Past Data and Forecasting Procedures.

C. The Forecast Methods.

1. Single Forecast Methods.
   - 1.1. Least Squares Method for Fitting an Exponential Model.
   - 1.2. Holt and Winters Method.
   - 1.3. The Box-Jenkins Method.
2. Combined Forecast Methods via Weighted Averages.
   - 2.1. Least Squares and Holt and Winters Methods.
   - 2.2. Least Squares and the Box-Jenkins Methods.
   - 2.3. Holt and Winters and the Box-Jenkins Methods.

D. Evaluation of the Accuracy of the Forecast Methods.

1. The Mean Square Error (MSE).
2. The Average Mean Sum Square Error (AMSE).
3. The Mean Absolute Percentage Error (MAPE).

A. Study Design

This study was emphasized on statistical methodology in forecasting techniques. The traffic accident data were used to compare the accuracy between single and combined forecasting methods.

B. Sizes of Past Data and Forecasting Procedures

This study used the traffic accident data in Bangkok from the Police Information System Center, the Royal Thai Police Department. The data were collected and separated into 2 groups. The first group was the traffic accident data in 60 months during January, 1992 to December, 1996. The second group was the data in 12 months during January to December 1997. The natures of the data in the first group were monthly time series with trend, seasonal effect and nonstationary as shown in Figure 5.

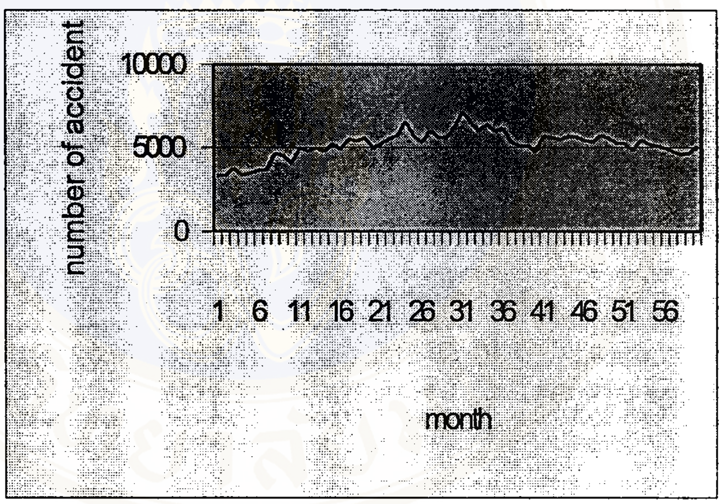


Figure 5. Trend and Seasonal Data Pattern of Traffic Accident in Bangkok during January, 1992 to December, 1996

The data in the first group were divided into 4 subgroups with different sizes as follows:

1. Small Size.

   - 5 months during January, 1992 to May, 1992.

2. Medium Size

   - 12 months during January, 1992 to December, 1992.

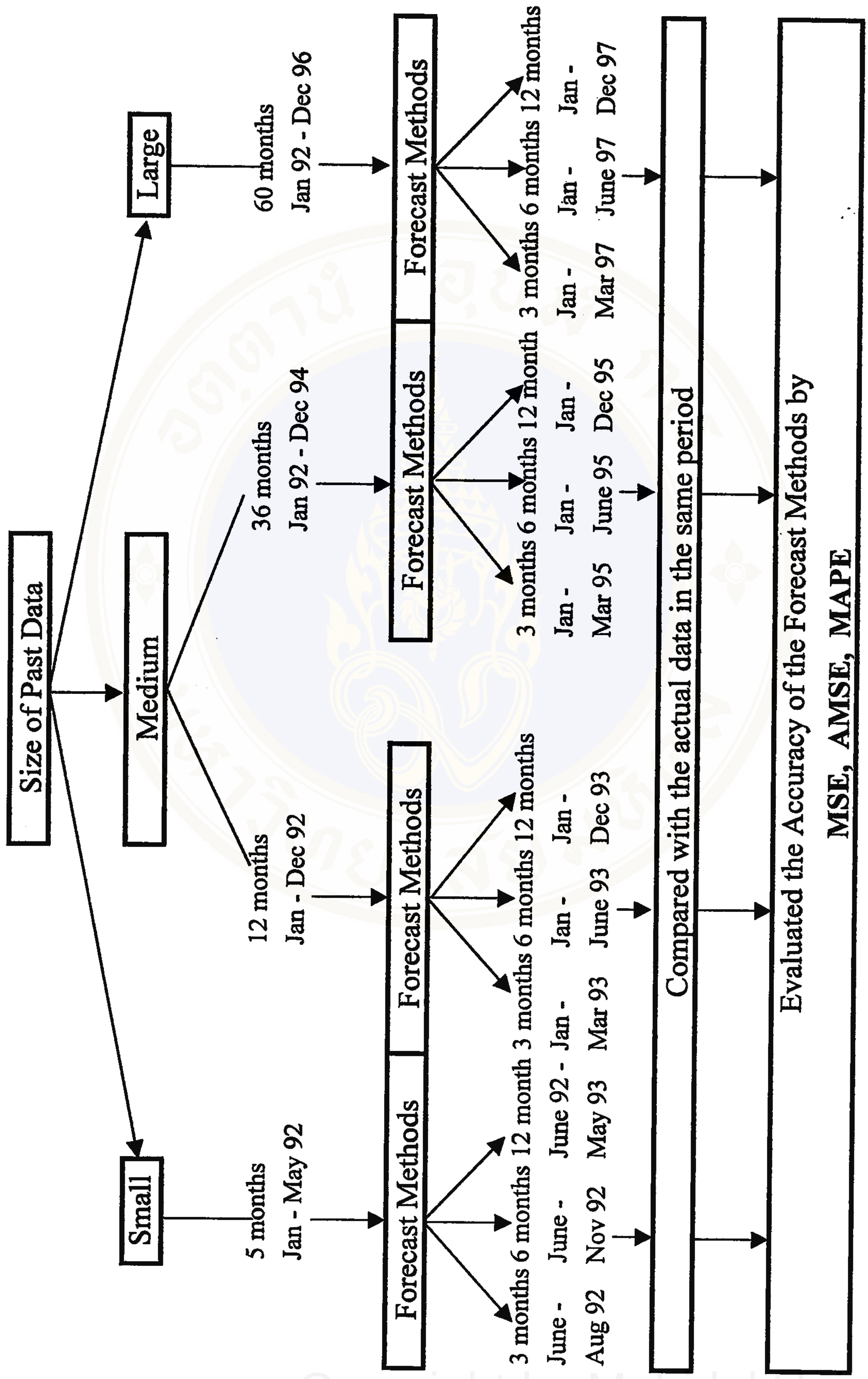
   - 36 months during January, 1992 to December, 1994.

3. Large Size

   - 60 months during January, 1992 to December, 1996.

For each size of past data, the single forecast and combined forecast techniques, using SPSS program and Microsoft Excel for Windows 97, were applied to forecast forward for next 3 months, 6 months and 12 months, respectively. For the large size data of 60 months during January 1992 to December 1996, the data were forecast to the year 1997 and were compared with the actual data in 1997, which were collected in the second group as mention earlier. The correctness of the forecast as measured against actual events in the same period was used as the criterion for evaluating the performance of the forecasting methods. In this study, three statistical indicators: mean square error (MSE), average mean sum square error (AMSE) and mean absolute percentage error (MAPE), were used to evaluate the accuracy of the methods.

In overall, the major steps of the forecasting procedures were shown in the following diagram.


Size of Past Data
Small
Medium
Large
5 months
Jan - May 92
12 months
Jan - Dec 92
36 months
Jan 92 - Dec 94
60 months
Jan 92 - Dec 96
Forecast Methods
Forecast Methods
Forecast Methods
Forecast Methods
3 months
June - Aug 92
6 months
June - Nov 92
12 months
June 92 - May 93
3 months
Jan - Mar 93
6 months
Jan - June 93
12 months
Jan - Dec 93
3 months
Jan - Mar 95
6 months
Jan - June 95
12 month
Jan - Dec 95
3 months
Jan - Mar 97
6 months
Jan - June 97
12 months
Jan - Dec 97
Compared with the actual data in the same period
Evaluated the Accuracy of the Forecast Methods by
MSE, AMSE, MAPE

C. The Forecast Methods

Two major forecast methods used in this study were single and combined techniques. The theoretical models and examples of these methods were explained in the next section.

1. Single Forecast Methods

The single forecast methods were concerned with 3 approaches.

1.1. Least Squares Method for Fitting an Exponential Model.

1.2. Holt and Winters Method.

1.3. The Box – Jenkins Method.

1.1. Least Squares Method

The least squares method can be used to estimate the data which linear or curvilinear. The general equation for equation for estimating a straight line is

$$\hat{y} = a + bx \tag{1}$$

Where $\hat{y}$ = estimated value of the dependent variable.

x = independent variable (time in trend analysis).

a = y - intercept (the value of y when $x = 0$).

b = slope of the trend line.

If a set of data appears to be best represented by a nonlinear, the regression curve must then try to determine the form of the curve and estimate the parameters. Sometimes a scatter diagram indicates that the means, $\mu_{Y/x}$,will probably be best represented by an exponential curve of the for

$$\mu_{Y/x} = \gamma\delta^{x} \tag{2}$$

where γ and δ are parameters to be estimated from the data. Denoting these estimates by c and d, respectively, $\mu_{Y/x}$ can be estimated by $\bar{y}_x$ from the sample regression curve

$$\bar{y}_x = cd^{x} \quad (3)$$

Taking logarithms to the base 10, obtain the regression curve

$$\log \bar{y}_x = \log c + (\log d)x \quad (4)$$

and each pair of observations in the sample satisfies the relation

$$\log y_i = \log c + (\log d)\, x_i + e_i$$

$$= a + bx_i + e_i \quad (5)$$

where $a = \log c$ and $b = \log d$. Therefore, it is possible to find a and b in this formulas. Estimated from a sample of size n by the line $\bar{y}_x = a + bx$

where

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \quad (6)$$

$$a = \bar{y} - b\bar{x} \quad (7)$$

using the points (x_i , $\log y_i$), and then determine c and d by taking antilogarithms.

The least squares procedure for fitting an exponential curve to a set of data was illustrated in Example 1.

Example 1. The following data represent traffic accident in Bangkok during the past 5 months:

x (months)	1	2	3	4	5
y (number of accident)	3,285	3,328	3,759	3,333	3,440

The method of least squares was used to estimate a curve of the form $\mu_{Y/x} = \gamma\delta^{x}$ and to predict the number of accident during next 2 months from the data of five months.

Solution. First, the y values were taken by logarithms. The logarithms of y values in each month were 3.5165, 3.5221, 3.5750, 3.5228, and 3.5365, respectively. Second, the values of a and b were estimated.

$$\sum_{i=1}^{5} x_i = 15, \qquad \sum_{i=1}^{5} \log y_i = 17.6729, \qquad \sum_{i=1}^{5} x_i^2 = 55$$

$$\sum_{i=1}^{5} x_i \log y_i = 53.0594, \qquad \bar{x} = 3, \qquad \overline{\log y} = 3.53458$$

Substituting the values of x and y in the formulas (6) and (7) for estimating a and b.

$$b = \frac{(5)(53.0594) - (15)(17.6729)}{(5)(55) - (15)^2} = 0.00407$$

and then

$$a = 3.53458 - (0.00407)(3) = 3.52237$$

Determine c and d by taking antilogarithms.

$$c = 10^{3.52237} = 3{,}329.43$$

$$d = 10^{0.00407} = 1.009415$$

and least squares regression curve from equation (3) was

$$\bar{y}_x = (3{,}329.43)(1.009415)^{x}$$

Based on this rate of growth, the expected numbers of accident in the next 2 months from now (x = 7) should be $\bar{y}_x = (3{,}329.43)(1.009415)^7 = 3{,}555$

1.2. Holt and Winters Method

Holt and Winters method has been extended to accommodate a seasonal component. At any time, t, the seasonal factor is given by the ratio of the observed value, x_t, to the smoothed series, y_t is:

$$y_t = a \,.\, x_t + (1\text{-}a) \,.\, (y_{t-1} + h_{t-1}) \quad (8)$$

where $0 \le a \le 1$

An exponentially smoothed estimate, h_t, of the trend at time t is:

$$h_t = b \,.\, (y_t - y_{t-1}) + (1 - b) \,.\, h_{t-1} \quad (9)$$

where $0 \le b \le 1$ is a second smoothing parameter. A smoothed estimate of this seasonal factor is given by

$$s_t = c \,.\, (x_t / y_t) + (1 - c) \,.\, s_{t-12} \quad (10)$$

where $0 \le c \le 1$ is a third smoothing parameter. This equation assumes a twelve period seasonality (monthly data). If quarterly data were used the last term of equation (10) would be $(1 - c) \,.\, s_{t-4}$.

The smoothed estimate of the series in equation (8) must now be altered so that it is the weighted average of the trend estimate and the current value modified now by the most recently available seasonal factor:

$$y_t = a \,.\, (x_t / s_{t-12}) + (1\text{-}a) \,.\, (y_{t-1} + h_{t-1}) \quad (11)$$

The equation (10) and (11) together with (9) describe the Holt – Winters process. Forecasts m periods ahead are given by

$$\hat{Y}_{t+m} = (y_t + m \,.\, h_t) \,.\, s_{t+m-12} \quad (12)$$

An example of parameter estimates is shown in Example 2 and Table 2.

Example 2. The calculations of the traffic accident data using Holt and Winters method were shown in Table 2. Since this is quite a long series, periods of a year to allow the smoothing to "settle down" have been allowed so that the initial calculations are in the following blocks:

1992 Use Holt's method and allow trend estimate to begin to settle.

1993 Trend continues to settle. First estimates of seasonal factors as ratios.

1994 Use Holt – Winters. First smoothed estimates of seasonal factors.

1995 Begin forecasts.

So, with starting conditions $y_1 = x_1$, $y_2 = x_2$ and $h_2 = x_2 - x_1$ the values for y_t and h_t until and including t = 24 are given by Holt's method using equation (8) and (9) as before. From t = 13 to t = 24 the first estimates of s_t are simple x_t / y_t. From t = 25 the Holt – Winters equations are used.

From equation (11),

y_{25} = (0.1 * 5,720/1.086) + 0.9 * (6,066.773+123.105) = 6,097.593

From equation (9),

h_{25} = 0.3 * (6,097.593 – 6,066.773) + (0.7 * 123.105) = 95.415

From equation (10),

s_{25} = 0.4 * (5,720/6,097.593) + (0.6* 1.086) = 1.027

Table 2. Traffic Accident Data forecast by Holt and Winters method.

	t	x_t	Smoothed a = 0.1 y_t	Trend b = 0.3 h_t	Season c = 0.4 s_t	$\hat{y}_t$
1992	1	3,285	3,285			
	2	3,328	3,328	43.000		
	3	3,759	3,409.800	54.640		
	4	3,333	3,451.296	50.697		
	5	3,440	3,495.794	48.837		
	6	3,622	3,552.367	51.158		
	7	3,649	3,608.073	52.522		
	8	4,633	3,757.836	81.694		
	9	4,488	3,904.377	101.148		
	10	4,046	4,009.573	102.363		
	11	4,950	4,195.742	127.505		
	12	4,756	4,366.522	140.487		
1993	13	4,942	4,550.509	153.537	1.086	
	14	4,763	4,709.941	155.305	1.011	
	.	.	.	.	.	
	.	.	.	.	.	
	.	.	.	.	.	
	24	6,548	6,066.773	123.105	1.079	
1994	25	5,720	6,097.593	95.415	1.027	
	26	5,276	6,095.415	66.142	0.952	
	.	.	.	.	.	
	.	.	.	.	.	
	.	.	.	.	.	
	35	5,959	6,627.821	75.908	0.938	
	36	6,117	6,600.101	44.820	1.018	
1995	37	5,253	6,491.993	- 1.059	0.940	6,823.340
	38	5,089	6,375.847	-35.585	0.891	6,375.233
	39	5,083	6,201.212	-77.300	0.944	6,915.842
	40	4,743	6,014.047	-110.259	0.882	6,398.584
	41	5,659	5,881.196	-117.037	0.983	6,801.519
	42	5,566	5,745.760	-122.557	0.985	6,851.575
	43	5,449	5,576.975	-136.425	1.024	7,299.775
	44	5,640	5,503.543	-117.527	0.967	6,465.195

(The calculated values in the table have been rounded to three decimal places)

Forecasts are made, as in the previous section, for one year ahead so that, for instance, the forecasts for 1995($37 \le t \le 48$) use y_{36} , h_{36} and s_{25} to s_{36} .

From equation (12).

$$\hat{y}_{37} = 1.027^* (6,600.101 + 44.820) = 6,823.340$$

$$\hat{y}_{38} = 0.952^* [6,600.101 + (2^* 44.820)] = 6,375.233$$

and so on.

1.3. The Box- Jenkins Method

The procedure consists basically of fitting a mixed autoregressive integrated moving average (ARIMA) model to a given set of data and then taking conditional expectations. The main stages in setting up a Box – Jenkins forecasting model are as follows:

1). Model Identification

Examine the data to see which member of the class of ARIMA processes appears to be most appropriate and make the series stationary. Then, consider transformations to stabilize the variance, and differencing.

2). Estimation

Estimate the parameters of the chosen model by least squares, and maximum likelihood estimation of Box – Jenkins models.

3). Diagnostic Checking

Examine the residuals from the fitted model to see if it is adequate in particular check the autocorrelation coefficients (ACF) of the residuals for white noise (calculates standard error assuming that the underlying process is white noise).

4). Consider Alternative Models if Necessary

If the first model appears to be inadequate for some reasons, then other ARIMA models may be tried until a satisfactory model is found.

Stationarity and Nonstationarity

Stationarity signifies that the probability of structure of the series does not change with time. In particular, a stationary series has a constant mean and variance and a covariance structure which depends only on the difference between two time points. If the roots of the polynomial $\phi(B)$ lie outside the unit circle, it may be shown that an ARMA(p, q) process is stationary.

If a time series is not stationary, then it can be made more nearly stationary by taking the first difference of the series.

First Difference

The first difference of the series is expressed as

$$X'_t = X_t - X_{t-1}$$

A very useful notational device is the backward shift operator, B, which is used as follows:

$$BX_t = X_{t-1}$$

In other words, B, operating on X_t, has the effect of shifting the data back 1 period.

Two applications of B to X_t shifts the data back 2 periods, as follows:

$$B(BX_t) = B^2X_t = X_{t-2.}$$

Using the backward, shift operator can rewritten.

