Ekaterina Gerasimova: **Analysis of Forecasting Models and Time-Series Modelling of Network Traffic**

Approval of the Graduate School of Applied and Social Sciences
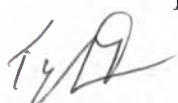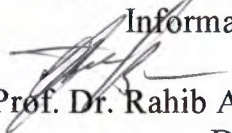
Prof. Dr. Fakhraddin Mamedov
Director

We certify this thesis is satisfactory for the award of the Degree of Master of Science in Computer Engineering

Examining Committee in charge:

Assoc. Prof. Dr. Doğan İbrahim, Chairman , Chairman of
Computer Engineering Department, NEU

Assoc. Prof. Dr. Faeq Radwan, Member , Computer Engineering
Department, NEU

Assoc. Prof. Dr. Ilham Huseynov, Member, Computer
Information System Department, NEU

Assoc. Prof. Dr. Rahib Abiyev, Supervisor, Computer Engineering
Department, NEU

DEPARTMENT OF COMPUTER ENGINEERING
DEPARTMENTAL DECISION

Date: 04.06.2004

Subject: Completion of M.Sc. Thesis

Participants: Assoc. Prof. Dr. Doğan İbrahim, Assoc.Prof. Dr. Faeq Radwan, Assoc.Prof. Dr. Ilham Huseynov, Assoc. Prof. Rahib Abiyev, Ibaid Alsoud, Shadi Jundy.

DECISION

We certify that the student whose number and name are given below, has fulfilled all the requirements for a M .S. degree in Computer Engineering.

CGPA

20011456   Ekaterina Gerasimova   3.214

Assoc. Prof. Dr. Doğan İbrahim, Committee Chairman , Chairman of Computer
Engineering Department, NEU

Assoc. Prof. Dr. Faeq Radwan, Committee Member , Computer Engineering
Department, NEU

Assoc. Prof. Dr. Ilham Huseynov, Committee Member, Computer Information System
Department, NEU

Assoc. Prof. Dr. Rahib Abiyev, Supervisor, Computer Engineering
Department, NEU

Chairman of Department
Assoc. Prof. Dr. Doğan İbrahim

NEU

JURY REPORT

DEPARTMENT OF
COMPUTER ENGINEERING

Academic Year: 2003-2004

STUDENT INFORMATION

Full Name	Ekaterina Gerasimova		
Undergraduate degree	BSc.	Date Received	Spring 2001
University	Ivanovo State University	CGPA	3.34

THESIS

Title	Analysis of Forecasting Models and Time-Series Modelling of Network Traffic

Description The aim of this thesis was analyzing of forecasting models and applying them for solving of economical and technical problems. The forecasting model for network traffic has been developed by using time-series analysis.

Supervisor	Assoc.Prof.Dr.Rahib Abiyev	Department	Computer Engineering

DECISION OF EXAMINING COMMITTEE

The jury has decided to accept / ~~reject~~ the student's thesis.
The decision was taken unanimously / ~~by majority~~.

COMMITTEE MEMBERS

Number Attending	3	Date	4.06.2004

Name	Signature
Assoc. Prof. Dr. Doğan İbrahim, Chairman of the jury	
Assoc. Prof. Dr. Faeq Radwan, Member	
Assoc. Prof. Dr. Ilham Huseynov, Member	

APPROVALS

Date	Chairman of Department
04.06.2004	Assoc. Prof. Dr. Doğan İbrahim

EAST UNIVERSITY

GRADUATE SCHOOL OF APPLIED AND SOCIAL SCIENCES

ANALYSIS OF FORECASTING MODELS AND TIME-SERIES MODELLING OF NETWORK TRAFFIC

EKATERINA GERASIMOVA

MASTER THESIS

DEPARTMENT OF COMPUTER ENGINEERING

Nicosia - 2004

ACKNOWLEDGEMENTS

ABSTRACT

To make accurate decision and improve the efficiency of the complicated system one of the effective way is predicting future behavior of these systems and making adequate control strategy. It gives us chance to make effective planning and managing of the process. For this reason forecasting plays a major role in most of our activities for the future.

The present work gives consideration of the Forecasting models and Time Series Analysis. Analysis of forecasting models and the use of those models in different industrial and non industrial areas are considered. As an example of the application of forecasting models to World Petroleum Production, Computer Engineering Department students number, and Network Traffic is considered. The simulation of these models has been done. Simulation results demonstrate that one of effective methodology for forecasting of future is time series analysis. The analysis and development of the time series models are considered. Using time series analysis the ARIMA models are developed for forecasting world petroleum production for year monthly, number of students for 2004-2005 academic year, and Network Traffic for a week. The simulations of ARIMA model for these problems have been done. By the comparison the results of forecasting by using ARIMA model with results of other models, demonstrate that first one gives more accurate results. The model can make reasonable predictions for one or more years to the future, suggesting that ARIMA modeling has great promise as a tool for short or long-range forecasting and planning.

TABLE OF CONTENTS

3

INTRODUCTION

Forecasting plays major roles in most of our activities and in all we do concerning the future. It is a branch of the anticipatory sciences used for identifying and projecting alternative possible futures. It is a conduit leading to plans for the development of "better" futures. Forecasted visions of possible futures open our freedom of choice over which future to encourage or discourage. In our fast-paced, rapidly changing world, the futures that we will experience will tend to be vastly different from our present reality in a growing number of ways. Furthermore, because of constant development of new knowledge and advances in the scientific (and ensuing technological advances), sociological, political, economic, and business areas, our global society has an ever increasing ability to shape (for better or worse) the futures we will eventually achieve.

As a result, society and each institution in it finds that more knowledge about possible futures and the consequences of today's decisions and actions is required. Thus, it is increasingly imperative that we have better forecasting tools and that we apply them in responsible ways. It is more and more important to forecast, ahead of time, with longer lead times, the possible futures implied by the changes produced by this new knowledge generation. To these ends, forecasting has become an essential tool for all participants in society to use in their attempts to decide, plan, design, steer, manage, implement, and control change by identifying preferable futures with forecasts.

In general terms, a forecast is simply a statement, based upon some criteria, concerning the future condition of something. A major purpose of forecasting is to give us choice over which future, either the trend path we are on, or an alternative, to plan, design, create, and to back with our resources.

Forecasting aids in identifying which futures to bring into fruition (the preferable) and which to forestall, or attempt to eliminate (the undesirables). Furthermore, forecasts are useful in assisting our intuitions about our plans, outlooks, investments, and so on, whenever we would like to have a better idea of the possible or most probable outcome.

5

Information from forecasts relative to:

• identifying:

> possible futures

> probable futures

> preferable futures

• providing a basis for understanding the process and dynamics of change;

• providing notions of where change may take us into the future;

• providing a systematic methodology, based upon a set of supportive assumptions, for the discovery of possible futures.

Since we know that individuals and society in general have the means and knowledge to shape major elements of our future, to grow new opportunities as well to set in motion means for avoiding or lessening the impact of negative future threats, forecasting again becomes an increasingly important tool for all of us to understand and use.

Forecasting has many applications, some of which are to:

• identify the trend path we are traveling into the future;

• identify alternative possible futures (alternatives to the trends);

• provide views of possible futures;

• raise awareness of possible futures so that we have choice over which future we support;

• generate "future histories" that we can study to determine our role in shaping them before they become a reality;

• provide information about possible futures so that realistic planning can occur;

• provide information on possible futures to aid decision-making and planning;

• justify the decisions and plans we make;

• discover possible breakthroughs;

• discover possible life, societal, scientific, political, social, technological, and institutional future turning points or paradigm shifts;

• track evolving change and advances;

- provide managers with information for choosing their organization's vision, mission, purpose, goals, objectives, strategies, plans, and tactics;
- provide information on possible futures for assessment relative to their possible future impacts and consequences.

One of important problem in complex systems such as technical, economical systems is to increase the efficiency of used control systems for these systems. Forecasting plays an important role in increasing the efficiency of the technological and economical systems. It is acquired by the predicting the future conditions of these system and making appropriate control decision.

Forecasting is a highly "noisy" application, because we are typically dealing with systems that receive thousands of inputs, which interact in a complex nonlinear fashion. Usually with a very small subset of inputs or measurements available from the system, the system behavior is to be estimated and extrapolated into the future.

The aim of this thesis is the analysis of forecasting models and application of time series analysis for solving forecasting of economical and technical problems. The use of forecasting methods to three different problems is represented. These are forecasting world petroleum production, number of registered students in Engineering Department and network traffic forecasting.

Thesis consists of introduction, five chapters, conclusion and appendix.

Introduction represents the actuality of the problem studied and the brief description of each chapter.
Chapter one gives the review of the usage of forecasting models for different problems.

Chapter two introduces the description of various forecasting models and calculation of forecast accuracy. The formulation and explanation of moving average, regression models, exponential smoothing and time series forecasting models are described.

In chapter three Box-Jenkins analysis and forecasting procedure have been represented. Auto Regressive Integrated Moving Average model, introduced by Box and Jenkins is described. There is explanation of structure and three primer stages in building a Box-Jenkins time series model. The different structure time-series models and their application have been analyzed.

Chapter four is the application of forecasting methods to two problems, world petroleum production and number of registered students of Engineering Department in University. For this purpose program, which was developed in Delphi has been used. The results and comparison of different forecasting methods were given.

Chapter five is devoted to the development network traffic forecasting by using time series analysis. The Autoregressive Integrated Moving Average model is used for forecasting of Network traffic. For analysis of the data and modeling of this process S plus package was applied. To make forecasting the program written in Delphi programming has been developed. The developed program allows to forecast the workload in network traffic and to plan capacity requirements of the network. One ARIMA model was chosen, and was used for forecasting.

In conclusion the obtained important results from the thesis are given

Appendix includes the tables with statistical data, listing of Delphi program and guide to program.

8

CHAPTER 1
THE USAGE OF MATHEMATICAL MODELS FOR SOLVING INDUSTRIAL PROBLEMS

1.1. Overview

Administrators in all organizations make plans to cope with future changes. The planning means to make decisions in advance about the future course of action. Obviously, then, planning and decision making are based on forecasts or expectations of what the future holds. Several applications of forecasting will be considered in this chapter, like a weather forecasting and business forecasting.

1.2. Forecasting Process

Forecasting process includes the following

a. Collect appropriate data

b. Examine data patterns

c. Choose a forecasting method (model)

d. Apply the model to past periods (ex post)

e. Examine the accuracy of model by examining ex post errors

f. If adequate (errors random and sufficiently small) use the model to forecast the future

g. Periodically check the accuracy of forecasts with actual experience

1.3. Forecasting in Engineering

On the base of statistical data the models for forecasting industrial and non industrial process has been created. In the world there are many works for solving forecasting problems in different fields, like a engineering, industrial, business, management.

In [5] there is introduced a methodology to predict *when and where* link additions/upgrades have to take place in an IP backbone network. Using SNMP statistics, collected continuously since 1999, aggregate demand is computed between any two adjacent PoPs and look at its evolution at time scales larger than one hour. It is shown that IP backbone traffic exhibits visible long term trends, strong periodicities, and variability at multiple time scales. This methodology relies on the wavelet multiresolution analysis and linear time series models. Using wavelet multiresolution analysis, we smooth the collected measurements until we identify the *overall long-term trend*. The fluctuations around the obtained trend are further analyzed at multiple time scales. There shown that the largest amount of variability in the original signal is due to its fluctuations at the 12 hour time scale. We model inter-PoP aggregate demand as a multiple linear regression model, consisting of the two identified components. It is shown that this model accounts for 98% of the total energy in the original signal, while explaining 90% of its variance. Weekly approximations of those components can be accurately modeled with low-order AutoRegressive Integrated Moving Average (ARIMA) models. Forecasting the long term trend and the fluctuations of the traffic at the 12 hour time scale yields accurate estimates for at least six months in the future.

In [6] non-linear threshold autoregressive models are examined for use in modeling the temporal variation in the byte-rate in Ethernet traffic. The model is comprised of a number of autoregressive processes each of which is to be used in a specified range of amplitude of the byte-rate. The local dynamics within each threshold range are captured by an autoregressive process. The switching between each submodel is conditioned on the amplitude of a lagged value of the time-series. To develop the model the Bellcore Ethernet LAN data is used. It is shown that non-linear threshold autoregressive processes can be used to capture the dynamics of Ethernet LAN traffic. This model also provides for both short and long term prediction capability and allows us to quantitatively identify the sources of long-range-dependence features in the traffic. When the aggregate traffic is partitioned into classes based on packet sizes, certain classes of traffic follow deterministic cyclical patterns. These periodic components arise from the process switching between different amplitude regimes. Superposed on this fundamental period are longer cycles that

10

can be localized either below or above the mean byte-rate. By constructing amplitude thresholds associated with a finite set of delay parameters, the dynamics within each threshold are captured by locally linear autoregressive processes. The aggregate process is globally nonlinear. This model is shown to provide good agreement with the marginal distributions and the correlation functions derived from the Ethernet traffic data. In addition, simulation experiments demonstrate that the loss statistics observed in finite buffer queues agree favorably with those generated by the measurements.

1.4. Business forecasting

Business organizations, public organizations, and individuals thus have the common goal of allocating available time among competing resources in some optimal manner. This goal is accomplished by making forecast of future activities and taking the proper actions as suggested by these forecast.

In business and public administration the organization with both short-term and long-term forecasts. The short-term forecast usually looks no more than one year into the future and involves forecasting sales, price changes, and customer demand, which, in turn, reflect the need for seasonal employment, short-term forecast usually looks from 2 to 10 years into the future and is used as planning model for product line and capital investment decisions, as indicated by changing demand patterns.

Naturally, the further a forecast is projected into the future, the more speculative it becomes. But since the future is always uncertain, we cannot expect complete accuracy any forecast. The time series underlying the process to be forecast is bound to be influenced by many causal factors- some forcing the time series up while conflicting factors act to force the series down. Nevertheless, business people must make forecasts of future business activity in order to budget their time and resources efficiently. They cannot hope to account for every possible factor that may cause the response of interest to rise or fall over time. All that can be expected is that the benefits gained by forecasting offset the opportunity cost for not forecasting. Note that such benefits are not limited to real monetary savings but may

11

imply a sharpening of the businessperson's thinking to consider the interplay of the events that affect the movement of the time series.

1.5. Weather forecasting

How often do you watch the weather on TV or listen on the radio for the weather forecast? The weather affects everything from afternoon swim practice to attacks on enemy forces during wars.

Weather forecasting used to be thought of as witchcraft. Today, we rely on weather forecasters to help us plan our days and prepare for life-threatening conditions.

1.5.1. Forecasting methods

a. Persistence Method (today equals tomorrow):
There are several different methods that can be used to create a forecast. The method a forecaster chooses depends upon the experience of the forecaster, the amount of information available to the forecaster, the level of difficulty that the forecast situation presents, and the degree of accuracy or confidence needed in the forecast.

The first of these methods is the Persistence Method; the simplest way of producing a forecast. The persistence method assumes that the conditions at the time of the forecast will not change. For example, if it is sunny and 87 degrees today, the persistence method predicts that it will be sunny and 87 degrees tomorrow. If two inches of rain fell today, the persistence method would predict two inches of rain for tomorrow.

The persistence method works well when weather patterns change very little and features on the weather maps move very slowly. It also works well, where summertime weather conditions vary little from day to day. However, if weather conditions change significantly from day to day, the persistence method usually breaks down and it is not the best forecasting method to use.

It may also appear that the persistence method would work only for shorter-term forecasts (e.g. a forecast for a day or two), but actually one of the most useful roles of the persistence forecast is predicting long range weather conditions or making climate forecasts. For example, it is often the case that one hot and dry month will be followed by another hot and dry month. So, making persistence forecasts for monthly and seasonal weather conditions can have some skill. Some of the other forecasting methods, such as numerical weather prediction, lose all their skill for forecasts longer than 10 days. This makes persistence a "hard to beat" method for forecasting longer time periods.

b. Trends Method (using mathematics)

The trends method involves determining the speed and direction of movement for fronts, high and low pressure centers, and areas of clouds and precipitation. Using this information, the forecaster can predict where he or she expects those features to be at some future time. For example, if a storm system is 1000 miles west of your location and moving to the east at 250 miles per day, using the trends method you would predict it to arrive in your area in 4 days.

Mathematics

(1000 miles / 250 miles per day = 4 days)

Using the trends method to forecast only a few hours into the future is known as "Nowcasting" and this method is frequently used to forecast precipitation. For example, if a line of thunderstorms is located 60 miles to your northwest and moving southeast at 30 miles per hour, you would predict the storms to arrive in your area in 2 hours.

The trends method works well when systems continue to move at the same speed in the same direction for a long period of time. If they slow down, speed up, change intensity, or change direction, the trends forecast will probably not work as well.

c. Other Forecasting Methods (climatology, analogue and numerical weather prediction)

Climatology: The Climatology Method is another simple way of producing a forecast. This method involves averaging weather statistics accumulated over many years to make the forecast. If you were making a forecast for temperature and precipitation, then you would use this recorded weather data to compute the averages for temperature and precipitation.

If these averages were 87 degrees with 0.18 inches of rain, then the weather forecast, using the climatology method, would call for a high temperature of 87 degrees with 0.18 inches of rain. The climatology method only works well when the weather pattern is similar to that expected for the chosen time of year. If the pattern is quite unusual for the given time of year, the climatology method will often fail.

Analog Method

The Analog Method is a slightly more complicated method of producing a forecast. It involves examining today's forecast scenario and remembering a day in the past when the weather scenario looked very similar (an analog). The forecaster would predict that the weather in this forecast will behave the same as it did in the past.

For example, suppose today is very warm, but a cold front is approaching your area. You remember similar weather conditions one last week, also a warm day with cold front approaching. You also remember how heavy thunderstorms developed in the afternoon as the cold front pushed through the area. Therefore, using the analog method, you would predict that this cold front will also produce thunderstorms in the afternoon.

The analog method is difficult to use because it is virtually impossible to find a perfect analog. Various weather features rarely align themselves in the same locations they were in the previous time. Even small differences between the current time and the analog can lead to very different results. However, as time passes and more weather data is archived, the

chances of finding a "good match" analog for the current weather situation should improve, and so should analog forecasts.

Numerical Weather Prediction

Numerical Weather Prediction (NWP) uses the power of computers to make a forecast. Complex computer programs, also known as forecast models, run on supercomputers and provide predictions on many atmospheric variables such as temperature, pressure, wind, and rainfall. A forecaster examines how the features predicted by the computer will interact to produce the day's weather.

The NWP method is flawed in that the equations used by the models to simulate the atmosphere are not precise. This leads to some error in the predictions. In addition, there are many gaps in the initial data since we do not receive many weather observations from areas in the mountains or over the ocean. If the initial state is not completely known, the computer's prediction of how that initial state will evolve will not be entirely accurate.

Despite these flaws, the NWP method is probably the best of the five discussed here at forecasting the day-to-day weather changes. Very few people, however, have access to the computer data. In addition, the beginning forecaster does not have the knowledge to interpret the computer forecast, so the simpler forecasting methods, such as the trends or analogue method, are recommended for the beginner.

1.5.2. Example of forecasting temperature

Effects of Cloud Cover:
During the day, the earth is heated by the sun. If skies are clear, more heat reaches the earth's surface. This leads to warmer temperatures.
However, if skies are cloudy, some of the sun's rays are reflected off the cloud droplets back into space. Therefore, less of the sun's energy is able to reach the earth's surface, which causes the earth to heat up more slowly. This leads to cooler temperatures.

Forecast Tip

When forecasting daytime temperatures, if cloudy skies are expected, forecast lower temperatures than you would predict if clear skies were expected. At night cloud cover has the opposite effect. If skies are clear, heat emitted from the earth's surface freely escapes into space, resulting in colder temperatures. However, if clouds are present, some of the heat emitted from the earth's surface is trapped by the clouds and reemitted back towards the earth. As a result, temperatures decrease more slowly than if the skies were clear.

When forecasting nighttime temperatures, if cloudy skies are expected, forecast warmer temperatures than you would predict if clear skies were expected.

1.6. Summary

Chapter gives examples of the usage of forecasting for solving different industrial and non industrial problems, like a Network traffic forecasting, business forecasting and weather forecasting.

CHAPTER 2
FORECASTING MODELS

2.1. Overview

Forecasting models are generally classified as naive models, econometric models and time-series models. An assumption underlying both econometric models and time series models is that sample observations from a random process provide reliable evidence of future activity. The difference between these two classes of models is that econometric models use auxiliary variables as predictors and time series models do not. Time series models are "pattern fitters", relying on an extension of inherent components.

There is no such thing as a single best forecasting model to use in all instances. A forecasting model that may be appropriate for estimating future levels of sales for an established product may be totally inappropriate for forecasting the sales of a new product not yet introduced to the marketplace. Thus, one of the primary tasks associated with forecasting is a matching an appropriate forecasting model to the time series to be forecast. The forecaster becomes more proficient at this task through experience gained from the study of time series behavior and from trial and error in the use of various forecasting procedures.

The purpose throughout this chapter will be explained various forecasting models and introduce the concept of forecast accuracy.

2.2. Naive models

There are some forecasting models that are very intuitive and easy to apply to any time series. These models are usually examined prior to beginning the search for a more sophisticated forecasting model. Because of their simplicity, the models we discuss in this section are often referred to as naive models. Naive models provide a baseline against which we can compare the forecast generated by the more sophisticated models. At least,

for example, we would want the forecasting accuracy provided by a more complicated model to be substantially better than the accuracy provided by a naïve model. We would have little justification for using a sophisticated, time-consuming analysis if it did not yield better forecast accuracy than a simple, inexpensive model.

The two of the most commonly used naive models, the no-change and percent-change models are presented.

- **No-change forecasting model**

The no-change forecasting model simply uses y_t as the forecast for y_{t+1}.

- **Percent-change model**

The percent-change model forecasts y_{t+1} will increase or decrease by some percentage of y_t. That is $y_{t+1}=(1+k)y_t$, where k represent the percentage change expressed in decimal form.

Many industrial companies use these simple models, sometimes without realizing that they are doing so. For example, if a company believes next year's productions (sales) ought to be about the same as this year's, then the company is implicitly using the no-change forecasting model. If the company believes that productions (sales) ought to increase by about 10% per year, then it is using a percent-change model with $k=0.10$.

The no-change forecasting model can be used as a basis of comparison for those time series that do not exhibit any long term growth or decline. The percent-change model, on the other hand, can be used for evaluating models that grow or decline in an exponential fashion.

2.3 Econometric forecasting models

An econometric model is a system of one or more equations that describe the relationship among several economic and time series variables and time series variables. Econometric

18

models are probabilistic models and capitalize on the probabilistic relationship that exists between a dependent variable representing the time series and any of a number of independent variables.

The primary feature that distinguishes econometric forecasting models from time series models is their use of economic and demographic variables that are thought to be causally related to y. Econometric models attempt to describe the relationship among such variables by use of one or more regression equations, but time series models ignore these causal variables and rely on a projection of the time series components inherent in y.

In building an econometric forecasting model, we usually begin with a large number of variables that might be closely related to the response. We then combine these variables to form models that are fitted to the sample data by using the method of least squares. However, a model that fits past data very well may be insensitive to the uncertainties associated with future events and may lead to inaccurate forecasts. Since forecasting is concerned with future events, we should select a forecasting model that demonstrates the best ability to forecast the future, not fit the past.

The linear regression model is a model that sometimes provides a suitable probabilistic model for establishing the long-term trend for a time series. For example, a long-term upward or downward trend might be isolated to fit the straight line

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

where the independent variable x represent time. A curvilinear long-term trend could be modeled by using a second-order function such as

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

The corresponding prediction equation could be determined by using the method of least squares.

The assumption of independence of the random error ε associated with successive measurements will not usually be satisfied. We would suspect that they would be conservative and that knowledge of the actual pattern of correlation would permit more accurate estimation and prediction. If the response is an average over a period of time, the correlation of adjacent response measurements will be reduced and will quite possibly satisfy adequately the assumption of independence implied in the least squares inferential procedures.

Other regression models can be constructed and fitted to data generated from economic time series by using the method of least squares. For example, the yearly production y of steel is a function of its price, the price of competitive structural materials, the production of competitive products during the preceding year, the amount of steel purchased during the immediately preceding years (to measure current inventory), and other variables. A linear model relating these independent variables to steel production might be

$$y = \beta_0 + \beta_1 x + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + \varepsilon,$$

where

x_1=time

x_2=price of steel

x_3=allowing curvature in the response curve as a function of price)

x_4= production of aluminum during previous year

x_5=price of aluminum

x_6=steel of production during previous year

$\dots$

$x_k = x_2 \, x_5$(an interaction effect between steel and aluminum prices)

The variable x_7= time could be included to capture a possible curvilinear long-term trend.

2.4. Forecasting based on time series models

2.4.1. Introduction to Time Series Analysis

Definition of Time Series: An ordered sequence of values of a variable at equally spaced time intervals.

Applications: The usage of time series models is twofold:

- Obtain an understanding of the underlying forces and structure that produced the observed data
- Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control.

Time Series Analysis is used for many applications such as:

- Economic Forecasting
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Yield Projections
- Process and Quality Control
- Inventory Studies
- Workload Projections
- Utility Studies
- Census Analysis

and many, many more...

There are many methods used to model and forecast time series

Techniques: The fitting of time series models can be an ambitious undertaking. There are many methods of model fitting including the following:

- Box-Jenkins ARIMA models

- Box-Jenkins Multivariate Models
- Holt-Winters Exponential Smoothing (single, double, triple)

The user's application and preference will decide the selection of the appropriate technique. It is beyond the realm and intention of the authors of this handbook to cover all these methods. The overview presented here will start by looking at some basic smoothing techniques:

- Averaging Methods
- Exponential Smoothing Techniques.

2.4.2. Moving average forecasting models

Simple (equally-weighted) Moving Average:

$$Y(t) = \frac{(Y(t-1) + Y(t-2) + \ldots + Y(t-k))}{k}$$

Here, the forecast equals the simple average of the last k observations. This average is "centered" at period $t-(k+1)/2$, which implies that the estimate of the local mean will tend to lag behind the true value of the local mean by about $(k+1)/2$ periods. Thus, we say the average age of the data in the simple moving average is $(k+1)/2$ relative to the period for which the forecast is computed: this is the amount of time by which forecasts will tend to *lag behind turning points* in the data. For example, if we are averaging the last 5 values, the forecasts will be about 3 periods late in responding to turning points. Note that if $k=1$, the simple moving average (SMA) model is equivalent to the random walk model (without growth). If k is very large (comparable to the length of the estimation period), the SMA model is equivalent to the mean model. As with any parameter of a forecasting model, you should choose the value of k in order to obtain the best "fit" to your data.

Here is an example of a series which appears to exhibit random fluctuations around a slowly-varying mean. First, let's try to fit it with a random walk model, which is equivalent to a simple moving average of 1 term:

Figure of Random walk.

Figure 2.1.

The random walk model responds very quickly to changes in the series, but in so doing it picks much of the "noise" in the data (the random fluctuations) as well as the "signal" (the local mean). If we instead try a simple moving average of 5 terms, we get a smoother-looking set of forecasts:

Simple moving average of 5 terms

Figure 2.2.

23

The 5-term simple moving average yields significantly smaller errors than the random walk model in this case. The average age of the data in this forecast is $3=(5+1)/2$, so that it tends to lag behind turning points by about three periods. (For example, a downturn seems to have occurred at period 21, but the forecasts do not turn around until several periods later.)

Notice that the *long-term forecasts* from the SMA model are a *horizontal straight line*, just as in the random walk model. Thus, the SMA model assumes that there is *no trend* in the data. However, whereas the forecasts from the random walk model are simply equal to the last observed value, the forecasts from the SMA model are equal to a *weighted average of recent values*.

Interestingly, the confidence limits computed by Statgraphics for the long-term forecasts of the simple moving average do *not* get wider as the forecasting horizon increases. This is obviously not correct! Unfortunately, there is no underlying statistical theory that tells us how the confidence intervals ought to widen for this model. If you were going to use this model in practice, you would be well advised to use an *empirical estimate* of the confidence limits for the longer-horizon forecasts. For example, you could set up a spreadsheet in which the SMA model would be used to forecast 2 steps ahead, 3 steps ahead, etc., within the historical data sample. You could then compute the sample standard deviations of the errors at each forecast horizon, and then construct confidence intervals for longer-term forecasts by adding and subtracting multiples of the appropriate standard deviation.

If we try a 9-term simple moving average, we get even smoother forecasts and more of a lagging effect:

Simple moving average of 9 terms

24

Figure 2.3.

The average age is now 5 periods (=(9+1)/2). If we take a 19-term moving average, the average age increases to 10:

Simple moving average of 19 terms

Figure 2.4.

Notice that, indeed, the forecasts are now lagging behind turning points by about 10 periods.

Moving average models, which function to generate a new series by computing moving averages of the original series, are oriented primarily toward removing the seasonal and

25

irregular components or isolating the trend-cycle components of a time series. The newly generated series is a "smoothed" version of the original series.

a. The Smoothing Process

Moving average models function to smooth the original time series by averaging a rolling subset of elements of the original series. The subset of the original series consists of an arbitrarily selected number of consecutive observations. The subset "rolls" or "moves" forward through the series starting from the earliest observation in the series, adding a new element at the leading edge while deleting the earliest element at the trailing edge, with each successive averaging process.

The effect of the moving average process is to ameliorate the degree of variation within the original series by composing the new smoothed series. It is possible to follow a first smoothing of a series with another smoothing of the successor series. The second smoothing may be followed by yet other smoothing. The moving average process may be used for two purposes, to remove unwanted variation from a time series, and as a forecasting model.

b. Removing Unwanted Variation

Moving average routines may be designed to remove the seasonal and random noise variation within a time series. If the moving average routine is used repeatedly on each newly-generated series, it may succeed in removing most of any cyclical variation present. What is left of the original series after early smoothing to remove seasonal and random or irregular components is a successor series retaining some combination of trend and cyclical behavior. If no trend or cyclical behavior are present in the time series, the smoothing may leave a successor series which plots as a nearly horizontal line against time on the horizontal axis. Assuming the presence of trend and cyclical behavior in the original series, the moving average process provides a method of isolating it.

While successive applications of an efficient moving-average routine may result in filtering out all variation other than the trend and cyclical behavior from an original series, this may

not be the objective. Rather, the analyst may wish to filter out only the seasonal or only the irregular variation. Either may be targeted by judiciously selecting the number of elements to be included in the moving average subset, and by designing an appropriate weighting system to accomplish his objective.

An unweighted moving average with a relatively small number of elements (say five to seven) will have its smoothing effect without destroying the seasonality present in a series. A moving average with a larger number of elements (eleven or more) with weights designed to emphasize the elements toward the center of the subset will likely be even more efficient in removing the irregular variation, but will tend also to destroy any seasonality still present.

If the analyst's intention is to deseasonalize a time series, a number of moving-average elements in the neighborhood of eleven to thirteen is called for. An odd number of elements is more easily handled than is an even number due to the need to center the moving averages relative to the object series. Also, an appropriately-designed weighting scheme applied to the elements of the moving average may serve to improve the efficiency of the seasonality removal process.

c. Unweighted Moving Average Models

We shall designate all unweighted moving average models with number of elements to be specified by the analyst as Class U.k models. The general form of the unweighted, centered moving average model with an odd number of subset elements may be specified as,

Model U.k: $MA_t = \Sigma(y_j)/k$,

j from $t-((k-1)/2)$ to $t+((k-1)/2)$,

where y is an observation in the original series at row t, k is the number of elements in the moving average, and j is the subset element counter.

Subjectively-Designed Weighting Factors

To this point we have made only passing references to the possibility of applying weighting factors to the elements of the moving average subset. If no explicit weights are used, then implicit weights of unity (value 1) are applied to each element in the subset, and the sum of the subset values must be divided by the sum of the weights (the number of elements times the weight of each) in computing each average.

The analyst may choose to use subjectively-determined, non-unitary weights to be applied to the subset elements in computing the averages. A typical scheme is to design the element weighting system so that the sum of the weights is unity (or 100 percent). In this case, each element is multiplied by its assigned fractional (or decimal value) weight, and it is unnecessary to divide the sum of the weighted values by the sum of the weights in order to compute the average, unless toward the end of the series the number of elements is diminishing.

For our purposes, all weighted moving average (WMA) models where the analyst both specifies the number of elements and subjectively determines the weights will be designated as Class W.k models. The general format of the Class W.k models may be specified as,

Model W.k: $WMA_t = \Sigma(y_j W_p)/k$,

j from $t-((k-1)/2)$ to $t+((k-1)/2)$,
p from 1 to k,

where W is an element weighting factor applied to the jth element in the moving average, and p is the element counter subscript.

Moving Averages as Forecasting Models

Any of the moving average routines described in this section may be used as forecasting models with a variable forecasting gap (i.e., lag between the value forecasted and the base value upon which it is constructed). Using the symbol i for the forecast gap, t for the subscript of the observation upon which the forecast is based, y to represent the forecasted value of the original series, and MA to represent any of the moving averages described in this chapter, the forecast model may be specified as

$y_{t+i} = MA_t,$

or if seasonality is thought to be present in the series being forecasted,

$y_{t+i} = MA_{t+i-12}.$

2.4.3. Brown's Simple Exponential Smoothing (exponentially weighted moving average)

The simple moving average model described above has the undesirable property that it treats the last k observations equally and completely ignores all preceding observations. Intuitively, past data should be discounted in a more gradual fashion for example, the most recent observation should get a little more weight than 2nd most recent, and the 2nd most recent should get a little more weight than the 3rd most recent, and so on. The simple exponential smoothing (SES) model accomplishes this. Let α denote a "smoothing constant" (a number between 0 and 1) and let $S(t)$ denote the value of the smoothed series at period t. The following formula is used recursively to update the smoothed series as new observations are recorded:

$$S(t) = \alpha y(t) + (1 - \alpha)S(t - 1)$$

Thus, the current smoothed value is an interpolation between the previous smoothed value and the current observation, where α controls the closeness of the interpolated value to the most recent observation. The *forecast* for the *next* period is simply the current smoothed value:

$$\hat{y}(t + 1) = S(t)$$

Equivalently, we can express the next forecast directly in terms of previous forecasts and previous observations, in any of the following ways:

$\hat{y}(t + 1) = \alpha y(t) + (1 - \alpha)\hat{y}(t)$ forecast=interpolation between previous *forecast* and previous *observation*

$\hat{y}(t+1) = \hat{y}(t) + \alpha\varepsilon(t)$ forecast=previous *forecast* plus fraction α of previous error, where

$\varepsilon(t) = y(t) - \hat{y}(t)$

$\hat{y}(t+1) = \alpha\left[y(t) + (1-\alpha)y(t-1) + ((1-\alpha)^2)y(t-2) + ((1-\alpha)^3)y(t-3) + ...\right]$

forecast=*exponentially weighted (i.e. discounted) moving average* with discount factor $1-\alpha$

The preceding four equations are all *mathematically equivalent* any one of them can be obtained by rearrangement of any of the others. The first equation above is probably the easiest to use if you are implementing the model on a spreadsheet: the forecasting formula fits in a single cell and contains cell references pointing to the previous forecast, the previous observation, and the cell where the value of is stored.

Note that if $\alpha = 1$, the SES model is equivalent to a random walk model (without growth). If $\alpha = 0$, the SES model is equivalent to the mean model, assuming that the first smoothed value is set equal to the mean.

The average age of the data in the simple-exponential-smoothing forecast is $1/\alpha$ relative to the period for which the forecast is computed. (This is not supposed to be obvious, but it can easily be shown by evaluating an infinite series.) Hence, the simple moving average forecast tends to lag behind turning points by about $1/\alpha$ periods. For example, when $\alpha = 0.5$ the lag is 2 periods; when $\alpha = 0.2$ the lag is 5 periods; when $\alpha = 0.1$ the lag is 10 periods, and so on.

For a given average age (i.e., amount of lag), the simple exponential smoothing (SES) forecast is somewhat superior to the simple moving average (SMA) forecast because *it places relatively more weight on the most recent observation* i.e., it is slightly more "responsive" to changes occurring in the recent past.

Another important advantage of the SES model over the SMA model is that the SES model uses a smoothing parameter which is continuously variable, so it can easily optimized by using a "solver" algorithm to minimize the mean squared error. The optimal value of α in the SES model for this series turns out to be 0.2961, as shown in Figure 2.5.

30

Simple exponential smoothing with alpha=0.2961

Figure 2.5. SES model

The average age of the data in this forecast is 1/0.2961 = 3.4 periods, which is similar to that of a 6-term simple moving average.

The long-term forecasts from the SES model are a *horizontal straight line*, as in the SMA model and the random walk model without growth. However, note that the confidence intervals computed by Statgraphics now diverge in a reasonable-looking fashion, and that they are substantially narrower than the confidence intervals for the random walk model. The SES model assumes that the series is somewhat "more predictable" than does the random walk model.

An SES model is actually a special case of an ARIMA model, so the statistical theory of ARIMA models provides a sound basis for calculating confidence intervals for the SES model. In particular, an SES model is an ARIMA model with one nonseasonal difference, an MA(1) term, and no constant term, otherwise known as an "ARIMA(0,1,1) model without constant". The MA(1) coefficient in the ARIMA model corresponds to the quantity $1-\alpha$ in the SES model.

It is possible to add the assumption of a non-zero constant trend to an SES model. To do this in Statgraphics, just specify an ARIMA model with one nonseasonal difference and an

MA(1) term *with* a constant. The long-term forecasts will then have a trend which is equal to the average trend observed over the entire estimation period. However, you cannot do this in conjunction with seasonal adjustment, because the seasonal adjustment options are disabled when the model type is set to ARIMA.

2.4.4. Brown's Linear (i.e., double) Exponential Smoothing

If the *trend* as well as the mean is varying slowly over time, a higher-order smoothing model is needed to track the varying trend. The simplest time-varying trend model is Brown's linear exponential smoothing (LES) model, which uses *two* different smoothed series that are centered at different points in time. The forecasting formula is based on an extrapolation of a line through the two centers.

The algebraic form of the linear exponential smoothing model, like that of the simple exponential smoothing model, can be expressed in a number of different but equivalent forms. The "standard" form of this model is usually expressed as follows: Let S' denote the *singly-smoothed* series obtained by applying simple exponential smoothing to series Y. That is, the value of S' at period t is given by:

$$S'(t) = \alpha y(t) + (1 - \alpha)S'(t-1)$$

(Recall that, under simple exponential smoothing, we would just let $\acute{Y}(t+1) = S'(t)$ at this point.) Then let S" denote the *doubly-smoothed* series obtained by applying simple exponential smoothing (using the same α) to series S':

$$S''(t) = \alpha S'(t) + (1 - \alpha)S''(t-1)$$

Finally, the forecast $\hat{y}(t+1)$ is given by:

$$\hat{y}(t+1) = a(t) + b(t)$$

where:

$$a(t) = 2S'(t) - S''(t) \text{...the estimated } \textit{level} \text{ at period t}$$

$b(t) = (\alpha/(1-\alpha))(S'(t) - S''(t))$ ...the estimated *trend* at period t.

Forecasts with longer lead times made at period t are obtained by adding multiples of the trend term. For example, the k-period-ahead forecast (i.e., the forecast for $y(t+k)$ made at period t) would be equal to a(t)+kb(t). For purposes of model-fitting (i.e., calculating forecasts, residuals, and residual statistics over the estimation period), the model can be started up by setting S'(1)=S''(1)= $y(1)$, i.e., set both smoothed series equal to the observed value at t=1.

A mathematically equivalent form of Brown's linear exponential smoothing model, which emphasizes its non-stationary character and is easier to implement on a spreadsheet, is the following:

$$\hat{y}(t) = 2y(t-1) - y(t-2) - 2(1-\alpha)\varepsilon(t-1) + ((1-\alpha)^2)\varepsilon(t-2)$$

or equivalently:

$$\hat{y}(t) - y(t-1) = y(t-1) - y(t-2) - 2(1-\alpha)\varepsilon(t-1) + ((1-\alpha)^2)\varepsilon(t-2)$$

In other words, the *predicted difference* at period t (namely $\hat{y}(t) - y(t-1)$) is equal to the previous observed difference (namely Y(t-1) - Y(t-2)) minus a weighted difference of the two previous forecast errors.

Caution: this form of the model is rather tricky to start up at the beginning of the estimation period. The following convention is recommended: first set $\hat{y}(1) = y(1)$, which yields $\varepsilon(1)$ = 0 (i.e., cheat a bit, and let the first forecast equal the actual first observation), then also set $\hat{y}(2) = y(1)$, which yields $\varepsilon(2) = \hat{y}(2) - y(1)$, then continue from this point using the equation above. This would yield the same fitted values as the formula based on S' and S'' if the latter were started up using S'(1) = S''(1) = y(1).

Once again, you can use your spreadsheet the "solver" or any nonlinear least squares algorithm to optimize the value of α. The optimal value of α in the LES model fitted to this series by Statgraphics is 0.1607. Note that the long-term forecasts of the LES model for

33

this time series appear to track the local trend observed in the last 10 periods. Also, the confidence intervals for the LES model extend faster than those of the SES model.

Brown linear exp. smoothing with alpha =0.1607

Figure 2.6. Brown LES model

2.5. Four common measures of forecast accuracy

When choosing between competing forecasting models or when evaluating an existing model, we need to use measures that summarize the overall accuracy provided by the model(s). Generally speaking, the closer the forecasts are to the actual values of the series, the more accurate the forecasting model is. Thus, the quality of a model can be evaluated by examining the series of forecast errors. All of the measures that follow are based on some simple function of the forecast errors.

The most commonly used measures of forecast accuracy are the mean absolute deviation (MAD), the mean square error (MSE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE). Formulas for the computation of these measures:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right|$$

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

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

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| (100\%)$$

The basic difference between the MAD and the MSE (or RMSE) is that by squaring the amount of the error, the MSE (and RMSE) penalizes extreme errors more heavily than does the MAD. Thus, the MAD is an appropriate measure of forecast accuracy when the costs of forecast error increase linearly with the size of his error. The MSE (and RMSE) is best if costs for large errors are disproportionately expensive.

Since the MAPE is measured as a percentage and is therefore "unitless", it is particularly useful for comparing the performance of the model on many different time series. One drawback to using the MAPE arises when a series has any extremely small terms since the division by those terms will tend to seriously inflate the MAPE. For this reason, it is also not wise to use the MAPE as a model selection tool.

When we use the MAD, MSE, or RMSE to select a good forecasting model, the time series data is usually split into two parts. The first part of the data is used to estimate the parameters of the particular model. Then with these estimates the model is used to forecast the remaining data points. The MAD, MSE, and RMSE are calculated from the forecast errors for this second part of the series.

2.6. Summary

This chapter presents the classification of forecasting models for three groups: naive models, econometric models, time series models, where time series models were considered more detailed. Also four common measures of forecast accuracy were considered.

CHAPTER 3

THE BOX-JENKINS FORECASTING PROCEDURE

3.1. Overview

ARIMA (Auto Regressive Integrated Moving Average) model was introduced by Box and Jenkins in 1976 includes three types of parameters: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q). In this chapter it will be explained structure and three primer stages in building a Box-Jenkins time series model.

3.2. Stochastic and Deterministic Dynamic Mathematical Models

The idea of using a mathematical model to describe the behavior of a physical phenomenon is well established. In particular, it is sometimes possible to derive a model based on physical laws, which enables us to calculate the value of some time-dependent quantity nearly exactly at any instant of time. Thus, we might calculate the trajectory of a missile launched in a known direction with known velocity. If exact calculation were possible, such a model would be entirely *deterministic*.

Probably no phenomenon is totally deterministic, however, because unknown factors can occur such as a variable wind velocity that can throw a missile slightly off course. In many problems we have to consider a time-dependent phenomenon, such as monthly sales of newsprint, in which there are many unknown factors and for which it is not possible to write a deterministic model that allows exact calculation of the future behavior of the phenomenon. Nevertheless, it may be possible to derive a model that can be used to calculate the *probability* of a future value lying between two specified limits. The models for time series that are needed, for example to achieve optimal forecasting and control are in fact stochastic models. It is necessary in what follows to distinguish between the probability

36

model or *stochastic process*, as it is sometimes called, and the observed time series. Thus a time series $z_1, z_2, ..., z_n$ of N successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by the stochastic process. Very often we shall omit the word stochastic from "stochastic process" and talk about the "process."

Stationary and Nonstationary Stochastic Models for Forecasting and Control

An important class of stochastic models for describing time series, which has received a great deal of attention, comprises what are called *stationary'* models, which assume that the process remains in *equilibrium* about a *constant mean level.* However, forecasting has been of particular importance in industry, business, and economics, where many time series are often better represented as nonstationary and, in particular, as having no natural constant mean level over time. It is not surprising, therefore, that many of the economic forecasting methods originally proposed by Holt, Winters, Brown, and the Id Monograph that used exponentially weighted moving averages can be shown to be appropriate for a particular type of *nonstationary* process. Although such methods are too narrow to deal efficiently with all time series, the fact that they give the right kind of forecast function supplies a clue to the *kind* of *nonstationary* model that might be useful in these problems.

The stochastic model for which the exponentially weighted moving average forecast yields minimum mean square error is a member of a class of *nonstationary* processes called autoregressive integrated moving average processes (ARIMA processes). This wider class of processes provides a range of models, stationary and nonstationary, that adequately represent many of the time series met in practice. Once an appropriate model has been determined for the series, the optimal forecasting procedure follows immediately. These forecasting procedures include the exponentially weighted moving average forecast as a special case.

Some simple operators:

It should be employed extensively the *shift operator B*, which is defined by $Bz_t = z_{t-1}$; hence $B^m z_t = z_{t-m}$. The inverse operation is performed by the *forward shift operator F* $= B^{-1}$ given by $Fz_t = z_{t+1}$; hence $F^m z_t = z_{t+m}$. Another important operator is the *backward difference operator* ∇, which can be written in terms of B, since

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_t$$

Linear filter model

The stochastic models we employ are based on the idea that a time series in which successive values are highly dependent can frequently be regarded as generated from a series of *independent* "shocks" a_t. These shocks are *random* drawings from a fixed distribution, usually assumed Normal and having mean zero and variance σ_a^2. Such a sequence of random variables $a_t, a_{t-1}, a_{t-2} \ldots$ is called a *white noise* process.

The white noise process a_t is supposed transformed to the process z_t what is called a *linear filter*. The linear filtering operation simply takes a weighted sum of previous random shocks a_t, so that

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + .. = \mu + \psi(B)a_t$$

3.3. Box-Jenkins Approach

The Box-Jenkins' modeling procedure involves identifying an appropriate ARMA model, fitting it to the dataset, and using the fitted model for forecasting. One of the attractive features of the Box-Jenkins method to forecasting is that ARMA processes are a very rich class of possible models and it is usually possible to find a model which provides an adequate description for the data.

The Box-Jenkins' modeling procedure involves iterative steps of model identification, parameter estimation, model diagnosis and forecasting. A modified procedure for searching

an adequate ARMA model. The modified procedure mainly involves four steps: identification, pre-estimation / estimation, diagnosis and forecast. Each of these steps will be discussed and illustrated in the sections that follow.

Figure 3.1. Block scheme of model construction

In contrast to other techniques, Box-Jenkins is a procedure, which uses a variable's past behavior to select the best forecasting model from a general class of models. It assumes that any time series pattern can be represented by one of three categories of models. These categories include:

• Autoregressive models: forecasts of a variable based on linear function of its past values

• Moving Average models: forecasts based on linear combination of past errors

• Autoregressive-Moving Average models: combination of the previous two categories

Note that one of the key questions is how many past values (the focal variable and/or its errors) should be included in the model.

Generally speaking, there are essentially three stages to a Box-Jenkins procedure:

1. Identifying the tentative model. Which of the three categories listed above is identified as the appropriate category is determined by first making the data stationary (usually by differencing the data) and then analyzing the autocorrelations and partial autocorrelations of the stationary data. Note that there are theoretical autocorrelation and partial autocorrelation profiles for each of the possible models. Therefore, determining the appropriate type of model for a specific situation is mainly a matter of matching the observed correlations to the theoretical correlations.

2. Determining the parameters of the model. This is similar to estimating the parameters in regression analysis.

3. Application of the model.

The Box-Jenkins ARMA model is a combination of the AR and MA models :

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

where the terms in the equation have the same meaning as given for the AR and MA model.

Autoregressive models

A stochastic model that can be extremely useful in the representation of certain practically occurring series is the *autoregressive* model. In this model, the current value of the process is expressed as a finite, linear aggregate of *previous values of the process* and a shock. Let us denote the values of a process at equally spaced times $t, t - 1, t - 2, \ldots$ by $z_t, z_{t-1}, z_{t-2} \ldots$ Also let $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots$ be deviations from μ; for example, $\tilde{z}_t = z_t - \mu$. Then

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + ... + \phi_p \tilde{z}_{t-p} + a_t \tag{1}$$

is called an *autoregressive* (AR) *process of order p*. The reason for this name is that a

linear model

$$\widetilde{z} = \phi_1 \widetilde{x}_1 + \phi_2 \widetilde{x}_2 + ... + \phi_p \widetilde{x}_p + a$$

relating a "dependent" variable z to a set of "independent" variables $x_1, x_2, ..., x_p$, plus an error term a, is often referred to as a regression model, and z is said to be "regressed" on $x_1, x_2, ..., x_p$. In (1) the variable z is regressed on previous values of itself; the model is autoregressive. If an autoregressive operator of order p is defined by:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$$

the autoregressive model may be written economically as

$$\phi(B)\widetilde{z}_t = a_t.$$

The model contains $p+2$ unknown parameters μ, $\phi_1, \phi_2, ..., \phi_p, \sigma_a^2$, which in practice have to be estimated from the data. The additional parameter σ_a^2 is the variance of the white noise process a_t.

Moving average models

The autoregressive model (1) expresses the deviation $\widetilde{z}_t$ of the process as *a finite* weighted sum of p previous deviations $\widetilde{z}_{t-1}, \widetilde{z}_{t-2}, ..., \widetilde{z}_{t-p}$ of the process, plus a random shock a_t. Equivalently it expresses $\widetilde{z}_t$ as infinite weighted sum of a's.

Another kind of model, of great practical importance in the representation of observed time series, is the finite *moving average* process. Here $\widetilde{z}_t$ linearly dependent on a *finite* number q of previous a's. Thus

$$\widetilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}$$

is called a *moving average* (MA) *process of order q*. The name "moving average" is somewhat misleading because the weights $1, -\theta_1, -\theta_2, ..., \theta_q$, which multiply the a's, need not total unity nor need they be positive. However, this nomenclature is in common use, and there fore we employ it.

41

If we define a *moving average operator* of order q by

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta q\, B^p$$

the moving average model may be written economically as

$$\tilde{z}_t = \theta\,(B)a_1$$

It contains q + 2 unknown parameters $\mu, \theta_1, \ldots, \theta_q, \sigma_a^2$, which in practice have to be estimated from the data.

Mixed autoregressive-moving average models

To achieve greater flexibility in fitting of actual time series, it is sometimes advantageous to include both autoregressive and moving average terms in the model. This leads to the mixed autoregressive-moving average model

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

which employs $p + q + 2$ unknown parameters $\mu, \phi_1, \ldots, \phi_p; \theta_1, \ldots, \theta_q; \sigma_a^2$, that are estimated from the data. This model may also be written in the form of the linear filter as $\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t$. In practice, it is frequently true that adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average, or mixed models, in which p and q are not greater than 2 and often less than 2.

Comments on Box-Jenkins Model

Box-Jenkins model characteristics are

1. The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for "Integrated".

2. Some formulations transform the series by subtracting the mean of the series from each data point. This yields a series with a mean of zero. Whether you need to do this or not is dependent on the software you use to estimate the model.

3. Box-Jenkins models can be extended to include seasonal autoregressive and seasonal moving average terms. Although this complicates the notation and mathematics of the model, the underlying concepts for seasonal autoregressive and seasonal moving average terms are similar to the non-seasonal autoregressive and moving average terms.

4. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms. As with modeling in general, however, only necessary terms should be included in the model.

There are three primary stages in building a Box-Jenkins time series model.

1. Model Identification
2. Model Estimation
3. Model Validation.

The following remarks regarding Box-Jenkins models should be noted.

1. Box-Jenkins models are quite flexible due to the inclusion of both autoregressive and moving average terms.

2. Based on the World decomposition theorem (not discussed in the Handbook), a stationary process can be approximated by an ARMA model. In practice, finding that approximation may not be easy.

3. Chatfield (1996) recommends decomposition methods for series in which the trend and seasonal components are dominant.

4. Building good ARIMA models generally requires more experience than commonly used statistical methods such as regression.

3.4. Box-Jenkins Model Identification

3.4.1. Stationarity

A common assumption in many time series techniques is that the data are stationary.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations (seasonality).

For practical purposes, stationarity can usually be determined from a run sequence plot.

The following plots are from a data set of monthly CO2 concentrations

Example of Run Sequence Plot

Figure 3.2.

The initial run sequence plot of the data indicates a rising trend. A visual inspection of this plot indicates that a simple linear fit should be sufficient to remove this upward trend.

If the time series is not stationary, we can often transform it to stationarity with one of the following techniques.

1. We can difference the data. That is, given the series Z_t, we create the new series

$$Y_t = Z_t - Z_{t-1}$$

44

The differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.

2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

3. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

The above techniques are intended to generate series with constant location and scale. Although seasonality also violates stationarity, this is usually explicitly incorporated into the time series model.

3.4.2. Seasonality

Many time series display seasonality. By seasonality, we mean periodic fluctuations. For example, retail sales tend to peak for the Christmas season and then decline after the holidays. So time series of retail sales will typically show increasing sales from September through December and declining sales in January and February.

Seasonality is quite common in economic time series. It is less common in engineering and scientific data.

If seasonality is present, it must be incorporated into the time series model. In this section, we discuss techniques for detecting seasonality.

a. Detecting Seasonality

The following graphical techniques can be used to detect seasonality.

1. A run sequence plot will often show seasonality.
2. A seasonal subseries plot is a specialized technique for showing seasonality.
3. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.
4. The autocorrelation plot can help identify seasonality.

The run sequence plot is a recommended first step for analyzing any time series. Although seasonality can sometimes be indicated with this plot, seasonality is shown more clearly by the seasonal subseries plot or the box plot. The seasonal subseries plot does an excellent job of showing both the seasonal differences (between group patterns) and also the within-group patterns. The box plot shows the seasonal difference (between group patterns) quite well, but it does not show within group patterns. However, for large data sets, the box plot is usually easier to read than the seasonal subseries plot.

Both the seasonal subseries plot and the box plot assume that the seasonal periods are known. In most cases, the analyst will in fact know this. For example, for monthly data, the period is 12 since there are 12 months in a year. However, if the period is not known, the autocorrelation plot can help. If there is significant seasonality, the autocorrelation plot should show spikes at lags equal to the period.

The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled.

b. Differencing to achieve stationarity

Box and Jenkins recommend the differencing approach to achieve stationarity. However, fitting a curve and subtracting the fitted values from the original data can also be used in the context of Box-Jenkins models.

c. Seasonal differencing

At the model identification stage, our goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. For example,

for monthly data we would typically include either a seasonal AR 12 term or a seasonal MA 12 term. For Box-Jenkins models, we do not explicitly remove seasonality before fitting the model. Instead, we include the order of the seasonal terms in the model specification to the ARIMA estimation software. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model idenfitication of the non-seasonal component of the model. In some cases, the seasonal differencing may remove most or all of the seasonality effect.

d. Identify p and q

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e., the p and q) of the autoregressive and moving average terms.

The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known.

Order of Autoregressive Process (p)

Specifically, for an AR(1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components.

For higher-order autoregressive processes, the sample autocorrelation needs to be supplemented with a partial autocorrelation plot. The partial autocorrelation of an AR(p) process becomes zero at lag $p+1$ and greater, so we examine the sample partial autocorrelation function to see if there is evidence of a departure from zero. This is usually determined by placing a 95% confidence interval on the sample partial autocorrelation plot (most software programs that generate sample autocorrelation plots will also plot this confidence interval). If the software program does not generate the confidence band, it is approximately $\pm 2/\sqrt{N}$, with N denoting the sample size.

47

Order of Moving Average Process (q)

The autocorrelation function of a MA(q) process becomes zero at lag $q+1$ and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero. We do this by placing the 95% confidence interval for the sample autocorrelation function on the sample autocorrelation plot. Most software that can generate the autocorrelation plot can also generate this confidence interval.

The autocorrelation function

The dependence structure of a stationary time series is characterized by the autocorrelation function. The autocorrelation function is defined as the correlation between z_t and z_{t+k}; ρ_k =correlation (z_t, z_{t+k}), k is called the time lag. The autocorrelation function is estimated by the empirical autocorrelation function: $r_k = c_k / c_0$, where k=0, 1, 2, ... ; c_k are the empirical autocovariances.

The empirical autocorrelation- and cross-correlation functions are the main tools for the identification of the ARIMA model. The autocorrelation function of the basic processes has a typical shape. The autocorrelation function of the AR(1) process decays exponentially. The autocorrelation function of the MA(1) process has only $\rho 1$ nonzero and of the MA(2) process only $\rho 1$ and $\rho 2$ nonzero.

Autocorrelation plots are a commonly-used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero.

In addition, autocorrelation plots are used in the model identification stage for Box-Jenkins autoregressive, moving average time series models.

Autocorrelation plots are formed by

- Vertical axis: Autocorrelation coefficient

$$r_n = c_n / c_0$$

where c_n is the autocovariance function

$$c_n = (1/N) \sum_{T=1}^{N-n} (y_t - \bar{y})(y_{t+h} - \bar{y})$$

and c_0 is the variance function

$$c_0 = \frac{\sum_{t=1}^{N} (y_t - \bar{y})^2}{N}$$

Note r_n is between -1 and +1.

- Horizontal axis: Time lag n (n = 1, 2, 3, ...)

Shape of Autocorrelation Function

The following table summarizes how we use the sample autocorrelation function for model identification.

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting	Mixed autoregressive and

49

after a few lags	moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

Examples of the autocorrelation plot:

Autocorrelation Plot for Strong Autocorrelation:

Figure 3.3.