First Difference

$X'_t = X_t - BX_t = (1-B)X_t$

Second-order Difference

$$
\begin{aligned}
X''_t &= X'_t - X'_{t-1} \\
&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\
&= X_t - 2X_{t-1} + X_{t-2} \\
&= (1-2B + B^2)X_t \\
&= (1-B)^2 X_t
\end{aligned}
$$

The purpose of taking differences is to achieve stationarity, and in general, if it takes a *d*th - order difference to achieve stationarity, we will write

*d*th - order difference = $(1-B)^d X_t$

Higher – Order Combination: ARIMA (p, d, q)

There is no limit to the variety of ARIMA models. The general model, which covers all the cases mentioned above and many, is known as ARIMA (p, d, q):

AR: p = order of the autoregressive process

I: d = degree of differencing involved

MA: q = order of the moving average process

Autoregressive Process

AR(1) model is a first-order autoregressive models. In general, for a *p*th-order AR process, it can be designated as follows:

ARIMA (*p*,0,0)

$$X_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + e_t,$$

Where μ' = constant term,

ϕ_p = pth autoregressive parameter,

e_t = the error term at time t

These two cases are defined as follows:

ARIMA (1,0,0)

$X_t = \mu' + \phi_1 X_{t-1} + e_t$

ARIMA (2,0,0)

$X_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t$

Moving Average Process

The general moving average (MA) process of order q can be written as follows:

ARIMA (0,0,q) or MA(q)

$X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q},$

Where μ = constant term,

θ_q = qth moving average parameter,

e_{t-q} = the error term at time $t-q$

ARIMA (0,0,1) or MA (1)

$X_t = \mu + (1 - \theta_1 B) e_t$

ARIMA (0,0,2) or MA (2)

$X_t = \mu + (1 - \theta_1 B - \theta_2 B^2) e_t$

Mixture: ARIMA Process

The general ARIMA (p, d, q) model is implied. The equation for the simplest case, ARIMA (1,1,1) is as follows:

ARIMA (1,1,1)

$$(1-B)(1-\phi_1 B)X_t = \mu' + (1-\theta_1 B)e_t$$

↑ ↑ ↑

First Difference AR(1) MA(1)

The terms can be multiplied out and rearranged as follows:

$$[1 - B(1+\phi_1) + \phi_1 B^2]X_t = \mu' + e_t - \theta_1 e_{t-1},$$

$$X_t = (1+\phi_1)X_{t-1} - \phi_1 X_{t-2} + \mu' + e_t - \theta_1 e_{t-1},$$

Seasonality and ARIMA models

Consider a data series that is collected quarterly. Then seasonal differences could be computed as follows:

$$X'_t = X_t - X_{t-4} = (1-B^4)X_t$$

The new data series, represented by X'_t, would now deal with differences between on quarter's data and the data four quarters ago. For data collected monthly, a full season's (year's) difference would be computed as follows:

$$X'_t = X_t - X_{t-12} = (1-B^{12})X_t$$

The ARIMA notation can be extended readily to handle seasonal aspects, and the general shorthand notation is

ARIMA (p, d, q) (P, D, Q) ^{S}

Where (p, d, q) = Nonseasonal part of the model.

(P, D, Q) = Seasonal part of the model.

s = number of periods per season

The algebra is simple but can get lengthy. So, for illustrative purposes, consider the following general ARIMA (1, 1, 1)(1, 1, 1) ^{4} model.

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)X_t = (1-\theta_1 B)(1-\theta_1 B^4)e_t \qquad (13)$$

All the factors can be multiplied out and the general model written in what is called "unscrambled form". Multiplying out equation (13) yields the following:

$$X_t = (1+\phi_1)X_{t-1} + (1+\Phi_1)X_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)X_{t-5}$$

$$+ (\phi_1+\phi_1\Phi_1)X_{t-6} - \Phi_1 X_{t-8} + (\Phi_1+\phi_1\Phi_1)X_{t-9} - \phi_1\Phi_1 X_{t-10}$$

$$+ e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + \theta_1\Theta_1 e_{t-5}$$

All this form, once the coefficients ϕ_1, Φ_1, θ_1, and Θ_1 have been estimated from the data, can be used for forecasting as shown in Table 3.and Example 3.

Table 3. Estimating the Parameters

(p,d,q)	parameter	equations	boundary
ARIMA (0,d,1) MA (1)	θ_1	$\rho_1 = \dfrac{-\theta_1}{1+\theta_1^2}$	$-1<\theta_1<1$
ARIMA(0,d,2) MA (2)	θ_1 , θ_2	$\rho_1 = \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$	$-1<\theta_2<1$
			$\theta_1+\theta_2<1$
		$\rho_2 = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\theta_2-\theta_1<1$
ARIMA(1,d,0) AR (1)	ϕ_1	$\rho_1=\phi_1$	$-1<\phi_1<1$
ARIMA(2,d,0) AR (2)	ϕ_1, ϕ_2	$\rho_1=\phi_1+\phi_2\rho_1$	$\phi_1+\phi_2<1$
		$\rho_2=\phi_1\rho_1+\phi_2$	$\phi_2-\phi_1<1$
			$-1<\phi_2<1$

Table 3.(conts)

(p,d,q)	parameter	equations	boundary
ARIMA(1,d,1) ARMA (1,1)	ϕ_1, θ_1	$\rho_1 = \dfrac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta^2_1-2\phi_1\theta_1}$	$-1 < \phi_1 < 1$
		$\rho_2 = \rho_1\phi_1$	$-1 < \theta_1 < 1$

To estimate the parameters ϕ, θ replacing ρ_1 and ρ_2 with r_1 and r_2 from the autocorrelation diagram.

Diagnostic Checking

The Box-Pierce Q statistic is used to test whether the autocorrelations for these residuals are significantly different from zero. If no model is fitted, then the autocorrelations can be considered to come from an ARIMA (0,0,0) model.

The Q statistic is computed as follows:

$$Q = n\sum_{k=1}^{m} r^2_k,$$

where m = maximum lag considered,
n = N – d
N = original number of observations, and
r_k = autocorrelation for lag k

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_t-\bar{Y})(Y_{t-k}-\bar{Y})}{\sum_{t=1}^{n}(Y_t-\bar{Y})^2}$$

and Q is distributed approximately as a chi-square statistic with (m-p-q) degrees of freedom.

Example 3. Forecasting 3 Months using Past Data of 60 Months during January, 1992 to December, 1996.

The procedure begins with the transform data of time series with nonstationary to stationary by taking the first difference of the series. The first difference of the series is expressed as $X'_t = X_t - X_{t-1}$. Model Identification is the important step to select appropriate model to forecast. The model selection to forecast will be considered from decreasing of Autocorrelation Coefficients and Partial Autocorrelation Coefficients of time series data and difference of data, as shown in the following outputs.

Autocorrelations: FIT_1 Fit for ACCIDENT from ARIMA

```
     Auto- Stand.
Lag  Corr.  Err. -1  -.75 -.5  -.25  0   .25  .5   .75   1  Box-Ljung  Prob.
                  +----+----+----+----+----+----+----+----+
  1   .184  .127                 .    I****.                    2.106  .147
  2  -.267  .126                 *****I    .                    6.601  .037
  3   .020  .125                 .    *    .                    6.626  .085
  4  -.018  .124                 .    *    .                    6.647  .156
  5  -.041  .122                 .   *I    .                    6.759  .239
  6   .053  .121                 .    I*   .                    6.950  .326
  7   .122  .120                 .    I**  .                    7.983  .334
  8  -.039  .119                 .   *I    .                    8.092  .425
  9  -.060  .118                 .   *I    .                    8.354  .499
 10   .031  .117                 .    I*   .                    8.424  .587
 11   .037  .115                 .    I*   .                    8.529  .665
 12   .138  .114                 .    I*** .                    9.980  .618
 13   .080  .113                 .    I**  .                   10.483  .654
 14   .002  .112                  .   *   .                    10.483  .726
 15  -.081  .111                  . **I   .                    11.017  .751
 16  -.048  .109                  .  *I   .                    11.214  .796

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .
Total cases:   60     Computable first lags:   58
```

Partial Autocorrelations: FIT_1 Fit for ACCIDENT from ARIMA

```
   Pr-Aut- Stand.
Lag  Corr.  Err. -1  -.75 -.5  -.25  0   .25  .5   .75   1
                  +----+----+----+----+----+----+----+----+
  1   .184  .130                 .    I****.
  2  -.311  .130                *.****I    .
  3   .162  .130                 .    I*** .
  4  -.173  .130                 . ***I    .
  5   .075  .130                 .    I*   .
  6  -.015  .130                 .    *    .
  7   .146  .130                 .    I*** .
  8  -.120  .130                 .  **I    .
  9   .075  .130                 .    I*   .
 10  -.050  .130                 .   *I    .
 11   .088  .130                 .    I**  .
 12   .122  .130                 .    I**  .
 13   .025  .130                 .    *    .
 14   .063  .130                 .    I*   .
 15  -.105  .130                 .  **I    .
 16   .053  .130                 .    I*   .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .
Total cases:   60     Computable first lags:   58
```

The data have slight seasonal variation. The Box–Jenkins method with ARIMA model identification is (1,1,0). Estimation is the step to analyse parameter estimation value of the model from time series data, which is the output of the model identification and will be used to analyse the best estimation value of the parameter. This value will be the parameter estimation, which effects sum square of residual to be the minimum value. An example of parameter estimation of ARIMA model is shown as follow.

Parameter	**Estimater**	**Standard Error**	**T-Ratio**	**Approx. Prob.**
Constant	-3.3197563	48.919865	-0.0678611	0.94613369
AR1	-0.5755701	0.108071	-5.3258451	0.00000178

Equation of model ARIMA (1,1,0) is

$$X_t = \mu' + (1 + \phi_1) * X_{t-1} - \phi_1 * X_{t-2} + e_t$$

Where μ' = constant term,

ϕ_p = pth autoregressive parameter,

e_t = the error term at time t

From output μ' = -3.3197563, ϕ_1 = -0.5755701

Results for ARIMA(1,1,0) process defined as follows

$$X_t = -3.3197563 + 0.42443 * X_{t-1} + 0.5755701 * X_{t-2} + e_t$$

An example is to forecast in 3 months using the past data in 60 months. The forecast data in each month will be computed by the following equations.

$$\hat{X}_{t+1} = -3.3197563 + 0.42443 * X_t + 0.5755701 * X_{t-1}$$

4812.39 = -3.3197563 + 0.42443* 5,035 + 0.5755701 * 4,654

4932.79 = -3.3197563 + 0.42443* 4,802 + 0.5755701 * 5,035

4732.47 = -3.3197563 + 0.42443* 4,646 + 0.5755701 * 4,802

2. Combined Forecast Methods via Weighted Averages.

This study emphasized only the combined forecast using weighted average. The combined forecast for period t is of the form

$$C\hat{y}_t = \sum_{i=1}^{m} w_i \hat{y}_{it}$$

where

$C\hat{y}_t$ is the combined forecast for period t

$\hat{y}_{it}$ is the forecast for period t from forecasting method i

w_i is the weight assigned to method i

m is the number of methods

t is period t ; t = 1, 2, …, n.

A weighted average of forecasts that minimizes the variance of the combined forecast error can be found in Newbold and Granger (3, 8).

$$w_i = \frac{\left[\sum_{t=2}^{n} e_{it}^2\right]^{-1}}{\sum_{i=1}^{m}\left[\sum_{t=2}^{n} e_{it}^2\right]^{-1}},$$

$$e_{it} = \frac{y_t - \hat{y}_{it}}{y_t}$$

where w_i is the weight assigned to method i

e_{it} is error from the forecast for period t and method i

$\hat{y}_{it}$ is the forecast for period t from forecasting method i

y_t is data time series for period t

i is forecasting method i ; i = 1, 2, …,m

All steps of calculation in a combined forecast method are shown in Example 4, 5 and 6.

Example 4. A Combined Forecast Method using the Least – Squares and Holt and Winters methods.

In this example, the procedure aims to forecast in 12 months using the past data of 60 months during January, 1992 to December, 1996.

In order to receive the accurate results Winkler (8) has suggested to use large size of data in a combined forecast approach. Therefore this study used the past data in 60 months as the base for forecasting by the combined methods.

The procedure consists of following three main steps.

Step 1 is to calculate sum square error of each single forecast method as follows.

$$\Sigma e^2_{it} = \Sigma \left[\frac{y_t - \hat{y}_{it}}{y_t} \right]^2$$

where w_i is the weight assigned to method i

e_{it} is error from the forecast for period t and method i

$\hat{y}_{it}$ is the forecast for period t from forecasting method i

y_t is data time series for period t

i is forecasting method i ; $i = 1, 2, \ldots, m$

Sum square error of the Least Squares method (Σe^2_{1t}) = 1.4570

Sum square error of Holt and Winters method (Σe^2_{2t}) = 0.0799

Sum square error of the Box –Jenkins method (Σe^2_{3t}) = 0.1251

Step 2 is to calculate a weighted average of forecasts from formula in Newbold and Granger (3,8).

A weighted average of the Least Squares method $(w_1) = \dfrac{(1.4570)^{-1}}{(1.4570 + 0.0799)^{-1}}$

$= 1.055$

A weighted average of Holt and Winters method $(w_2) = \dfrac{(0.0799)^{-1}}{(0.0799 + 1.4570)^{-1}}$

$= 19.235$

Step 3 is to compute the combined forecast for month t using the following equation.

$$\hat{Cy}_t = [w_1 * \hat{y}_{1t} + w_2 * \hat{y}_{2t}] / (w_1 + w_2)$$

where

$\hat{Cy}_t$ is the combined forecast for period t

$\hat{y}_{it}$ is the forecast value for period t from forecasting method i

w_i is the weight assigned to method i

t is period t ; t = 1, 2, …, n.

An example is to forecast in 3 months. The combined forecast for each month is computed as follow.

Month 1 $\hat{Cy}_1 = [1.055 * 5828.85 + (19.235) * 5003.56] / 20.29 = 5046.47$

5828.85 is the forecast value for month 61 from the Least Square method using the following equation $\hat{y}_{61} = (4405.548)*(1.0046)^{61}$

5003.56 is the forecast value for month 61 from the Holt and Winters method using the following equation $\hat{y}_{60+1} = 4990.063+(1*(-72.094))*1.017$

20.29 are the weight assigned to the combined method $w_1 + w_2 = 1.055 + 19.235$

Month 2 $\hat{Cy}_2 = [1.055* 5855.67 + (19.235) * 4576.20] / 20.29 = 4642.72$

5855.67 is the forecast value for month 62 from the Least Square method using the following equation $\hat{y}_{62} = (4405.548)*(1.0046)^{62}$

4576.20 is the forecast value for month 62 from the Holt and Winters method using the following equation $\hat{y}_{60+2}$ = 4990.063+(2*(-72.094))*1.017

20.29 are the weight assigned to the combined method $w_1 + w_2$=1.055 + 19.235

Month 3 $\hat{Cy}_3$ = [1.055 * 5882.60 + (19.235) * 4597.56]/ 20.29 = 4664.38

5882.60 is the forecast value for month 63 from the Least Square method using the following equation $\hat{y}_{63}$ = (4405.548)*(1.0046)^{63}

4597.56 is the forecast value for month 63 from the Holt and Winters method using the following equation $\hat{y}_{60+3}$ = 4990.063+(3*(-72.094))*1.017

20.29 are the weight assigned to the combined method $w_1 + w_2$=1.055 + 19.235

Example 5. A Combined Forecast Method using the Least – Squares and the Box - Jenkins methods.

The procedure is the same as describe in Example 4.

Step 1 is to calculate sum square error of each single forecast method as follows.

Sum square error of the Least Squares method (Σe^2_{1t}) = 1.4570

Sum square error of Holt and Winters method (Σe^2_{2t}) = 0.0799

Sum square error of the Box –Jenkins method (Σe^2_{3t}) = 0.1251

Step 2 is to calculate a weighted average of forecasts from formula in Newbold and Granger (3,8).

A weighted average of the Least Squares method (w_1) = $\dfrac{(1.4570)^{-1}}{(1.4570 + 0.1251)^{-1}}$

= 1.086

A weighted average of the Box - Jenkins method (w_2) = $\dfrac{(0.1251)^{-1}}{(0.1251 + 1.4570)^{-1}}$

= 12.647

Step 3 is to compute the combined forecast for month t using the following equation.

$$\hat{Cy}_t = w_1 * \hat{y}_{1t} + w_2 * \hat{y}_{2t}$$

For forecasting 3 months, each month is computed as follows.

Month 1 $\hat{Cy}_1$ = [1.086 * 5828.86 + 12.647 * 4812.39] /13.733 = 4892.77

Month 2 $\hat{Cy}_2$ = [1.086 * 5855.67 + 12.647 * 4932.79] /13.733 = 5005.77

Month 3 $\hat{Cy}_3$ = [1.086 * 5882.61 + 12.647 * 4732.47] /13.733 = 4823.42

Example 6. A Combined Forecast Method using Holt and Winters and the Box - Jenkins methods.

The procedure is the same as describe in previous example.

Step 1 is to calculate sum square error of each single forecast method as follows.

Sum square error of the Least Squares method (Σe^2_{1t}) = 1.4570

Sum square error of Holt and Winters method (Σe^2_{2t}) = 0.0799

Sum square error of the Box –Jenkins method (Σe^2_{3t}) = 0.1251

Step 2 is to calculate a weighted average of forecasts from formula in Newbold and Granger (3,8).

A weighted average of Holt and Winters method (w_1) = $\dfrac{(0.0799)^{-1}}{(0.0799+0.1251)^{-1}}$

= 2.566

A weighted average of the Box - Jenkins method (w_2) = $\dfrac{(0.1251)^{-1}}{(0.1251+0.0799)^{-1}}$

= 1.639

Step 3 is to calculate the combined forecast for month t using the following equation.

$$\hat{Cy}_t = w_1 * \hat{y}_{1t} + w_2 * \hat{y}_{2t}$$

For forecasting 3 months, each month is computed as follows.

Month 1 $\hat{Cy}_1 = [2.566 * 5003.56 + 1.639 * 4812.39] / 4.205 = 4929.05$

Month 2 $\hat{Cy}_2 = [2.566 * 4576.20 + 1.639 * 4932.79] / 4.205 = 4715.19$

Month 3 $\hat{Cy}_3 = [2.566 * 4597.56 + 1.639 * 4732.47] / 4.205 = 4650.14$

D. Evaluation of the Accuracy of the Forecast Methods

The mean square error (MSE), the average mean sum square error (AMSE) and the mean absolute percentage error (MAPE), were calculated for evaluating the performance of single forecasting methods and combined forecasting method (1).

Mean square error: MSE

$$MSE_j = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{n}$$

Average mean sum square error: AMSE

$$AMSE = \frac{\sum_{j=1}^{n} MSE_j}{n}$$

Mean absolute percentage error: MAPE

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{y_t}}{n} * 100$$

An example of computations of the MSE, AMSE and MAPE is shown in Example7.

Example 7. Computations of the MSE, AMSE and MAPE of the Combined Forecast in 3 Months using the Past Data of 60 Months during Jan 92 to Dec 96.

No.	Month of Forecast	Actual	Combined Forecast	Error	Absolute Percentage Error	Squared Error
61.	Jan 97	4,802	5,046.47	-244.47	5.09	59,766.67
62.	Feb 97	4,646	4,642.72	3.28	0.07	10.73
63.	Mar 97	5,091	4,664.38	426.62	8.38	182,006.68
	Sum				13.54	241,784.08

$$\text{MSE} = \frac{241,784.08}{3} = 80,594.69$$

$$\text{AMSE} = \frac{80,594.69}{3} = 26,864.89$$

$$\text{MAPE} = \frac{13.54}{3} = 4.51\ \%$$

CHAPTER IV

RESULTS

This chapter describes outcomes of data analysis using the single and combined forecast methods in the following topics.