We can make the following conclusions from the above plot.

1. The data come from an underlying autoregressive model with strong positive autocorrelation.

The plot starts with a high autocorrelation at lag 1 (only slightly less than 1) that slowly declines. It continues decreasing until it becomes negative and starts showing an incresing negative autocorrelation. The decreasing autocorrelation is generally linear with little noise. Such a pattern is the autocorrelation plot signature of "strong autocorrelation", which in turn provides high predictability if modeled properly.

The next step would be to estimate the parameters for the autoregressive model:

50

$$y_t = A_0 + A_1 y_{t-1} + E_t$$

Such estimation can be performed by using least squares linear regression or by fitting a Box-Jenkins autoregressive (AR) model.

Partial Autocorrelation Plot:

Partial autocorrelation plots are commonly used tool for model identification in Box-Jenkins models.

The partial autocorrelation at lag k is the autocorrelation between y_t and y_{t-k} that is not accounted for by lags 1 through k-1.

There are algorithms for computing the partial autocorrelation based on the sample autocorrelations.

Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR(p) process is zero at lag $p+1$ and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order. We look for the point on the plot where the partial autocorrelations essentially become zero. Placing a 95% confidence interval for statistical significance is helpful for this purpose.

The approximate 95% confidence interval for the partial autocorrelations are at $\pm 2/\sqrt{N}$

3.5. Box-Jenkins Model Estimation

Estimating the parameters for the Box-Jenkins models is a quite complicated nonlinear estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models. Fortunately, many commercial statistical software programs now fit Box-Jenkins models.

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation.

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation.

The identification process having led to a tentative formulation for the model, we then need to obtain efficient estimates *of* the parameters. After the parameters have been estimated, the fitted model will be subjected to diagnostic checks and tests of goodness of fit. As pointed out by R. A. Fisher, for tests of goodness of fit to be relevant, it is necessary that efficient use of the data should have been made in the fitting process. If this is not so, inadequacy of fit may simply arise because of the inefficient fitting and not because the form *of* the model is inadequate.

Likelihood Function

Suppose that we have a sample *of* N observations z with which we associate an N-dimensional random variable, whose known probability distribution depends on some unknown parameters ξ. We use the vector ξ to denote a general set *of* parameters and, in particular, it could refer to the $p + q + 1$ parameters $(\varphi, \theta, \sigma)$ of the ARIMA model.

Before the data are available, $p(z/\xi)$ will associate a density with each different outcome z of the experiment, for fixed ξ. After the data have come to hand, we are led to contemplate the various values of ξ that might have given rise to the 'fixed set *of* observations z actually

52

obtained. The appropriate function for this purpose is the *likelihood function* $L(\xi /z)$, which is *of* the same form as $p(z/ \xi)$, but in which z is now fixed but ξ is variable. It is only the relative value *of* $L(\xi /z)$ which is *of* interest, so that the likelihood function is usually regarded as containing an *arbitrary multiplicative constant.*

It is *often* convenient to work with the log-likelihood function $ln[L(\xi /z)] = l(\xi /z)$, which contains an *arbitrary additive constant.* One reason that the likelihood function is fundamental importance in estimation theory is because of the "likelihood principle'. This principle says that (given that the assumed model is correct) all that the *data* have to tell us about the parameters is contained in the likelihood function, all other aspects *of* the data being irrelevant. From a Bayesian point of view, the likelihood function is equally important, since it is the component in the posterior distribution of the parameters which comes from the data.

For a complete understanding of the estimation situation, it is necessary to make a through analytical and graphical study of the likelihood function, or in the Bayesian framework, the posterior distribution of the parameters, which in the situations we consider, is dominated by the likelihood. In many examples, for moderate and large samples, the log-likelihood function will be unimodal and can be approximated adequately over a sufficiently extensive region near the maximum by a quadratic function.

In such cases the log-likelihood function can be described by its maximum and its second derivatives at the maximum. The values of the parameters that maximize the likelihood function, or equivalently the log-likelihood function, are called *maximum likelihood (ML) estimates.* The second derivatives of the log-likelihood function provide measures of "spread" of the likelihood function and can be used to calculate approximate standard errors for the estimates.

The limiting properties of maximum likelihood estimates are usually established for independent observations. But as was shown by Whittle, they may be extended to cover stationary time series.

3.6. Box-Jenkins Model Diagnostics

Model diagnostics for Box-Jenkins models is similar to model validation for non-linear least squares fitting.

That is, the error term A_t is assumed to follow the assumptions for a stationary univariate process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance. If the Box-Jenkins model is a good model for the data, the residuals should satisfy these assumptions.

If these assumptions are not satisfied, we need to fit a more appropriate model. That is, we go back to the model identification step and try to develop a better model. Hopefully the analysis of the residuals can provide some clues as to a more appropriate model.

3.7. Examples of Forecast Functions and Their Updating

3.7.1. Forecasting a General IMA(0,d,q) Process

As an example, consider the process of order $(0, 1, 3)$,

$$(1 - B)z_{t+1} = \left(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3\right)a_{t+1}$$

Taking conditional expectations at time we obtain

$$\hat{z}_t - z_t = -\theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2}$$

$$\hat{z}_t(2) - \hat{z}_t(1) = -\theta_2 a_t - \theta_3 a_{t-1}$$

$$\hat{z}_t(3) - \hat{z}_t(2) = -\theta_3 a_t$$

$$\hat{z}_t(l) - \hat{z}_t(l-1) = 0 \qquad l = 4,5,6,\dots$$

Hence $\hat{z}_t(l) = \hat{z}_t(3) = b_0^{(t)}$ for all $l > 2$, as expected, since q-p-d=2.

3.7.2. Forecasting Autoregressive Process

Consider a process of order $(p, d, 0)$

$$\phi(B)z_t = a_t$$

54

The eventual forecast function is the solution of $\phi(B)\hat{z}_t(l) = 0$. It applies for all lead times and passes through the last $p + d$ available values of the series.

The best forecast for all future time is very nearly the current value of the stock. The weight function for $\hat{z}_t(l)$ is a spike at time t and there is no averaging over past history.

Nonstationary autoregressive models of order (p, d, 0)

For the model

$$\phi(B)\nabla^d z_t = a_t$$

it will be the dth difference of the process that decays back to its mean when projected several steps ahead. The mean of $\nabla^d z_t$ will usually be assumed to be zero unless contrary evidence is available. When needed, it is possible to introduce a nonzero mean by replacing $\nabla^d z_t$ by $\nabla^d z_t - \mu_\omega$ in the model. For example, consider the model

$$(1 - \phi B)(\nabla z_t - \mu_\omega) = a_t$$

After substituting $t+j$ for t and taking conditional expectations at origin t,

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu_\omega = \phi^j(z_t - z_{t-1} - \mu_\omega)$$

or $\hat{\omega}_t(j) - \mu_\omega = \phi^j(\omega_t - \mu_\omega)$, where $\omega_t = \nabla z_t$, which shows how the forecasted difference decays exponentially from the initial value $\omega_t = z_t - z_{t-1}$ to its mean value μ_ω. On summing this expression from $j = 1$ to $j = l$, that is, using $\hat{z}_t(l) = \hat{\omega}_t(l) + ... + \hat{\omega}_t(1) + z_t$, we obtain the forecast function

$$\hat{z}_t(l) = z_t + \mu_w l + (z_t - z_{t-1} - \mu_w)\frac{\phi(1-\phi^l)}{1-\phi}, \quad l \geq 1$$

3.8. Summary

This chapter presents the detailed description of the Box-Jenkins forecasting procedure. The explanation of three primary stages in building a Box-Jenkins is given: model identification, model estimation, model diagnostic.

CHAPTER 4

IMPLEMENTATION OF FORECASTING METHODS

4.1. Overview

At the present chapter two examples of implementation of forecasting methods will be considered. One of them is prediction of world petroleum production and second is prediction of the number of Computer Engineering Department students. For this purpose program, which was written in Delphi, is applied. Also all the forecasting methods, which were used in program, are have done in Excel with purpose to check results.

4.2. World Petroleum Production

Central European Bank gives time series for World petroleum production for 8 years from 1995 to 2002 monthly. The original time series are shown in the Figure 4.1, where x axis is a time, and y axis is Petroleum production monthly in Mb/d.
Series title: Petroleum Production, World. The part of those data is given below, whole set is given in Table 1 in Appendix.

Year	Month	Value,(Mb/d)
2002	1	66.391,57
2002	2	66.523,81
2002	3	66.338,93
2002	4	65.780,53
2002	5	66.184,34
2002	6	66.058,17
2002	7	66.654,99
2002	8	66.362,88
2002	9	67.111,00
2002	10	68.769,01
2002	11	68.764,28
2002	12	67.008,72

Plot of actual values for World Petroleum Production

Figure 4.1.

4.2.1. Forecast for 2002

Three kind of forecasting methods are applied to present example: Moving average forecasting model, an exponential-smoothing forecasting model, Box-Jenkins forecasting procedure. Here appeared problem, which model will fit better to the present time series, the quality of model can be evaluated by examining the forecast errors. As far as actual data for 2003 aren't known, forecast errors can't be calculated, so at first, forecasting procedures are applied for 2002 year.

4.2.1.a. Moving Average Forecasting Model

One-step-ahead moving average forecast is applied for forecasting petroleum production in 2002, let k be equal to 5.
The results of forecasting which were gotten in Excell are shown below.

Implementation of Moving Average Forecasting Model in Excel

Table 4.1.

Year	Month	Actual Value,(Mb/d)	One-Step Forecast	One-step absolute deviation
2001	1	68.940,12		
2001	2	68.478,21		
2001	3	69.165,97		
2001	4	68.268,40		
2001	5	67.576,67		
2001	6	66.004,01		
2001	7	67.978,71		
2001	8	68.165,35		
2001	9	67.792,01		
2001	10	67.680,40		
2001	11	67.929,27		
2001	12	67.473,76		
2002	1	66.391,57	67.808,16	1.416,59
2002	2	66.523,81	67.453,40	929,59
2002	3	66.338,93	67.199,76	860,83
2002	4	65.780,53	66.931,47	1.150,94
2002	5	66.184,34	66.501,72	317,38
2002	6	66.058,17	66.243,84	185,67
2002	7	66.654,99	66.177,16	477,83
2002	8	66.362,88	66.203,39	159,49
2002	9	67.111,00	66.208,18	902,82
2002	10	68.769,01	66.474,28	2.294,73
2002	11	68.764,28	66.991,21	1.773,07
2002	12	67.008,72	67.532,43	523,71
		1.617.401,11		10.992,65
Mean absolute deviation				916,05

To check the accuracy of the model in Figure 4.2 are given two graphs: actual values, which are known and values, which were gotten by using present method.

Plot of forecasts by Moving Average model for 2002 and actual values in Excel
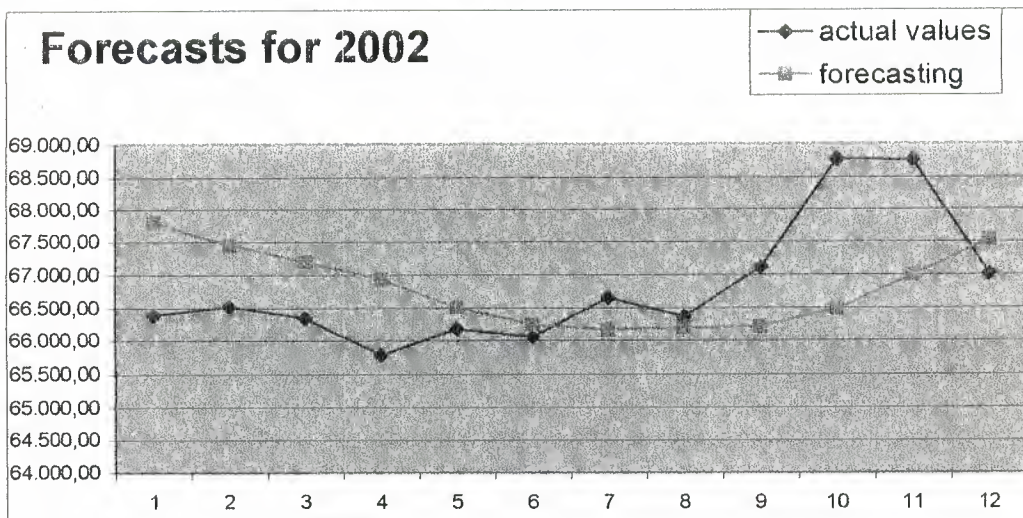

Figure 4.2.

According to the same purpose program gives following result:

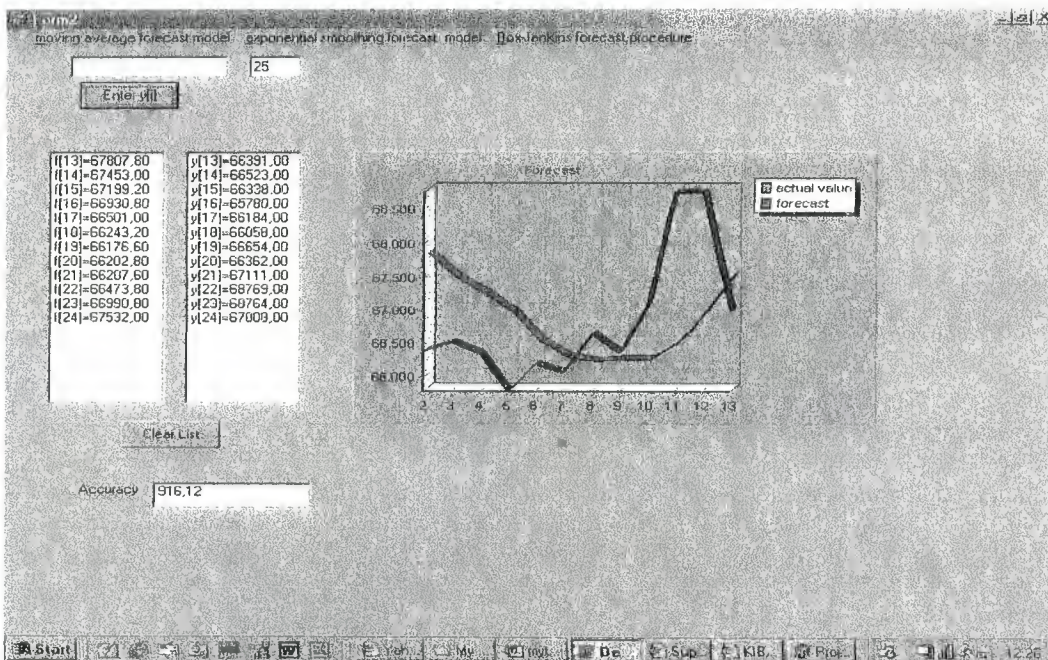**Plot of forecasts by Moving Average model for 2002 and actual values in Delphi**

Figure 4.3.

Forecasts by Moving Average model for 2002 and actual values in Delphi

Table 4.2.

forecast	Actual values
67807	66391
67453	66523
67199	66338
66930	65780
66501	66184
66243	66654
66202	66362
66207	67111
66473	68769
66990	68764
67532	67008

4.2.1.b. An Exponential-Smoothing Forecasting Model

The results of applying Brown's method of multiple smoothing are present by following table and Figure 4.4:

Implementation of Exponential-Smoothing Forecasting Model in Excel

Table 4.3.

Year	Month	Actual Value, (Mb/d)	St	St(2)	St(3)	Forecast, T=1, a=0,2	Forecast Error
2002	1	66.391,57	66.391,57	66.391,57	66.391,57		
2002	2	66.523,81	66.418,02	66.396,86	66.392,63	66.391,57	132,24
2002	3	66.338,93	66.402,20	66.397,93	66.393,69	66.470,91	131,98
2002	4	65.780,53	66.277,87	66.373,92	66.389,73	66.407,59	627,06
2002	5	66.184,34	66.259,16	66.350,96	66.381,98	66.032,44	151,90
2002	6	66.058,17	66.218,96	66.324,56	66.370,50	66.049,42	8,75
2002	7	66.654,99	66.306,17	66.320,89	66.360,57	65.993,73	661,26
2002	8	66.362,88	66.317,51	66.320,21	66.352,50	66.326,79	36,09
2002	9	67.111,00	66.476,21	66.351,41	66.352,28	66.360,37	750,63
2002	10	68.769,01	66.934,77	66.468,08	66.375,44	66.828,57	1.940,44
2002	11	68.764,28	67.300,67	66.634,60	66.427,27	68.102,58	661,70
2002	12	67.008,72	67.242,28	66.756,14	66.493,05	68.850,05	1.841,33
		801.948,23					6.943,38
MAD							631,22

60

Ekaterina Gerasimova: **Analysis of Forecasting Models and Time-Series Modelling of Network Traffic**

Approval of the Graduate School of Applied and Social Sciences

Prof. Dr. Fakhraddin Mamedov
Director

We certify this thesis is satisfactory for the award of the Degree of Master of Science in Computer Engineering

Examining Committee in charge:

Assoc. Prof. Dr. Doğan İbrahim, Chairman , Chairman of
Computer Engineering Department, NEU

Assoc. Prof. Dr. Faeq Radwan, Member , Computer Engineering
Department, NEU

Assoc. Prof. Dr. Ilham Huseynov, Member, Computer
Information System Department, NEU

Assoc. Prof. Dr. Rahib Abiyev, Supervisor, Computer Engineering
Department, NEU

DEPARTMENT OF COMPUTER ENGINEERING
DEPARTMENTAL DECISION

Date: 04.06.2004

Subject: Completion of M.Sc. Thesis

Participants: Assoc. Prof. Dr. Doğan İbrahim, Assoc.Prof. Dr. Faeq Radwan, Assoc.Prof. Dr. Ilham Huseynov, Assoc. Prof. Rahib Abiyev, Ibaid Alsoud, Shadi Jundy.

DECISION

We certify that the student whose number and name are given below, has fulfilled all the requirements for a M .S. degree in Computer Engineering.

CGPA

20011456	Ekaterina Gerasimova	3.214

Assoc. Prof. Dr. Doğan İbrahim, Committee Chairman , Chairman of Computer Engineering Department, NEU

Assoc. Prof. Dr. Faeq Radwan, Committee Member , Computer Engineering Department, NEU

Assoc. Prof. Dr. Ilham Huseynov, Committee Member, Computer Information System Department, NEU

Assoc. Prof. Dr. Rahib Abiyev, Supervisor, Computer Engineering Department, NEU

Chairman of Department
Assoc. Prof. Dr. Doğan İbrahim

DEPARTMENT OF
COMPUTER ENGINEERING

Academic Year: 2003-2004

STUDENT INFORMATION

Full Name	Ekaterina Gerasimova		
Undergraduate degree	BSc.	Date Received	Spring 2001
University	Ivanovo State University	CGPA	3.34

THESIS

Title	Analysis of Forecasting Models and Time-Series Modelling of Network Traffic

Description The aim of this thesis was analyzing of forecasting models and applying them for solving of economical and technical problems. The forecasting model for network traffic has been developed by using time-series analysis.

Supervisor	Assoc.Prof.Dr.Rahib Abiyev	Department	Computer Engineering

DECISION OF EXAMINING COMMITTEE

The jury has decided to accept / reject the student's thesis.
The decision was taken unanimously / by majority.

COMMITTEE MEMBERS

Number Attending	3	Date	4.06.2004

Name	Signature
Assoc. Prof. Dr. Doğan İbrahim, Chairman of the jury	
Assoc. Prof. Dr. Faeq Radwan, Member	
Assoc. Prof. Dr. Ilham Huseynov, Member	

APPROVALS

Date	Chairman of Department
04.06.2004	Assoc. Prof. Dr. Doğan İbrahim

EAST UNIVERSITY

GRADUATE SCHOOL OF APPLIED AND SOCIAL SCIENCES

ANALYSIS OF FORECASTING MODELS AND TIME-SERIES MODELLING OF NETWORK TRAFFIC

EKATERINA GERASIMOVA

MASTER THESIS

DEPARTMENT OF COMPUTER ENGINEERING

Nicosia - 2004

ACKNOWLEDGEMENTS

ABSTRACT

To make accurate decision and improve the efficiency of the complicated system one of the effective way is predicting future behavior of these systems and making adequate control strategy. It gives us chance to make effective planning and managing of the process. For this reason forecasting plays a major role in most of our activities for the future.

The present work gives consideration of the Forecasting models and Time Series Analysis. Analysis of forecasting models and the use of those models in different industrial and non industrial areas are considered. As an example of the application of forecasting models to World Petroleum Production, Computer Engineering Department students number, and Network Traffic is considered. The simulation of these models has been done. Simulation results demonstrate that one of effective methodology for forecasting of future is time series analysis. The analysis and development of the time series models are considered. Using time series analysis the ARIMA models are developed for forecasting world petroleum production for year monthly, number of students for 2004-2005 academic year, and Network Traffic for a week. The simulations of ARIMA model for these problems have been done. By the comparison the results of forecasting by using ARIMA model with results of other models, demonstrate that first one gives more accurate results. The model can make reasonable predictions for one or more years to the future, suggesting that ARIMA modeling has great promise as a tool for short or long-range forecasting and planning.

1

TABLE OF CONTENTS

INTRODUCTION

Forecasting plays major roles in most of our activities and in all we do concerning the future. It is a branch of the anticipatory sciences used for identifying and projecting alternative possible futures. It is a conduit leading to plans for the development of "better" futures. Forecasted visions of possible futures open our freedom of choice over which future to encourage or discourage. In our fast-paced, rapidly changing world, the futures that we will experience will tend to be vastly different from our present reality in a growing number of ways. Furthermore, because of constant development of new knowledge and advances in the scientific (and ensuing technological advances), sociological, political, economic, and business areas, our global society has an ever increasing ability to shape (for better or worse) the futures we will eventually achieve.

As a result, society and each institution in it finds that more knowledge about possible futures and the consequences of today's decisions and actions is required. Thus, it is increasingly imperative that we have better forecasting tools and that we apply them in responsible ways. It is more and more important to forecast, ahead of time, with longer lead times, the possible futures implied by the changes produced by this new knowledge generation. To these ends, forecasting has become an essential tool for all participants in society to use in their attempts to decide, plan, design, steer, manage, implement, and control change by identifying preferable futures with forecasts.

In general terms, a forecast is simply a statement, based upon some criteria, concerning the future condition of something. A major purpose of forecasting is to give us choice over which future, either the trend path we are on, or an alternative, to plan, design, create, and to back with our resources.

Forecasting aids in identifying which futures to bring into fruition (the preferable) and which to forestall, or attempt to eliminate (the undesirables). Furthermore, forecasts are useful in assisting our intuitions about our plans, outlooks, investments, and so on, whenever we would like to have a better idea of the possible or most probable outcome.

5

Information from forecasts relative to:
- identifying:
    - > possible futures
    - > probable futures
    - > preferable futures
- providing a basis for understanding the process and dynamics of change;
- providing notions of where change may take us into the future;
- providing a systematic methodology, based upon a set of supportive assumptions, for the discovery of possible futures.

Since we know that individuals and society in general have the means and knowledge to shape major elements of our future, to grow new opportunities as well to set in motion means for avoiding or lessening the impact of negative future threats, forecasting again becomes an increasingly important tool for all of us to understand and use.

Forecasting has many applications, some of which are to:
- identify the trend path we are traveling into the future;
- identify alternative possible futures (alternatives to the trends);
- provide views of possible futures;
- raise awareness of possible futures so that we have choice over which future we support;
- generate "future histories" that we can study to determine our role in shaping them before they become a reality;
- provide information about possible futures so that realistic planning can occur;
- provide information on possible futures to aid decision-making and planning;
- justify the decisions and plans we make;
- discover possible breakthroughs;
- discover possible life, societal, scientific, political, social, technological, and institutional future turning points or paradigm shifts;
- track evolving change and advances;

- provide managers with information for choosing their organization's vision, mission, purpose, goals, objectives, strategies, plans, and tactics;
- provide information on possible futures for assessment relative to their possible future impacts and consequences.

One of important problem in complex systems such as technical, economical systems is to increase the efficiency of used control systems for these systems. Forecasting plays an important role in increasing the efficiency of the technological and economical systems. It is acquired by the predicting the future conditions of these system and making appropriate control decision.

Forecasting is a highly "noisy" application, because we are typically dealing with systems that receive thousands of inputs, which interact in a complex nonlinear fashion. Usually with a very small subset of inputs or measurements available from the system, the system behavior is to be estimated and extrapolated into the future.

The aim of this thesis is the analysis of forecasting models and application of time series analysis for solving forecasting of economical and technical problems. The use of forecasting methods to three different problems is represented. These are forecasting world petroleum production, number of registered students in Engineering Department and network traffic forecasting.

Thesis consists of introduction, five chapters, conclusion and appendix.

Introduction represents the actuality of the problem studied and the brief description of each chapter.

Chapter one gives the review of the usage of forecasting models for different problems.

Chapter two introduces the description of various forecasting models and calculation of forecast accuracy. The formulation and explanation of moving average, regression models, exponential smoothing and time series forecasting models are described.

7

In chapter three Box-Jenkins analysis and forecasting procedure have been represented. Auto Regressive Integrated Moving Average model, introduced by Box and Jenkins is described. There is explanation of structure and three primer stages in building a Box-Jenkins time series model. The different structure time-series models and their application have been analyzed.

Chapter four is the application of forecasting methods to two problems, world petroleum production and number of registered students of Engineering Department in University. For this purpose program, which was developed in Delphi has been used. The results and comparison of different forecasting methods were given.

Chapter five is devoted to the development network traffic forecasting by using time series analysis. The Autoregressive Integrated Moving Average model is used for forecasting of Network traffic. For analysis of the data and modeling of this process S plus package was applied. To make forecasting the program written in Delphi programming has been developed. The developed program allows to forecast the workload in network traffic and to plan capacity requirements of the network. One ARIMA model was chosen, and was used for forecasting.

In conclusion the obtained important results from the thesis are given

Appendix includes the tables with statistical data, listing of Delphi program and guide to program.

CHAPTER 1

THE USAGE OF MATHEMATICAL MODELS FOR SOLVING INDUSTRIAL PROBLEMS

1.1. Overview

Administrators in all organizations make plans to cope with future changes. The planning means to make decisions in advance about the future course of action. Obviously, then, planning and decision making are based on forecasts or expectations of what the future holds. Several applications of forecasting will be considered in this chapter, like a weather forecasting and business forecasting.

1.2. Forecasting Process

Forecasting process includes the following

  a. Collect appropriate data
  b. Examine data patterns
  c. Choose a forecasting method (model)
  d. Apply the model to past periods (ex post)
  e. Examine the accuracy of model by examining ex post errors
  f. If adequate (errors random and sufficiently small) use the model to forecast the future
  g. Periodically check the accuracy of forecasts with actual experience

1.3. Forecasting in Engineering

On the base of statistical data the models for forecasting industrial and non industrial process has been created. In the world there are many works for solving forecasting problems in different fields, like a engineering, industrial, business, management.

9

In [5] there is introduced a methodology to predict *when and where* link additions/upgrades have to take place in an IP backbone network. Using SNMP statistics, collected continuously since 1999, aggregate demand is computed between any two adjacent PoPs and look at its evolution at time scales larger than one hour. It is shown that IP backbone traffic exhibits visible long term trends, strong periodicities, and variability at multiple time scales. This methodology relies on the wavelet multiresolution analysis and linear time series models. Using wavelet multiresolution analysis, we smooth the collected measurements until we identify the *overall long-term trend*. The fluctuations around the obtained trend are further analyzed at multiple time scales. There shown that the largest amount of variability in the original signal is due to its fluctuations at the 12 hour time scale. We model inter-PoP aggregate demand as a multiple linear regression model, consisting of the two identified components. It is shown that this model accounts for 98% of the total energy in the original signal, while explaining 90% of its variance. Weekly approximations of those components can be accurately modeled with low-order AutoRegressive Integrated Moving Average (ARIMA) models. Forecasting the long term trend and the fluctuations of the traffic at the 12 hour time scale yields accurate estimates for at least six months in the future.

In [6] non-linear threshold autoregressive models are examined for use in modeling the temporal variation in the byte-rate in Ethernet traffic. The model is comprised of a number of autoregressive processes each of which is to be used in a specified range of amplitude of the byte-rate. The local dynamics within each threshold range are captured by an autoregressive process. The switching between each submodel is conditioned on the amplitude of a lagged value of the time-series. To develop the model the Bellcore Ethernet LAN data is used. It is shown that non-linear threshold autoregressive processes can be used to capture the dynamics of Ethernet LAN traffic. This model also provides for both short and long term prediction capability and allows us to quantitatively identify the sources of long-range-dependence features in the traffic. When the aggregate traffic is partitioned into classes based on packet sizes, certain classes of traffic follow deterministic cyclical patterns. These periodic components arise from the process switching between different amplitude regimes. Superposed on this fundamental period are longer cycles that

can be localized either below or above the mean byte-rate. By constructing amplitude thresholds associated with a finite set of delay parameters, the dynamics within each threshold are captured by locally linear autoregressive processes. The aggregate process is globally nonlinear. This model is shown to provide good agreement with the marginal distributions and the correlation functions derived from the Ethernet traffic data. In addition, simulation experiments demonstrate that the loss statistics observed in finite buffer queues agree favorably with those generated by the measurements.

1.4. Business forecasting

Business organizations, public organizations, and individuals thus have the common goal of allocating available time among competing resources in some optimal manner. This goal is accomplished by making forecast of future activities and taking the proper actions as suggested by these forecast.

In business and public administration the organization with both short-term and long-term forecasts. The short-term forecast usually looks no more than one year into the future and involves forecasting sales, price changes, and customer demand, which, in turn, reflect the need for seasonal employment, short-term forecast usually looks from 2 to 10 years into the future and is used as planning model for product line and capital investment decisions, as indicated by changing demand patterns.

Naturally, the further a forecast is projected into the future, the more speculative it becomes. But since the future is always uncertain, we cannot expect complete accuracy any forecast. The time series underlying the process to be forecast is bound to be influenced by many causal factors- some forcing the time series up while conflicting factors act to force the series down. Nevertheless, business people must make forecasts of future business activity in order to budget their time and resources efficiently. They cannot hope to account for every possible factor that may cause the response of interest to rise or fall over time. All that can be expected is that the benefits gained by forecasting offset the opportunity cost for not forecasting. Note that such benefits are not limited to real monetary savings but may

11

imply a sharpening of the businessperson's thinking to consider the interplay of the events that affect the movement of the time series.

1.5. Weather forecasting

How often do you watch the weather on TV or listen on the radio for the weather forecast? The weather affects everything from afternoon swim practice to attacks on enemy forces during wars.

Weather forecasting used to be thought of as witchcraft. Today, we rely on weather forecasters to help us plan our days and prepare for life-threatening conditions.

1.5.1. Forecasting methods

a. Persistence Method (today equals tomorrow):
There are several different methods that can be used to create a forecast. The method a forecaster chooses depends upon the experience of the forecaster, the amount of information available to the forecaster, the level of difficulty that the forecast situation presents, and the degree of accuracy or confidence needed in the forecast.

The first of these methods is the Persistence Method; the simplest way of producing a forecast. The persistence method assumes that the conditions at the time of the forecast will not change. For example, if it is sunny and 87 degrees today, the persistence method predicts that it will be sunny and 87 degrees tomorrow. If two inches of rain fell today, the persistence method would predict two inches of rain for tomorrow.

The persistence method works well when weather patterns change very little and features on the weather maps move very slowly. It also works well, where summertime weather conditions vary little from day to day. However, if weather conditions change significantly from day to day, the persistence method usually breaks down and it is not the best forecasting method to use.

It may also appear that the persistence method would work only for shorter-term forecasts (e.g. a forecast for a day or two), but actually one of the most useful roles of the persistence forecast is predicting long range weather conditions or making climate forecasts. For example, it is often the case that one hot and dry month will be followed by another hot and dry month. So, making persistence forecasts for monthly and seasonal weather conditions can have some skill. Some of the other forecasting methods, such as numerical weather prediction, lose all their skill for forecasts longer than 10 days. This makes persistence a "hard to beat" method for forecasting longer time periods.

b. Trends Method (using mathematics)

The trends method involves determining the speed and direction of movement for fronts, high and low pressure centers, and areas of clouds and precipitation. Using this information, the forecaster can predict where he or she expects those features to be at some future time. For example, if a storm system is 1000 miles west of your location and moving to the east at 250 miles per day, using the trends method you would predict it to arrive in your area in 4 days.

Mathematics

(1000 miles / 250 miles per day = 4 days)

Using the trends method to forecast only a few hours into the future is known as "Nowcasting" and this method is frequently used to forecast precipitation. For example, if a line of thunderstorms is located 60 miles to your northwest and moving southeast at 30 miles per hour, you would predict the storms to arrive in your area in 2 hours.

The trends method works well when systems continue to move at the same speed in the same direction for a long period of time. If they slow down, speed up, change intensity, or change direction, the trends forecast will probably not work as well.

13

c. Other Forecasting Methods (climatology, analogue and numerical weather prediction)

Climatology: The Climatology Method is another simple way of producing a forecast. This method involves averaging weather statistics accumulated over many years to make the forecast. If you were making a forecast for temperature and precipitation, then you would use this recorded weather data to compute the averages for temperature and precipitation.

If these averages were 87 degrees with 0.18 inches of rain, then the weather forecast, using the climatology method, would call for a high temperature of 87 degrees with 0.18 inches of rain. The climatology method only works well when the weather pattern is similar to that expected for the chosen time of year. If the pattern is quite unusual for the given time of year, the climatology method will often fail.

Analog Method

The Analog Method is a slightly more complicated method of producing a forecast. It involves examining today's forecast scenario and remembering a day in the past when the weather scenario looked very similar (an analog). The forecaster would predict that the weather in this forecast will behave the same as it did in the past.

For example, suppose today is very warm, but a cold front is approaching your area. You remember similar weather conditions one last week, also a warm day with cold front approaching. You also remember how heavy thunderstorms developed in the afternoon as the cold front pushed through the area. Therefore, using the analog method, you would predict that this cold front will also produce thunderstorms in the afternoon.

The analog method is difficult to use because it is virtually impossible to find a perfect analog. Various weather features rarely align themselves in the same locations they were in the previous time. Even small differences between the current time and the analog can lead to very different results. However, as time passes and more weather data is archived, the

chances of finding a "good match" analog for the current weather situation should improve, and so should analog forecasts.

Numerical Weather Prediction

Numerical Weather Prediction (NWP) uses the power of computers to make a forecast. Complex computer programs, also known as forecast models, run on supercomputers and provide predictions on many atmospheric variables such as temperature, pressure, wind, and rainfall. A forecaster examines how the features predicted by the computer will interact to produce the day's weather.

The NWP method is flawed in that the equations used by the models to simulate the atmosphere are not precise. This leads to some error in the predictions. In addition, there are many gaps in the initial data since we do not receive many weather observations from areas in the mountains or over the ocean. If the initial state is not completely known, the computer's prediction of how that initial state will evolve will not be entirely accurate.

Despite these flaws, the NWP method is probably the best of the five discussed here at forecasting the day-to-day weather changes. Very few people, however, have access to the computer data. In addition, the beginning forecaster does not have the knowledge to interpret the computer forecast, so the simpler forecasting methods, such as the trends or analogue method, are recommended for the beginner.

1.5.2. Example of forecasting temperature

Effects of Cloud Cover:
During the day, the earth is heated by the sun. If skies are clear, more heat reaches the earth's surface. This leads to warmer temperatures.
However, if skies are cloudy, some of the sun's rays are reflected off the cloud droplets back into space. Therefore, less of the sun's energy is able to reach the earth's surface, which causes the earth to heat up more slowly. This leads to cooler temperatures.

Forecast Tip

When forecasting daytime temperatures, if cloudy skies are expected, forecast lower temperatures than you would predict if clear skies were expected. At night cloud cover has the opposite effect. If skies are clear, heat emitted from the earth's surface freely escapes into space, resulting in colder temperatures. However, if clouds are present, some of the heat emitted from the earth's surface is trapped by the clouds and reemitted back towards the earth. As a result, temperatures decrease more slowly than if the skies were clear.

When forecasting nighttime temperatures, if cloudy skies are expected, forecast warmer temperatures than you would predict if clear skies were expected.

1.6. Summary

Chapter gives examples of the usage of forecasting for solving different industrial and non industrial problems, like a Network traffic forecasting, business forecasting and weather forecasting.

CHAPTER 2
FORECASTING MODELS

2.1. Overview

Forecasting models are generally classified as naive models, econometric models and time-series models. An assumption underlying both econometric models and time series models is that sample observations from a random process provide reliable evidence of future activity. The difference between these two classes of models is that econometric models use auxiliary variables as predictors and time series models do not. Time series models are "pattern fitters", relying on an extension of inherent components.

There is no such thing as a single best forecasting model to use in all instances. A forecasting model that may be appropriate for estimating future levels of sales for an established product may be totally inappropriate for forecasting the sales of a new product not yet introduced to the marketplace. Thus, one of the primary tasks associated with forecasting is a matching an appropriate forecasting model to the time series to be forecast. The forecaster becomes more proficient at this task through experience gained from the study of time series behavior and from trial and error in the use of various forecasting procedures.

The purpose throughout this chapter will be explained various forecasting models and introduce the concept of forecast accuracy.

2.2. Naive models

There are some forecasting models that are very intuitive and easy to apply to any time series. These models are usually examined prior to beginning the search for a more sophisticated forecasting model. Because of their simplicity, the models we discuss in this section are often referred to as naive models. Naive models provide a baseline against which we can compare the forecast generated by the more sophisticated models. At least,

for example, we would want the forecasting accuracy provided by a more complicated model to be substantially better than the accuracy provided by a naïve model. We would have little justification for using a sophisticated, time-consuming analysis if it did not yield better forecast accuracy than a simple, inexpensive model.

The two of the most commonly used naive models, the no-change and percent-change models are presented.

- **No-change forecasting model**

The no-change forecasting model simply uses y_t as the forecast for y_{t+1}.

- **Percent-change model**

The percent-change model forecasts y_{t+1} will increase or decrease by some percentage of y_t. That is $y_{t+1}=(1+k)y_t$, where k represent the percentage change expressed in decimal form.

Many industrial companies use these simple models, sometimes without realizing that they are doing so. For example, if a company believes next year's productions (sales) ought to be about the same as this year's, then the company is implicitly using the no-change forecasting model. If the company believes that productions (sales) ought to increase by about 10% per year, then it is using a percent-change model with $k=0.10$.

The no-change forecasting model can be used as a basis of comparison for those time series that do not exhibit any long term growth or decline. The percent-change model, on the other hand, can be used for evaluating models that grow or decline in an exponential fashion.

2.3 Econometric forecasting models

An econometric model is a system of one or more equations that describe the relationship among several economic and time series variables and time series variables. Econometric

models are probabilistic models and capitalize on the probabilistic relationship that exists between a dependent variable representing the time series and any of a number of independent variables.

The primary feature that distinguishes econometric forecasting models from time series models is their use of economic and demographic variables that are thought to be causally related to y. Econometric models attempt to describe the relationship among such variables by use of one or more regression equations, but time series models ignore these causal variables and rely on a projection of the time series components inherent in y.

In building an econometric forecasting model, we usually begin with a large number of variables that might be closely related to the response. We then combine these variables to form models that are fitted to the sample data by using the method of least squares. However, a model that fits past data very well may be insensitive to the uncertainties associated with future events and may lead to inaccurate forecasts. Since forecasting is concerned with future events, we should select a forecasting model that demonstrates the best ability to forecast the future, not fit the past.

The linear regression model is a model that sometimes provides a suitable probabilistic model for establishing the long-term trend for a time series. For example, a long-term upward or downward trend might be isolated to fit the straight line

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

where the independent variable x represent time. A curvilinear long-term trend could be modeled by using a second-order function such as

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

The corresponding prediction equation could be determined by using the method of least squares.

19

The assumption of independence of the random error ε associated with successive measurements will not usually be satisfied. We would suspect that they would be conservative and that knowledge of the actual pattern of correlation would permit more accurate estimation and prediction. If the response is an average over a period of time, the correlation of adjacent response measurements will be reduced and will quite possibly satisfy adequately the assumption of independence implied in the least squares inferential procedures.

Other regression models can be constructed and fitted to data generated from economic time series by using the method of least squares. For example, the yearly production y of steel is a function of its price, the price of competitive structural materials, the production of competitive products during the preceding year, the amount of steel purchased during the immediately preceding years (to measure current inventory), and other variables. A linear model relating these independent variables to steel production might be

$$y = \beta_0 + \beta_1 x + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \varepsilon,$$

where

$\quad x_1 = $ time

$\quad x_2 = $ price of steel

$\quad x_3 = $ allowing curvature in the response curve as a function of price)

$\quad x_4 = $ production of aluminum during previous year

$\quad x_5 = $ price of aluminum

$\quad x_6 = $ steel of production during previous year

$\quad \ldots$

$\quad x_k = x_2 x_5$ (an interaction effect between steel and aluminum prices)

The variable $x_7 = $ time could be included to capture a possible curvilinear long-term trend.

2.4. Forecasting based on time series models

2.4.1. Introduction to Time Series Analysis

Definition of Time Series: An ordered sequence of values of a variable at equally spaced time intervals.

Applications: The usage of time series models is twofold:

- Obtain an understanding of the underlying forces and structure that produced the observed data
- Fit a model and proceed to forecasting, monitoring or even feedback and feedforward control.

Time Series Analysis is used for many applications such as:

- Economic Forecasting
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Yield Projections
- Process and Quality Control
- Inventory Studies
- Workload Projections
- Utility Studies
- Census Analysis

and many, many more...

There are many methods used to model and forecast time series

Techniques: The fitting of time series models can be an ambitious undertaking. There are many methods of model fitting including the following:

- Box-Jenkins ARIMA models

- Box-Jenkins Multivariate Models
- Holt-Winters Exponential Smoothing (single, double, triple)

The user's application and preference will decide the selection of the appropriate technique. It is beyond the realm and intention of the authors of this handbook to cover all these methods. The overview presented here will start by looking at some basic smoothing techniques:

- Averaging Methods
- Exponential Smoothing Techniques.

2.4.2. Moving average forecasting models

Simple (equally-weighted) Moving Average:

$$Y(t) = \frac{(Y(t-1) + Y(t-2) + \ldots + Y(t-k))}{k}$$

Here, the forecast equals the simple average of the last k observations. This average is "centered" at period $t-(k+1)/2$, which implies that the estimate of the local mean will tend to lag behind the true value of the local mean by about $(k+1)/2$ periods. Thus, we say the average age of the data in the simple moving average is $(k+1)/2$ relative to the period for which the forecast is computed: this is the amount of time by which forecasts will tend to *lag behind turning points* in the data. For example, if we are averaging the last 5 values, the forecasts will be about 3 periods late in responding to turning points. Note that if k=1, the simple moving average (SMA) model is equivalent to the random walk model (without growth). If k is very large (comparable to the length of the estimation period), the SMA model is equivalent to the mean model. As with any parameter of a forecasting model, you should choose the value of k in order to obtain the best "fit" to your data.

Here is an example of a series which appears to exhibit random fluctuations around a slowly-varying mean. First, let's try to fit it with a random walk model, which is equivalent to a simple moving average of 1 term:

Figure of Random walk.

Figure 2.1.

The random walk model responds very quickly to changes in the series, but in so doing it picks much of the "noise" in the data (the random fluctuations) as well as the "signal" (the local mean). If we instead try a simple moving average of 5 terms, we get a smoother-looking set of forecasts:

Simple moving average of 5 terms

Figure 2.2.

23

The 5-term simple moving average yields significantly smaller errors than the random walk model in this case. The average age of the data in this forecast is $3=(5+1)/2$, so that it tends to lag behind turning points by about three periods. (For example, a downturn seems to have occurred at period 21, but the forecasts do not turn around until several periods later.)

Notice that the *long-term forecasts* from the SMA model are a *horizontal straight line*, just as in the random walk model. Thus, the SMA model assumes that there is *no trend* in the data. However, whereas the forecasts from the random walk model are simply equal to the last observed value, the forecasts from the SMA model are equal to a *weighted average of recent values*.

Interestingly, the confidence limits computed by Statgraphics for the long-term forecasts of the simple moving average do *not* get wider as the forecasting horizon increases. This is obviously not correct! Unfortunately, there is no underlying statistical theory that tells us how the confidence intervals ought to widen for this model. If you were going to use this model in practice, you would be well advised to use an *empirical estimate* of the confidence limits for the longer-horizon forecasts. For example, you could set up a spreadsheet in which the SMA model would be used to forecast 2 steps ahead, 3 steps ahead, etc., within the historical data sample. You could then compute the sample standard deviations of the errors at each forecast horizon, and then construct confidence intervals for longer-term forecasts by adding and subtracting multiples of the appropriate standard deviation.

If we try a 9-term simple moving average, we get even smoother forecasts and more of a lagging effect:

Simple moving average of 9 terms

Figure 2.3.

The average age is now 5 periods (=(9+1)/2). If we take a 19-term moving average, the average age increases to 10:

Simple moving average of 19 terms
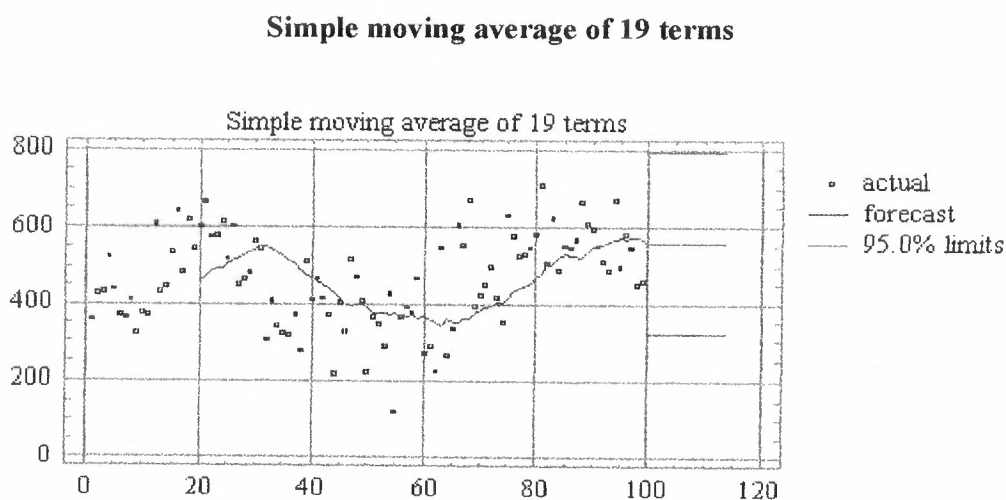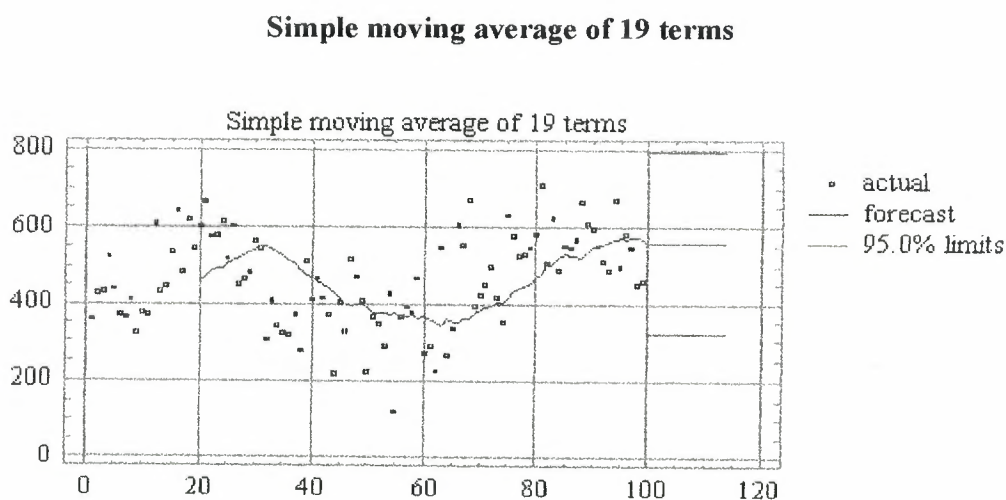
Figure 2.4.

Notice that, indeed, the forecasts are now lagging behind turning points by about 10 periods.

Moving average models, which function to generate a new series by computing moving averages of the original series, are oriented primarily toward removing the seasonal and

irregular components or isolating the trend-cycle components of a time series. The newly generated series is a "smoothed" version of the original series.

a. The Smoothing Process

Moving average models function to smooth the original time series by averaging a rolling subset of elements of the original series. The subset of the original series consists of an arbitrarily selected number of consecutive observations. The subset "rolls" or "moves" forward through the series starting from the earliest observation in the series, adding a new element at the leading edge while deleting the earliest element at the trailing edge, with each successive averaging process.

The effect of the moving average process is to ameliorate the degree of variation within the original series by composing the new smoothed series. It is possible to follow a first smoothing of a series with another smoothing of the successor series. The second smoothing may be followed by yet other smoothing. The moving average process may be used for two purposes, to remove unwanted variation from a time series, and as a forecasting model.

b. Removing Unwanted Variation

Moving average routines may be designed to remove the seasonal and random noise variation within a time series. If the moving average routine is used repeatedly on each newly-generated series, it may succeed in removing most of any cyclical variation present. What is left of the original series after early smoothing to remove seasonal and random or irregular components is a successor series retaining some combination of trend and cyclical behavior. If no trend or cyclical behavior are present in the time series, the smoothing may leave a successor series which plots as a nearly horizontal line against time on the horizontal axis. Assuming the presence of trend and cyclical behavior in the original series, the moving average process provides a method of isolating it.

While successive applications of an efficient moving-average routine may result in filtering out all variation other than the trend and cyclical behavior from an original series, this may

not be the objective. Rather, the analyst may wish to filter out only the seasonal or only the irregular variation. Either may be targeted by judiciously selecting the number of elements to be included in the moving average subset, and by designing an appropriate weighting system to accomplish his objective.

An unweighted moving average with a relatively small number of elements (say five to seven) will have its smoothing effect without destroying the seasonality present in a series. A moving average with a larger number of elements (eleven or more) with weights designed to emphasize the elements toward the center of the subset will likely be even more efficient in removing the irregular variation, but will tend also to destroy any seasonality still present.

If the analyst's intention is to deseasonalize a time series, a number of moving-average elements in the neighborhood of eleven to thirteen is called for. An odd number of elements is more easily handled than is an even number due to the need to center the moving averages relative to the object series. Also, an appropriately-designed weighting scheme applied to the elements of the moving average may serve to improve the efficiency of the seasonality removal process.

c. Unweighted Moving Average Models

We shall designate all unweighted moving average models with number of elements to be specified by the analyst as Class U.k models. The general form of the unweighted, centered moving average model with an odd number of subset elements may be specified as,

Model U.k: $MA_t = \Sigma(y_j)/k$,

j from $t-((k-1)/2)$ to $t+((k-1)/2)$,

where y is an observation in the original series at row t, k is the number of elements in the moving average, and j is the subset element counter.

Subjectively-Designed Weighting Factors

To this point we have made only passing references to the possibility of applying weighting factors to the elements of the moving average subset. If no explicit weights are used, then implicit weights of unity (value 1) are applied to each element in the subset, and the sum of the subset values must be divided by the sum of the weights (the number of elements times the weight of each) in computing each average.

The analyst may choose to use subjectively-determined, non-unitary weights to be applied to the subset elements in computing the averages. A typical scheme is to design the element weighting system so that the sum of the weights is unity (or 100 percent). In this case, each element is multiplied by its assigned fractional (or decimal value) weight, and it is unnecessary to divide the sum of the weighted values by the sum of the weights in order to compute the average, unless toward the end of the series the number of elements is diminishing.

For our purposes, all weighted moving average (WMA) models where the analyst both specifies the number of elements and subjectively determines the weights will be designated as Class W.k models. The general format of the Class W.k models may be specified as,

Model W.k: $WMA_t = \Sigma(y_j W_p)/k$,

j from $t-((k-1)/2)$ to $t+((k-1)/2)$,
p from 1 to k,

where W is an element weighting factor applied to the jth element in the moving average, and p is the element counter subscript.

Moving Averages as Forecasting Models

Any of the moving average routines described in this section may be used as forecasting models with a variable forecasting gap (i.e., lag between the value forecasted and the base value upon which it is constructed). Using the symbol i for the forecast gap, t for the subscript of the observation upon which the forecast is based, y to represent the forecasted value of the original series, and MA to represent any of the moving averages described in this chapter, the forecast model may be specified as

$y_{t+i} = MA_t,$

or if seasonality is thought to be present in the series being forecasted,

$y_{t+i} = MA_{t+i-12}.$

2.4.3. Brown's Simple Exponential Smoothing (exponentially weighted moving average)

The simple moving average model described above has the undesirable property that it treats the last k observations equally and completely ignores all preceding observations. Intuitively, past data should be discounted in a more gradual fashion for example, the most recent observation should get a little more weight than 2nd most recent, and the 2nd most recent should get a little more weight than the 3rd most recent, and so on. The simple exponential smoothing (SES) model accomplishes this. Let α denote a "smoothing constant" (a number between 0 and 1) and let S(t) denote the value of the smoothed series at period t. The following formula is used recursively to update the smoothed series as new observations are recorded:

$$S(t) = \alpha y(t) + (1 - \alpha)S(t - 1)$$

Thus, the current smoothed value is an interpolation between the previous smoothed value and the current observation, where α controls the closeness of the interpolated value to the most recent observation. The *forecast* for the *next* period is simply the current smoothed value:

$$\hat{y}(t + 1) = S(t)$$

Equivalently, we can express the next forecast directly in terms of previous forecasts and previous observations, in any of the following ways:

$\hat{y}(t + 1) = \alpha y(t) + (1 - \alpha)\hat{y}(t)$ forecast=interpolation between previous *forecast* and previous *observation*

29

$\hat{y}(t+1) = \hat{y}(t) + \alpha\varepsilon(t)$ forecast=previous *forecast* plus fraction α of previous error, where

$\varepsilon(t) = y(t) - \hat{y}(t)$

$\hat{y}(t+1) = \alpha\left[y(t) + (1-\alpha)y(t-1) + ((1-\alpha)^2)y(t-2) + ((1-\alpha)^3)y(t-3) + ...\right]$

forecast=*exponentially weighted (i.e. discounted) moving average* with discount factor $1-\alpha$

The preceding four equations are all *mathematically equivalent* any one of them can be obtained by rearrangement of any of the others. The first equation above is probably the easiest to use if you are implementing the model on a spreadsheet: the forecasting formula fits in a single cell and contains cell references pointing to the previous forecast, the previous observation, and the cell where the value of is stored.

Note that if $\alpha = 1$, the SES model is equivalent to a random walk model (without growth). If $\alpha = 0$, the SES model is equivalent to the mean model, assuming that the first smoothed value is set equal to the mean.

The average age of the data in the simple-exponential-smoothing forecast is $1/\alpha$ relative to the period for which the forecast is computed. (This is not supposed to be obvious, but it can easily be shown by evaluating an infinite series.) Hence, the simple moving average forecast tends to lag behind turning points by about $1/\alpha$ periods. For example, when $\alpha = 0.5$ the lag is 2 periods; when $\alpha = 0.2$ the lag is 5 periods; when $\alpha = 0.1$ the lag is 10 periods, and so on.

For a given average age (i.e., amount of lag), the simple exponential smoothing (SES) forecast is somewhat superior to the simple moving average (SMA) forecast because *it places relatively more weight on the most recent observation* i.e., it is slightly more "responsive" to changes occurring in the recent past.

Another important advantage of the SES model over the SMA model is that the SES model uses a smoothing parameter which is continuously variable, so it can easily optimized by using a "solver" algorithm to minimize the mean squared error. The optimal value of α in the SES model for this series turns out to be 0.2961, as shown in Figure 2.5.

30

Simple exponential smoothing with alpha=0.2961

Figure 2.5. SES model

The average age of the data in this forecast is $1/0.2961 = 3.4$ periods, which is similar to that of a 6-term simple moving average.

The long-term forecasts from the SES model are a *horizontal straight line*, as in the SMA model and the random walk model without growth. However, note that the confidence intervals computed by Statgraphics now diverge in a reasonable-looking fashion, and that they are substantially narrower than the confidence intervals for the random walk model. The SES model assumes that the series is somewhat "more predictable" than does the random walk model.

An SES model is actually a special case of an ARIMA model, so the statistical theory of ARIMA models provides a sound basis for calculating confidence intervals for the SES model. In particular, an SES model is an ARIMA model with one nonseasonal difference, an MA(1) term, and no constant term, otherwise known as an "ARIMA(0,1,1) model without constant". The MA(1) coefficient in the ARIMA model corresponds to the quantity $1-\alpha$ in the SES model.

It is possible to add the assumption of a non-zero constant trend to an SES model. To do this in Statgraphics, just specify an ARIMA model with one nonseasonal difference and an

31

MA(1) term *with* a constant. The long-term forecasts will then have a trend which is equal to the average trend observed over the entire estimation period. However, you cannot do this in conjunction with seasonal adjustment, because the seasonal adjustment options are disabled when the model type is set to ARIMA.

2.4.4. Brown's Linear (i.e., double) Exponential Smoothing

If the *trend* as well as the mean is varying slowly over time, a higher-order smoothing model is needed to track the varying trend. The simplest time-varying trend model is Brown's linear exponential smoothing (LES) model, which uses *two* different smoothed series that are centered at different points in time. The forecasting formula is based on an extrapolation of a line through the two centers.

The algebraic form of the linear exponential smoothing model, like that of the simple exponential smoothing model, can be expressed in a number of different but equivalent forms. The "standard" form of this model is usually expressed as follows: Let S' denote the *singly-smoothed* series obtained by applying simple exponential smoothing to series Y. That is, the value of S' at period t is given by:

$$S'(t) = \alpha y(t) + (1 - \alpha)S'(t - 1)$$

(Recall that, under simple exponential smoothing, we would just let $\acute{Y}(t+1) = S'(t)$ at this point.) Then let S" denote the *doubly-smoothed* series obtained by applying simple exponential smoothing (using the same α) to series S':

$$S''(t) = \alpha S'(t) + (1 - \alpha)S''(t - 1)$$

Finally, the forecast $\hat{y}(t + 1)$ is given by:

$$\hat{y}(t + 1) = a(t) + b(t)$$

where:

$$a(t) = 2S'(t) - S''(t) \text{...the estimated } level \text{ at period t}$$

32

$b(t) = (\alpha /(1-\alpha))(S'(t) - S''(t))$ ...the estimated *trend* at period t.