A Single Forecast Methods.

1. Least Squares Method for fitting an Exponential Curve.
2. Holt and Winters Method
3. The Box-Jenkins Method
4. A Comparison on the Accuracy of the Single Forecast Methods

B. The Combined Forecasts via Weighted Averages.

1. Least Squares and Holt and Winters Methods
2. Least Squares and the Box-Jenkins Methods
3. Holt and Winters and the Box-Jenkins Methods
4. A Comparison on the Accuracy of the Combined Forecast Methods

C. A Comparison on the Accuracy between the Single and Combined Forecast Methods.

A. Single Forecast Methods

Four groups of traffic accident data in Bangkok in 5, 12, 36 and 60 months were forecast to the next 3 periods of 3, 6 and 12 months, respectively. Then, the statistical indicators, MSE, AMSE and MAPE were calculated to evaluate the accuracy of each single forecast method.

1. **Least Squares Method for Fitting an Exponential Curve**

The results of forecasting the traffic accident data using the least square method for fitting an exponential curve were shown in Table 4, 5, 6 and 7.

Table 4. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 5 Months (January to May, 1992) using the Least Squares Method for Fitting an Exponential Curve.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 Months						
6.	June 92	3,622	3,521.67	100.32	2.76	10,064.42
7.	July 92	3,649	3,554.78	94.21	2.58	8,877.00
8.	Aug 92	4,633	3,588.19	1,044.80	22.55	1,091,613.08
	Sum				**27.89**	**1,110,554.50**
Forecast 6 Months						
6.	June 92	3,622	3,521.67	100.32	2.76	10,064.42
7.	July 92	3,649	3,554.78	94.21	2.58	8,877.00
8.	Aug 92	4,633	3,588.19	1,044.80	22.55	1,091,613.08
9.	Sep 92	4,488	3,621.92	866.07	19.29	750,083.89
10.	Oct 92	4,046	3,655.97	390.02	9.63	152,121.63
11.	Nov 92	4,950	3,690.33	1,259.66	25.44	1,586,747.33
	Sum				**82.25**	**3,599,507.35**
Forecast 12 Months						
6.	June 92	3,622	3,521.67	100.32	2.76	10,064.42
7.	July 92	3,649	3,554.78	94.21	2.58	8,877.00
8.	Aug 92	4,633	3,588.19	1,044.80	22.55	1,091,613.08
9.	Sep 92	4,488	3,621.92	866.07	19.29	750,083.89
10.	Oct 92	4,046	3,655.97	390.02	9.63	152,121.63
11.	Nov 92	4,950	3,690.33	1,259.66	25.44	1,586,747.33
12.	Dec 92	4,756	3,725.02	1,030.97	21.67	1,062,904.12
13.	Jan 93	4,942	3,760.04	1,181.95	23.91	1,397,022.71
14.	Feb 93	4,763	3,795.38	967.61	20.31	936,274.43
15.	Mar 93	5,183	3,831.06	1,351.93	26.08	1,827,731.25
16.	April 93	4,855	3,867.07	987.92	20.34	975,994.05
17.	May 93	5,486	3,903.42	1,582.57	28.84	2,504,539.19
	Sum				**223.40**	**12,303,973.10**

From Table 4 using the size of past data in 5 months forecast to 3, 6 and 12 months, the forecast data were clearly less than the actual data. The errors of the forecasting method, the least squares approach, trended to be increase relative to the increase of periods of forecast (3 to 6 months). As shown in Table 8, the value of mean absolute percentage error (MAPE) was 9.29 for forecasting in 3 months and increase to be 18.61 for forecasting in 12 months periods.

The result of using the past data in 12, 36 and 60 months forecast the data in 3, 6, and 12 months indicated that the forecast data were all greater than the actual data in the same periods (see Table 5, 6, 7). All indicators of errors MSE, AMSE and MAPE were increase as the same phenomena in using the size of past data in 5 months. The value of MAPE was 4.31 when using the past data in 12 months forecast to 3 months as compared to MAPE of 15.52 for forecasting to 12 months (Table 8). The errors (MAPE) of forecasting when using the past data in 36 months were about 36 to 41 % higher than the error of other periods. The range of errors (MAPE), using the size of past data in 60 months was about 21 to 31 % (Table 8).

In sum, the least square method for fitting an exponential curve gave high errors, especially when using the actual data in 36 and 60 months for forecasting. This technique was appropriate in forecasting the accident data in short time period in 3 months, using the past data in 12 months.

Table 5. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 12 Months (January to December,1992) using the Least Squares Method for Fitting an Exponential Curve.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
13.	Jan 93	4,942	4,982.24	-40.24	0.81	1,619.76
14.	Feb 93	4,763	5,171.57	-408.57	8.57	166,930.84
15.	Mar 93	5,183	5,368.09	-185.09	3.57	34,258.83
	Sum				**12.95**	**202,809.43**
Forecast 6 months						
13.	Jan 93	4,942	4,982.24	-40.24	0.81	1,619.76
14.	Feb 93	4,763	5,171.57	-408.57	8.57	166,930.84
15.	Mar 93	5,183	5,368.09	-185.09	3.57	34,258.83
16.	April 93	4,855	5,572.07	-717.07	14.76	514,202.16
17.	May 93	5,486	5,783.81	-297.81	5.42	88,695.50
18.	June 93	5,343	6,003.60	-660.60	12.36	436,396.30
	Sum				**45.49**	**1,242,103.39**
Forecast 12 months						
13.	Jan 93	4,942	4,982.24	-40.24	0.81	1,619.76
14.	Feb 93	4,763	5,171.57	-408.57	8.57	166,930.84
15.	Mar 93	5,183	5,368.09	-185.09	3.57	34,258.83
16.	April 93	4,855	5,572.07	-717.07	14.76	514,202.16
17.	May 93	5,486	5,783.81	-297.81	5.42	88,695.50
18.	June 93	5,343	6,003.60	-660.60	12.36	436,396.30
19.	July 93	5,548	6,231.73	-683.73	12.32	467,500.25
20.	Aug 93	4,916	6,468.54	-1,552.54	31.58	2,410,399.14
21.	Sep 93	5,223	6,714.35	-1,491.35	28.55	2,224,127.11
22.	Oct 93	5,510	6,969.49	-1,459.49	26.48	2,130,128.85
23.	Nov 93	5,689	7,234.33	-1,545.33	27.16	2,388,066.28
24.	Dec 93	6,548	7,509.24	-961.24	14.67	923,985.70
	Sum				**186.25**	**11,786,310.72**

Table 6. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 36 Months (January to December,1994) using the Least Squares Method for Fitting an Exponential Curve.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
37.	Jan 95	5,253	6,862.26	-1,609.26	30.63	2,589,747.31
38.	Feb 95	5,089	6,985.79	-1,896.79	37.27	3,597,812.42
39.	Mar 95	5,083	7,111.53	-2,028.53	39.90	4,114,951.21
	Sum				**107.80**	**10,302,510.94**
Forecast 6 months						
37.	Jan 95	5,253	6,862.26	-1,609.26	30.63	2,589,747.31
38.	Feb 95	5,089	6,985.79	-1,896.79	37.27	3,597,812.42
39.	Mar 95	5,083	7,111.53	-2,028.53	39.90	4,114,951.21
40.	April 95	4,743	7,239.54	-2,496.54	52.63	6,232,721.30
41.	May 95	5,659	7,369.85	-1,710.85	30.23	2,927,020.12
42.	June 95	5,566	7,502.51	-1,936.51	34.79	3,750,074.81
	Sum				**225.45**	**23,212,327.17**
Forecast 12 months						
37.	Jan 95	5,253	6,862.26	-1,609.26	30.63	2,589,747.31
38.	Feb 95	5,089	6,985.79	-1,896.79	37.27	3,597,812.42
39.	Mar 95	5,083	7,111.53	-2,028.53	39.90	4,114,951.21
40.	April 95	4,743	7,239.54	-2,496.54	52.63	6,232,721.30
41.	May 95	5,659	7,369.85	-1,710.85	30.23	2,927,020.12
42.	June 95	5,566	7,502.51	-1,936.51	34.79	3,750,074.81
43.	July 95	5,449	7,637.55	-2,188.55	40.16	4,789,778.18
44.	Aug 95	5,640	7,775.03	-2,135.03	37.85	4,558,362.48
45.	Sep 95	5,608	7,914.98	-2,306.98	41.13	5,322,169.53
46.	Oct 95	5,410	8,057.45	-2,647.45	48.93	7,009,004.56
47.	Nov 95	5,287	8,202.48	-2,915.48	55.14	8,500,062.18
48.	Dec 95	5,682	8,350.13	-2,668.13	46.95	7,118,925.01
	Sum				**495.61**	**60,510,629.11**

Table 7. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 60 Months (January to December, 1996) using the Least Squares Method for Fitting an Exponential Curve.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	5,828.85	-1,026.85	21.38	1,054,439.15
62.	Feb 97	4,646	5,855.67	-1,209.67	26.03	1,463,305.44
63.	Mar 97	5,091	5,882.60	-791.60	15.54	626,642.77
	Sum				**62.95**	**3,144,387.36**
Forecast 6 months						
61.	Jan 97	4,802	5,828.85	-1,026.85	21.38	1,054,439.15
62.	Feb 97	4,646	5,855.67	-1,209.67	26.03	1,463,305.44
63.	Mar 97	5,091	5,882.60	-791.60	15.54	626,642.77
64.	April 97	4,248	5,909.66	-1,661.66	39.11	2,761,139.58
65.	May 97	5,018	5,936.85	-918.85	18.31	844,289.33
66.	June 97	4,991	5,964.16	-973.16	19.49	947,043.70
	Sum				**139.86**	**7,696,859.97**
Forecast 12 months						
61.	Jan 97	4,802	5,828.85	-1,026.85	21.38	1,054,439.15
62.	Feb 97	4,646	5,855.67	-1,209.67	26.03	1,463,305.44
63.	Mar 97	5,091	5,882.60	-791.60	15.54	626,642.77
64.	April 97	4,248	5,909.66	-1,661.66	39.11	2,761,139.58
65.	May 97	5,018	5,936.85	-918.85	18.31	844,289.33
66.	June 97	4,991	5,964.16	-973.16	19.49	947,043.70
67.	July 97	4,828	5,991.59	-1,163.59	24.10	1,353,957.62
68.	Aug 97	4,615	6,019.15	-1,404.15	30.42	1,971,660.23
69.	Sep 97	5,220	6,046.84	-826.84	15.83	683,674.83
70.	Oct 97	4,060	6,074.66	-2,014.66	49.62	4,058,862.22
71.	Nov 97	4,123	6,102.60	-1,979.60	48.01	3,918,836.97
72.	Dec 97	3,682	6,130.67	-2,448.67	66.50	5,996,020.23
	Sum				**374.34**	**25,679,872.07**

Table 8. A Comparison on the Accuracies of the Least Squares Methods in Forecasting the Traffic Accident Data using Different Sizes of Past Data (5, 12, 36 and 60 Months).

Sizes of Past Data(months)	Periods of Forecast (months)	MSE	AMSE	MAPE (%)
5	3	370,184.83	123,394.94	9.29
	6	599,917.89	99,986.31	13.70
	12	1,025,331.09	85,444.25	18.61
12	3	67,603.14	22,534.38	4.31
	6	207,017.23	34,502.87	7.58
	12	982,192.56	818,49.38	15.52
36	3	3,434,170.31	1,144,723.43	35.93
	6	3,868,721.19	6,447,86.86	37.57
	12	5,042,552.42	420,212.70	41.30
60	3	1,048,129.12	349,376.37	20.98
	6	1,282,809.99	213,801.66	23.31
	12	2,139,989.33	178,332.44	31.19

2. Holt and Winters Method.

For Holt and Winters method, the period of forecast using the same sizes of past data were the same as those in the Least Square approach. Except for one period of forecast using the past data in 5 months, January to May 1992, since this period of data was too short to forecast when using the Holt and Winters method. This technique must have available data at least 12 months for estimating the smoothing parameters at many steps of forecast. Consequently, the results of forecasting by Holt and Winters method began with the past data in 12 months during January to December, 1992.

The outcomes of the analysis were shown in Table 9, 10 and 11. The results indicated that the forecast data were less than the actual data when using the size of data in 12 months forecast to 3, 6 and 12 months (Table 9). The error, MAPE were increased from 4.86 for short period of forecast in 3 months to 5.60 for long period of forecast in 12 months (Table 12).

When using the size of data in 36 and 60 months forecast to 3, 6 and 12 months the phenomena were opposite to the former forecasting pattern. That is the forecast data were greater than the actual data in the same period. The errors were monthly fluctuated (see Table 10 and 11). The statistical indicators of errors, MSE, AMSE and MAPE were reported in Table 12. It seemed that the accuracy of the Holt and Winters method varied with the size of past data to forecast, as illustrated by MAPE with value of 3.53 to 11.08 (Table 12). This technique was appropriate in forecasting the accident data in 6 months using the past data of 60 and 12 months which indicated the low values of MAPE at 3.53 % and 4.14 % respectively (Table 12).

Table 9. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 12 Months (January to December, 1992) using Holt and Winters Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
13.	Jan 93	4,942	4,905.48	36.52	0.74	1,333.45
14.	Feb 93	4,763	4,290.15	472.85	9.93	223,584.81
15.	Mar 93	5,183	4,980.14	202.86	3.91	41,151.50
	Sum				**14.58**	**266,069.73**
Forecast 6 months						
13.	Jan 93	4,942	4,905.48	36.52	0.74	1,333.45
14.	Feb 93	4,763	4,290.15	472.85	9.93	223,584.81
15.	Mar 93	5,183	4,980.14	202.86	3.91	41,151.50
16.	April 93	4,855	4,842.78	12.22	0.25	149.40
17.	May 93	5,486	5,314.69	171.32	3.12	29,348.89
18.	June 93	5,343	4,971.41	371.59	6.95	138,076.07
	Sum				**24.88**	**433,644.08**
Forecast 12 months						
13.	Jan 93	4,942	4,905.48	36.52	0.74	1,333.45
14.	Feb 93	4,763	4,290.15	472.85	9.93	223,584.81
15.	Mar 93	5,183	4,980.14	202.86	3.91	41,151.50
16.	April 93	4,855	4,842.78	12.22	0.25	149.40
17.	May 93	5,486	5,314.69	171.32	3.12	29,348.89
18.	June 93	5,343	4,971.41	371.59	6.95	138,076.07
19.	July 93	5,548	5,321.99	226.01	4.07	51,079.28
20.	Aug 93	4,916	5,005.50	-89.50	1.82	8,010.28
21.	Sep 93	5,223	5,824.27	-601.27	11.51	361,527.92
22.	Oct 93	5,510	5,518.25	-8.25	0.15	67.99
23.	Nov 93	5,689	5,618.61	70.39	1.24	4,954.50
24.	Dec 93	6,548	5,003.25	1,544.75	23.59	2,386,258.55
	Sum				**67.29**	**3,245,542.62**

Table 10. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 36 Months (January to December, 1994) using Holt and Winters Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
37.	Jan 95	5,253	5,754.49	-501.49	9.55	251,495.38
38.	Feb 95	5,089	5,546.48	-457.48	8.99	209,287.99
39.	Mar 95	5,083	5,831.43	-748.43	14.72	560,149.67
	Sum				**33.26**	**1,020,933.04**
Forecast 6 months						
37.	Jan 95	5,253	5,754.49	-501.49	9.55	251,495.38
38.	Feb 95	5,089	5,546.48	-457.48	8.99	209,287.99
39.	Mar 95	5,083	5,831.43	-748.43	14.72	560,149.67
40.	April 95	4,743	5,488.89	-745.89	15.73	556,353.81
41.	May 95	5,659	5,731.45	-72.45	1.28	5,248.36
42.	June 95	5,566	5,585.85	-19.83	0.36	393.16
	Sum				**50.63**	**1,582,928.37**
Forecast 12 months						
37.	Jan 95	5,253	5,754.49	-501.49	9.55	251,495.38
38.	Feb 95	5,089	5,546.48	-457.48	8.99	209,287.99
39.	Mar 95	5,083	5,831.43	-748.43	14.72	560,149.67
40.	April 95	4,743	5,488.89	-745.89	15.73	556,353.81
41.	May 95	5,659	5,731.45	-72.45	1.28	5,248.36
42.	June 95	5,566	5,585.85	-19.83	0.36	393.16
43.	July 95	5,449	5,632.99	-183.99	3.38	33,852.21
44.	Aug 95	5,640	5,148.83	491.17	8.71	241,250.58
45.	Sep 95	5,608	5,282.22	325.78	5.81	106,131.67
46.	Oct 95	5,410	5,419.94	-9.94	0.18	98.82
47.	Nov 95	5,287	5,315.42	-28.42	0.54	807.67
48.	Dec 95	5,682	5,494.37	187.63	3.30	35,203.92
	Sum				**72.54**	**2,000,273.23**

Table 11. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 60 Months (January to December, 1996) using Holt and Winters Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	5,003.56	-201.56	4.20	40,626.41
62.	Feb 97	4,646	4,576.20	69.80	1.50	4,872.55
63.	Mar 97	5,091	4,597.56	493.44	9.69	243,482.85
	Sum				**15.39**	**288,981.81**
Forecast 6 months						
61.	Jan 97	4,802	5,003.56	-201.56	4.20	40,626.41
62.	Feb 97	4,646	4,576.20	69.80	1.50	4,872.55
63.	Mar 97	5,091	4,597.56	493.44	9.69	243,482.85
64.	April 97	4,248	4,380.58	-132.58	3.12	17,577.64
65.	May 97	5,018	5,106.25	-88.25	1.76	7,788.89
66.	June 97	4,991	4,943.23	47.77	0.96	2,281.82
	Sum				**21.23**	**316,630.16**
Forecast 12 months						
61.	Jan 97	4,802	5,003.56	-201.56	4.20	40,626.41
62.	Feb 97	4,646	4,576.20	69.80	1.50	4,872.55
63.	Mar 97	5,091	4,597.56	493.44	9.69	243,482.85
64.	April 97	4,248	4,380.58	-132.58	3.12	17,577.64
65.	May 97	5,018	5,106.25	-88.25	1.76	7,788.89
66.	June 97	4,991	4,943.23	47.77	0.96	2,281.82
67.	July 97	4,828	4,875.01	-47.01	0.97	2,209.93
68.	Aug 97	4,615	4,620.95	-5.95	0.13	35.38
69.	Sep 97	5,220	4,164.80	1,055.20	20.21	1,113,455.60
70.	Oct 97	4,060	3,819.51	240.49	5.92	57,833.50
71.	Nov 97	4,123	3,809.96	313.04	7.59	97,996.76
72.	Dec 97	3,682	4,162.08	-480.08	13.04	230,476.50
	Sum				**69.10**	**1,818,637.90**

Table 12. A Comparison on the Accuracies of the Holt and Winters Method in Forecasting the Traffic Accident Data using Different Sizes of Past Data (12, 36 and 60 Months).