Forecasts with longer lead times made at period t are obtained by adding multiples of the trend term. For example, the k-period-ahead forecast (i.e., the forecast for $y(t+k)$ made at period t) would be equal to a(t)+kb(t). For purposes of model-fitting (i.e., calculating forecasts, residuals, and residual statistics over the estimation period), the model can be started up by setting S'(1)=S''(1)= $y(1)$, i.e., set both smoothed series equal to the observed value at t=1.

A mathematically equivalent form of Brown's linear exponential smoothing model, which emphasizes its non-stationary character and is easier to implement on a spreadsheet, is the following:

$$\hat{y}(t) = 2y(t-1) - y(t-2) - 2(1-\alpha)\varepsilon(t-1) + ((1-\alpha)^2)\varepsilon(t-2)$$

or equivalently:

$$\hat{y}(t) - y(t-1) = y(t-1) - y(t-2) - 2(1-\alpha)\varepsilon(t-1) + ((1-\alpha)^2)\varepsilon(t-2)$$

In other words, the *predicted difference* at period t (namely $\hat{y}(t) - y(t-1)$) is equal to the previous observed difference (namely Y(t-1) - Y(t-2)) minus a weighted difference of the two previous forecast errors.

Caution: this form of the model is rather tricky to start up at the beginning of the estimation period. The following convention is recommended: first set $\hat{y}(1) = y(1)$, which yields $\varepsilon(1)$ = 0 (i.e., cheat a bit, and let the first forecast equal the actual first observation), then also set $\hat{y}(2) = y(1)$, which yields $\varepsilon(2) = \hat{y}(2) - y(1)$, then continue from this point using the equation above. This would yield the same fitted values as the formula based on S' and S'' if the latter were started up using S'(1) = S''(1) = y(1).

Once again, you can use your spreadsheet the "solver" or any nonlinear least squares algorithm to optimize the value of α. The optimal value of α in the LES model fitted to this series by Statgraphics is 0.1607. Note that the long-term forecasts of the LES model for

33

this time series appear to track the local trend observed in the last 10 periods. Also, the confidence intervals for the LES model extend faster than those of the SES model.

Brown linear exp. smoothing with alpha =0.1607

Figure 2.6. Brown LES model

2.5. Four common measures of forecast accuracy

When choosing between competing forecasting models or when evaluating an existing model, we need to use measures that summarize the overall accuracy provided by the model(s). Generally speaking, the closer the forecasts are to the actual values of the series, the more accurate the forecasting model is. Thus, the quality of a model can be evaluated by examining the series of forecast errors. All of the measures that follow are based on some simple function of the forecast errors.

The most commonly used measures of forecast accuracy are the mean absolute deviation (MAD), the mean square error (MSE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE). Formulas for the computation of these measures:

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right|$$

34

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(y_t - \hat{y}_t \right)^2}$$

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| (100\%)$$

The basic difference between the MAD and the MSE (or RMSE) is that by squaring the amount of the error, the MSE (and RMSE) penalizes extreme errors more heavily than does the MAD. Thus, the MAD is an appropriate measure of forecast accuracy when the costs of forecast error increase linearly with the size of his error. The MSE (and RMSE) is best if costs for large errors are disproportionately expensive.

Since the MAPE is measured as a percentage and is therefore "unitless", it is particularly useful for comparing the performance of the model on many different time series. One drawback to using the MAPE arises when a series has any extremely small terms since the division by those terms will tend to seriously inflate the MAPE. For this reason, it is also not wise to use the MAPE as a model selection tool.

When we use the MAD, MSE, or RMSE to select a good forecasting model, the time series data is usually split into two parts. The first part of the data is used to estimate the parameters of the particular model. Then with these estimates the model is used to forecast the remaining data points. The MAD, MSE, and RMSE are calculated from the forecast errors for this second part of the series.

2.6. Summary

This chapter presents the classification of forecasting models for three groups: naive models, econometric models, time series models, where time series models were considered more detailed. Also four common measures of forecast accuracy were considered.

CHAPTER 3

THE BOX-JENKINS FORECASTING PROCEDURE

3.1. Overview

ARIMA (Auto Regressive Integrated Moving Average) model was introduced by Box and Jenkins in 1976 includes three types of parameters: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q). In this chapter it will be explained structure and three primer stages in building a Box-Jenkins time series model.

3.2. Stochastic and Deterministic Dynamic Mathematical Models

The idea of using a mathematical model to describe the behavior of a physical phenomenon is well established. In particular, it is sometimes possible to derive a model based on physical laws, which enables us to calculate the value of some time-dependent quantity nearly exactly at any instant of time. Thus, we might calculate the trajectory of a missile launched in a known direction with known velocity. If exact calculation were possible, such a model would be entirely *deterministic*.

Probably no phenomenon is totally deterministic, however, because unknown factors can occur such as a variable wind velocity that can throw a missile slightly off course. In many problems we have to consider a time-dependent phenomenon, such as monthly sales of newsprint, in which there are many unknown factors and for which it is not possible to write a deterministic model that allows exact calculation of the future behavior of the phenomenon. Nevertheless, it may be possible to derive a model that can be used to calculate the *probability* of a future value lying between two specified limits. The models for time series that are needed, for example to achieve optimal forecasting and control are in fact stochastic models. It is necessary in what follows to distinguish between the probability

36

model or *stochastic process,* as it is sometimes called, and the observed time series. Thus a time series $z_1, z_2, ..., z_n$ of N successive observations is regarded as a sample realization from an infinite population of such time series that could have been generated by the stochastic process. Very often we shall omit the word stochastic from "stochastic process" and talk about the "process."

Stationary and Nonstationary Stochastic Models for Forecasting and Control

An important class of stochastic models for describing time series, which has received a great deal of attention, comprises what are called *stationary'* models, which assume that the process remains in *equilibrium* about a *constant mean level.* However, forecasting has been of particular importance in industry, business, and economics, where many time series are often better represented as nonstationary and, in particular, as having no natural constant mean level over time. It is not surprising, therefore, that many of the economic forecasting methods originally proposed by Holt, Winters, Brown, and the Id Monograph that used exponentially weighted moving averages can be shown to be appropriate for a particular type of *nonstationary* process. Although such methods are too narrow to deal efficiently with all time series, the fact that they give the right kind of forecast function supplies a clue to the *kind* of *nonstationary* model that might be useful in these problems.

The stochastic model for which the exponentially weighted moving average forecast yields minimum mean square error is a member of a class of *nonstationary* processes called autoregressive integrated moving average processes (ARIMA processes). This wider class of processes provides a range of models, stationary and nonstationary, that adequately represent many of the time series met in practice. Once an appropriate model has been determined for the series, the optimal forecasting procedure follows immediately. These forecasting procedures include the exponentially weighted moving average forecast as a special case.

Some simple operators:

It should be employed extensively the *shift operator B*, which is defined by $Bz_t = z_{t-1}$; hence $B^m z_t = z_{t-m}$. The inverse operation is performed by the *forward shift operator F* $= B^{-1}$ given by $Fz_t = z_{t+1}$; hence $F^m z_t = z_{t+m}$. Another important operator is the *backward difference operator* ∇, which can be written in terms of B, since

$$\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$$

Linear filter model

The stochastic models we employ are based on the idea that a time series in which successive values are highly dependent can frequently be regarded as generated from a series of *independent* "shocks" a_t. These shocks are *random* drawings from a fixed distribution, usually assumed Normal and having mean zero and variance σ_a^2. Such a sequence of random variables $a_t, a_{t-1}, a_{t-2} \ldots$ is called a *white noise* process.

The white noise process a_t is supposed transformed to the process z_t what is called a *linear filter*. The linear filtering operation simply takes a weighted sum of previous random shocks a_t, so that

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + .. = \mu + \psi(B)a_t$$

3.3. Box-Jenkins Approach

The Box-Jenkins' modeling procedure involves identifying an appropriate ARMA model, fitting it to the dataset, and using the fitted model for forecasting. One of the attractive features of the Box-Jenkins method to forecasting is that ARMA processes are a very rich class of possible models and it is usually possible to find a model which provides an adequate description for the data.

The Box-Jenkins' modeling procedure involves iterative steps of model identification, parameter estimation, model diagnosis and forecasting. A modified procedure for searching

38

an adequate ARMA model. The modified procedure mainly involves four steps: identification, pre-estimation / estimation, diagnosis and forecast. Each of these steps will be discussed and illustrated in the sections that follow.

Figure 3.1. Block scheme of model construction

In contrast to other techniques, Box-Jenkins is a procedure, which uses a variable's past behavior to select the best forecasting model from a general class of models. It assumes that any time series pattern can be represented by one of three categories of models. These categories include:

• Autoregressive models: forecasts of a variable based on linear function of its past values

• Moving Average models: forecasts based on linear combination of past errors

• Autoregressive-Moving Average models: combination of the previous two categories

Note that one of the key questions is how many past values (the focal variable and/or its errors) should be included in the model.

Generally speaking, there are essentially three stages to a Box-Jenkins procedure:

1. Identifying the tentative model. Which of the three categories listed above is identified as the appropriate category is determined by first making the data stationary (usually by differencing the data) and then analyzing the autocorrelations and partial autocorrelations of the stationary data. Note that there are theoretical autocorrelation and partial autocorrelation profiles for each of the possible models. Therefore, determining the appropriate type of model for a specific situation is mainly a matter of matching the observed correlations to the theoretical correlations.

2. Determining the parameters of the model. This is similar to estimating the parameters in regression analysis.

3. Application of the model.

The Box-Jenkins ARMA model is a combination of the AR and MA models :

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}$$

where the terms in the equation have the same meaning as given for the AR and MA model.

Autoregressive models

A stochastic model that can be extremely useful in the representation of certain practically occurring series is the *autoregressive* model. In this model, the current value of the process is expressed as a finite, linear aggregate of *previous values of the process* and a shock. Let us denote the values of a process at equally spaced times t, $t - 1$, $t - 2$, ... by $z_t, z_{t-1}, z_{t-2} \ldots$ Also let $\tilde{z}_t$, $\tilde{z}_{t-1}$, $\tilde{z}_{t-2}, \ldots$ be deviations from μ; for example, $\tilde{z}_t = z_t - \mu$. Then

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t \tag{1}$$

is called an *autoregressive* (AR) *process of order p*. The reason for this name is that a

linear model

$$\tilde{z} = \phi_1 \tilde{x}_1 + \phi_2 \tilde{x}_2 + \ldots + \phi_p \tilde{x}_p + a$$

relating a "dependent" variable z to a set of "independent" variables $x_1, x_2, \ldots, x_p$, plus an error term a, is often referred to as a regression model, and z is said to be "regressed" on $x_1, x_2, \ldots, x_p$. In (1) the variable z is regressed on previous values of itself; the model is autoregressive. If an autoregressive operator of order p is defined by:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

the autoregressive model may be written economically as

$$\phi(B)\tilde{z}_t = a_t.$$

The model contains $p+2$ unknown parameters μ, $\phi_1, \phi_2, \ldots, \phi_p, \sigma_a^2$, which in practice have to be estimated from the data. The additional parameter σ_a^2 is the variance of the white noise process a_t.

Moving average models

The autoregressive model (1) expresses the deviation $\tilde{z}_t$ of the process as *a finite* weighted sum of p previous deviations $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \ldots, \tilde{z}_{t-p}$ of the process, plus a random shock a_t. Equivalently it expresses $\tilde{z}_t$ as infinite weighted sum of a's.

Another kind of model, of great practical importance in the representation of observed time series, is the finite *moving average* process. Here $\tilde{z}_t$ linearly dependent on a *finite* number q of previous a's. Thus

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

is called a *moving average* (MA) *process of order q*. The name "moving average" is somewhat misleading because the weights $1, -\theta_1, -\theta_2, \ldots, \theta_q$, which multiply the a's, need not total unity nor need they be positive. However, this nomenclature is in common use, and there fore we employ it.

If we define a *moving average operator* of order q by

$$\theta(B) = I - \theta_1 B - \theta_2 B^2 - \ldots - \theta q\, B^p$$

the moving average model may be written economically as

$$\tilde{z}_t = \theta(B)a_1$$

It contains q + 2 unknown parameters $\mu, \theta_1, \ldots, \theta_q, \sigma_a^2$, which in practice have to be estimated from the data.

Mixed autoregressive-moving average models

To achieve greater flexibility in fitting of actual time series, it is sometimes advantageous to include both autoregressive and moving average terms in the model. This leads to the mixed autoregressive-moving average model

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

or

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

which employs p + q + 2 unknown parameters $\mu, \phi_1, \ldots, \phi_p; \theta_1, \ldots, \theta_q; \sigma_a^2$, that are estimated from the data. This model may also be written in the form of the linear filter as $\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t$. In practice, it is frequently true that adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average, or mixed models, in which p and q are not greater than 2 and often less than 2.

Comments on Box-Jenkins Model

Box-Jenkins model characteristics are

1. The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for "Integrated".

2. Some formulations transform the series by subtracting the mean of the series from each data point. This yields a series with a mean of zero. Whether you need to do this or not is dependent on the software you use to estimate the model.

3. Box-Jenkins models can be extended to include seasonal autoregressive and seasonal moving average terms. Although this complicates the notation and mathematics of the model, the underlying concepts for seasonal autoregressive and seasonal moving average terms are similar to the non-seasonal autoregressive and moving average terms.

4. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms. As with modeling in general, however, only necessary terms should be included in the model.

There are three primary stages in building a Box-Jenkins time series model.

1. Model Identification
2. Model Estimation
3. Model Validation.

The following remarks regarding Box-Jenkins models should be noted.

1. Box-Jenkins models are quite flexible due to the inclusion of both autoregressive and moving average terms.

2. Based on the World decomposition theorem (not discussed in the Handbook), a stationary process can be approximated by an ARMA model. In practice, finding that approximation may not be easy.

3. Chatfield (1996) recommends decomposition methods for series in which the trend and seasonal components are dominant.

4. Building good ARIMA models generally requires more experience than commonly used statistical methods such as regression.

3.4. Box-Jenkins Model Identification

3.4.1. Stationarity

A common assumption in many time series techniques is that the data are stationary.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations (seasonality).

For practical purposes, stationarity can usually be determined from a run sequence plot.

The following plots are from a data set of monthly CO_2 concentrations

Example of Run Sequence Plot

Figure 3.2.

The initial run sequence plot of the data indicates a rising trend. A visual inspection of this plot indicates that a simple linear fit should be sufficient to remove this upward trend.

If the time series is not stationary, we can often transform it to stationarity with one of the following techniques.

1. We can difference the data. That is, given the series Z_t, we create the new series

$$Y_t = Z_t - Z_{t-1}$$

44

The differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.

2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

3. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

The above techniques are intended to generate series with constant location and scale. Although seasonality also violates stationarity, this is usually explicitly incorporated into the time series model.

3.4.2. Seasonality

Many time series display seasonality. By seasonality, we mean periodic fluctuations. For example, retail sales tend to peak for the Christmas season and then decline after the holidays. So time series of retail sales will typically show increasing sales from September through December and declining sales in January and February.

Seasonality is quite common in economic time series. It is less common in engineering and scientific data.

If seasonality is present, it must be incorporated into the time series model. In this section, we discuss techniques for detecting seasonality.

a. Detecting Seasonality

The following graphical techniques can be used to detect seasonality.

1. A run sequence plot will often show seasonality.
2. A seasonal subseries plot is a specialized technique for showing seasonality.
3. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.
4. The autocorrelation plot can help identify seasonality.

The run sequence plot is a recommended first step for analyzing any time series. Although seasonality can sometimes be indicated with this plot, seasonality is shown more clearly by the seasonal subseries plot or the box plot. The seasonal subseries plot does an excellent job of showing both the seasonal differences (between group patterns) and also the within-group patterns. The box plot shows the seasonal difference (between group patterns) quite well, but it does not show within group patterns. However, for large data sets, the box plot is usually easier to read than the seasonal subseries plot.

Both the seasonal subseries plot and the box plot assume that the seasonal periods are known. In most cases, the analyst will in fact know this. For example, for monthly data, the period is 12 since there are 12 months in a year. However, if the period is not known, the autocorrelation plot can help. If there is significant seasonality, the autocorrelation plot should show spikes at lags equal to the period.

The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled.

b. Differencing to achieve stationarity

Box and Jenkins recommend the differencing approach to achieve stationarity. However, fitting a curve and subtracting the fitted values from the original data can also be used in the context of Box-Jenkins models.

c.Seasonal differencing

At the model identification stage, our goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. For example,

for monthly data we would typically include either a seasonal AR 12 term or a seasonal MA 12 term. For Box-Jenkins models, we do not explicitly remove seasonality before fitting the model. Instead, we include the order of the seasonal terms in the model specification to the ARIMA estimation software. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model idenfitication of the non-seasonal component of the model. In some cases, the seasonal differencing may remove most or all of the seasonality effect.

d. Identify p and q

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e., the *p* and *q*) of the autoregressive and moving average terms.

The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known.

Order of Autoregressive Process (p)

Specifically, for an AR(1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components.

For higher-order autoregressive processes, the sample autocorrelation needs to be supplemented with a partial autocorrelation plot. The partial autocorrelation of an AR(*p*) process becomes zero at lag *p*+1 and greater, so we examine the sample partial autocorrelation function to see if there is evidence of a departure from zero. This is usually determined by placing a 95% confidence interval on the sample partial autocorrelation plot (most software programs that generate sample autocorrelation plots will also plot this confidence interval). If the software program does not generate the confidence band, it is approximately $\pm 2 / \sqrt{N}$, with N denoting the sample size.

47

The autocorrelation function of a MA(q) process becomes zero at lag $q+1$ and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero. We do this by placing the 95% confidence interval for the sample autocorrelation function on the sample autocorrelation plot. Most software that can generate the autocorrelation plot can also generate this confidence interval.

The autocorrelation function

The dependence structure of a stationary time series is characterized by the autocorrelation function. The autocorrelation function is defined as the correlation between z_t and z_{t+k}; ρ_k =correlation (z_t, z_{t+k}), k is called the time lag. The autocorrelation function is estimated by the empirical autocorrelation function: $r_k = c_k / c_0$, where k=0, 1, 2, ... ; c_k are the empirical autocovariances.

The empirical autocorrelation- and cross-correlation functions are the main tools for the identification of the ARIMA model. The autocorrelation function of the basic processes has a typical shape. The autocorrelation function of the AR(1) process decays exponentially. The autocorrelation function of the MA(1) process has only $\rho 1$ nonzero and of the MA(2) process only $\rho 1$ and $\rho 2$ nonzero.

Autocorrelation plots are a commonly-used tool for checking randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at varying time lags. If random, such autocorrelations should be near zero for any and all time-lag separations. If non-random, then one or more of the autocorrelations will be significantly non-zero.

In addition, autocorrelation plots are used in the model identification stage for Box-Jenkins autoregressive, moving average time series models.

Autocorrelation plots are formed by

- Vertical axis: Autocorrelation coefficient

  $$r_n = c_n / c_0$$

  where c_n is the autocovariance function

  $$c_n = (1/N)\sum_{T=1}^{N-n}(y_t - \bar{y})(y_{t+h} - \bar{y})$$

  and c_0 is the variance function

  $$c_0 = \frac{\sum_{t=1}^{N}(y_t - \bar{y})^2}{N}$$

  Note r_n is between -1 and +1.

  - Horizontal axis: Time lag n (n = 1, 2, 3, ...)

Shape of Autocorrelation Function

The following table summarizes how we use the sample autocorrelation function for model identification.

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting	Mixed autoregressive and

after a few lags	moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

Examples of the autocorrelation plot:

Autocorrelation Plot for Strong Autocorrelation:

Figure 3.3.

We can make the following conclusions from the above plot.

1. The data come from an underlying autoregressive model with strong positive autocorrelation.

The plot starts with a high autocorrelation at lag 1 (only slightly less than 1) that slowly declines. It continues decreasing until it becomes negative and starts showing an incresing negative autocorrelation. The decreasing autocorrelation is generally linear with little noise. Such a pattern is the autocorrelation plot signature of "strong autocorrelation", which in turn provides high predictability if modeled properly.

The next step would be to estimate the parameters for the autoregressive model:

50

$$y_t = A_0 + A_1 y_{t-1} + E_t$$

Such estimation can be performed by using least squares linear regression or by fitting a Box-Jenkins autoregressive (AR) model.

Partial Autocorrelation Plot:

Partial autocorrelation plots are commonly used tool for model identification in Box-Jenkins models.

The partial autocorrelation at lag k is the autocorrelation between y_t and y_{t-k} that is not accounted for by lags 1 through k-1.

There are algorithms for computing the partial autocorrelation based on the sample autocorrelations.

Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR(p) process is zero at lag $p+1$ and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order. We look for the point on the plot where the partial autocorrelations essentially become zero. Placing a 95% confidence interval for statistical significance is helpful for this purpose.

The approximate 95% confidence interval for the partial autocorrelations are at $\pm 2/\sqrt{N}$

3.5. Box-Jenkins Model Estimation

Estimating the parameters for the Box-Jenkins models is a quite complicated nonlinear estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models. Fortunately, many commercial statistical software programs now fit Box-Jenkins models.

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation.

The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation.

The identification process having led to a tentative formulation for the model, we then need to obtain efficient estimates *of* the parameters. After the parameters have been estimated, the fitted model will be subjected to diagnostic checks and tests of goodness of fit. As pointed out by R. A. Fisher, for tests of goodness of fit to be relevant, it is necessary that efficient use of the data should have been made in the fitting process. If this is not so, inadequacy of fit may simply arise because of the inefficient fitting and not because the form *of* the model is inadequate.

Likelihood Function

Suppose that we have a sample *of* N observations z with which we associate an N-dimensional random variable, whose known probability distribution depends on some unknown parameters ξ. We use the vector ξ to denote a general set *of* parameters and, in particular, it could refer to the $p + q + 1$ parameters $(\varphi, \theta, \sigma)$ of the ARIMA model.

Before the data are available, $p(z/\xi)$ will associate a density with each different outcome z of the experiment, for fixed ξ. After the data have come to hand, we are led to contemplate the various values of ξ that might have given rise to the 'fixed set *of* observations z actually

obtained. The appropriate function for this purpose is the *likelihood function* $L(\xi /z)$, which is *of* the same form as $p(z/ \xi)$, but in which z is now fixed but ξ is variable. It is only the relative value *of* $L(\xi /z)$ which is *of* interest, so that the likelihood function is usually regarded as containing an *arbitrary multiplicative constant.*

It is *often* convenient to work with the log-likelihood function $ln[L(\xi /z)] = l(\xi /z)$, which contains an *arbitrary additive constant.* One reason that the likelihood function is fundamental importance in estimation theory is because of the "likelihood principle'. This principle says that (given that the assumed model is correct) all that the *data* have to tell us about the parameters is contained in the likelihood function, all other aspects *of* the data being irrelevant. From a Bayesian point of view, the likelihood function is equally important, since it is the component in the posterior distribution of the parameters which comes from the data.

For a complete understanding of the estimation situation, it is necessary to make a through analytical and graphical study of the likelihood function, or in the Bayesian framework, the posterior distribution of the parameters, which in the situations we consider, is dominated by the likelihood. In many examples, for moderate and large samples, the log-likelihood function will be unimodal and can be approximated adequately over a sufficiently extensive region near the maximum by a quadratic function.

In such cases the log-likelihood function can be described by its maximum and its second derivatives at the maximum. The values of the parameters that maximize the likelihood function, or equivalently the log-likelihood function, are called *maximum likelihood (ML) estimates.* The second derivatives of the log-likelihood function provide measures of "spread" of the likelihood function and can be used to calculate approximate standard errors for the estimates.

The limiting properties of maximum likelihood estimates are usually established for independent observations. But as was shown by Whittle, they may be extended to cover stationary time series.

3.6. Box-Jenkins Model Diagnostics

Model diagnostics for Box-Jenkins models is similar to model validation for non-linear least squares fitting.

That is, the error term A_t is assumed to follow the assumptions for a stationary univariate process. The residuals should be white noise (or independent when their distributions are normal) drawings from a fixed distribution with a constant mean and variance. If the Box-Jenkins model is a good model for the data, the residuals should satisfy these assumptions.

If these assumptions are not satisfied, we need to fit a more appropriate model. That is, we go back to the model identification step and try to develop a better model. Hopefully the analysis of the residuals can provide some clues as to a more appropriate model.

3.7. Examples of Forecast Functions and Their Updating

3.7.1. Forecasting a General IMA(0,d,q) Process

As an example, consider the process of order $(0, 1, 3)$,

$$(1 - B)z_{t+1} = \left(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3\right)a_{t+1}$$

Taking conditional expectations at time we obtain

$$\hat{z}_t - z_t = -\theta_1 a_t - \theta_2 a_{t-1} - \theta_3 a_{t-2}$$

$$\hat{z}_t(2) - \hat{z}_t(1) = -\theta_2 a_t - \theta_3 a_{t-1}$$

$$\hat{z}_t(3) - \hat{z}_t(2) = -\theta_3 a_t$$

$$\hat{z}_t(l) - \hat{z}_t(l-1) = 0 \quad l = 4,5,6,\ldots$$

Hence $\hat{z}_t(l) = \hat{z}_t(3) = b_0^{(t)}$ for all $l > 2$, as expected, since q-p-d=2.

3.7.2. Forecasting Autoregressive Process

Consider a process of order $(p, d, 0)$

$$\phi(B)z_t = a_t$$

54

The eventual forecast function is the solution of $\phi(B)\hat{z}_t(l) = 0$. It applies for all lead times and passes through the last $p + d$ available values of the series.

The best forecast for all future time is very nearly the current value of the stock. The weight function for $\hat{z}_t(l)$ is a spike at time t and there is no averaging over past history.

Nonstationary autoregressive models of order (p, d, 0)

For the model

$$\phi(B)\nabla^d z_t = a_t$$

it will be the dth difference of the process that decays back to its mean when projected several steps ahead. The mean of $\nabla^d z_t$ will usually be assumed to be zero unless contrary evidence is available. When needed, it is possible to introduce a nonzero mean by replacing $\nabla^d z_t$ by $\nabla^d z_t - \mu_\omega$ in the model. For example, consider the model

$$(1 - \phi B)(\nabla z_t - \mu_\omega) = a_t$$

After substituting $t+j$ for t and taking conditional expectations at origin t,

$$\hat{z}_t(j) - \hat{z}_t(j-1) - \mu_\omega = \phi^j(z_t - z_{t-1} - \mu_\omega)$$

or $\hat{\omega}_t(j) - \mu_\omega = \phi^j(\omega_t - \mu_\omega)$, where $\omega_t = \nabla z_t$, which shows how the forecasted difference decays exponentially from the initial value $\omega_t = z_t - z_{t-1}$ to its mean value μ_ω. On summing this expression from j = 1 to j = l, that is, using $\hat{z}_t(l) = \hat{\omega}_t(l) + ... + \hat{\omega}_t(1) + z_t$, we obtain the forecast function

$$\hat{z}_t(l) = z_t + \mu_w l + (z_t - z_{t-1} - \mu_w)\frac{\phi(1 - \phi^l)}{1 - \phi}, \quad l \geq 1$$

3.8. Summary

This chapter presents the detailed description of the Box-Jenkins forecasting procedure. The explanation of three primary stages in building a Box-Jenkins is given: model identification, model estimation, model diagnostic.

CHAPTER 4

IMPLEMENTATION OF FORECASTING METHODS

4.1. Overview

At the present chapter two examples of implementation of forecasting methods will be considered. One of them is prediction of world petroleum production and second is prediction of the number of Computer Engineering Department students. For this purpose program, which was written in Delphi, is applied. Also all the forecasting methods, which were used in program, are have done in Excel with purpose to check results.

4.2. World Petroleum Production

Central European Bank gives time series for World petroleum production for 8 years from 1995 to 2002 monthly. The original time series are shown in the Figure 4.1, where x axis is a time, and y axis is Petroleum production monthly in Mb/d.

Series title: Petroleum Production, World. The part of those data is given below, whole set is given in Table 1 in Appendix.

Year	Month	Value,(Mb/d)
2002	1	66.391,57
2002	2	66.523,81
2002	3	66.338,93
2002	4	65.780,53
2002	5	66.184,34
2002	6	66.058,17
2002	7	66.654,99
2002	8	66.362,88
2002	9	67.111,00
2002	10	68.769,01
2002	11	68.764,28
2002	12	67.008,72

Plot of actual values for World Petroleum Production

Figure 4.1.

4.2.1. Forecast for 2002

Three kind of forecasting methods are applied to present example: Moving average forecasting model, an exponential-smoothing forecasting model, Box-Jenkins forecasting procedure. Here appeared problem, which model will fit better to the present time series, the quality of model can be evaluated by examining the forecast errors. As far as actual data for 2003 aren't known, forecast errors can't be calculated, so at first, forecasting procedures are applied for 2002 year.