Sizes of Past Data(months)	Period of Forecast (months)	MSE	AMSE	MAPE (%)
12	3	88,689.91	29,563.30	4.86
	6	72,274.01	12,045.66	4.14
	12	270,461.90	22,538.49	5.60
36	3	340,311.01	113,437.00	11.08
	6	263,821.39	43,970.23	8.43
	12	166,689.44	13,890.78	6.05
60	3	96,327.27	32,109.09	5.13
	6	52,771.69	8,795.28	3.53
	12	151,553.20	12,629.43	5.76

3. The Box – Jenkins Method

The forecast data of traffic accident in 3, 6 and 12 months using the Box-Jenkins methods were based on the past data in 5, 12, 36 and 60 months. The results were illustrated in Tables 13, 14, 15 and 16, respectively. The error pattern of this method was rather different from previous two single forecast techniques. The forecast data were monthly fluctuated and swung, i.e. some were greater and some were less than the actual data (Table 13, 14, 15, 16) . However, the statistical indicators of the accuracy of this method, MSE, AMSE and MAPE showed the trend of decrease errors of forecast when increased the size of past data in months. As shown in Table 17, the range of error, MAPE in forecasting the data in 3, 6 and 12 months using the size of past data in 5 months were 12.95 and 10.80 % as compared with the ranges of 4.47 and 8.85 % when using the size of data in 60 months (Table 17).

Table 13. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 5 Months (January to May,1992) using the Box-Jenkins Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
6.	June 92	3,622	3,682.76	-60.76	1.67	3,692.53
7.	July 92	3,649	3,262.12	386.88	10.60	149,677.53
8.	Aug 92	4,633	3,402.61	1,230.39	26.56	1,513,864.10
	Sum				**38.84**	**1,667,234.17**
Forecast 6 months						
6.	June 92	3,622	3,682.76	-60.76	1.67	3,692.53
7.	July 92	3,649	3,262.12	386.88	10.60	149,677.53
8.	Aug 92	4,633	3,402.61	1,230.39	26.56	1,513,864.10
9.	Sep 92	4,488	3,389.49	1,098.50	24.47	1,206,702.63
10.	Oct 92	4,046	3,675.98	370.01	9.14	136,908.65
11.	Nov 92	4,950	4,675.96	274.03	5.53	75,094.10
	Sum				**77.97**	**3,085,939.55**
Forecast 12 months						
6.	June 92	3,622	3,682.76	-60.76	1.67	3,692.53
7.	July 92	3,649	3,262.12	386.88	10.60	149,677.53
8.	Aug 92	4,633	3,402.61	1,230.39	26.56	1,513,864.10
9.	Sep 92	4,488	3,389.49	1,098.50	24.47	1,206,702.63
10.	Oct 92	4,046	3,675.98	370.01	9.14	136,908.65
11.	Nov 92	4,950	4,675.96	274.03	5.53	75,094.10
12.	Dec 92	4,756	4,253.56	502.43	10.56	252,445.92
13.	Jan 93	4,942	4,079.65	862.35	17.45	743,650.31
14.	Feb 93	4,763	4,866.93	-103.93	2.18	10,801.35
15.	Mar 93	5,183	4,759.49	423.51	8.17	179,356.94
16.	April 93	4,855	4,813.23	41.77	0.86	1,745.07
17.	May 93	5,486	4,804.82	681.18	12.42	464,008.00
	Sum				**129.61**	**4,737,947.13**

Table 14. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 12 Months (January to December,1992) using the Box-Jenkins Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
13.	Jan 93	4,942	4,277.46	664.54	13.45	441,616.41
14.	Feb 93	4,763	4,927.59	-164.59	3.46	27,089.81
15.	Mar 93	5,183	4,819.46	363.54	7.01	132,160.77
	Sum				**23.92**	**600,866.99**
Forecast 6 months						
13.	Jan 93	4,942	4,277.46	664.54	13.45	441,616.41
14.	Feb 93	4,763	4,927.59	-164.59	3.46	27,089.81
15.	Mar 93	5,183	4,819.46	363.54	7.01	132,160.77
16.	April 93	4,855	4,925.10	-70.10	1.44	4,913.33
17.	May 93	5,486	4,879.98	606.02	11.05	367,254.81
18.	June 93	5,343	5,133.01	210.00	3.93	44,098.81
	Sum				**40.34**	**1,017,133.94**
Forecast 12 months						
13.	Jan 93	4,942	4,277.46	664.54	13.45	441,616.41
14.	Feb 93	4,763	4,927.59	-164.59	3.46	27,089.81
15.	Mar 93	5,183	4,819.46	363.54	7.01	132,160.77
16.	April 93	4,855	4,925.10	-70.10	1.44	4,913.33
17.	May 93	5,486	4,879.98	606.02	11.05	367,254.81
18.	June 93	5,343	5,133.01	210.00	3.93	44,098.81
19.	July 93	5,548	5,022.97	525.03	9.46	275,659.39
20.	Aug 93	4,916	5,475.69	-559.69	11.39	313,256.98
21.	Sep 93	5,223	5,407.05	-184.05	3.52	33,874.97
22.	Oct 93	5,510	5,424.06	85.94	1.56	7,385.75
23.	Nov 93	5,689	5,011.76	677.24	11.90	458,650.37
24.	Dec 93	6,548	5,313.16	1,234.84	18.86	1,524,830.53
	Sum				**97.03**	**3,630,791.93**

Table 15. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 36 Months (January to December,1994) using the Box-Jenkins Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
37.	Jan 95	5,253	6,007.40	-754.40	14.36	569,125.69
38.	Feb 95	5,089	5,724.51	-635.51	12.49	403,874.68
39.	Mar 95	5,083	5,162.49	-79.49	1.56	6,319.10
	Sum				**28.41**	**979,319.47**
Forecast 6 months						
37.	Jan 95	5,253	6,007.40	-754.40	14.36	569,125.69
38.	Feb 95	5,089	5,724.51	-635.51	12.49	403,874.68
39.	Mar 95	5,083	5,162.49	-79.49	1.56	6,319.10
40.	April 95	4,743	5,066.65	-323.65	6.82	104,752.09
41.	May 95	5,659	4,916.57	742.43	13.12	551,208.24
42.	June 95	5,566	5,118.41	447.59	8.04	200,339.57
	Sum				**56.39**	**1,835,619.37**
Forecast 12 months						
37.	Jan 95	5,253	6,007.40	-754.40	14.36	569,125.69
38.	Feb 95	5,089	5,724.51	-635.51	12.49	403,874.68
39.	Mar 95	5,083	5,162.49	-79.49	1.56	6,319.10
40.	April 95	4,743	5,066.65	-323.65	6.82	104,752.09
41.	May 95	5,659	4,916.57	742.43	13.12	551,208.24
42.	June 95	5,566	5,118.41	447.59	8.04	200,,339.57
43.	July 95	5,449	5,599.12	-150.12	2.76	22,536.70
44.	Aug 95	5,640	5,495.77	144.23	2.56	20,802.68
45.	Sep 95	5,608	5,511.64	96.36	1.72	9,285.16
46.	Oct 95	5,410	5,606.44	-196.44	3.63	38,587.82
47.	Nov 95	5,287	5,502.83	-215.83	4.08	46,580.47
48.	Dec 95	5,682	5,337.18	344.82	6.07	118,900.67
	Sum				**77.21**	**2,092,312.87**

Table 16. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 60 Months (January to December,1996) using the Box-Jenkins Method.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	4,812.39	-10.39	0.22	107.92
62.	Feb 97	4,646	4,932.79	-286.79	6.17	82,247.68
63.	Mar 97	5,091	4,732.47	358.53	7.04	128,544.02
	Sum				**13.43**	**210,899.62**
Forecast 6 months						
61.	Jan 97	4,802	4,812.39	-10.39	0.22	107.92
62.	Feb 97	4,646	4,932.79	-286.79	6.17	82,247.68
63.	Mar 97	5,091	4,732.47	358.53	7.04	128,544.02
64.	April 97	4,248	4,831.55	-583.55	13.74	340,533.00
65.	May 97	5,018	4,729.89	288.11	5.74	83,009.53
66.	June 97	4,991	4,571.49	419.51	8.41	175,987.16
	Sum				**41.32**	**810,429.31**
Forecast 12 months						
61.	Jan 97	4,802	4,812.39	-10.39	0.22	107.92
62.	Feb 97	4,646	4,932.79	-286.79	6.17	82,247.68
63.	Mar 97	5,091	4,732.47	358.53	7.04	128,544.02
64.	April 97	4,248	4,831.55	-583.55	13.74	340,533.00
65.	May 97	5,018	4,729.89	288.11	5.74	83,009.53
66.	June 97	4,991	4,571.49	419.51	8.41	175,987.16
67.	July 97	4,828	5,003.22	-175.22	3.63	30,702.45
68.	Aug 97	4,615	4,918.50	-303.50	6.58	92,111.43
69.	Sep 97	5,220	4,734.28	485.72	9.31	235,926.70
70.	Oct 97	4,060	4,868.46	-808.46	19.91	653,608.95
71.	Nov 97	4,123	4,724.34	-601.34	14.59	361,612.16
72.	Dec 97	3,682	4,083.42	-401.42	10.90	161,137.81
	Sum				**106.24**	**2,345,528.81**

Table 17. A Comparison on the Accuracies of the Box-Jenkins Methods in Forecasting the Traffic Accident Data using Different Sizes of Past Data (5, 12, 36 and 60 Months).

Sizes of Past Data(months)	Periods of Forecast (months)	MSE	AMSE	MAPE (%)
5	3	555,744.72	185,248.24	12.95
	6	514,323.26	85,720.54	12.99
	12	94,828.93	32,902.41	10.80
12	3	200,288.99	66,762.99	7.97
	6	169,522.32	28,253.72	6.72
	12	302,565.99	25,213.83	8.08
36	3	326,439.82	108,813.27	9.47
	6	305,936.56	50,989.43	9.39
	12	174,359.40	14,529.95	6.43
60	3	70,299.87	23,433.29	4.47
	6	135,071.55	22,511.93	6.88
	12	195,460.73	16,288.39	8.85

4. A Comparison on the Accuracy of the Single Forecast Methods.

In general, the accuracy of each forecast method can be evaluated by three statistical indicators of errors, MSE, AMSE and MAPE. The less values of errors the more accuracy of the forecast method is indicated. Among these error indices, the MAPE reports as percentage which is more easier to interpret and understand its meaning than the other indices. Therefore, this study used the MAPE as the error index to evaluate and compare the accuracies among the Least Squares, Holt and Winters and the Box-Jenkins methods.

Table 18. A Comparison on the Mean Absolute Percentage Error (MAPE) of Three Single Forecast Methods.

Sizes of Past Data (months)	Periods of Forecast (months)	Least Squares method MAPE %	Holt and Winters method MAPE %	Box-Jenkins method MAPE %
5	3	9.29 *	-	12.95
	6	13.70	-	12.99
	12	18.61	-	10.80
12	3	4.31	4.86	7.97
	6	7.58	4.14 *	6.72
	12	15.52	5.60	8.08
36	3	35.93	11.08	9.47
	6	37.57	8.43	9.39
	12	41.30	6.05 *	6.43
60	3	20.98	5.13	4.47
	6	23.31	3.53 *	6.88
	12	31.19	5.76	8.85

* means minimum MAPE for each size of data

Table 18 compared the MAPE indices among three forecast methods classified by sizes of past data to forecast, 5, 12, 36 and 60 months, and period of forecast, 3, 6 and 12 months. Based on the same size of past data and same periods of forecast, except the periods of forecast in 3 months using past data in 5 and 12 months to forecast, the Least Squares method yielded the errors higher than the other two methods. In particular, the forecast during 3 to 12 months using the past data in 36 months gave the MAPE higher than 35 % and when using the past data in 60 months the MAPE was between 21 and 31 % (Table 18). The low values of MAPE of the Least Squares methods were 4.3 % and 9.3 % for forecasting 3 months from the past data in 12 and 5 months , respectively.

Considering the Holt and Winters and the Box-Jenkins methods, the error as represented by MAPE of the forecast by the Holt and Winters method, on the average were lower than those of the Box-Jenkins and of the Least Squares methods. The forecast data, by the Holt and Winters approach, during 3 to 6 months using the past data in 60 months gave the errors ranging 3.5 % to 5.8 % as compared with the errors in the same event of the Box-Jenkins technique ranging 4.5 % to 8.9 %. The accuracy of the Box-Jenkins method was better than that of the Holt and Winters method in the forecast period in 3 months using the past data in 36 months. The error was about 9.5% (Table 18).

In conclusion, the Least Squares method was appropriate for forecasting in short period in 3 months using the past data in 5 and 12 months. Also, the Box-Jenkins method was appropriate for forecasting in 3 months using the past data in 60 months, as reported by the MAPE at 4.5 %. The Holt and Winters method was appropriate for forecasting in long period in 6 and 12 months using the past data in 12 and 60 months (Table 18).

B. The Combined Forecasts via Weighted Averages.

The combined forecasts in this study were three pairs of the single forecasts methods as follow:

1. The Combined Method between the Least – Squares and Holt and Winters Methods.
2. The Combined Method between the Least – Squares and the Box – Jenkins Methods.
3. The Combined Method between Holt and Winters and the Box – Jenkins Methods.

The combined methods used three periods of forecast in 3, 6 and 12 months, from the large size of past data in 60 months as suggested by Winkler (8). The results of forecasting by each pair of the combined techniques were described in the next section.

1. The Combined Method between the Least Squares and Holt and Winters Methods.

Table 19 reported the forecast data and errors of the combined forecast between the Least Squares and Holt and Winters methods. The error pattern was fluctuated. Some were greater and some were less than the actual data. The MAPE values were between 3.5 % and 5.6 % (Table 22).

2. The Combined Method between the Least Squares and the Box-Jenkins Methods.

The combined forecast between the Least Squares and the Box-Jenkins via weighted averages gave the errors quite similar to the error pattern of the combined forecast between the Least Squares and Holt and Winters techniques (see Table 20).

The ranges of error (MAPE) were somewhat higher than those of the former method. They were about 5.0 to 9.7 % (Table 22).

Table 19. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 60 Months (January to December,1996) using the Combined Method between the Least Squares and Holt and Winters Methods.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	5,046.47	-244.47	5.09	59,766.67
62.	Feb97	4,646	4,642.72	3.28	0.07	10.73
63.	Mar 97	5,091	4,664.38	426.62	8.38	182,006.68
	Sum				**13.54**	**241,784.08**
Forecast 6 months						
61.	Jan 97	4,802	5,046.47	-244.47	5.09	59,766.67
62.	Feb97	4,646	4,642.72	3.28	0.07	10.73
63.	Mar 97	5,091	4,664.38	426.62	8.38	182,006.68
64.	April 97	4,248	4,460.09	-212.09	4.99	44,980.98
65.	May 97	5,018	5,149.44	-131.44	2.62	17,277.13
66.	June97	4,991	4,996.32	-5.32	0.11	28.26
	Sum				**21.26**	**304,070.45**
Forecast 12 months						
61.	Jan 97	4,802	5,046.47	-244.47	5.09	59,766.67
62.	Feb 97	4,646	4,642.72	3.28	0.07	10.73
63.	Mar 97	5,091	4,664.38	426.62	8.38	182,006.68
64.	April 97	4,248	4,460.09	-212.09	4.99	44,980.98
65.	May 97	5,018	5,149.44	-131.44	2.62	17,277.13
66.	June 97	4,991	4,996.32	-5.32	0.11	28.26
67.	July 97	4,828	4,933.07	-105.07	2.18	11,039.29
68.	Aug 97	4,615	4,693.65	-78.65	1.70	6,185.70
69.	Sep 97	5,220	4,262.66	957.34	18.34	916,509.20
70.	Oct 97	4,060	3,936.77	123.28	3.04	15,184.93
71.	Nov 97	4,123	3,929.16	193.84	4.70	37,572.24
72.	Dec 97	3,682	4,264.44	-582.44	15.82	339,235.18
	Sum				**67.04**	**1,629,796.99**

Table 20. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data in 60 Months (January to December,1996) using the Combined Method between the Least Squares and the Box-Jenkins Methods.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	4,892.77	-90.77	1.89	8,239.30
62.	Feb 97	4,646	5,005.77	-359.77	7.74	129,434.29
63.	Mar 97	5,091	4,823.42	267.58	5.26	71,597.94
	Sum				**14.89**	**209,271.53**
Forecast 6 months						
61.	Jan 97	4,802	4,892.77	-90.77	1.89	8,239.30
62.	Feb 97	4,646	5,005.77	-359.77	7.74	129,434.29
63.	Mar 97	5,091	4,823.42	267.58	5.26	71,597.94
64.	April 97	4,248	4,916.81	-668.81	15.74	447,305.48
65.	May 97	5,018	4,825.33	192.67	3.84	37,120.72
66.	June 97	4,991	4,681.62	309.38	6.20	95,713.80
	Sum				**40.67**	**789,411.53**
Forecast 12 months						
61.	Jan 97	4,802	4,892.77	-90.77	1.89	8,239.30
62.	Feb 97	4,646	5,005.77	-359.77	7.74	129,434.29
63.	Mar 97	5,091	4,823.42	267.58	5.26	71,597.94
64.	April 97	4,248	4,916.81	-668.81	15.74	447,305.48
65.	May 97	5,018	4,825.33	192.67	3.84	37,120.72
66.	June 97	4,991	4,681.62	309.38	6.20	95,713.80
67.	July 97	4,828	5,081.38	-253.38	5.25	64,202.18
68.	Aug 97	4,615	5,005.54	-390.54	8.46	152,520.21
69.	Sep 97	5,220	4,838.07	381.93	7.32	145,867.03
70.	Oct 97	4,060	4,963.85	-903.85	22.26	816,938.92
71.	Nov 97	4,123	4,833.33	-710.33	17.23	504,575.05
72.	Dec 97	3,682	4,245.32	-563.32	15.30	317,324.94
	Sum				**116.49**	**2,790,839.86**

3. The Combined Method between Holt and Winters and the Box-Jenkins Methods.

Data analysis of the combined forecast between Holt and Winters and the Box-Jenkins methods was shown in Table 21.

Table 21. Forecasting the Traffic Accident Data in Bangkok during 3, 6 and 12 Months from the Past Data is 60 Months (January to December,1996) using the Combined Method between the Least Squares and the Box-Jenkins methods.