4.2.1.a. Moving Average Forecasting Model

One-step-ahead moving average forecast is applied for forecasting petroleum production in 2002, let k be equal to 5.
The results of forecasting which were gotten in Excell are shown below.

57

Implementation of Moving Average Forecasting Model in Excel

Table 4.1.

Year	Month	Actual Value,(Mb/d)	One-Step Forecast	One-step absolute deviation
2001	1	68.940,12		
2001	2	68.478,21		
2001	3	69.165,97		
2001	4	68.268,40		
2001	5	67.576,67		
2001	6	66.004,01		
2001	7	67.978,71		
2001	8	68.165,35		
2001	9	67.792,01		
2001	10	67.680,40		
2001	11	67.929,27		
2001	12	67.473,76		
2002	1	66.391,57	67.808,16	1.416,59
2002	2	66.523,81	67.453,40	929,59
2002	3	66.338,93	67.199,76	860,83
2002	4	65.780,53	66.931,47	1.150,94
2002	5	66.184,34	66.501,72	317,38
2002	6	66.058,17	66.243,84	185,67
2002	7	66.654,99	66.177,16	477,83
2002	8	66.362,88	66.203,39	159,49
2002	9	67.111,00	66.208,18	902,82
2002	10	68.769,01	66.474,28	2.294,73
2002	11	68.764,28	66.991,21	1.773,07
2002	12	67.008,72	67.532,43	523,71
		1.617.401,11		10.992,65
Mean absolute deviation				916,05

To check the accuracy of the model in Figure 4.2 are given two graphs: actual values, which are known and values, which were gotten by using present method.

Plot of forecasts by Moving Average model for 2002 and actual values in Excel

Figure 4.2.

According to the same purpose program gives following result:

Plot of forecasts by Moving Average model for 2002 and actual values in Delphi

Figure 4.3.

Forecasts by Moving Average model for 2002 and actual values in Delphi

Table 4.2.

forecast	Actual values
67807	66391
67453	66523
67199	66338
66930	65780
66501	66184
66243	66654
66202	66362
66207	67111
66473	68769
66990	68764
67532	67008

4.2.1.b. An Exponential-Smoothing Forecasting Model

The results of applying Brown's method of multiple smoothing are present by following table and Figure 4.4:

Implementation of Exponential-Smoothing Forecasting Model in Excel

Table 4.3.

Year	Month	Actual Value, (Mb/d)	St	St(2)	St(3)	Forecast, T=1, a=0,2	Forecast Error
2002	1	66.391,57	66.391,57	66.391,57	66.391,57		
2002	2	66.523,81	66.418,02	66.396,86	66.392,63	66.391,57	132,24
2002	3	66.338,93	66.402,20	66.397,93	66.393,69	66.470,91	131,98
2002	4	65.780,53	66.277,87	66.373,92	66.389,73	66.407,59	627,06
2002	5	66.184,34	66.259,16	66.350,96	66.381,98	66.032,44	151,90
2002	6	66.058,17	66.218,96	66.324,56	66.370,50	66.049,42	8,75
2002	7	66.654,99	66.306,17	66.320,89	66.360,57	65.993,73	661,26
2002	8	66.362,88	66.317,51	66.320,21	66.352,50	66.326,79	36,09
2002	9	67.111,00	66.476,21	66.351,41	66.352,28	66.360,37	750,63
2002	10	68.769,01	66.934,77	66.468,08	66.375,44	66.828,57	1.940,44
2002	11	68.764,28	67.300,67	66.634,60	66.427,27	68.102,58	661,70
2002	12	67.008,72	67.242,28	66.756,14	66.493,05	68.850,05	1.841,33
		801.948,23					6.943,38
MAD							631,22

60

Plot of forecasts by Exponential-Smoothing model for 2002 and actual values in Excel

Figure 4.4.

Program gives following results:

Plot of forecasts by Exponential-Smoothing model for 2002 and actual values in Delphi
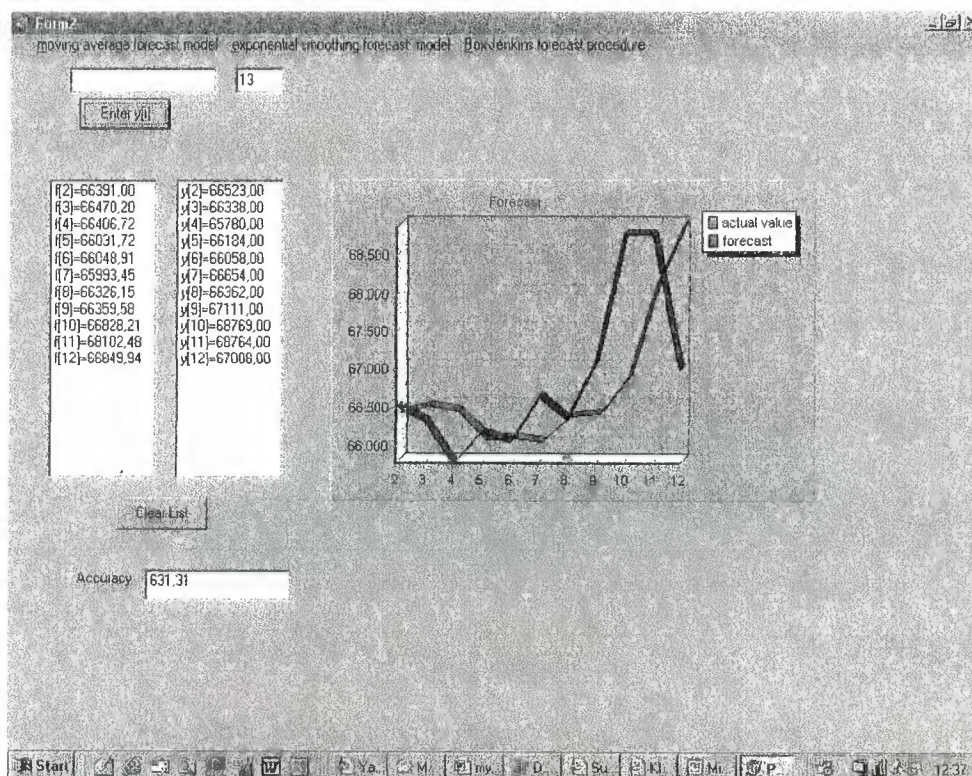
Figure 4.5.

Forecasts by Exponential-Smoothing model for 2002 and actual values in Delphi

Table 4.4.

Forecasts	Actual values
66391	66523
66470	66338
66406	65780
66031	66184
66048	66058
65993	66654
66326	66362
66359	67111
66828	68769
68102	68764
68849	67008

4.2.1.c.The Box-Jenkins Forecasting Procedure

As clear from Figure1. there is trend in present data. In order to choose the best ARIMA model for given set of data, first it should be removed trend by performing an appropriate amount of differencing. After first order differencing trend is removed. See Figure 4.6. The order of differencing required is denoted by d. So first-order differencing is needed, then $d=1$.

First order differencing plot
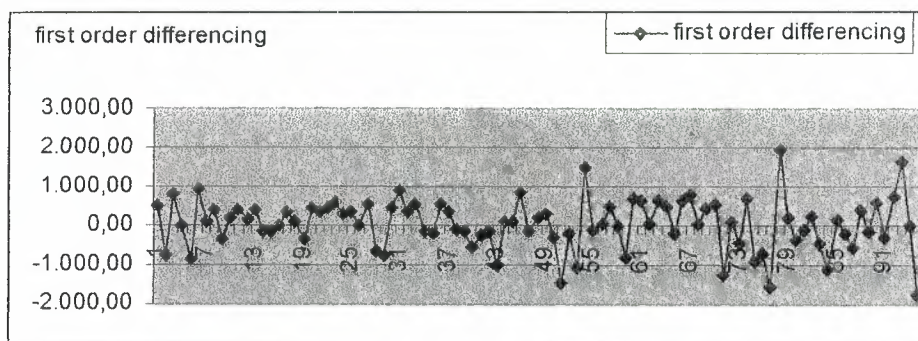
Figure 4.6.

After that, the appropriate number p of autoregressive terms and/or the number q of moving average terms must be selected. The resulting model is then referred to as an ARIMA(p,d,q) model.

One of primary tools used in determining the parameters p and q in an ARIMA model is the autocorrelation function (acf) of the time series. The acf summarize the various correlations that a time series has with itself.

The acf was calculated by using the package S-PLUS. Figure 4.7 shows the output of this data.

Autocorrelation plot
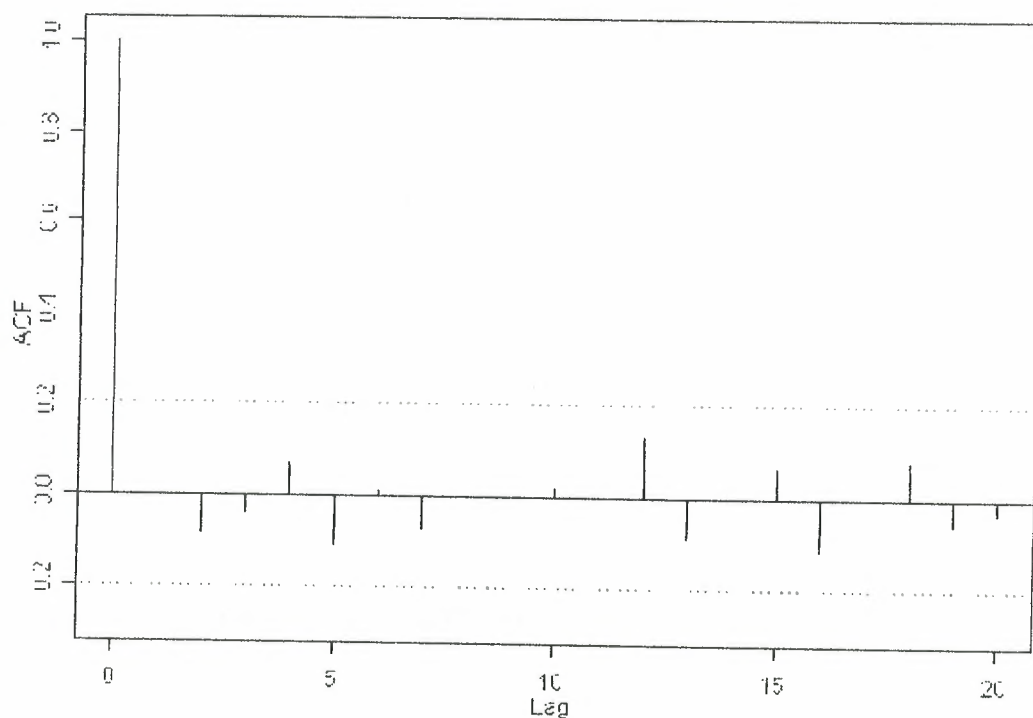
Series : Katia1$prod

Figure 4.7.

To interpret the acf, is needed the method for testing which of the r_k's are statistically different from zero.

Figure 4.8 presents the components of pacf.

Partial autocorrelation plot

Series : Katia1$prod
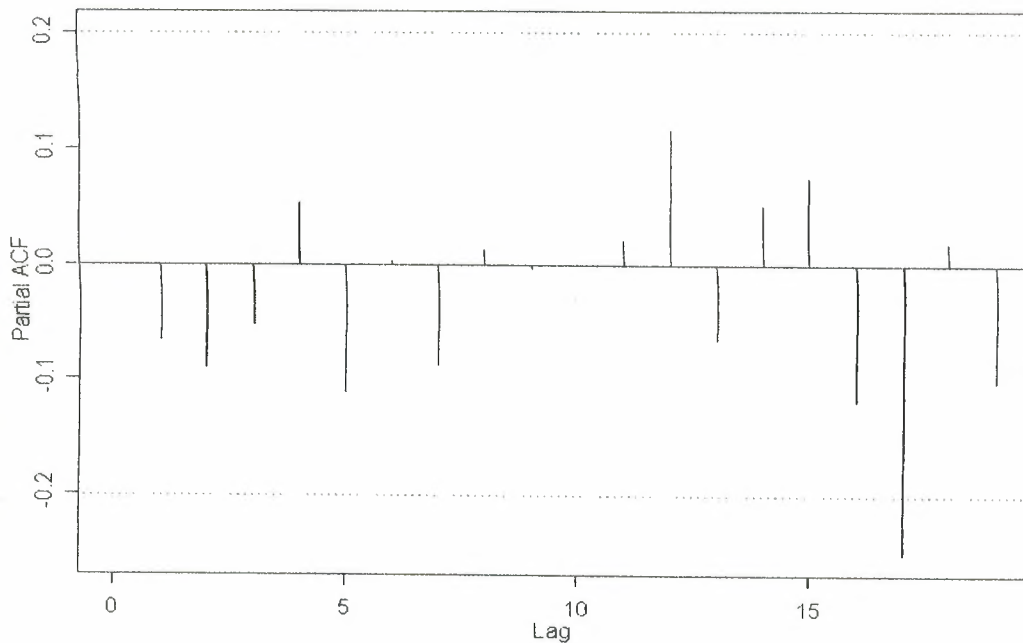
Figure 4.8.

As clear from the plot parameter $q=0$ is suitable for creating model, but also $q=1$ can be checked.

Test concerning autocorrelation coefficients

1. Null hypothesis: $H_0 : \rho_k = 0$.

2. Altenative hypothesis: $H_a : \rho_k \neq 0$, where ρ_k is the population autocorrelation coefficient of lag k.

3. Test statistic: r_k, the sample autocorrelation coefficient of lag k.

4.Rejection region: At an approximate significance level of $\alpha \approx 0,05$,

reject H_0 if $| r_k | > 2/\sqrt{n}$.

In this case n=96, $2/\sqrt{96}=0,2$, so $| r_k | > 0,2$, (-0,2; 0,2) is interval of confidence.

To check which model fits better present data package S-PLUS was used, result are presented by following Table 4.5.

Results of Akaike Diagnostic Test

Table 4.5.

Model	AIC
AR(1)	1484,57
MA(1)	1499,93
ARMA(1,1)	1485,92

Where AIC is Akaike information criterion, which shows fitness of model.

$AIC=nln(S/n)+2p$, where S is the sum of squared residuals up to time T(the residuals are the within-sample one-step-ahead forecast errors), n is number of observations, p denotes the number of parameters fitted in the model. As lower the value of AIC, as better model fits to the data. AIC for AR(1) is lowest value, so AR(1) suits better than other models. The results of forecasting by using program in Delphi are shown below:

Plot of forecasts by ARIMA(1,0,0) in Delphi

Figure 4.9.

65
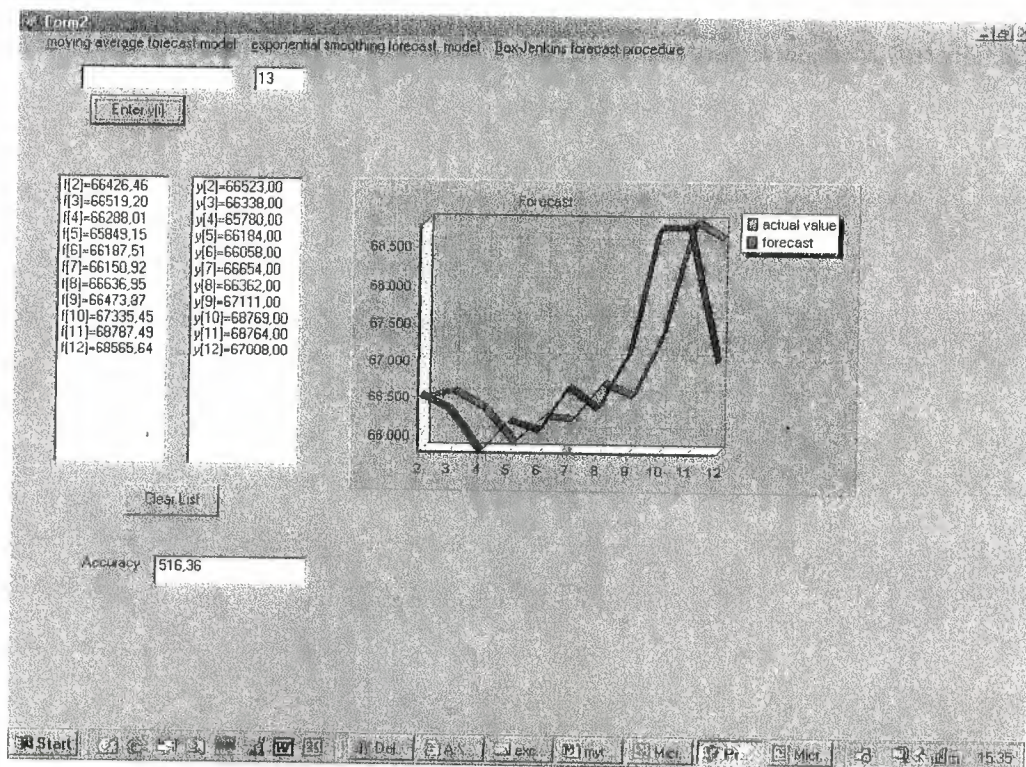
Forecasts by ARIMA(1,0,0) in Delphi

Table 4.6.

Forecasts	Actual values
66426	66523
66519	66338
66288	65780
65849	66184
66187	66058
66150	66654
66636	66362
66473	67111
67335	68769
68787	68764
68565	67008

4.2.2. Forecast for 2003

To make a forecasting for 2003 year, the best model should be chosen, it is a model, which is not very complicated and at the same time gives least error. By comparing results, which were gotten before, forecast accuracy of three applied models is equal to values presented by Table 4.7.

Table for errors of forecasting

Table 4.7.

Model	MAD
Moving Average Forecasting Model	916,12
An Exponential-Smoothing Forecasting Model	631,31
AR(1)	516,36

From this results follows that AR(1) model fits better then other applied models. So this model will be applied for forecasting 2003.

66

Plot of forecasted values by ARIMA(1,0,0) in Delphi
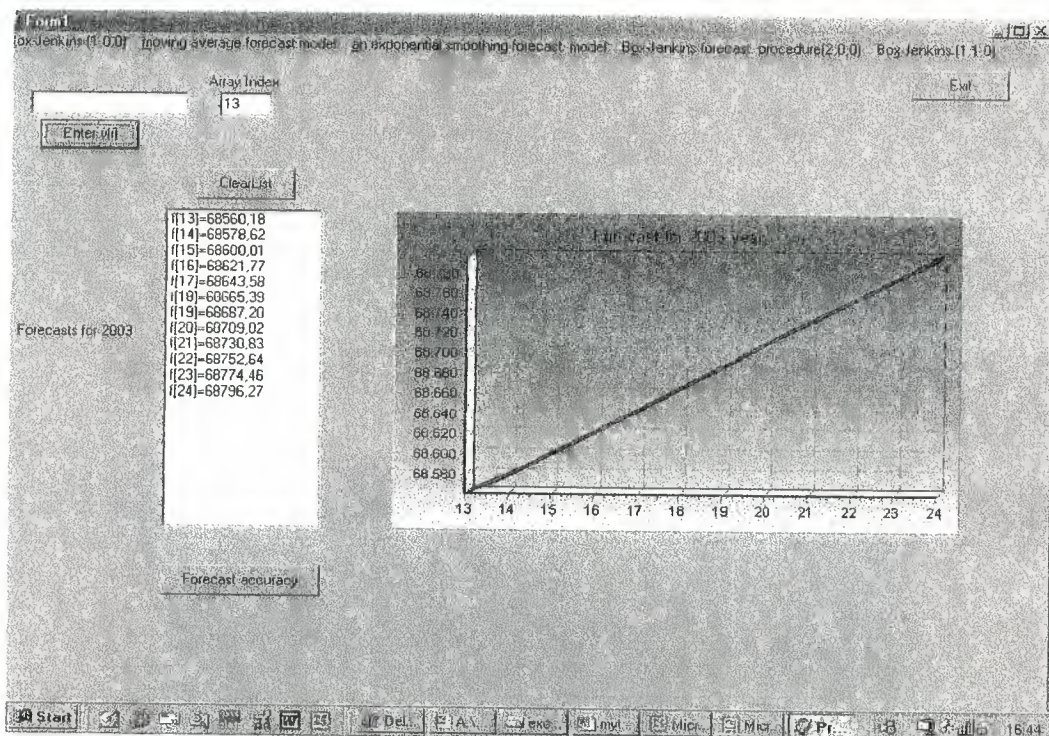
Figure 4.10.

Forecasted values by ARIMA(1,0,0) in Delphi
Table 4.8.

68560
68579
68600
68622
68644
68665
68687
68709
68731
68753
68774
68796

4.3. The Number of Engineering Department's students

Total number of students at the Faculty of Engineering according to academic years for period from 1994 to 2004 is represented by following table and Figure 4.11.

Academic Year	Number of students
1994-1995	1203
1995-1996	835
1996-1997	835
1997-1998	788
1998-1999	896
1999-2000	951
2000-2001	1230
2001-2002	1288
2002-2003	1331
2003-2004	1450

Plot for data of Engineering Department students number
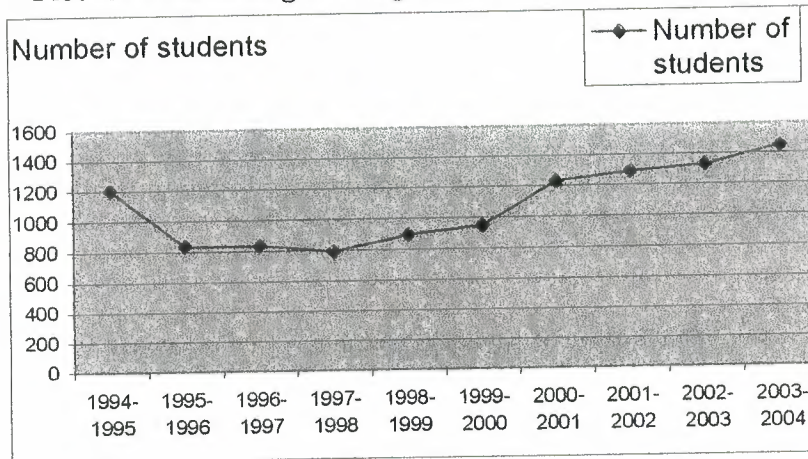
Figure 4.11.

4.3.1. Forecast for academic year 2004-2005 by Exponential-Smoothing Model

As clear from the Table 4.9. the number of Engineering Department students in academic year 2004-2005 is equal to 1453,79 with forecasting error $\varepsilon= 209,785$.

Implementation of forecasting by Exponential-Smoothing method in Excell

Table 4.9.

Year	Number of students	St	St(2)	St(3)	Forecast, T=1, a=0,2	Forecast Error
1994-1995	1203	1203	1203	1203		
1995-1996	835	1129,4	1188,28	1200,056	1203	368
1996-1997	835	1070,52	1164,728	1192,99	982,2	147,2
1997-1998	788	1014,016	1134,586	1181,309	849,72	61,72
1998-1999	896	990,4128	1105,751	1166,198	747,92	148,08
1999-2000	951	982,5302	1081,107	1149,18	760,472	190,528
2000-2001	1230	1032,024	1071,29	1133,602	811,647	418,353
2001-2002	1288	1083,219	1073,676	1121,617	1018,95	269,0503
2002-2003	1331	1132,775	1085,496	1114,392	1184,967	146,0334
2003-2004	1450	1196,22	1107,641	1113,042	1310,9	139,1002
forcec. 2004-2005					1453,79	
MAD						209,785

4.3.2. Forecast for academic year 2004-2005 by using model AR(1)

Implementation of forecasting with ARIMA(1,0,0) in Delphi

Figure 4.12.

69
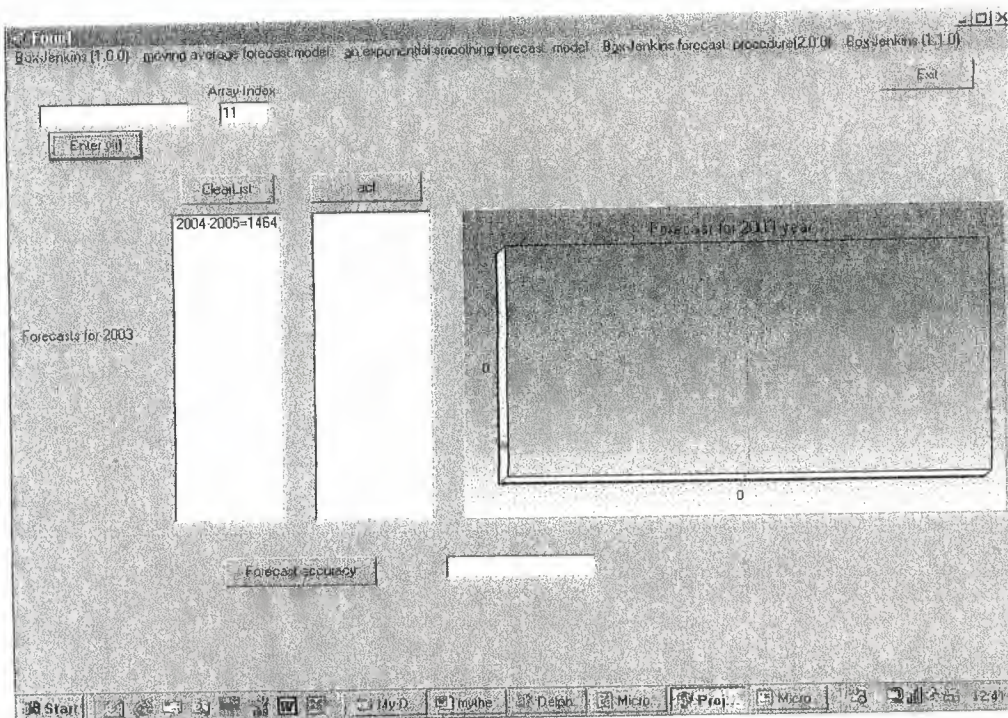
Program gives forecast for academic year 2004-2005, which is equal to 1464 students.

4.4. Summary

Two examples of implementation of forecasting methods are considered. For the first one which is world petroleum production forecasting, three forecasting methods were used, they are Moving Average method, Exponential-Smoothing forecasting method and ARIMA model. The implementation was done by two ways in Excel and with Delphi program help. Also S-PLUS applied package was used for model diagnostic, like a result ARIMA(1,1,0) model was chosen for forecasting.

The second example is prediction of Engineering Department students number. Forecasting was implemented by using Exponential-Smoothing model and AR(1) model.

CHAPTER 5

NETWORK TRAFFIC FORECASTING

5.1. Overview

In planning for the future needs of any complicated system, accurate forecasting of the workload is important to access future capacity requirements, and to plan for changes.

A detailed forecast of traffic patterns allows for more accurate planning and better decisions. If hardware changes are required at some point in a coming year, the model could estimate when traffic levels are likely to be lowest. If a particular backbone link is approaching maximum capacity, a detailed model of the traffic patterns on that link could predict when the capacity is likely to be exceeded. The ability to make forecasts two or more years in advance has great advantages for planning for future requirements.

One of the key issues in measurement-based network control is to predict the bandwidth requirement in a next control time interval based on the online measurement of traffic characteristics. Therefore, traffic prediction plays an important role in network control. The goal is to forecast future traffic variations as precisely as possible, based on the measured traffic history. Traffic predictability denotes the possibility for prediction to satisfy some precision requirement over desired prediction control time interval. On one hand, a large prediction interval is needed to provide sufficient time for control actions, and to offset the inevitable delays caused by traffic measurement (sampling/ smoothing) and traffic prediction (modeling/computing). On the other hand, a small prediction error is desirable for the following reason: Control actions based on erroneous prediction may inadvertently compromise the control performance. To overcome this problem, one must usually be conservative and set aside network resources, specifically the bandwidth resource in this paper. Thus, in order to achieve high resource utilization, a network controller would prefer precise prediction. Unfortunately, prediction accuracy deteriorates quickly as the prediction interval increases.

There are many factors affecting the amount of traffic, most of which cannot be measured or identified. To predict probable future traffic, the best available basis is an analysis of previously observed traffic patterns. The first requirement for an adequate model is that the

71

odel must be a time series model; also it should be model which can accept non-ationary data. A time series model that fits these criteria is the autoregressive integrated moving average process (ARIMA).

.2. Data

he data used in this paper was taken from Near East University's computer center and onsists of daily packet totals, between January 13, 2004 and Mart 8, 2004. The data is resented by Table 2, represented in Appendix and Figure 5.1. And the part of this table is hown below:

Date	Traffic IN (bytes)	Traffic OUT (bytes)	Traffic (bytes)
08.03.2004	43.977.484	115.642.494	159.619.978
07.03.2004	66.971.974	150.424.195	217.396.169
06.03.2004	71.250.793	218.398.210	289.649.003
05.03.2004	87.176.803	303.743.898	390.920.701
04.03.2004	94.620.977	247.238.424	341.859.401
03.03.2004	73.123.069	314.037.940	387.161.009
02.03.2004	591.336.560	306.554.361	897.890.921
01.03.2004	61.119.094	266.353.130	327.472.224

Figure 5.1.

5.3 . Missing Data

The data for February 22 is missed. There are two options for dealing with missing data: ignore them and omit those data points, or estimate the missing points. Because the ARIMA model examines the pattern of data over time, estimating the missing points and preserving the overall pattern was judged to be more desirable.

72

5.4. Model Identification

The purpose of the model identification step is to determine likely values for p, d and q. Once one or more promising sets of (p,d,q) have been identified, the model parameters for those orders can be estimated, diagnostic tests can be run, and the resulting model forecasts can be examined.

Differencing is required to make data stationary. Figure 5.2 shows the data without differencing and Figure 5.3 after first-order differencing.

Plot of actual data for Network Traffic
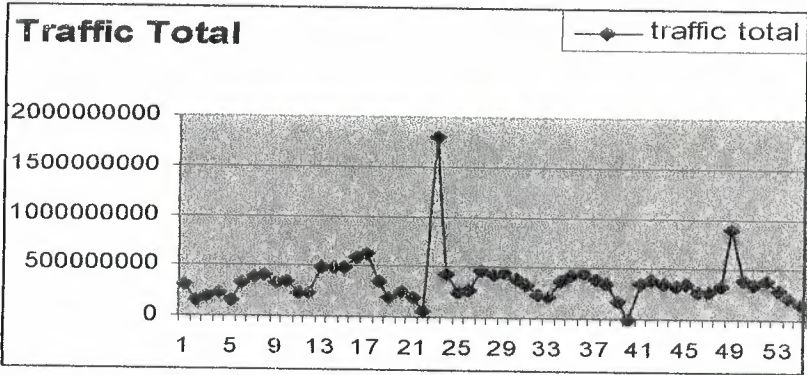
Figure 5.2.

Plot of Network Traffic data after first-order differencing
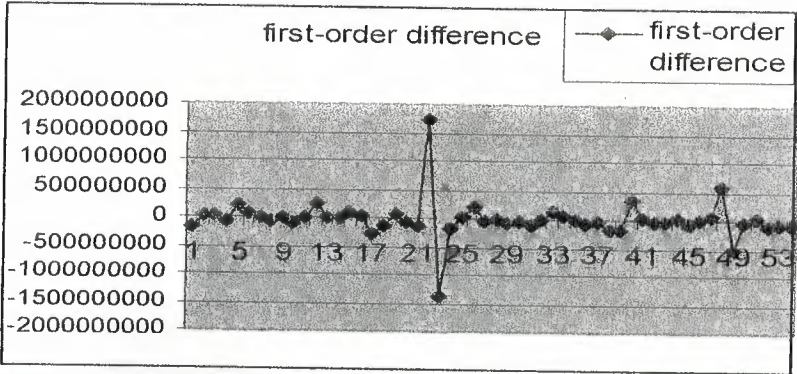
Figure 5.3.

After differencing of first order stationarity is reached, so $d=1$.

Once stationarity have been addressed, the next step is to identify the order (i.e., the p and q) of the autoregressive and moving average terms.

The acf and partial acf was calculated by using the package S-PLUS. Figure 5.4 and Figure 5.5 show the output of this data.

Autoregression is determined by examining the pattern of autocorrelations. If all values except the first are 0, p should be 0. Continually decreasing values indicate a p of 1, and a mixture of increasing and decreasing values mean a p of 2.

The interval of confidence is $|r_k|>2/\sqrt{n}$, or (-0,26;0,26).

Let's assume that null hypothesis is 'all the correlations are equal to zero', by the test concerning autocorrelations coefficients, if $|r_k|>2/\sqrt{n}$, then reject null hypothesis. All autocorrelations components accept first inside of confidence interval, so there are no correlations between actual data.

Autocorrelation plot

Series : kat$V1
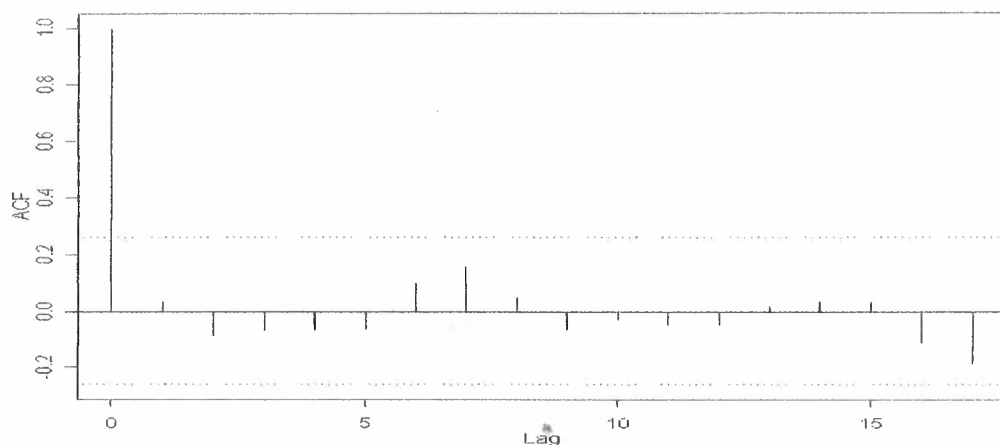
Figure 5.4.

The moving average is determined in a similar fashion, by examining the pattern of the partial autocorrelations, again after the data has been differenced. The meaning of the patterns is the same as for the autoregression.

Partial Autocorrelation plot

Series : kat$V1
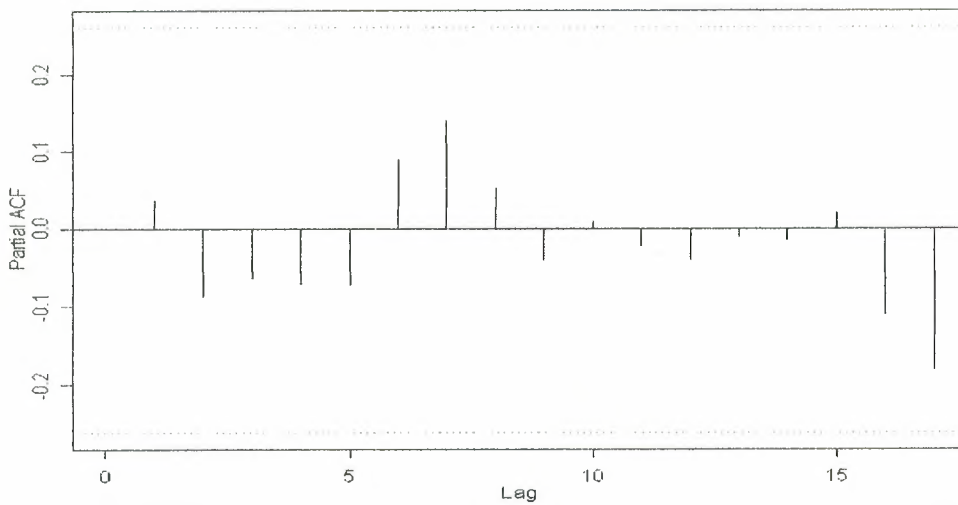
Figure 5.5.

As seen in the Figure 5.4, Figure 5.5 possibly indicating *p=1*, *q=0 or possibly q=1*.

5.5. The Model Parameters

Two models were chosen by the previous step ARIMA(1,1,0) and ARIMA(1,1,1).
By using the program in Delphi parameters of model ARIMA(1,1,0) and ARIMA(1,1,1)
also errors of forecasting were found. Following table shows estimated parameters and
error of forecasting.

Table of forecasting errors
Table 5.1.

Model	Estimated Parameters	MAD
ARIMA(1,1,0)	$\phi_1 = -0,385$	265.940.692
ARIMA(1,1,1)	$\phi_1 = -0,38, \theta_1 = -0,91$	340.966.206

5.6. Simulation Results

The available data is from 13th January through 8th March. To test the ARIMA model's
ability to predict across a week, the model was given only the data from February through

75

March, and was asked to forecast the Network traffic for last week from 1st March to 7th March, which is known.

This prediction was then compared to the actual data for that week.

First of all ARIMA(1,1,0) was checked. Forecasting for one week from 1st March to 7th March was implemented by using present model. Figure 5.6 shows the comparison between actual data and forecasted values. Where blue line indicates actual values of Network traffic totals at the present week, and accordingly red line is the forecast for the same week. The forecast calculated by Delphi program is presented by Table 5.2.

The mean absolute deviation of this model is equal to 265.940.692.

Forecasting with ARIMA(1,1,0) for last week from 1 March to 7 March in Delphi
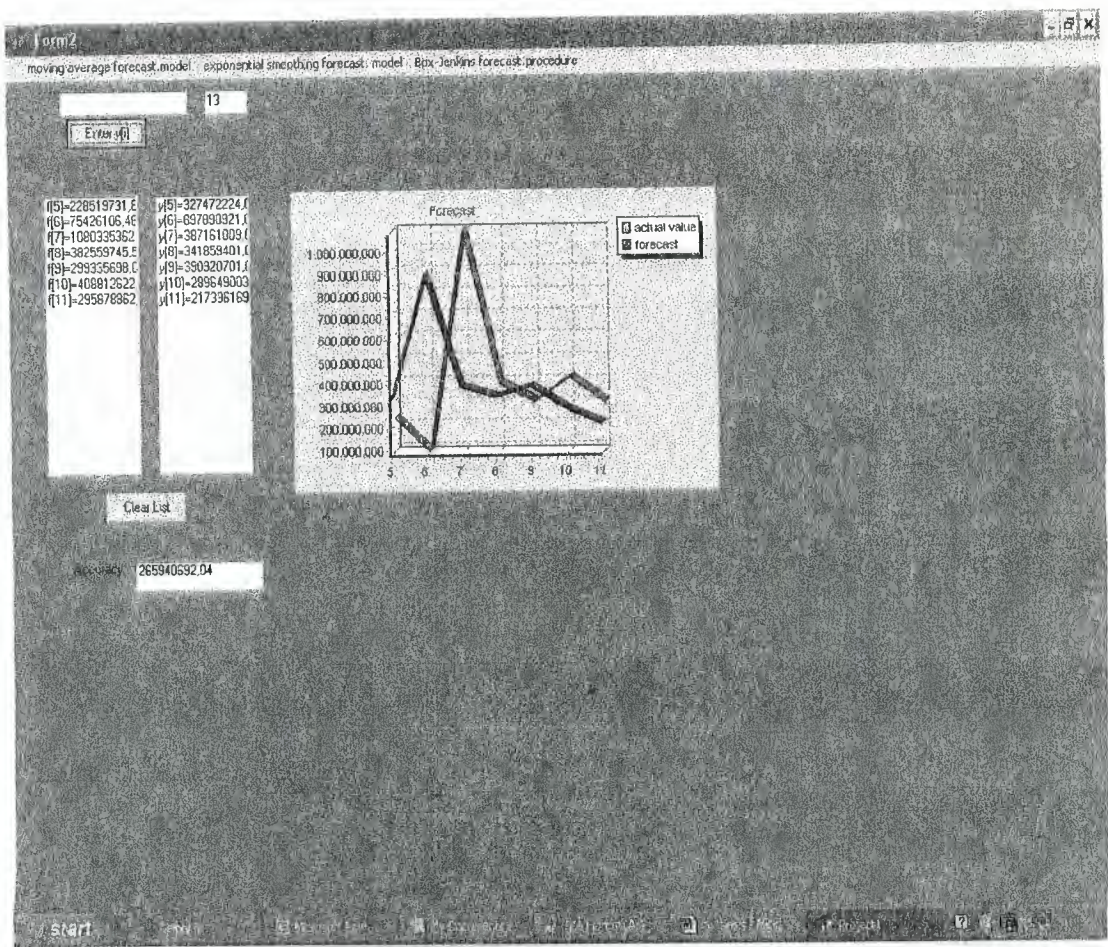
Figure 5.6.

Forecasting with ARIMA(1,1,0) for last week from 1 March to 7 March in Delphi

Table 5.2.

Forecasts	Actual values
228.519.731	327.472.224
75.426.106	897.890.921
1080.335.362	387.161.009
382.559.745	341.859.401
299.335.698	390.920.701
408.812.622	289.649.003
295.878.862	217.396.169

Second step is estimation of model ARIMA(1,1,1), results are represented by Figure 5.7. Like at the previous case program output gives comparison of actual data for a week and forecasted values for the same week. Accordantly there is blue line indicating actual data and red line indicating forecasts. The forecast is presented by the Table 5.3 and the accuracy is 340.966.206.

Forecasting with ARIMA(1,1,1) for last week from 1 March to 7 March in Delphi

Figure 5.7.

77

Table 5.3.

Forecasts	Actual values
281.278.882	327.472.224
196.536.344	897.890.921
1.832.290.325	387.161.009
1.026.132.798	341.859.401
359.368.666	390.920.701
512.153.341	289.649.003
424.257.941	217.396.169

Analysis of the Figures 5.6 and 5.7 can be seen that the ARIMA(1,1,0) model's prediction was quite accurate. So from the previous model estimation follows that ARIMA(1,1,0) fits more than ARIMA(1,1,1) to the considering process. So forecast for next week promise to be quite accurate too. The Network traffic forecast for one week from 9th March to 15th March was calculated by using ARIMA(1,0,0) model and presented by Figure 5.8.

One week prediction from 9 March to 15 March with ARIMA(1,1,0) in Delphi
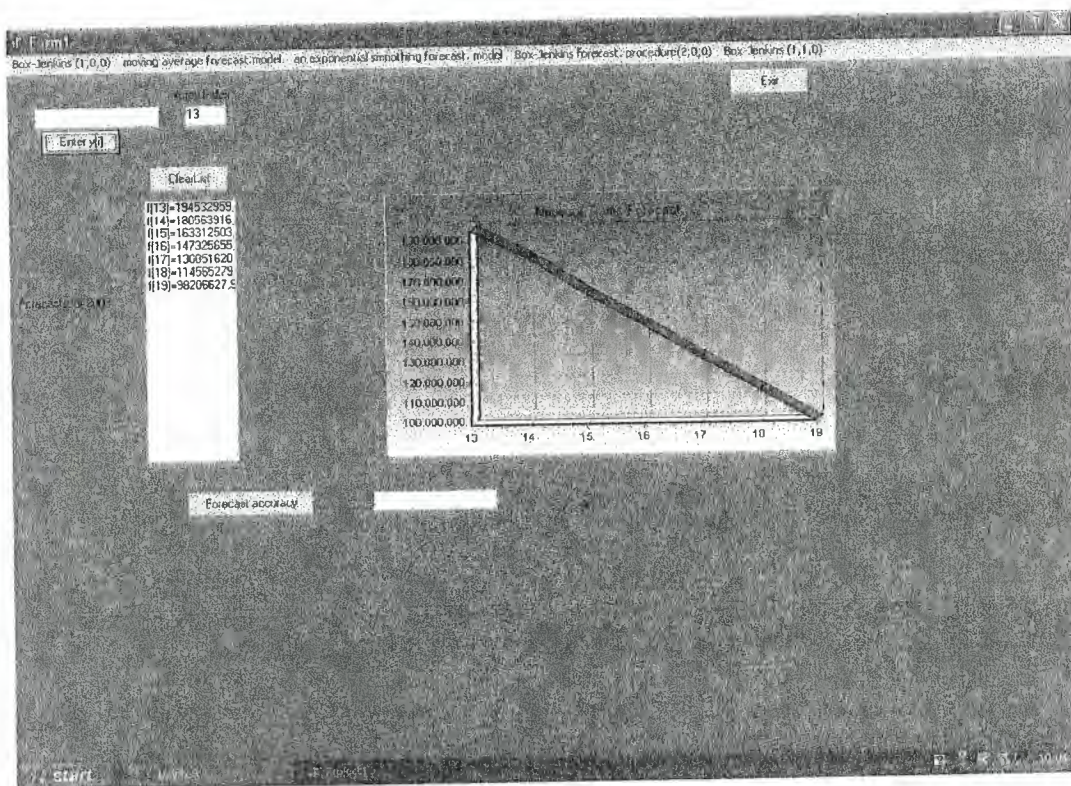
Figure 5.8.

78

The following Table 5.4 is the Delphi program's output, i.e. forecast for the week from 9 March to 15 March.

One week prediction from 9 March to 15 March with ARIMA(1,1,0) in Delphi
Table 5.4.

194.534.959
180.563.916
163.312.503
147.325.655
130.851.620
114.565.279
98.206.627

5.7. Summary

The purpose of the work was to examine the feasibility of using time series analysis to make detailed prediction about network traffic. The close match between predicted and observed traffic levels suggests that this approach can be used for forecast and planning with some confidence. The stages of building ARIMA model were considered in detail, and ARIMA(1,1,0) was chosen for forecasting, like a result one week network traffic prediction was gotten.

CONCLUSION

The analysis of application of different forecasting models to industrial and non-industrial systems has been presented. The results of analysis demonstrate that the forecasting models based on time-series analysis are more accurate models.

The description of naive models, econometric models and time-series models has been given. As an example Moving Average forecasting model, Exponential Smoothing forecasting model and Box-Jenkins forecasting procedure are represented. The applications of these forecasting models to world petroleum production, number of students of Engineering faculty and network traffic analysis are given. The statistical data for 8 years for world petroleum production forecasting are taken. In the result of application of Moving Average model, Exponential-Smoothing model and Box-Jenkins forecasting procedure it was clear that time series models gives more accurate results. The accuracy of ARIMA model was better than other models. In the result identification and estimation the structure and parameters values of ARIMA models has been found. The S plus package was used to select structure of ARIMA models. Different structures of ARIMA models for given problems have been tested. The ARIMA(1,1,0) model was chosen for world petroleum production forecasting to obtain more accurate results. Also in the result of investigation ARIMA(1,0,0) model was chosen for forecasting number of students of Engineering faculty.

The analysis of Network Traffic shows that it has high order nonlinearity. For prediction model statistical data for one week has been taken. First-order differencing is applied to avoid from the trend. The ARIMA model was chosen to make short-term forecast. The S plus package was used to select structure of ARIMA models. For these models the values of parameters (p,d,q) were identified. The analysis of autocorrelation function and partial autocorrelation function plots show that ARIMA(1,1,0) is most suitable model for given network traffic problem.

All forecasting model have been realized using Delphi programming. Simulation results of models have been gotten in numerical and graphical formats.

Results of this research work allow to learn theory of time series forecasting, practical part is necessary to realize the main ideas, and the creating of program for implementation of used forecasting methods makes this realizing more complete.

REFERENCES:

1. Box G.E.P., Jenkins G.M., Reinsel G.C., Time Series Analysis, Forecasting and Control, 1994, 3rd. Ed., Prentice-Hall

2. Mendenball W., Reinmuth J. E., Beaver R.J., Statistics for Management and Economics, 1993, 7th Ed., Wadsworth, Inc.

3. McClave J.T., Benson P. G., Sincich T., Statisticsfor Business and Economics, 1998, 7th Ed., Prentice-Hall, Inc.

4. Sanders D. H. Statistics a Fresh Approach, 1994, 4th Ed., McGraw Hill.

5. Konstantina Papagiannaki, Nina Taft, Zhi-Li Zhang, Christophe Diot, Long-Term Forecasting of Internet Backbone Traffic: Observation and Initial Models

6. Kavitha Chandra, Chun You, Gbenga Olowoyeye and Charles Thompson Non-Linear Time-Series Models of Ethernet Traffic

7. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Probability and Statistics for Engineers and Scientists, 1998, 6th Ed., Prentice-Hall

8. Akaike H., A New Look at the Statistical Model Identification,IEEE Trans

9. AndersonI. I.W., The Statistical Analysis of Time Series,1971, Wiley

10. Box G.E.P. , G.M. Jenkins, Some Recent Advances in Forecasting and Control Part I, 1968, Appl. Statist.

11. Box G.E.P. , G.M. Jenkins, D.W. Bacon, Models for Forecasting Seasonal and Nonseasonal Time Series, 1967, Wiley

12. Brockwell P.J., R.A. Davis, Time series: Theory and Methods, 1987, Springer-Verlag

13. Durbin J., The Fitting of the Time Series Models, 1960, Internat. Statist. Inst.

14. Fuller W.A., Introduction to Statistical Time Series, 1976, Wiley

15. Hannan E. J., Time Series Analysis, 1960, Methuen

16. Gersch W., G. Kitagawa, The Prediction of Time Series with Trends and Seasonalities, 1983, Business Econom. Statist.

17. Wilks S.S., Mathematical Statistics, 1962, Wiley

18. Godfrey L.G., Testing the Adequacy of a Time Series Model, 1979, Biometrica

19. Box G.E.P., G.M. Jenkins, Some Recent Advances in Forecasting and Control Part II, 1974 Appl. Statist.

20. Kendall M.G., Stuart A., The Advanced Theory of Statistics, Vol. 3, 1996, Charles Griffin

APPENDIX

World Petroleum Production
Table 1

Year	Month	Value,(Mb/d)
1995	1	61.845,09
1995	2	62.366,37
1995	3	61.622,49
1995	4	62.422,38
1995	5	62.427,01
1995	6	61.563,90
1995	7	62.522,78
1995	8	62.641,34
1995	9	63.041,18
1995	10	62.705,11
1995	11	62.903,93
1995	12	63.292,91
1996	1	63.454,63
1996	2	63.855,80
1996	3	63.704,33
1996	4	63.559,26
1996	5	63.557,57
1996	6	63.885,49
1996	7	63.975,96
1996	8	63.645,90
1996	9	64.110,92
1996	10	64.467,82
1996	11	64.925,84
1996	12	65.501,23
1997	1	65.781,99
1997	2	66.143,83
1997	3	66.119,91
1997	4	66.665,64
1997	5	66.000,64
1997	6	65.244,31
1997	7	65.694,78
1997	8	66.569,91
1997	9	66.917,73
1997	10	67.453,45
1997	11	67.307,54
1997	12	67.125,03
1998	1	67.655,63
1998	2	68.020,28
1998	3	67.897,29

84

1998	4	67.762,26
1998	5	67.219,59
1998	6	66.952,63
1998	7	66.821,78
1998	8	65.807,03
1998	9	65.904,14
1998	10	66.024,94
1998	11	66.870,20
1998	12	66.696,52
1999	1	66.891,00
1999	2	67.211,18
1999	3	66.888,37
1999	4	65.445,83
1999	5	65.253,04
1999	6	64.202,45
1999	7	65.725,35
1999	8	65.602,84
1999	9	65.641,80
1999	10	66.156,27
1999	11	66.143,10
1999	12	65.337,35
2000	1	66.032,24
2000	2	66.658,74
2000	3	66.689,07
2000	4	67.353,56
2000	5	67.856,83
2000	6	67.646,48
2000	7	68.273,49
2000	8	69.078,70
2000	9	69.115,68
2000	10	69.556,65
2000	11	70.101,60
2000	12	68.838,87
2001	1	68.940,12
2001	2	68.478,21
2001	3	69.165,97
2001	4	68.268,40
2001	5	67.576,67
2001	6	66.004,01
2001	7	67.978,71
2001	8	68.165,35
2001	9	67.792,01
2001	10	67.680,40
2001	11	67.929,27
2001	12	67.473,76

2002	1	66.391,57
2002	2	66.523,81
2002	3	66.338,93
2002	4	65.780,53
2002	5	66.184,34
2002	6	66.058,17
2002	7	66.654,99
2002	8	66.362,88
2002	9	67.111,00
2002	10	68.769,01
2002	11	68.764,28
2002	12	67.008,72

Network Traffic Data

Table 2

Date	Traffic IN (bytes)	Traffic OUT (bytes)	Traffic (bytes)
08.03.2004	43.977.484	115.642.494	159.619.978
07.03.2004	66.971.974	150.424.195	217.396.169
06.03.2004	71.250.793	218.398.210	289.649.003
05.03.2004	87.176.803	303.743.898	390.920.701
04.03.2004	94.620.977	247.238.424	341.859.401
03.03.2004	73.123.069	314.037.940	387.161.009
02.03.2004	591.336.560	306.554.361	897.890.921
01.03.2004	61.119.094	266.353.130	327.472.224
29.02.2004	34.054.227	239.101.257	273.155.484
28.02.2004	48.987.496	217.398.944	266.386.440
27.02.2004	54.519.150	291.121.079	345.640.229
26.02.2004	56.549.595	265.479.849	322.029.444
25.02.2004	48.656.454	308.929.846	357.586.300
24.02.2004	55.225.952	324.819.334	380.045.286
23.02.2004	56.353.281	296.938.553	353.291.834
22.02.2004	0	0	0
21.02.2004	28.450.965	144.569.591	173.020.556
20.02.2004	56.033.432	297.806.356	353.839.788
19.02.2004	62.726.126	321.770.941	384.497.067
18.02.2004	60.328.349	376.715.149	437.043.498
17.02.2004	56.446.547	375.811.676	432.258.223
16.02.2004	63.256.382	311.134.077	374.390.459
15.02.2004	32.296.557	190.217.226	222.513.783
14.02.2004	32.984.730	190.391.856	223.376.586
13.02.2004	54.032.262	275.605.620	329.637.882
12.02.2004	72.233.950	308.485.016	380.718.966

11.02.2004	88.434.497	354.494.690	442.929.187
10.02.2004	83.432.097	351.884.646	435.316.743
09.02.2004	115.284.571	351.229.601	466.514.172
08.02.2004	64.316.774	212.269.229	276.586.003
07.02.2004	39.530.078	206.322.251	245.852.329
06.02.2004	80.604.214	339.269.690	419.873.904
05.02.2004	76.473.146	1.723.415.694	1.799.888.840
04.02.2004	10.651.823	51.488.605	62.140.428
03.02.2004	44.826.113	154.210.739	199.036.852
02.02.2004	54.008.561	204.131.147	258.139.708
01.02.2004	41.710.374	149.557.453	191.267.827
31.01.2004	63.081.704	277.447.720	340.529.424
30.01.2004	148.776.609	465.177.689	613.954.298
29.01.2004	174.748.660	413.560.191	588.308.851
28.01.2004	100.104.084	375.684.153	475.788.237
27.01.2004	101.974.610	381.099.552	483.074.162
26.01.2004	140.087.247	349.704.220	489.791.467
25.01.2004	96.467.065	138.140.750	234.607.815
24.01.2004	72.303.507	162.504.409	234.807.916
23.01.2004	72.724.544	281.010.757	353.735.301
22.01.2004	58.474.910	280.300.800	338.775.710
21.01.2004	72.148.271	337.049.574	409.197.845
20.01.2004	71.867.001	322.171.043	394.038.044
19.01.2004	58.046.457	280.724.121	338.770.578
18.01.2004	43.094.012	103.329.044	146.423.056
17.01.2004	70.618.571	159.265.023	229.883.594
16.01.2004	56.674.180	130.417.571	187.091.751
15.01.2004	58.810.983	90.523.477	149.334.460
14.01.2004	87.570.569	231.312.341	318.882.910
13.01.2004	3.065.917	4.128.884	7.194.801

THE GUIDE TO APPLIED PROGRAM

First step:

If suitable forecasting method is chosen, Form 1 makes calculation of forecasts, according to chosen method.

Second step:

On the left top corner are situated two windows. The left one with button 'Enter data' under it for entering data, second one shows index of time series, so it let us to be sure that data was entered right.

Third step:

1. The panel on the top of screen includes menu with titles of forecasting methods: Moving Average model, Exponential- Smoothing forecasting model, Box-Jenkins forecasting procedure, where last one has substructure: ARIMA(1,0,0), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA(2,0,0). Choose suitable method from menu.

2. Results of forecasting according to chosen method will be written in the window with title 'Forecasts'.

3. There is a chart to the right side from the window with forecasted values. It shows graph, which is created according to forecasted values.

Forth step:

If suitable model for given time series is not known, Form 2 is created to solve this problem. Form 2 includes menu with forecasting methods titles and windows for entering data like a Form 1. The aim of two windows below to show results, left one forecasted values, right one actual data. This interface was done for easiest comparing actual data, which are known with forecasted values. Also there is one small window with label 'accuracy' on the lower part of screen. It shows error of forecasting, which was calculated accordantly to chosen method.

Chart on the right side shows two graphs, where blue colored curve is actual data, and red colored curve is forecasted data. So it can be seen how much forecast is accurate.

Fifth step:

After several methods were checked and accuracy was known, model can be chosen by following parameters: firstly model shouldn't be complicated, secondly, it should have least error.

Sixth step:

First Form gives forecasting by using method which was chosen on the previous steps.