No.	Month	Actual	Forecast	Error	Absolute Percentage Error	Squared Error
Forecast 3 months						
61.	Jan 97	4,802	4,929.05	-127.05	2.65	16,140.76
62.	Feb 97	4,646	4,715.19	-69.19	1.49	4,786.80
63.	Mar 97	5,091	4,650.14	440.86	8.66	194,353.67
	Sum				**12.80**	**215,281.23**
Forecast 6 months						
61.	Jan 97	4,802	4,929.05	-127.05	2.65	16,140.76
62.	Feb 97	4,646	4,715.19	-69.19	1.49	4,786.80
63.	Mar 97	5,091	4,650.14	440.86	8.66	194,353.67
64.	April 97	4,248	4,556.36	-308.36	7.26	95,084.45
65.	May 97	5,018	4,959.56	58.44	1.16	3,415.69
66.	June 97	4,991	4,798.34	192.66	3.86	37,119.01
	Sum				**25.08**	**350,900.38**
Forecast 12 months						
61.	Jan 97	4,802	4,929.05	-127.05	2.65	16,140.76
62.	Feb 97	4,646	4,715.19	-69.19	1.49	4,786.80
63.	Mar 97	5,091	4,650.14	440.86	8.66	194,353.67
64.	April 97	4,248	4,556.36	-308.36	7.26	95,084.45
65.	May 97	5,018	4,959.56	58.44	1.16	3,415.69
66.	June 97	4,991	4,798.34	192.66	3.86	37,119.01
67.	July 97	4,828	4,924.98	-96.98	2.01	9,405.76
68.	Aug 97	4,615	4,736.93	-121.93	2.64	14,865.81
69.	Sep 97	5,220	4,386.76	833.24	15.96	694,280.68
70.	Oct 97	4,060	4,228.37	-168.37	4.15	28,347.22
71.	Nov 97	4,123	4,166.36	-43.36	1.05	1,880.07
72.	Dec 97	3,682	4,131.42	-449.42	12.21	201,978.39
	Sum				**63.10**	**1,301,658.31**

The forecast data gave fluctuate errors with some greater and less than the actual data. The MAPE indicator of the accuracy of the combined forecast method was quite low at 4 to 5 % (Table 22).

4. A Comparison on the Accuracy of the Combined Forecast Methods.

Table 22 showed the values of mean absolute percentage errors (MAPE) of three combined forecast techniques. For the periods of forecast in 3 and 12 months using the past data of 60 months, the combined method between Holt and Winters and the Box-Jenkins gave the low errors (MAPE) at 4.3 to 5.3 %, where as for the forecast in 6 months, a pair of the Least Squares and Holt and Winters showed the lowest error at 3.5 % (Table 22). The combined forecast between the Least Squares and the Box-Jenkins yielded the errors, on the average, about 5.0 to 9.7 % higher than the other two methods.

Table 22. A Comparison on the Accuracy of Three Combined Forecast Methods.

Combined Forecasts	Period of Forecast (months)	MSE	AMSE	MAPE %
Least Square + Holt -Winters	3	80,594.69	26,864.89	4.51
	6	50,678.41	8,446.40	3.54
	12	135,816.42	11,318.04	5.59
Least Square + Box -Jenkins	3	69,757.17	23,252.39	4.96
	6	131,568.58	21,928.09	6.77
	12	232,569.99	19,380.83	9.71
Holt-Winters + Box -Jenkins	3	71,760.41	23,920.14	4.26
	6	58,483.40	9,747.23	4.18
	12	108,471.53	9,039.29	5.26

C. A Comparison on the Accuracy between the Single and Combined Forecast Method.

The important indicator of MAPE was used to evaluate and compare the accuracy between the single and combined forecast methods. Based on the same size of past traffic accident data in 60 months for forecasting, two combined forecasts, i.e. a pair of Holt and Winters and the Box-Jenkins method (HWS + BJ) and a pair of the Least Squares and Holt and Winters (LS + HWS) , displays the accuracy of forecast methods (Table 23).

Table 23. A Comparison on the Accuracy between Single and Combined Forecast Methods using the Past Data in 60 Months.

Forecast in months	Stat Ind.	Single Forecasts LS	HWS	BJ	Combined Forecasts LS +HWS	LS+BJ	HWS+BJ
3	MSE	1,048,129.12	96,327.27	70,299.87	80,594.69	69,757.17	71,760.41
	AMSE	349,376.37	32,109.09	23,433.29	26,864.89	23,252.39	23,920.14
	MAPE	20.98 %	5.13 %	4.47 %	4.51 %	4.96 %	4.26 %
6	MSE	1,282,809.99	52,771.69	135,071.55	50,678.41	131,568.58	58,483.40
	AMSE	213,801.66	8,795.28	22,511.93	8,446.40	21,928.09	9,747.23
	MAPE	23.31 %	3.53 %	6.88 %	3.54 %	6.77 %	4.18 %
12	MSE	2,139,989.33	151,553.20	195,460.73	135,816.42	232,569.99	108,471.53
	AMSE	178,332.44	12,629.43	16,288.39	11,318.04	19,380.83	9,039.29
	MAPE	31.19 %	5.76 %	8.85 %	5.59 %	9.71 %	5.26 %

LS Least-Squares method
HWS Holt and Winters method
BJ Box –Jenkins method

In considering the period of forecast in short term in 3 months, the accuracy of the methods were rank in subsequent order; HWS + BJ with MAPE at 4.26 %, the Box-Jenkins method with MAPE at 4.47 % and LS + HWS with MAPE at 4.51 %. The Least Squares procedure gave high error of forecast with MAPE more than 20 % (Table 23).

In the medium term of forecast in 6 months, the single method of Holt and Winters gave the error at 3.53 % close to that of the combined method of LS + HWS with MAPE at 3.54 % as compared with the HWS + BJ with MAPE at 4.18 % (Table 23). For this case, the accuracy of the single method of HWS were the same as the combined forecast of LS + HWS. So, the single approach of HWS is more practically easier recommended to use in forecast during this period.

In the long term of forecast in 12 months, the combined forecasts of HWS + BJ and LS + HWS and the single forecast of HWS reported the errors at 5.26 %, 5.59 % and 5.76 % respectively (Table 23). However, it was seem that the combined forecast of LS + BJ gave the error higher than that of the other two combined methods and also higher than that of some of the single forecast (LS).

CHAPTER V

DISCUSSION

The findings of this study were discussed emphasizely on the main objectives.

A. The Single Forecast Methods.

In comparison among three methods of forecasts, the results revealed that, on the average, Holt and Winters technique gave highest accuracy followed by the Box-Jenkins and Regression using Least Square method, respectively. It did appear that the Least Squares approach yielded the results of forecast nearly accurate as those of the Box-Jenkins when used the size of data in 5 and 12 months forecast in short periods during 3 to 6 months. This finding was also supported by Kunaporn's study (16) which used the Least Squares technique to analyze the trends of the numbers of injured and killed persons and the assurance of road traffic accidents in Bangkok. The trends of incidence and death rates were applicable for making a relatively short period forecast only.

In the long forecast period with large size of past data during 36 to 60 months, the Least Squares method gave the errors so high as 4 to 6 times of those from the other two methods, as reported by MAPE around 20 to 41 %. This outcome may be due to the weak point of the Least Squares technique in this study that the traffic accident data in long period did not fit an exponential model of the Least Squares approach. This evidence, as shown by large errors of forecast, can be seen in Table 6 and 7.

In Newbold and Granger (3)'s study as well as in Ngamaugpak study (4), the authors had found that the Box-Jenkins method produced superior forecasts of economic time series in the short period. The performance of the Box-Jenkins in this study also showed the same phenomena as in Newbold - Granger's and Ngamaugpak's studies. It yielded the lowest errors of forecasting the traffic accidents during three months of forecast. But in 6 and 12 months of forecasts, the Holt and Winters method gave better accuracy than the Box-Jenkins. The reasons behind this result were that the outperformance of the Holt-Winters technique depends on the settle down of smoothed estimates of seasonal factors of the traffic accident data. The author has examined crossectional traffic accident data yearly in 1992 to 1997. It was found that the nature of the data movement in time series increased quite smoothly and stationary, as can be seen in Table 9, 10, 11. This corresponds to the fit model of the Holt-Winters approach, which could produce the high accuracy of the forecast results.

B. The Combined Forecast Methods

This study used the combined forecasts by weighted averages with large size of past data as suggested by Winkler (8), Newbold and Granger (3). The findings indicated that the combined forecast between Holt-Winters and the Box-Jenkins was more efficient in forecasting both short term and long term seasonal time series.

Since the traffic accident data were better fit to the Holt-Winters model than the Least Squares with exponential model, as previous discussed in topic A. Consequently, the combination of Holt-Winters and the Box-Jenkins would yield the results more accurate than those from other two pairs of combined forecasts, i.e. the Least Squares + Holt-Winters, and the Least Squares + Box-Jenkins. So, it could be concluded from this study that the combined forecast of Holt-Winters and the Box-Jenkins was the efficient method for forecasting the traffic accident in Bangkok.

C. A Comparison on the Accuracy between the Single and Combined Forecast Methods.

The main conclusions from many studies in the past (3)(4)(8)(11)(14) revealed that the combined forecast procedures provided forecasts that were more accurate overall than forecasts from individual methods. This is also quite true for the results of this study. All combined forecasts on the average performed more efficient forecasts than those from the single forecast approach, as reported by low percentage of error (MAPE) in Table 23. However, among the combined forecast methods, it should confirm the previous conclusion that the combined technique of Holt-Winters and the Box-Jenkins was the appropriate method used to forecast the traffic accident in Bangkok.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

This study was purposively conducted to compare the accuracy between the single and combined forecast methods. The single methods consist of the Least Squares, the Box-Jenkins and Holt and Winters techniques. For the combined forecasts via weight averages, they include three pairs of the single methods, i.e. the Least Squares and Holt and Winters, the Least Squares, and the Box-Jenkins and Holt and Winters and the Box-Jenkins Methods.

Each forecast method used the traffic accident data in Bangkok during January, 1992 to December, 1997. The data were collected from a police information center, the royal Thai police department. The traffic accident data were monthly time series with trend, seasonal effect and nonstationary. The past accident data were divided into 4 groups: 5 months, 12 months, 36 months and 60 months.

For each size of past data, the single forecast and combined forecast techniques, using SPSS program and Microsoft Excel for Windows 97, were applied to forecast forward for next 3 months, 6 months and 12 months, respectively. For the large size data of 60 months during January, 1992 to December, 1996, the data were forecast to the year 1997 and were compared with the actual data in 1997. The correctness of the forecast as measured against actual events in the same period was used as the criterion for evaluating the performance of the forecasting methods. In this study, three statistical indicators, i.e. mean square error (MSE), average mean sum square error (AMSE) and mean absolute percentage error (MAPE) were used to evaluate the accuracy of the methods.

The results could be summarized as follows.

1. The Single Forecast Methods

The Least Squares method was appropriate for forecasting in short period in 3 months using the past data in 5 and 12 months. Also, the Box-Jenkins method was appropriate for forecasting in 3 months using the past data in 60 months, as reported by the MAPE at 4.5 %. The Holt and Winters method was appropriate for forecasting in long period in 6 and 12 months using the past data in 12 and 60 months.

2. The Combined Forecast Methods

For the periods of forecast in 3 and 12 months using the past data of 60 months, the combined method between Holt and Winters and the Box-Jenkins gave the low errors (MAPE) at 4.3 to 5.3 %, where as for the forecast in 6 months, a pair of Least Squares and Holt and Winters showed the lowest error at 3.5 %. The combined forecast between the Least Squares and the Box-Jenkins yielded the errors, on the average, about 5.0 to 9.7 % higher than the other two methods.

3. A Comparison on the Accuracy between the Single and Combined Forecast Methods.

Based on the same size of past accident data in 60 months for forecasting, two combined forecasts, i.e. a pair of Holt and Winters and the Box-Jenkins method (HWS + BJ) and a pair of the Least Squares and Holt and Winters (LS +HWS), displays the accuracy of forecast methods as indicated by the percentages of MAPE.

In considering the period of forecast in short term in 3 months, the accuracy of the methods were rank in subsequent order; HWS + BJ with MAPE at 4.26 % the Box-Jenkins method with MAPE at 4.47 % and LS +HWS with MAPE at 4.51 %. The Least Squares procedure gave high error of forecast with MAPE more than 20 %.

In the medium term of forecast in 6 months, the single method of Holt and Winters gave the error at 3.53 % close to that of the combined method of LS + HWS with MAPE at 3.54 % as compared with the HWS +BJ with MAPE at 4.18 %.

In the long term of forecast in 12 months, the combined forecasts of HWS +BJ and LS + HWS and the single forecast of HWS reported the error at 5.26 %, 5.59 % and 5.76 %, respectively.

In sum, two combined forecast techniques, HWS + BJ and LS + HWS yielded the accuracies of forecasting the traffic accident in Bangkok better than those of the single forecast methods. Among the single forecasts, HWS revealed the low errors of its forecasting results.

B. Recommendations

From the results of this study, there were some recommendations that the forecaster should carefully select the efficient forecasting technique in predicting reliable traffic accident data. These data would be very useful to the policeman, government officials and other organizations for making policy and administrative planning in order to reduce the traffic accident in Bangkok.

The guidelines for selection of the appropriate forecasting methods were as follows.

1. The forecaster should divide the past traffic accident data to be forecast into 3 groups, small (3-6 months), medium (12-36 months) and large sizes (60 months or more). Each group of the data would be considered relative to the nature of the data, time series, and the length of forecasting times, 3,6 and 12 months.

2. For the small and medium size of past accident data, the single forecasts should be selected. Among the single forecast techniques, the Least Squares method was

appropriate for forecasting the traffic accident during 3 to 6 months using small size of past data. Whereas for the medium size of past data, the Holt and Winters method should be selected to forecast the events covered both short and long periods of forecast, 3 to 12 months.

3. The combined forecast approaches, especially the Holt and Winters + the Box-Jenkins method should be used to forecast the traffic accident in Bangkok during short and long periods by using the large size of past data.

C. Recommendations for Further Studies

This research study has provided significant insights into the comparison of different efficient forecasting techniques. Although numerous questions have been answered, opportunities for further research still exist.

1. Future research needs to continue to study the accuracy of the combined forecasts using Baysian technique as compare to that using weighted averages approach.

2. Since this research has used the limited traffic accident data only time series in nature, future study should continue to use different nature of movement data for forecasting, such as horizontal and cyclical.

3. Further study should be conducted to compare the efficient single and combined forecast methods using the traffic accident data covering all provinces in Thailand.

REFERENCES

1. Makridakis, S., Wheelwright, S.C., McGee, V.E. **Forecasting: Method and Application. 2nd ed.** New York: John Wiley & Son, 1983.
2. Taesombut, S **Quantitative Forecasting Techniques**. Bangkok: Physic Center, 1996.
3. Newbold, P , Granger, C.W.J. Experience with Forecasting Univariate Time Series and the Combination of Forecasts. **J. R. Statistics Society, Series A**, 1974; 137: 131-145.
4. Ngamaugpak W. Comparison in Forecast Methods. [M.S. Thesis in Applies Statistics]. Bangkok: Faculty of science, Thamasart University, 1995.
5. Helfenstein, U. Box-Jenkins Modeling in Medical Research. **Statistic Method in Medical Research**, 1996; 5: 3-22.
6. Milner, P.C. Ten- Year Follow-up of Arima Forecasts of Attendances at Accident and Emergency Departments in The Trent Region. **Statistics in Medicine**, 1997; 16: 2117-2125.
7. Bates, J M., Granger, C.W.J. The Combination of Forecasts. **Operation Research Quarterly**, 1969; 20: 451-468.
8. Winkler, R.L, Makridakis, S. The Combination of Forecasts. **J. R. Statistics Society, Series A**, 1983; 146: 150-157.
9. Police Information System Center. **Traffic Accident Situation in Thailand**, Annual Report, 1992-1997.

10. Steering Committee on Future Health Scenarios. **Accident in the year 2000.** London: Kluwer Academic Publisher, 1989.

11. Deutsch, M. The Combination of Forecasts using Changing Weight. **International Journal of Forecasting**, 1994 ;10: 47-57.

12. Chatfield, C. **The Analysis of Time Series: Theory and Practice.** London: Chapman and Hall, 1975.

13. Makridakis, S., Hibon M. Arma Models and the Box-Jenkins Methodology. **Journal of Forecasting** 1997; 16: 147-163.

14. Clemen, RT. Combination Economic Forecasts. **Journal of Business & Economic Statistics**, January 1986; 4: 39-46.

15. Luengprateep, J.Accuracy Comparison in Forecasting Time Series with Deseasonalized Data via Holt (1960) and Holt Adding A Damping Parameter Methods. [M.S.Thesis in Applied Statistics]. Bangkok: Faculty of Science, Thamasart University, 1992.

16. Kunaporn, T. Statistical Analysis of Road Traffic Accidents in Krungthep Maha Nakorn. [M.S.Thesis in Biostatistics]. Bangkok: Faculty of Graduate Studies, Mahidol University, 1986.

17. Johansson, P. Speed limitation and motorway casualties: a time series count data regression approach. **Accident Analysis & Prevention,** 1996; 28(1): 73 - 87.

18. Rock, SM. Impact of the 65 mph speed limit on accidents, deaths, and injuries in Illinois. **Accident Analysis & Prevention,** 1995; 27(2): 207-214.

19. Martinez-Schnell B., Zaidi A. Time series analysis of injuries. **Statistics in Medicine,** 1989; 8(12): 1497- 1508.

20. Karmonrut, K. A Comparative Study on forecasting Techniques Suitable for Data Characteristics [M.S. Thesis in Statistics] Bangkok: Chulalongkorn University,1986.

21. Box, G. E. P. and Jenkins, G. M. **Time series Analysis: Forecasting and Control.** San Francisco: Holden-Day, 1976.

22. Garcia – Ferrer, A. Hoyo, JD. Analysis of the car Accident Indexes in Spain: A Multiple Time Series Approach. **Journal of Business & Economic Statistics**, January 1987; 5: 27-37.

23. Na Ayuthya, R S., Bohning, D. Risk Factors for Traffic Accidents in Bangkok Metropolis: a case reference study. **Southeast Asian Journal of Tropical Medicine & Public Health**, 1997; 28(4): 881-885.

24. Jessop, A. **Decision and Forecasting Models with Transport Applications.** New York: Ellis Horwood, 1990.

Appendix A

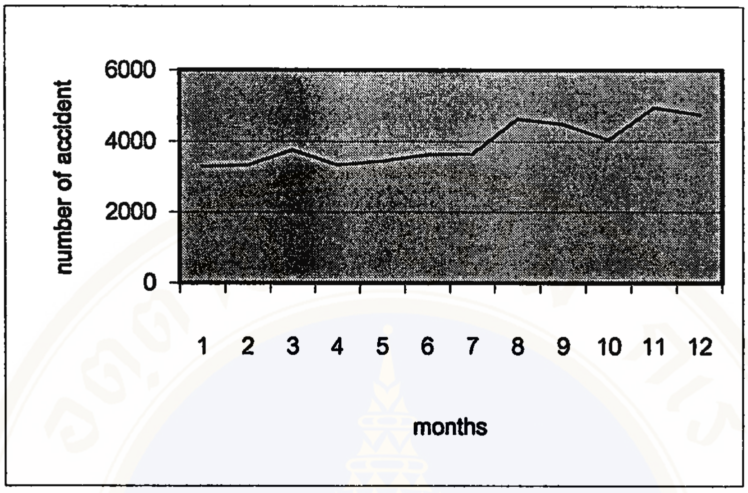
The Traffic Accident Data in Bangkok

MONTH \ YEAR	1992	1993	1994	1995	1996	1997
JANUARY	3285	4942	5720	5253	5545	4802
FEBRUARY	3328	4763	5276	5089	5196	4646
MARCH	3759	5183	5987	5083	5214	5091
APRIL	3333	4855	5607	4743	4805	4248
MAY	3440	5486	5584	5659	5437	5018
JUNE	3622	5343	6152	5566	5158	4991
JULY	3649	5548	7109	5449	5061	4828
AUGUST	4633	4916	6491	5640	4926	4615
SEPTEMBER	4488	5223	5976	5608	4734	5220
OCTOBER	4046	5510	6381	5410	4543	4060
NOVEMBER	4950	5689	5959	5287	4654	4123
DECEMBER	4756	6548	6117	5682	5035	3682

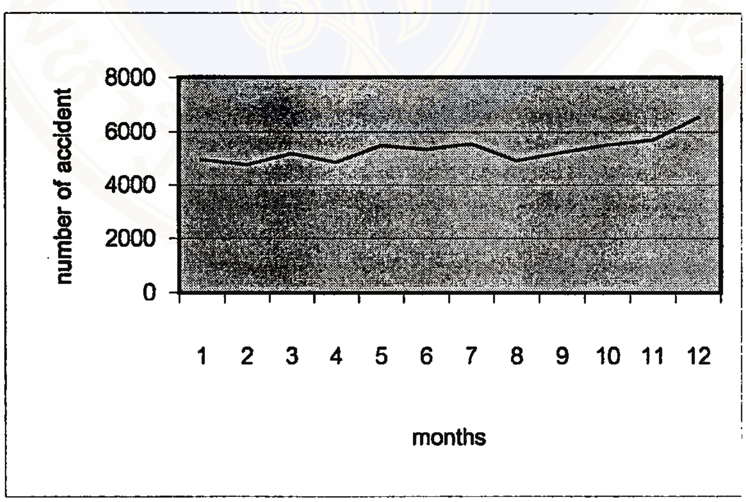
Unit: times

Source: Police Information System Center.

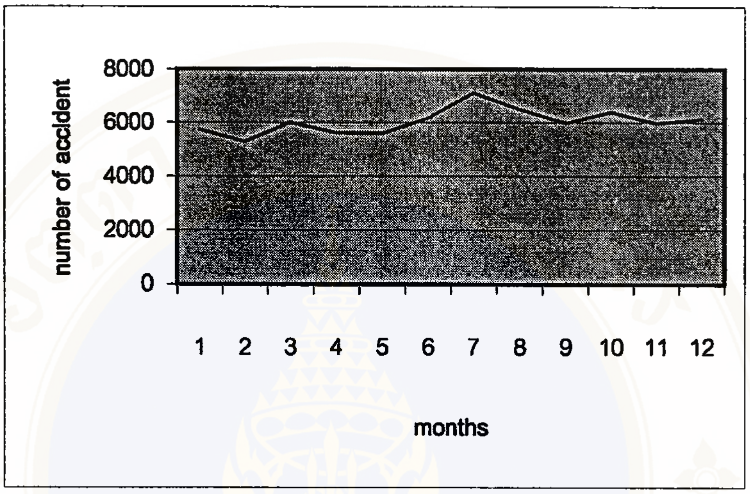
Graph of the Traffic Accident data until January 1992 to December 1992



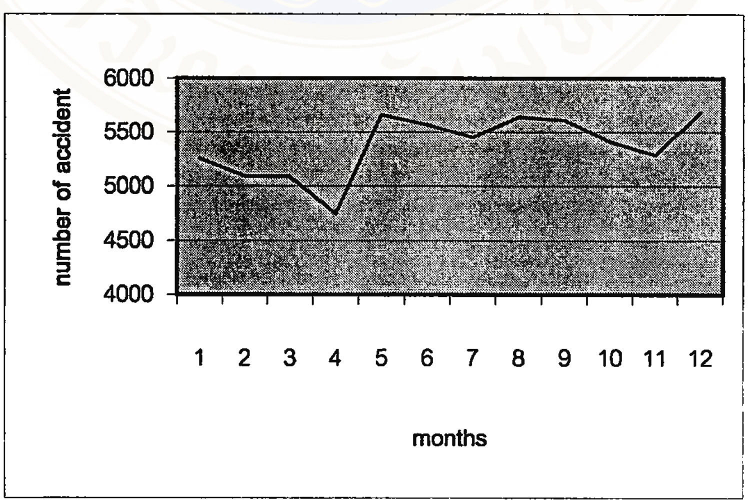
Graph of the Traffic Accident data until January 1993 to December 1993



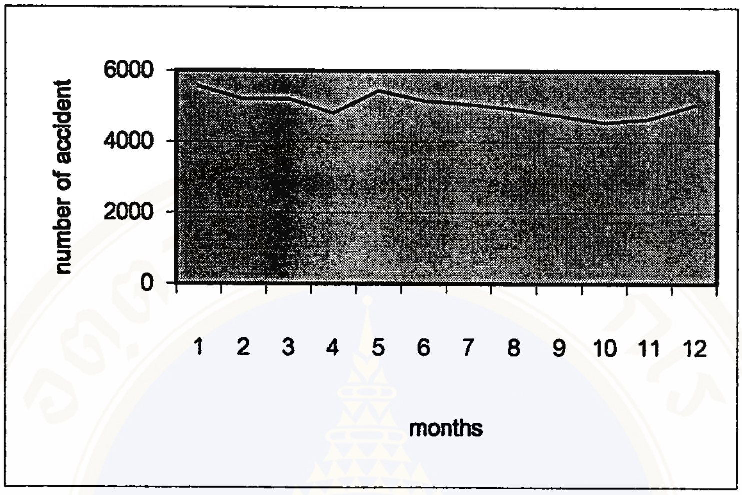
Graph of the Traffic Accident data until January 1994 to December 1994



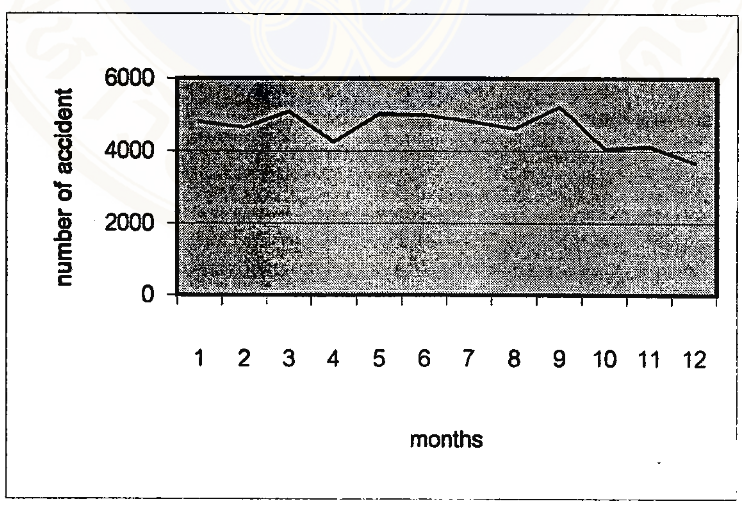
Graph of the Traffic Accident data until January 1995 to December 1995



Graph of the Traffic Accident data until January 1996 to December 1996



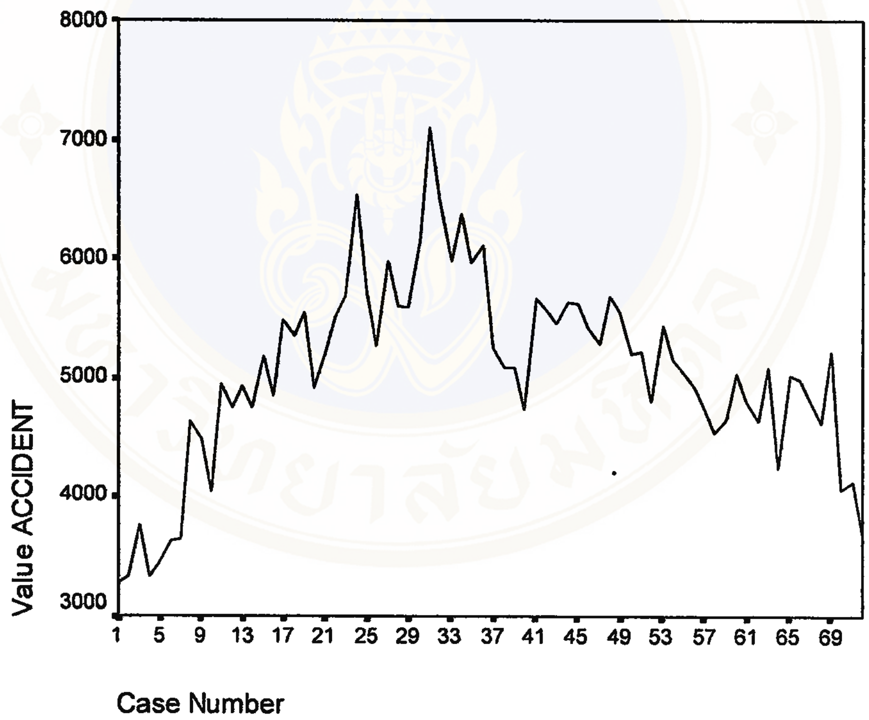
Graph of the Traffic Accident data until January 1997 to December 1997



Appendix B

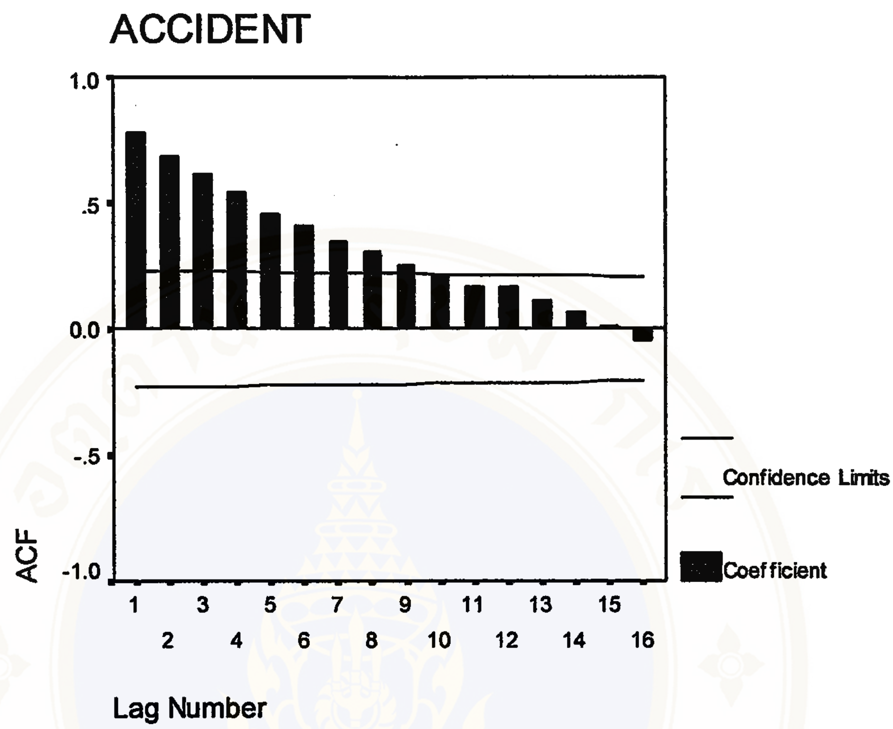
Graph and Autocorrelation Before Transformation of Data

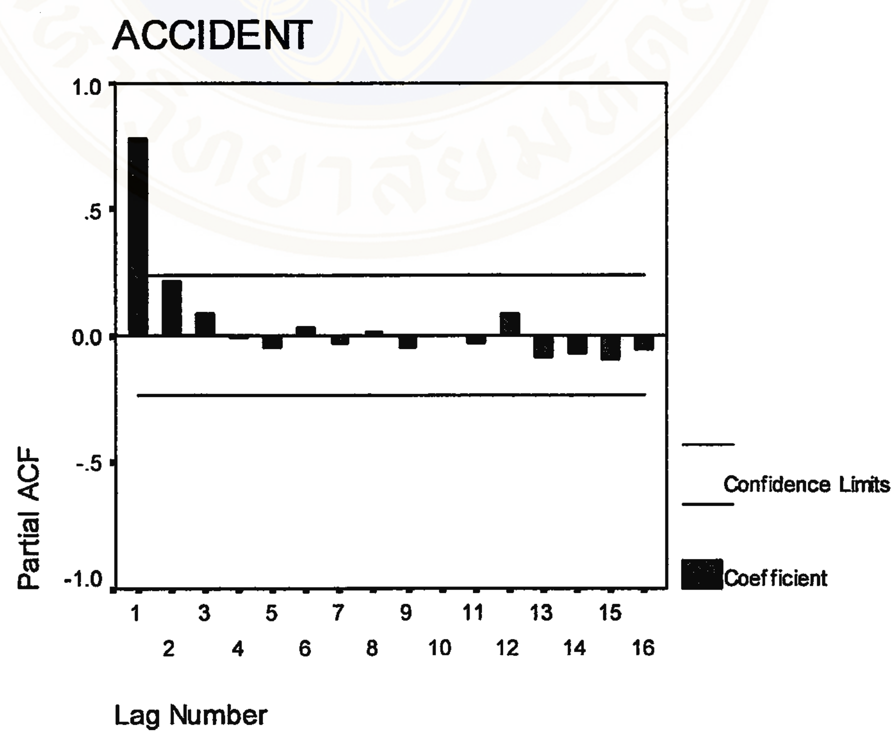
Graph



ACF

```
MODEL:  MOD_1.

Autocorrelations:   ACCIDENT

       Auto- Stand.
Lag    Corr.   Err. -1   -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                     +----+----+----+----+----+----+----+----+
  1    .776   .115              .    I****.***********            45.185   .000
  2    .688   .115              .    I****.*********              81.172   .000
  3    .618   .114              .    I****.*******               110.666   .000
  4    .541   .113              .    I****.******                133.573   .000
  5    .461   .112               .   I***.*****                  150.445   .000
  6    .412   .111               .   I***.****                   164.174   .000
  7    .350   .110               .   I***.***                    174.192   .000
  8    .308   .110               .   I***.**                     182.080   .000
  9    .251   .109               .   I***.*                      187.400   .000
 10    .214   .108               .   I****                       191.337   .000
 11    .167   .107               .   I***.                       193.782   .000
 12    .168   .106               .   I***.                       196.284   .000
 13    .114   .105               .   I**  .                      197.453   .000
 14    .064   .104               .   I*   .                      197.831   .000
 15    .005   .103               .   *    .                      197.834   .000
 16   -.044   .103               .  *I    .                      198.019   .000

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:   72     Computable first lags:   71


Partial Autocorrelations:   ACCIDENT

     Pr-Aut- Stand.
Lag    Corr.   Err. -1   -.75  -.5 -.25   0   .25  .5   .75   1
                     +----+----+----+----+----+----+----+----+
  1    .776   .118              .    I****.***********
  2    .215   .118              .    I****.
  3    .085   .118              .    I**  .
  4   -.007   .118              .    *    .
  5   -.044   .118              .   *I    .
  6    .028   .118              .    I*   .
  7   -.031   .118              .   *I    .
  8    .015   .118              .    *    .
  9   -.046   .118              .   *I    .
 10    .003   .118              .    *    .
 11   -.034   .118              .   *I    .
 12    .085   .118              .    I**  .
 13   -.085   .118              .  **I    .
 14   -.074   .118              .   *I    .
 15   -.092   .118              .  **I    .
 16   -.057   .118              .   *I    .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:   72     Computable first lags:   71
```

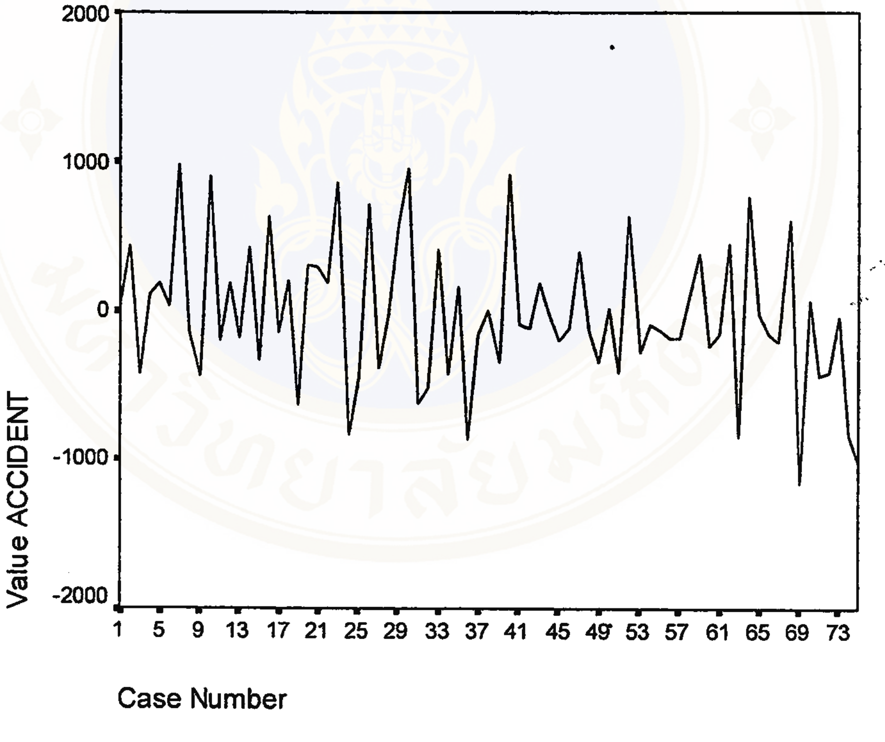

ACCIDENT
1.0
.5
0.0
-.5
-1.0
ACF
Confidence Limits
Coefficient
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number


ACCIDENT
1.0
.5
0.0
-.5
-1.0
Partial ACF
Confidence Limits
Coefficient
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number

Appendix C

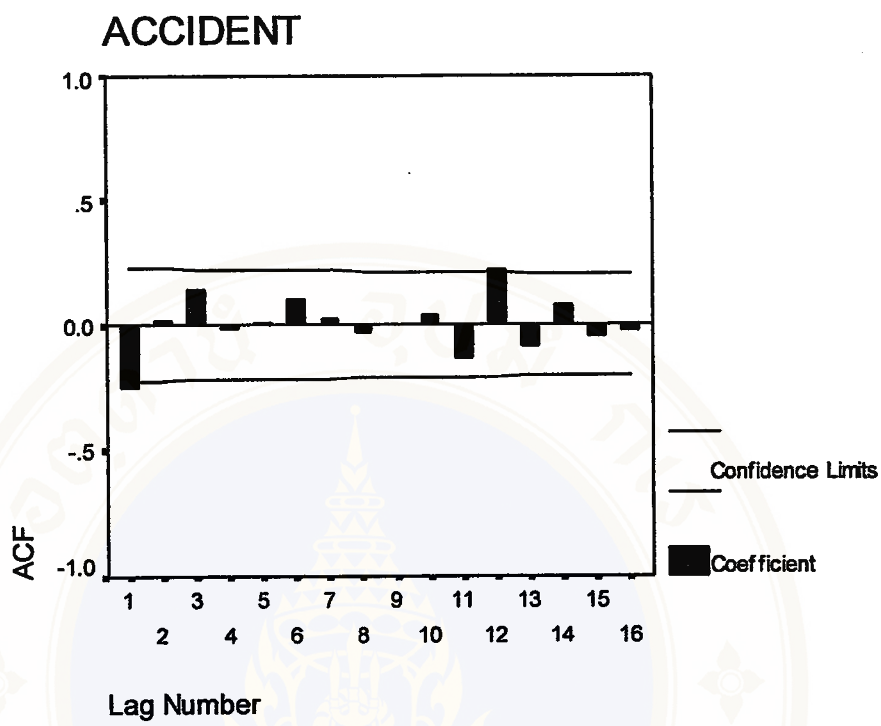
Graph and Autocorrelation After Transformation of Data

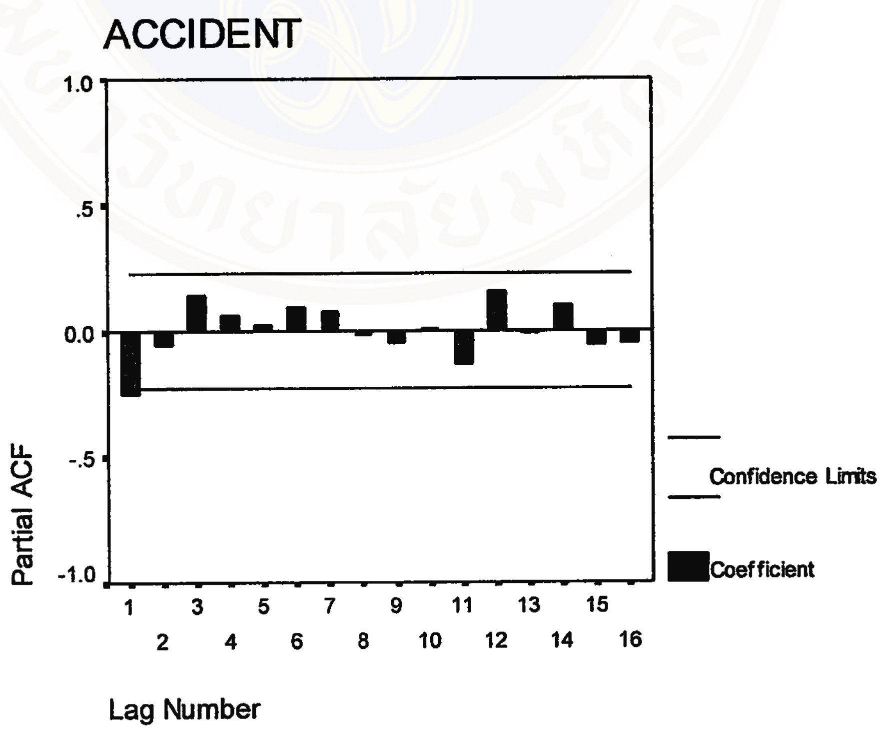
Graph



ACF

```
MODEL:  MOD_2.
Autocorrelations:   ACCIDENT

     Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1  Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+
  1  -.256   .113               *****I    .                    5.095   .024
  2   .012   .112                  .  *   .                    5.106   .078
  3   .145   .112                  .  I***.                    6.791   .079
  4  -.019   .111                  .  *   .                    6.819   .146
  5   .010   .110                  .  *   .                    6.827   .234
  6   .100   .109                  .  I** .                    7.663   .264
  7   .025   .109                  .  I*  .                    7.717   .358
  8  -.035   .108                  . *I   .                    7.824   .451
  9  -.001   .107                  .  *   .                    7.824   .552
 10   .040   .106                  .  I*  .                    7.963   .632
 11  -.135   .105                  .***I  .                    9.609   .566
 12   .217   .104                  .  I****                   13.915   .306
 13  -.090   .104                  . **I  .                   14.664   .329
 14   .075   .103                  .  I** .                   15.199   .365
 15  -.044   .102                  . *I   .                   15.389   .424
 16  -.025   .101                  .  *   .                   15.449   .492

Plot Symbols:      Autocorrelations *    Two Standard Error Limits .

Total cases:  75     Computable first lags:  74


Partial Autocorrelations:   ACCIDENT

   Pr-Aut- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+
  1  -.256   .115               *****I    .
  2  -.057   .115                 .  *I    .
  3   .143   .115                 .   I*** .
  4   .062   .115                 .   I*   .
  5   .023   .115                 .   *    .
  6   .093   .115                 .   I**  .
  7   .076   .115                 .   I**  .
  8  -.016   .115                 .   *    .
  9  -.050   .115                 .  *I    .
 10   .008   .115                 .   *    .
 11  -.133   .115                 . ***I   .
 12   .161   .115                 .   I*** .
 13  -.007   .115                 .   *    .
 14   .104   .115                 .   I**  .
 15  -.052   .115                 .  *I    .
 16  -.044   .115                 .  *I    .

Plot Symbols:      Autocorrelations *    Two Standard Error Limits .

Total cases:  75     Computable first lags:  74
```


ACCIDENT
1.0
.5
0.0
-.5
-1.0
ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient


ACCIDENT
1.0
.5
0.0
-.5
-1.0
Partial ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient

Appendix D

Output Analysis (ARIMA)

Arima (Size of Past Data in 5 months)

```
MODEL:  MOD_1

Model Description:

Variable:   ACCIDENT
Regressors: NONE

Non-seasonal differencing: 1
No seasonal component in model.

Parameters:
AR1       ________  < value originating from estimation >
AR2       ________  < value originating from estimation >
CONSTANT  ________  < value originating from estimation >

95.00 percent confidence intervals will be generated.

Split group number: 1  Series length: 5
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:

AR1       -.92602
AR2       -.46540
CONSTANT -27.5170

Marquardt constant = .001
Adjusted sum of squares = 498831.27

              Iteration History:

  Iteration    Adj. Sum of Squares    Marquardt Constant

        1              485923.81             10.000000
        2              402912.47              1.000000
        3              391813.49             10.000000
        4              298110.95              1.000000
        5              285977.61             10.000000
```

6	172054.99	1.000000
7	159923.60	10.000000
8	151897.56	1.000000
9	132587.30	10.000000

Conclusion of estimation phase.
Estimation terminated at iteration number 10 because:
Maximum number of iterations was exceeded.

FINAL PARAMETERS:

Number of residuals	4
Standard error	178.14056
Log likelihood	-26.467142
AIC	58.934283
SBC	57.093166

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	1	130813.88	31734.059

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-1.197742	.105947	-11.305079	.05616654
AR2	-.962134	.073901	-13.019293	.04880236
CONSTANT	-38.944426	39.626988	-.982775	.50553028

Covariance Matrix:

	AR1	AR2
AR1	.01122482	.00424591
AR2	.00424591	.00546130

Correlation Matrix:

	AR1	AR2
AR1	1.0000000	.5422917
AR2	.5422917	1.0000000

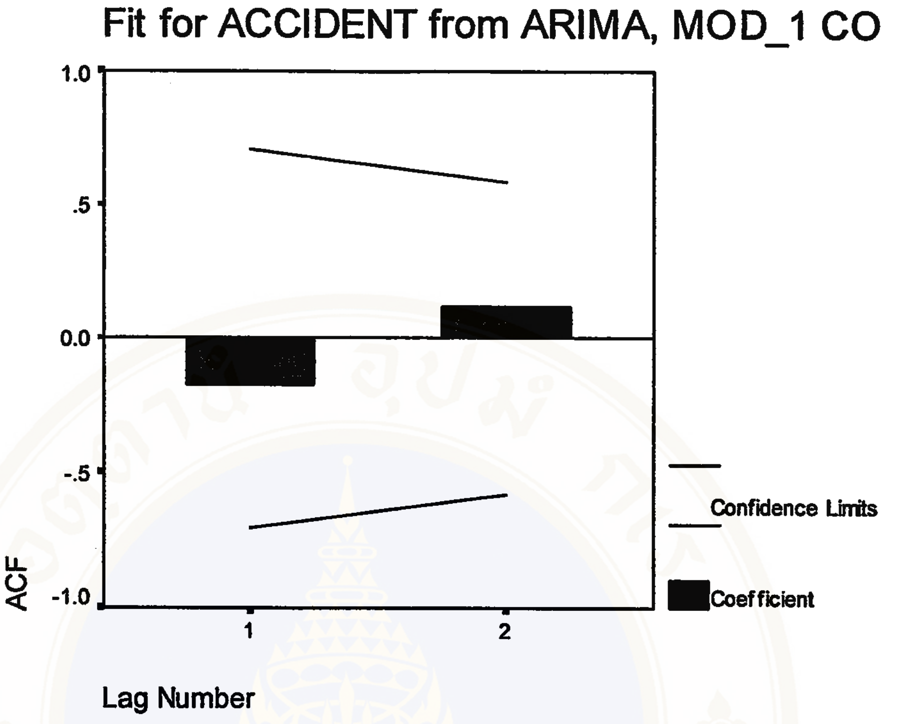
Regressor Covariance Matrix:

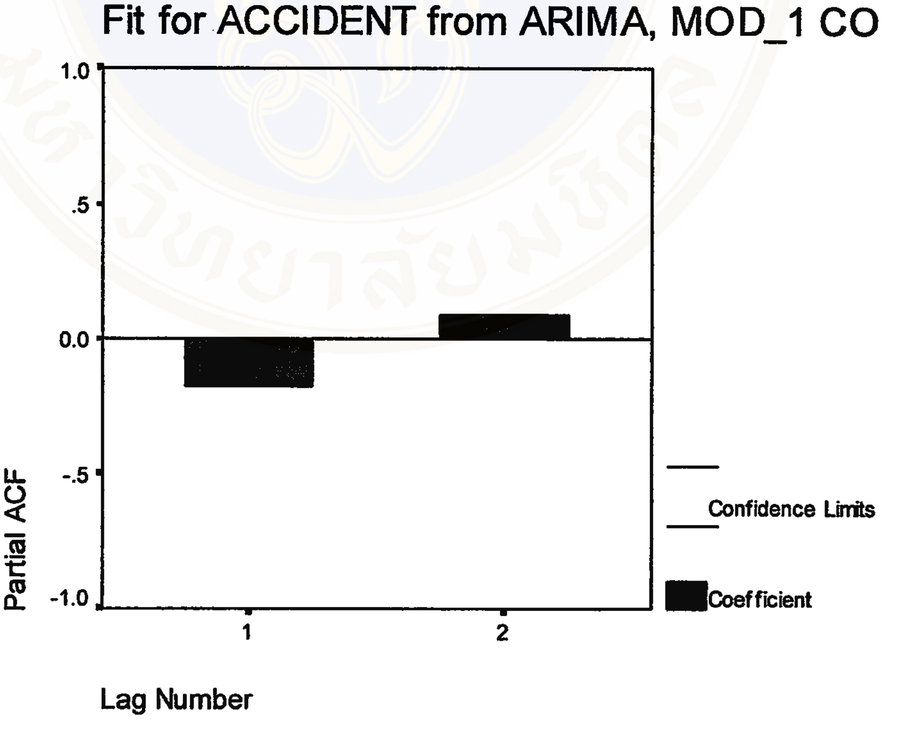
	CONSTANT
CONSTANT	1570.2982

```
Regressor Correlation Matrix:

               CONSTANT

CONSTANT     1.0000000


The following new variables are being created:

  Name         Label

  FIT_1        Fit for ACCIDENT from ARIMA, MOD_1 CON
  ERR_1        Error for ACCIDENT from ARIMA, MOD_1 CON
  LCL_1        95% LCL for ACCIDENT from ARIMA, MOD_1 CON
  UCL_1        95% UCL for ACCIDENT from ARIMA, MOD_1 CON
  SEP_1        SE of fit for ACCIDENT from ARIMA, MOD_1 CON
```

ACF

```
MODEL:  MOD_2.

Variable:  FIT_1       Missing cases:  1    Valid cases:  4


Autocorrelations:   FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

     Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+
  1  -.178   .354        .        ****I          .              .254   .614
  2   .119   .289          .          I**      .                .425   .809

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  5     Computable first lags:  3


Partial Autocorrelations:    FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

   Pr-Aut- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+
  1  -.178   .500  .              ****I                    .
  2   .090   .500  .                  I**                  .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  5     Computable first lags:  3
```


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
ACF
1
2
Lag Number
Confidence Limits
Coefficient


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
Partial ACF
1
2
Lag Number
Confidence Limits
Coefficient

Arima (Size of Past Data in 12 Months)

MODEL: MOD_3

Model Description:

Variable: ACCIDENT
Regressors: NONE

Non-seasonal differencing: 1
No seasonal component in model.

Parameters:
AR1 ________ < value originating from estimation >
AR2 ________ < value originating from estimation >
CONSTANT ________ < value originating from estimation >

95.00 percent confidence intervals will be generated.

Split group number: 1 Series length: 12
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:

AR1 -1.01087
AR2 -.70231
CONSTANT 17.80119

Marquardt constant = .001
Adjusted sum of squares = 2141023.8

Iteration History:

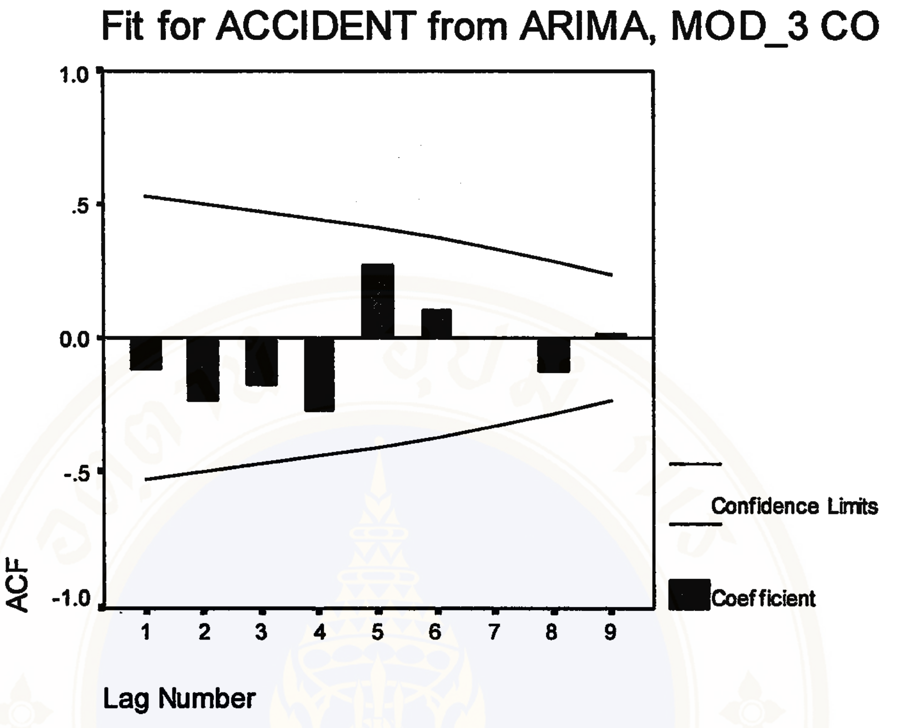
Iteration	Adj. Sum of Squares	Marquardt Constant
1	2099868.6	.00100000
2	2098051.9	.00010000
3	2097831.3	.00001000
4	2097803.8	.00000100

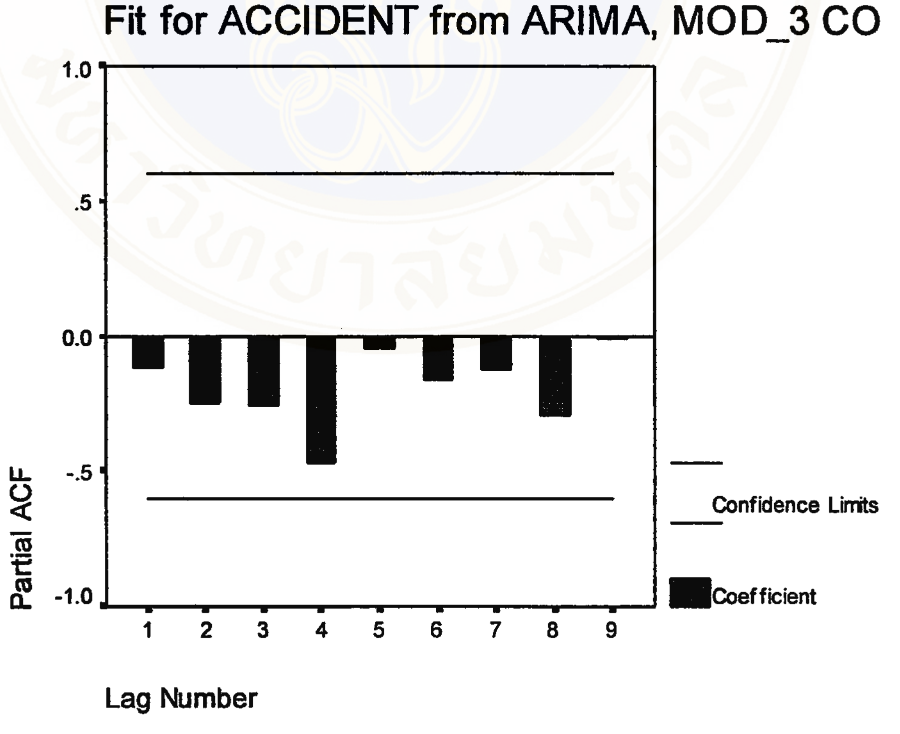
Conclusion of estimation phase.
Estimation terminated at iteration number 5 because:
Sum of squares decreased by less than .001 percent.

```
FINAL PARAMETERS:
Number of residuals  11
Standard error       464.50855
Log likelihood       -82.52034
AIC                  171.04068
SBC                  172.23437

            Analysis of Variance:

             DF  Adj. Sum of Squares    Residual Variance

Residuals     8            2097800.7            215768.19

           Variables in the Model:

                        B          SEB       T-RATIO   APPROX. PROB.

AR1              -.991483      .202439    -4.8976783       .00119728
AR2              -.765843      .214507    -3.5702523       .00729099
CONSTANT        21.432114    55.632002      .3852479       .71009463

Covariance Matrix:

                AR1          AR2

AR1       .04098167    .02439803
AR2       .02439803    .04601314

Correlation Matrix:

                AR1          AR2

AR1       1.0000000     .5618481
AR2        .5618481    1.0000000

Regressor Covariance Matrix:

             CONSTANT

CONSTANT    3094.9196

Regressor Correlation Matrix:

             CONSTANT

CONSTANT    1.0000000

The following new variables are being created:

  Name         Label

  FIT_1        Fit for ACCIDENT from ARIMA, MOD_3 CON
  ERR_1        Error for ACCIDENT from ARIMA, MOD_3 CON
  LCL_1        95% LCL for ACCIDENT from ARIMA, MOD_3 CON
  UCL_1        95% UCL for ACCIDENT from ARIMA, MOD_3 CON
  SEP_1        SE of fit for ACCIDENT from ARIMA, MOD_3 CON
```

ACF

```
MODEL:  MOD_4.

Variable:  FIT_1        Missing cases:  1     Valid cases:  11


Autocorrelations:    FIT_1    Fit for ACCIDENT from ARIMA, MOD_3 CON

     Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+
  1  -.116  .264             .      **I        .                   .193   .661
  2  -.234  .251              .  *****I       .                   1.060   .589
  3  -.176  .237               .  ****I      .                    1.615   .656
  4  -.274  .221               . *****I      .                    3.153   .533
  5   .274  .205                .     I*****  .                   4.948   .422
  6   .103  .187                 .    I**   .                     5.254   .512
  7  -.001  .167                 .     *    .                     5.254   .629
  8  -.123  .145                  .  **I   .                      5.970   .651
  9   .014  .118                   .   *   .                      5.984   .742

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  12     Computable first lags:  10


Partial Autocorrelations:    FIT_1    Fit for ACCIDENT from ARIMA, MOD_3 CON

   Pr-Aut- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+
  1  -.116  .302            .       **I       .
  2  -.250  .302            .    *****I       .
  3  -.259  .302            .    *****I       .
  4  -.474  .302            .*********I       .
  5  -.045  .302            .        *I       .
  6  -.159  .302            .      ***I       .
  7  -.126  .302            .      ***I       .
  8  -.295  .302            .   ******I       .
  9  -.004  .302            .         *       .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  12     Computable first lags:  10
```


Fit for ACCIDENT from ARIMA, MOD_3 CO
1.0
.5
0.0
-.5
-1.0
ACF
1
2
3
4
5
6
7
8
9
Lag Number
Confidence Limits
Coefficient


Fit for ACCIDENT from ARIMA, MOD_3 CO
1.0
.5
0.0
-.5
-1.0
Partial ACF
1
2
3
4
5
6
7
8
9
Lag Number
Confidence Limits
Coefficient

Arima (Size of Past Data in 36 Months)

MODEL: MOD_1

Model Description:

Variable: ACCIDENT
Regressors: NONE

Non-seasonal differencing: 1
No seasonal component in model.

Parameters:
AR1 ________ < value originating from estimation >
CONSTANT ________ < value originating from estimation >

95.00 percent confidence intervals will be generated.

Split group number: 1 Series length: 36
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:

AR1 -.55570
CONSTANT -19.8490

Marquardt constant = .001
Adjusted sum of squares = 15638494.4

Iteration History:

Iteration	Adj. Sum of Squares	Marquardt Constant
1	15634625.7	.00100000

Conclusion of estimation phase.
Estimation terminated at iteration number 2 because:
Sum of squares decreased by less than .001 percent.

FINAL PARAMETERS:

Number of residuals	35
Standard error	684.48466
Log likelihood	-277.3513
AIC	558.70261
SBC	561.81331

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	33	15634617.8	468519.25

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-.568598	.145222	-3.9153676	.00042737
CONSTANT	-19.757319	74.535504	-.2650726	.79260234

Covariance Matrix:

	AR1
AR1	.02108946

Correlation Matrix:

	AR1
AR1	1.0000000

Regressor Covariance Matrix:

	CONSTANT
CONSTANT	5555.5413

Regressor Correlation Matrix:

	CONSTANT
CONSTANT	1.0000000

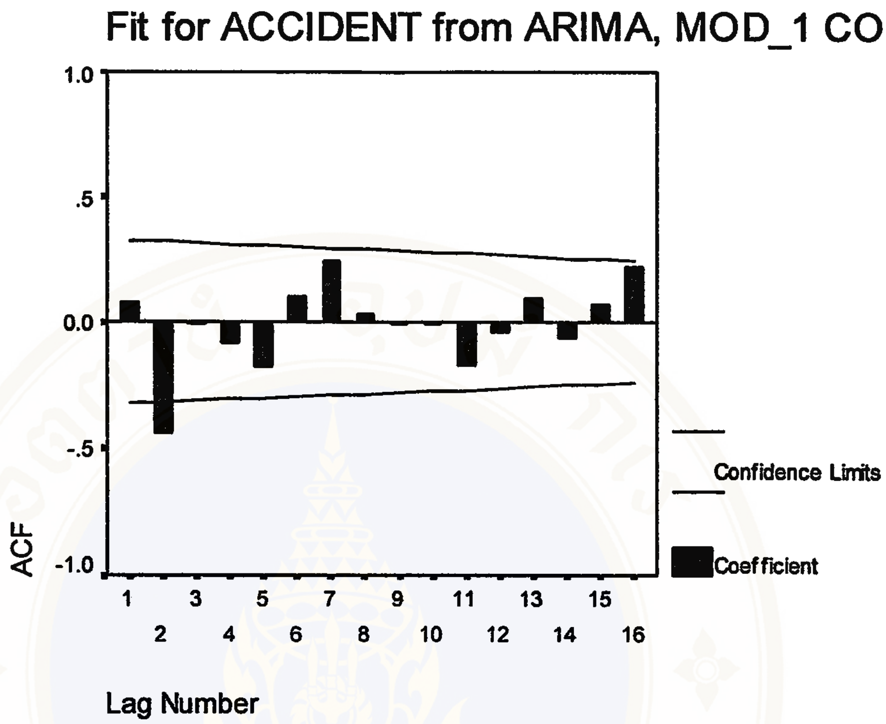
The following new variables are being created:

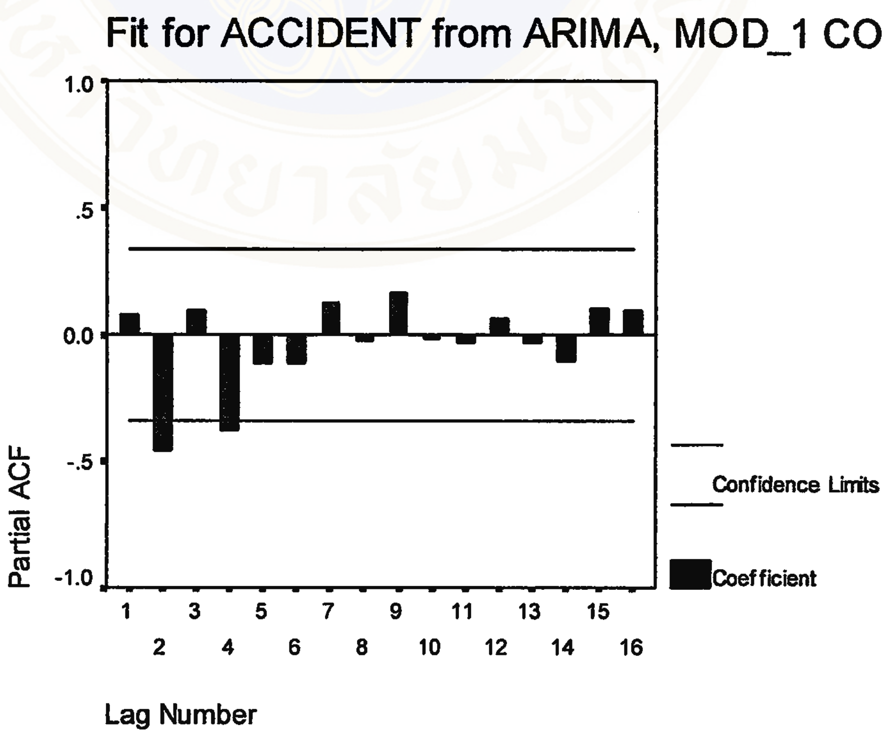
Name	Label
FIT_1	Fit for ACCIDENT from ARIMA, MOD_1 CON
ERR_1	Error for ACCIDENT from ARIMA, MOD_1 CON
LCL_1	95% LCL for ACCIDENT from ARIMA, MOD_1 CON
UCL_1	95% UCL for ACCIDENT from ARIMA, MOD_1 CON
SEP_1	SE of fit for ACCIDENT from ARIMA, MOD_1 CON

ACF

```
MODEL:  MOD_2.

Variable:   FIT_1        Missing cases:  1     Valid cases:  35


Autocorrelations:   FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

     Auto- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                   +----+----+----+----+----+----+----+----+
  1   .078   .162                .     I**   .                      .234   .628
  2  -.444   .160             ***.*****I     .                     7.977   .019
  3  -.010   .157                .     *     .                     7.981   .046
  4  -.086   .155                .   **I     .                     8.291   .081
  5  -.181   .152                . ****I     .                     9.706   .084
  6   .101   .150                .     I**   .                    10.163   .118
  7   .245   .147                .     I*****.                    12.943   .074
  8   .035   .144                .     I*    .                    13.002   .112
  9  -.011   .142                .     *     .                    13.008   .162
 10  -.006   .139                .     *     .                    13.010   .223
 11  -.174   .136                 . ***I    .                     14.639   .200
 12  -.039   .133                 .   *I    .                     14.722   .257
 13   .096   .130                 .    I**  .                     15.267   .291
 14  -.059   .127                 .   *I    .                     15.483   .346
 15   .072   .124                 .    I*   .                     15.823   .394
 16   .220   .121                 .    I****.                     19.132   .262

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  36     Computable first lags:  34
```

```
Partial Autocorrelations:   FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

   Pr-Aut- Stand.
Lag  Corr.   Err. -1  -.75  -.5 -.25   0   .25  .5   .75   1
                   +----+----+----+----+----+----+----+----+
  1   .078   .169               .      I**    .
  2  -.453   .169             **.******I      .
  3   .096   .169               .      I**    .
  4  -.381   .169              *.******I      .
  5  -.107   .169               .    **I      .
  6  -.114   .169               .    **I      .
  7   .128   .169               .      I***   .
  8  -.025   .169               .     *I      .
  9   .164   .169               .      I***   .
 10  -.020   .169               .      *      .
 11  -.030   .169               .     *I      .
 12   .063   .169               .      I*     .
 13  -.029   .169               .     *I      .
 14  -.100   .169               .    **I      .
 15   .100   .169               .      I**    .
 16   .094   .169               .      I**    .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  36     Computable first lags:  34
```


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
Partial ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient

Arima (Size of Past Data in 60 Months)

```
MODEL:  MOD_1

Model Description:

Variable:   ACCIDENT
Regressors: NONE

Non-seasonal differencing: 1
No seasonal component in model.

Parameters:
AR1        ________   < value originating from estimation >
CONSTANT   ________   < value originating from estimation >

95.00 percent confidence intervals will be generated.

Split group number: 1  Series length: 60
No missing data.
Melard's algorithm will be used for estimation.

Termination criteria:
Parameter epsilon: .001
Maximum Marquardt constant: 1.00E+09
SSQ Percentage: .001
Maximum number of iterations: 10

Initial values:

AR1       -.57505
CONSTANT -3.32055

Marquardt constant = .001
Adjusted sum of squares = 19866591.1


Conclusion of estimation phase.
Estimation terminated at iteration number 1 because:
   Sum of squares decreased by less than .001 percent.

FINAL PARAMETERS:

Number of residuals   59
Standard error        588.36005
Log likelihood        -459.17542
AIC                   922.35084
SBC                   926.50591
```

Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	57	19866583.3	346167.55

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
AR1	-.5755701	.108071	-5.3258451	.00000178
CONSTANT	-3.3197563	48.919865	-.0678611	.94613369

Covariance Matrix:

	AR1
AR1	.01167937

Correlation Matrix:

	AR1
AR1	1.0000000

Regressor Covariance Matrix:

	CONSTANT
CONSTANT	2393.1532

Regressor Correlation Matrix:

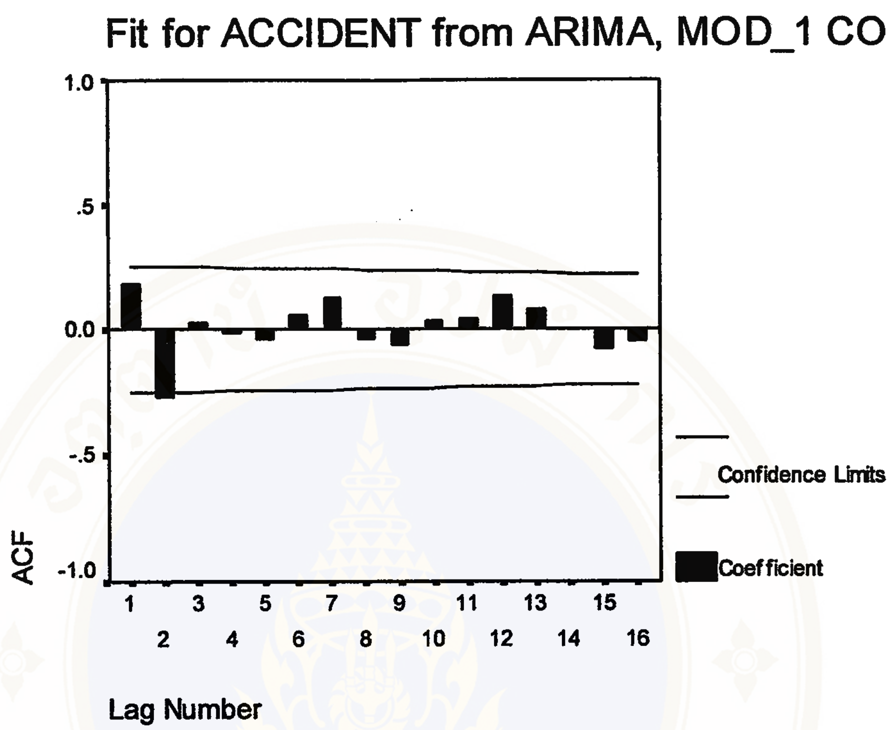
	CONSTANT
CONSTANT	1.0000000

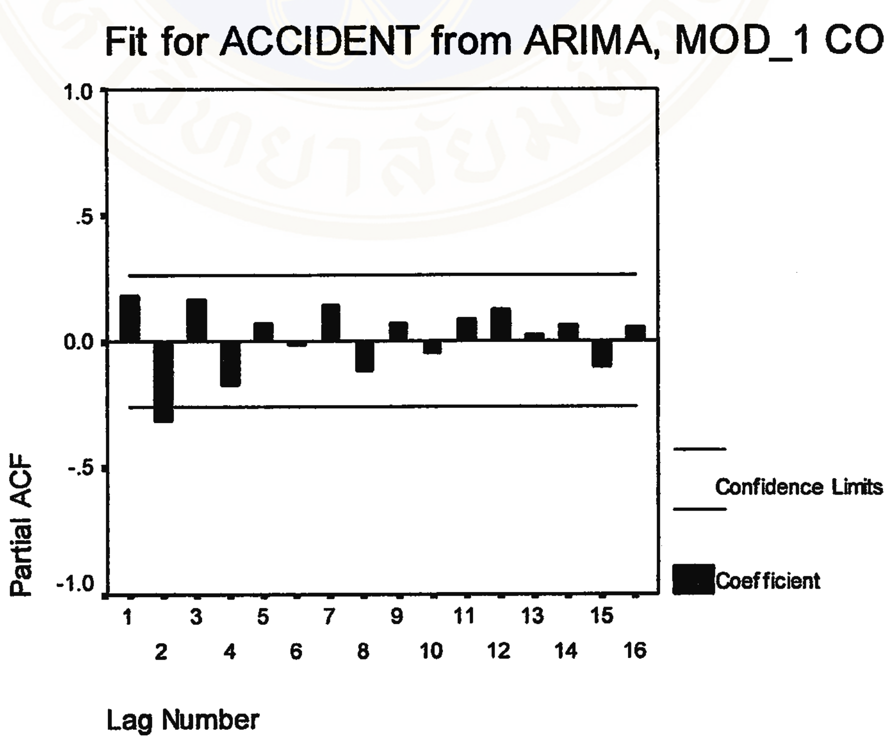
The following new variables are being created:

Name	Label
FIT_1	Fit for ACCIDENT from ARIMA, MOD_1 CON
ERR_1	Error for ACCIDENT from ARIMA, MOD_1 CON
LCL_1	95% LCL for ACCIDENT from ARIMA, MOD_1 CON
UCL_1	95% UCL for ACCIDENT from ARIMA, MOD_1 CON
SEP_1	SE of fit for ACCIDENT from ARIMA, MOD_1 CON

ACF

```
MODEL:  MOD_2.

Variable:  FIT_1        Missing cases:  1     Valid cases:  59

Autocorrelations:   FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

     Auto- Stand.
Lag  Corr.   Err. -1   -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.
                  +----+----+----+----+----+----+----+----+
  1   .184   .127               .    I****.                        2.106  .147
  2  -.267   .126              *****I    .                         6.601  .037
  3   .020   .125               .    *    .                        6.626  .085
  4  -.018   .124               .    *    .                        6.647  .156
  5  -.041   .122               .   *I    .                        6.759  .239
  6   .053   .121               .    I*   .                        6.950  .326
  7   .122   .120               .    I**  .                        7.983  .334
  8  -.039   .119               .   *I    .                        8.092  .425
  9  -.060   .118               .   *I    .                        8.354  .499
 10   .031   .117               .    I*   .                        8.424  .587
 11   .037   .115               .    I*   .                        8.529  .665
 12   .138   .114               .    I*** .                        9.980  .618
 13   .080   .113               .    I**  .                       10.483  .654
 14   .002   .112                .   *   .                        10.483  .726
 15  -.081   .111                . **I   .                        11.017  .751
 16  -.048   .109                .  *I   .                        11.214  .796

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  60     Computable first lags:  58


Partial Autocorrelations:   FIT_1   Fit for ACCIDENT from ARIMA, MOD_1 CON

   Pr-Aut- Stand.
Lag  Corr.   Err. -1   -.75  -.5 -.25   0   .25  .5   .75   1
                  +----+----+----+----+----+----+----+----+
  1   .184   .130               .    I****.
  2  -.311   .130              *.****I    .
  3   .162   .130               .    I*** .
  4  -.173   .130               . ***I    .
  5   .075   .130               .    I*   .
  6  -.015   .130               .    *    .
  7   .146   .130               .    I*** .
  8  -.120   .130               .  **I    .
  9   .075   .130               .    I*   .
 10  -.050   .130               .   *I    .
 11   .088   .130               .    I**  .
 12   .122   .130               .    I**  .
 13   .025   .130               .    *    .
 14   .063   .130               .    I*   .
 15  -.105   .130               .  **I    .
 16   .053   .130               .    I*   .

Plot Symbols:      Autocorrelations *     Two Standard Error Limits .

Total cases:  60     Computable first lags:  58
```


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient


Fit for ACCIDENT from ARIMA, MOD_1 CO
1.0
.5
0.0
-.5
-1.0
Partial ACF
1 3 5 7 9 11 13 15
2 4 6 8 10 12 14 16
Lag Number
Confidence Limits
Coefficient

BIOGRAPHY

NAME Miss Juntanee Vongspanich

DATE OF BIRTH 4 February 1972

PLACE OF BIRTH Bangkok, Thailand

INSTITITIONS ATTENDED

Prince of Songkla University,

Pattani. 1990-1994:

Bachelor of Science (Education)

Mahidol University,1995-1999:

Master of Science (Biostatistics).