If model of forecasting is

Go to Form 1
Forecast Ahead

Go to Form 2
Calculation

Enter Data

Enter Data

Chose Method

Choose Method

Forecasting
Ahead

Forecasting of known data
Calculation accuracy

Error Quite Small
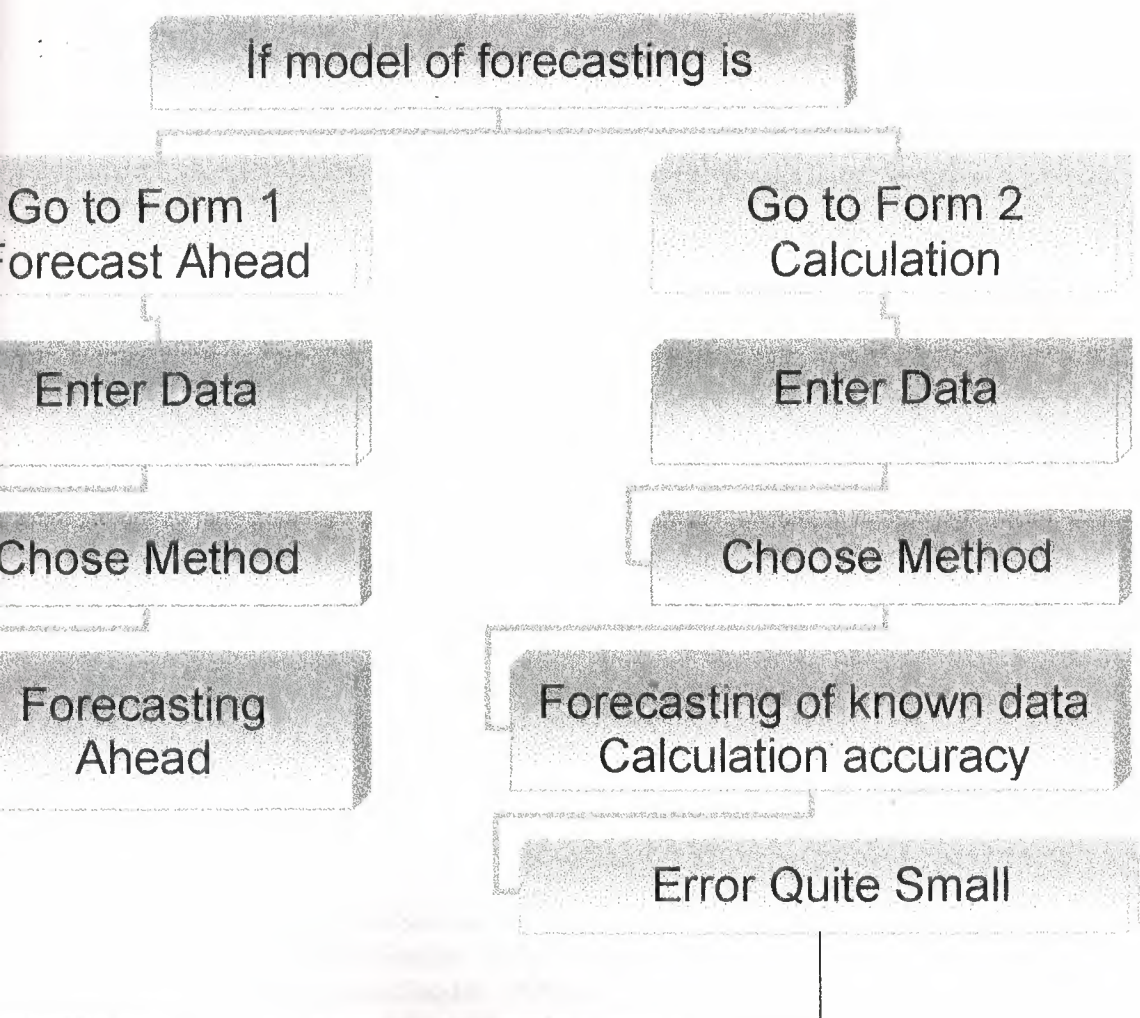
```
unit Unit1;

interface

uses
  Windows, Messages, SysUtils,{ Variants,} Classes, Graphics, Controls, Forms,
  Dialogs, StdCtrls, ExtCtrls, TeeProcs, TeEngine, Chart, Series, Menus, Unit2;
const suntitled=1;
     soverwrite=2;
type
 TForm1 = class(TForm)
   Edit1: TEdit;
   Button1: TButton;
   Edit2: TEdit;
   Label1: TLabel;
   Chart1: TChart;
   Button2: TButton;
   Series1: TLineSeries;
   MainMenu1: TMainMenu;
   movingavera1: TMenuItem;
   anexponentialsmoothingforecastmodel1: TMenuItem;
   ListBox1: TListBox;
   Button3: TButton;
   Label2: TLabel;
   Button5: TButton;
   Edit3: TEdit;
   File1: TMenuItem;
   New1: TMenuItem;
   Open1: TMenuItem;
   Save1: TMenuItem;
   N1: TMenuItem;
   BoxJenkis1: TMenuItem;
   BoxJenkis1001: TMenuItem;
   BoxJenkinsforecast1: TMenuItem;
   BoxJenkins1102: TMenuItem;
   Memo1: TMemo;
   OpenDialog1: TOpenDialog;
   SaveDialog1: TSaveDialog;
   Saveas1: TMenuItem;
   Exit1: TMenuItem;
   Button4: TButton;
   procedure Button1Click(Sender: TObject);
   procedure FormCreate(Sender: TObject);
   procedure Button2Click(Sender: TObject);
   procedure movingavera1Click(Sender: TObject);
   procedure Button5Click(Sender: TObject);
   procedure Button3Click(Sender: TObject);
```

```
procedure anexponentialsmoothingforecastmodel1Click(Sender: TObject);
procedure BoxJenkinsforecstprocedure1Click(Sender: TObject);
procedure New1Click(Sender: TObject);
procedure Open1Click(Sender: TObject);
procedure Save1Click(Sender: TObject);
procedure BoxJenkis1001Click(Sender: TObject);
procedure BoxJenkinsforecast1Click(Sender: TObject);
procedure BoxJenkins1102Click(Sender: TObject);
procedure Saveas1Click(Sender: TObject);
procedure Exit1Click(Sender: TObject);
procedure Button4Click(Sender: TObject);
private
ffilename:string;
file11:Text;
y  : array[1..12]of real;
f: array[13..24]of real;
 { Private declarations }
public
  { Public declarations }
end;

var
 Form1: TForm1;
 code:integer;
 result: real;
 s:string;
 apos,n:integer;

implementation

{$R *.dfm}


procedure TForm1.new1Click(Sender: TObject);
begin
 Memo1.visible:=true;
 Memo1.Lines.Clear;
 Memo1.Modified := False;
 end;

procedure TForm1.open1Click(Sender: TObject);
var  FileExt:String[4];
begin
If opendialog1.execute then begin
   FileExt:=ExtractFileExt(OpenDialog1.Filename);
       Memo1.visible:=True;
       Memo1.Lines.LoadFromFile(OpenDialog1.FileName);
```

```
        FFileName := OpenDialog1.FileName;
        Memo1.SetFocus;
        Memo1.Modified := False;
        Memo1.ReadOnly := ofReadOnly in OpenDialog1.Options;

  end;
end;

procedure TForm1.Save1Click(Sender: TObject);
begin
   if FFileName = LoadStr(sUntitled) then
     Save1Click(Sender)
   else
 begin
   Memo1.Lines.SaveToFile(FFileName);
   Memo1.Modified := True{False};
 end;
end;

procedure TForm1.Saveas1Click(Sender: TObject);
begin
   if SaveDialog1.Execute then
    begin
      if FileExists(SaveDialog1.FileName) then
      if MessageDlg(FmtLoadStr(sOverwrite, [SaveDialog1.FileName]),
      mtConfirmation, mbYesNoCancel, 0) <> idYes then Exit;
    Memo1.Lines.SaveToFile(SaveDialog1.FileName);
    FFileName := SaveDialog1.FileName;
    Memo1.Modified := False;
    end;
  end;


procedure TForm1.Button4Click(Sender: TObject);
var
 i: integer;
 s1:string;
begin
If opendialog1.execute then begin
   FFileName := OpenDialog1.FileName;
   assignfile(file11,ffilename);
   reset(file11);
   i:=1;
   while not eof(file11) do
   begin
     readln(file11,y[i]);
     str(y[i],s1);
```

92

```
       memo1.text:=memo1.text+s1+#13#10;
       i:=i+1;
   end;
          end;
end;

procedure TForm1.Button2Click(Sender: TObject);
begin
Halt;
end;

procedure TForm1.Button1Click(Sender: TObject);
var
 t:integer;
begin

  val(edit1.Text,t,code);
  y[apos]:=t;
  apos:=apos+1;
  edit2.Text:=inttostr(apos);
  edit1.text:='';

end;


procedure TForm1.FormCreate(Sender: TObject);
begin
apos:=1;
n:=1;
edit2.Text:=inttostr(apos);
end;


procedure TForm1.movingavera1Click(Sender: TObject);
var
i:integer;
begin
  for i:=12 to 23 do
    begin
     f[i+1]:=(y[i]+y[i-1]+y[i-2]+y[i-3]+y[i-4])/5;
    end;

  for i:= 13 to 24 do

  begin
  series1.AddXY(i-11,f[i],'',clRed);
  listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
```

```
end;

end;

procedure TForm1.Button5Click(Sender: TObject);
begin
 Form1.Hide;
 Form2.Show;
end;

procedure TForm1.Button3Click(Sender: TObject);
begin
 listbox1.Clear;
 apos:=1;
 edit2.Text:=inttostr(apos);
end;

procedure TForm1.anexponentialsmoothingforecastmodel1Click(
 Sender: TObject);
var
i:integer;
s,ss,sss:array[1..12] of real;
begin
s[1]:=y[1];
ss[1]:=y[1];
sss[1]:=y[1];
 for i:=2 to 12 do
  begin
   s[i]:=0.2*y[i]+0.8*s[i-1];
   ss[i]:=0.2*s[i]+0.8*ss[i-1];
   sss[i]:=0.2*ss[i]+0.8*sss[i-1];
  end;

 for i:=1 to 11 do
  begin
   f[i]:=3.8125*s[i]-4.375*ss[i]+1.5625*sss[i];
  end;

 for i:=1 to 12 do
  begin
   series1.AddXY(i,f[i],'',clRed);
   listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
  end;

end;
```

94

```pascal
procedure TForm1.BoxJenkinsforecstprocedure1Click(Sender: TObject);
var
i:integer;
v,w,z,vv,ww,vw,vz,wz:array[1..10] of Double;
d,d1,d2,d3,a,b,c,sumv,sumw,sumz,sumvv,sumww,sumvw,sumvz,sumwz: real;
begin
sumv:=0;
sumw:=0;
sumz:=0;
sumvv:=0;
sumww:=0;
sumvw:=0;
sumvz:=0;
sumwz:=0;
for i:=1 to 10 do
   begin
    v[i]:=y[i];
    w[i]:=y[i+1];
    z[i]:=y[i+2];
   end;
for i:=1 to 10 do
   begin
    vv[i]:=v[i]*v[i];
    ww[i]:=w[i]*w[i];
    vw[i]:=v[i]*w[i];
    vz[i]:=v[i]*z[i];
    wz[i]:=w[i]*z[i];
    sumv:=sumv+v[i];
    sumw:=sumw+w[i];
    sumz:=sumz+z[i];
    sumvv:=sumvv+vv[i];
    sumww:=sumww+ww[i];
    sumvw:=sumvw+vw[i];
    sumvz:=sumvz+vz[i];
    sumwz:=sumwz+wz[i];
   end;
d:=12*sumvv*sumww-12*sumvw*sumvw-
sumv*sumv*sumww+sumv*sumvw*sumw+sumw*sumv*sumvw-sumw*sumvv*sumw;
d1:=sumz*sumvv*sumww-sumz*sumvw*sumvw-
sumv*sumvz*sumww+sumv*sumwz*sumvw+sumw*sumvz*sumvw-
sumw*sumwz*sumvv;
d2:=12*sumvz*sumww-12*sumvw*sumwz-
sumz*sumv*sumww+sumz*sumvw*sumw+sumw*sumv*sumwz-sumw*sumvz*sumw;
d3:=12*sumvv*sumwz-12*sumvz*sumvw-
sumv*sumv*sumwz+sumv*sumw*sumvz+sumz*sumv*sumvw-sumz*sumvv*sumw;
a:=d1/d;
b:=d2/d;
```

95

```
c:=d3/d;
f[13]:=a+b*y[11]+c*y[12];
f[14]:=a+b*y[12]+c*f[13];
for i:=15 to 24 do
   begin
     f[i]:=a+b*f[i-2]+c*f[i-1];
   end;
for i:= 13 to 24 do
   begin
     series1.AddXY(i-11,f[i],'',clRed);
     listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
   end;
end;


procedure TForm1.BoxJenkins1102Click(Sender: TObject);
var
v,w,yy,xx,xy:array[1..10]of real;
y1:array[1..11] of real;
f:array[12..24] of real;
f1:array[12..24] of real;
i:integer;
xaverage,yaverage,sumx,sumy,sumxx,sumyy,sumxy,sxx,sxy,b1,b2:real;
begin
  for i:=1 to 11 do
    begin
      y1[i]:=y[i+1]-y[i];
    end;
  sumy:=0;
  sumx:=0;
  sumyy:=0;
  sumxx:=0;
  sumxy:=0;
  for i:=1 to 10 do
    begin
      v[i]:=y1[i];
      w[i]:=y1[i+1];
    end;
  for i:=1 to 10 do
    begin
      yy[i]:=w[i]*w[i];
      xx[i]:=v[i]*v[i];
      xy[i]:=v[i]*w[i];
      sumy:=sumy+w[i];
      sumx:=sumx+v[i];
      sumxx:=sumxx+xx[i];
      sumyy:=sumyy+yy[i];
```

```
     sumxy:=sumxy+xy[i];
   end;
xaverage:=sumx/10;
yaverage:=sumy/10;
sxx:=sumxx-(sumx*sumx)/10;
sxy:=sumxy-(sumx*sumy)/10;
b1:=sxy/sxx;
b2:=yaverage-b1*xaverage;
f[12]:=b2+b1*y1[11];
f1[12]:=y[11]+f[12];
  for i:=13 to 24 do
    begin
      f[i]:=b2+b1*f[i-1];
      f1[i]:=f1[i-1]+f[i];
    end;

    for i:=13 to 24 do
    begin
      series1.AddXY(i,f1[i],'',clRed);
      listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f1[i]));
    end;
end;


procedure TForm1.BoxJenkinsforecast1Click(Sender: TObject);
var
i:integer;
v,w,z,vv,ww,vw,vz,wz:array[1..10] of Double;
d,d1,d2,d3,a,b,c,sumv,sumw,sumz,sumvv,sumww,sumvw,sumvz,sumwz: real;
begin
sumv:=0;
sumw:=0;
sumz:=0;
sumvv:=0;
sumww:=0;
sumvw:=0;
sumvz:=0;
sumwz:=0;
for i:=1 to 10 do
    begin
      v[i]:=y[i];
      w[i]:=y[i+1];
      z[i]:=y[i+2];
    end;
  for i:=1 to 10 do
    begin
      vv[i]:=v[i]*v[i];
```

```
     ww[i]:=w[i]*w[i];
     vw[i]:=v[i]*w[i];
     vz[i]:=v[i]*z[i];
     wz[i]:=w[i]*z[i];
      sumv:=sumv+v[i];
      sumw:=sumw+w[i];
      sumz:=sumz+z[i];
      sumvv:=sumvv+vv[i];
      sumww:=sumww+ww[i];
      sumvw:=sumvw+vw[i];
      sumvz:=sumvz+vz[i];
      sumwz:=sumwz+wz[i];
    end;
d:=12*sumvv*sumww-12*sumvw*sumvw-
sumv*sumv*sumww+sumv*sumvw*sumw+sumw*sumv*sumvw-sumw*sumvv*sumw;
d1:=sumz*sumvv*sumww-sumz*sumvw*sumvw-
sumv*sumvz*sumww+sumv*sumwz*sumvw+sumw*sumvz*sumvw-
sumw*sumwz*sumvv;
d2:=12*sumvz*sumww-12*sumvw*sumwz-
sumz*sumv*sumww+sumz*sumvw*sumw+sumw*sumv*sumwz-sumw*sumvz*sumw;
d3:=12*sumvv*sumwz-12*sumvz*sumvw-
sumv*sumv*sumwz+sumv*sumw*sumvz+sumz*sumv*sumvw-sumz*sumvv*sumw;
a:=d1/d;
b:=d2/d;
c:=d3/d;
f[13]:=a+b*y[11]+c*y[12];
f[14]:=a+b*y[12]+c*f[13];
for i:=15 to 24 do
    begin
      f[i]:=a+b*f[i-2]+c*f[i-1];
    end;
  for i:= 13 to 24 do
    begin
      series1.AddXY(i-11,f[i],'',clRed);
      listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
    end;
end;

procedure TForm1.BoxJenkis1001Click(Sender: TObject);
var
v,w,yy,xx,xy:array[1..12]of real;
i:integer;
xaverage,yaverage,sumx,sumy,sumxx,sumyy,sumxy,sxx,sxy,b1,b2:real;
begin
sumy:=0;
sumx:=0;
sumyy:=0;
```

```pascal
    sumxx:=0;
    sumxy:=0;
    for i:=1 to 11 do
      begin
        v[i]:=y[i];
        w[i]:=y[i+1];
      end;
    for i:=1 to 11 do
      begin
        yy[i]:=w[i]*w[i];
        xx[i]:=v[i]*v[i];
        xy[i]:=v[i]*w[i];
        sumy:=sumy+w[i];
        sumx:=sumx+v[i];
        sumxx:=sumxx+xx[i];
        sumyy:=sumyy+yy[i];
        sumxy:=sumxy+xy[i];
      end;
xaverage:=sumx/12;
yaverage:=sumy/12;
sxx:=sumxx-(sumx*sumx)/12;
sxy:=sumxy-(sumx*sumy)/12;
b1:=sxy/sxx;
b2:=yaverage-b1*xaverage;
  for i:=13 to 24 do
    begin
      f[i]:=b2+b1*y[i-1];
    end;
  for i:= 13 to 24 do
    begin
      series1.AddXY(i-11,f[i],'',clRed);
      listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
    end;
end;

procedure TForm1.Exit1Click(Sender: TObject);
begin
close;
end;

end.
```

```pascal
unit Unit2;

interface

uses
  Windows, Messages, SysUtils, Classes, Graphics, Controls, Forms, Dialogs,
  Menus, ExtCtrls, TeeProcs, TeEngine, Chart, StdCtrls, Series;
const suntitled=1;
    soverwrite=2;


type
  TForm2 = class(TForm)
    MainMenu1: TMainMenu;
    ListBox1: TListBox;
    ListBox2: TListBox;
    Button1: TButton;
    Edit1: TEdit;
    Edit2: TEdit;
    Chart1: TChart;
    movingaverage1: TMenuItem;
    movingaverageforecastmodel1: TMenuItem;
    exponentialsmoothingforecastmodel1: TMenuItem;
    BoxJenkinsforecastprocedure1: TMenuItem;
    N1001: TMenuItem;
    N2001: TMenuItem;
    Button2: TButton;
    Label1: TLabel;
    Series1: TLineSeries;
    Series2: TLineSeries;
    ListBox3: TListBox;
    N1101: TMenuItem;
    N1111: TMenuItem;
    Edit3: TEdit;
    Edit4: TEdit;
    Open1: TMenuItem;
    Save1: TMenuItem;
    Saveas1: TMenuItem;
    Exit1: TMenuItem;
    New1: TMenuItem;
    OpenDialog1: TOpenDialog;
    SaveDialog1: TSaveDialog;
    Memo1: TMemo;
    Button3: TButton;
    procedure Button1Click(Sender: TObject);
    procedure Button2Click(Sender: TObject);
    procedure FormCreate(Sender: TObject);
    procedure movingaverageforecastmodel1Click(Sender: TObject);
```

```
   procedure exponentialsmoothingforecastmodel1Click(Sender: TObject);
   procedure N1001Click(Sender: TObject);
   procedure N2001Click(Sender: TObject);
   procedure N1101Click(Sender: TObject);
   procedure N1111Click(Sender: TObject);
   procedure Open1Click(Sender: TObject);
   procedure New1Click(Sender: TObject);
   procedure Save1Click(Sender: TObject);
   procedure Saveas1Click(Sender: TObject);
   procedure Exit1Click(Sender: TObject);
   procedure Button3Click(Sender: TObject);
 private
   { Private declarations }
   ffilename:string;
   file1:Text;
   y  : array[1..24]of real;
 public
   { Public declarations }
 end;

var
 Form2: TForm2;
 apos2,n:integer;
 code:integer;
 dev:array[1..12]of real;
 mad: real;
implementation

{$R *.DFM}

procedure TForm2.New1Click(Sender: TObject);
begin
 Memo1.visible:=true;
 Memo1.Lines.Clear;
 Memo1.Modified := False;
end;

procedure TForm2.Open1Click(Sender: TObject);
var  FileExt:String[4];
begin
If opendialog1.execute then begin
  FileExt:=ExtractFileExt(OpenDialog1.Filename);
      Memo1.visible:=True;
      Memo1.Lines.LoadFromFile(OpenDialog1.FileName);
      FFileName := OpenDialog1.FileName;
      Memo1.SetFocus;
      Memo1.Modified := False;
```

101

```
      Memo1.ReadOnly := ofReadOnly in OpenDialog1.Options;

  end;
end;

procedure TForm2.Save1Click(Sender: TObject);
begin
  if FFileName = LoadStr(sUntitled) then
    Save1Click(Sender)
  else
   begin
     Memo1.Lines.SaveToFile(FFileName);
     Memo1.Modified := True{False};
   end;
  end;

procedure TForm2.Saveas1Click(Sender: TObject);
begin
if SaveDialog1.Execute then
  begin
    if FileExists(SaveDialog1.FileName) then
      if MessageDlg(FmtLoadStr(sOverwrite, [SaveDialog1.FileName]),
        mtConfirmation, mbYesNoCancel, 0) <> idYes then Exit;
    Memo1.Lines.SaveToFile(SaveDialog1.FileName);
    FFileName := SaveDialog1.FileName;
    Memo1.Modified := False;
   end;
  end;

procedure TForm2.Exit1Click(Sender: TObject);
begin
close;
end;

procedure TForm2.Button3Click(Sender: TObject);
var
  i: integer;
  s1:string;
begin
If opendialog1.execute then begin
    FFileName := OpenDialog1.FileName;
    assignfile(file1,ffilename);
    reset(file1);
    i:=1;
    while not eof(file1) do
    begin
      readln(file1,y[i]);
```

```
      str(y[i],s1);
      memo1.text:=memo1.text+s1+#13#10;
      i:=i+1;
    end;
        end;
        end;

procedure TForm2.Button1Click(Sender: TObject);
var
t:integer;
begin
  val(edit1.Text,t,code);
  y[apos2]:=t;
  apos2:=apos2+1;
  edit2.Text:=inttostr(apos2);
  edit1.text:='';
end;

procedure TForm2.Button2Click(Sender: TObject);
begin
listbox1.Clear;
listbox2.Clear;
apos2:=1;
edit2.Text:=inttostr(apos2);

end;

procedure TForm2.FormCreate(Sender: TObject);
begin
apos2:=1;
n:=1;
edit2.Text:=inttostr(apos2);
end;

procedure TForm2.movingaverageforecastmodel1Click(Sender: TObject);
var
i: integer;
acc: real;
f:array[13..24] of real;
begin
mad:=0;
  for i:=13 to 24 do
    begin
      f[i]:=(y[i-1]+y[i-2]+y[i-3]+y[i-4]+y[i-5])/5;
      dev[i]:=Abs(f[i]-y[i]);
      mad:=mad+dev[i];
    end;
```

103

```
acc:=mad/12;
  for i:= 13 to 24 do
   begin
    series1.AddXY(i-11,f[i],'',clRed);
    series2.AddXY(i-11,y[i],'',clBlue);
    listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
    listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0.00',y[i]));
   end;
listbox3.Items.Add(formatfloat('0.00',acc));
end;

procedure TForm2.exponentialsmoothingforecastmodel1Click(Sender: TObject);
var
i:integer;
mad,acc:real;
s,ss,sss:array[1..12] of real;
f:array[2..12] of real;
begin
s[1]:=y[1];
ss[1]:=y[1];
sss[1]:=y[1];
  for i:=2 to 12 do
   begin
    s[i]:=0.2*y[i]+0.8*s[i-1];
    ss[i]:=0.2*s[i]+0.8*ss[i-1];
    sss[i]:=0.2*ss[i]+0.8*sss[i-1];
   end;

  mad:=0;
  for i:=2 to 12 do
   begin
    f[i]:=3.8125*s[i-1]-4.375*ss[i-1]+1.5625*sss[i-1];
    dev[i]:=Abs(f[i]-y[i]);
    mad:=mad+dev[i];
   end;
acc:=mad/11;

  for i:=2 to 12 do
   begin
    series1.AddXY(i,f[i],'',clRed);
    series2.AddXY(i,y[i],'',clblue);
    listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
    listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0.00',y[i]));
   end;
listbox3.Items.Add(formatfloat('0.00',acc));

end;
```

104

```
procedure TForm2.N1001Click(Sender: TObject);
var
v,w,yy,xx,xy,f,dev:array[1..12]of real;
i:integer;
xaverage,yaverage,sumx,sumy,sumxx,sumyy,sumxy,sxx,sxy,b1,b2,acc,mad:real;
begin
sumy:=0;
sumx:=0;
sumyy:=0;
sumxx:=0;
sumxy:=0;
for i:=1 to 11 do
  begin
   v[i]:=y[i];
   w[i]:=y[i+1];
  end;
  for i:=1 to 11 do
   begin
    yy[i]:=w[i]*w[i];
    xx[i]:=v[i]*v[i];
    xy[i]:=v[i]*w[i];
    sumy:=sumy+w[i];
    sumx:=sumx+v[i];
    sumxx:=sumxx+xx[i];
    sumyy:=sumyy+yy[i];
    sumxy:=sumxy+xy[i];
   end;
xaverage:=sumx/12;
yaverage:=sumy/12;
sxx:=sumxx-(sumx*sumx)/12;
sxy:=sumxy-(sumx*sumy)/12;
b1:=sxy/sxx;
b2:=yaverage-b1*xaverage;
mad:=0;
f[1]:=y[1];
  for i:=2 to 12 do
   begin
    f[i]:=b2+b1*y[i-1];
    dev[i]:=Abs(f[i]-y[i]);
    mad:=mad+dev[i];
   end;
acc:=mad/12;

  for i:=1 to 12 do
   begin
    series1.AddXY(i,f[i],'',clRed);
```

```
      series2.AddXY(i,y[i],'',clBlue);
      listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
      listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0.00',y[i]));
    end;
listbox3.Items.Add(formatfloat('0.00',acc));

end;

procedure TForm2.N2001Click(Sender: TObject);
var
i:integer;
v,w,z,vv,ww,vw,vz,wz:array[1..10] of Double;
dev,f:array[1..12] of real;
d,d1,d2,d3,a,b,c,sumv,sumw,sumz,sumvv,sumww,sumvw,sumvz,sumwz,acc,mad: real;
begin
sumv:=0;
sumw:=0;
sumz:=0;
sumvv:=0;
sumww:=0;
sumvw:=0;
sumvz:=0;
sumwz:=0;
for i:=1 to 10 do
   begin
     v[i]:=y[i];
     w[i]:=y[i+1];
     z[i]:=y[i+2];
   end;
for i:=1 to 10 do
   begin
     vv[i]:=v[i]*v[i];
     ww[i]:=w[i]*w[i];
     vw[i]:=v[i]*w[i];
     vz[i]:=v[i]*z[i];
     wz[i]:=w[i]*z[i];
     sumv:=sumv+v[i];
     sumw:=sumw+w[i];
     sumz:=sumz+z[i];
     sumvv:=sumvv+vv[i];
     sumww:=sumww+ww[i];
     sumvw:=sumvw+vw[i];
     sumvz:=sumvz+vz[i];
     sumwz:=sumwz+wz[i];
   end;
d:=12*sumvv*sumww-12*sumvw*sumvw-
sumv*sumv*sumww+sumv*sumvw*sumw+sumw*sumv*sumvw-sumw*sumvv*sumw;
```

```
d1:=sumz*sumvv*sumww-sumz*sumvw*sumvw-
sumv*sumvz*sumww+sumv*sumwz*sumvw+sumw*sumvz*sumvw-
sumw*sumwz*sumvv;
d2:=12*sumvz*sumww-12*sumvw*sumwz-
sumz*sumv*sumww+sumz*sumvw*sumw+sumw*sumv*sumwz-sumw*sumvz*sumw;
d3:=12*sumvv*sumwz-12*sumvz*sumvw-
sumv*sumv*sumwz+sumv*sumw*sumvz+sumz*sumv*sumvw-sumz*sumvv*sumw;
a:=d1/d;
b:=d2/d;
c:=d3/d;
mad:=0;
f[1]:=y[1];
f[2]:=y[2];
for i:=3 to 12 do
   begin
     f[i]:=a+b*y[i-2]+c*y[i-1];
     dev[i]:=Abs(f[i]-y[i]);
     mad:=mad+dev[i];
   end;
acc:=mad/12;
for i:= 1 to 12 do
   begin
     series1.AddXY(i,f[i],'',clRed);
     series2.AddXY(i,y[i],'',clBlue);
     listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f[i]));
     listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0.00',y[i]));
   end;
listbox3.Items.Add(formatfloat('0.00',acc));
end;

procedure TForm2.N1101Click(Sender: TObject);
var
v,w,yy,xx,xy,dev:array[1..10]of real;
y1:array[1..11] of real;
f:array[1..12] of real;
f1:array[2..12] of real;
i:integer;
xaverage,yaverage,sumx,sumy,sumxx,sumyy,sumxy,sxx,sxy,b1,b2,acc,mad:real;
begin
  for i:=1 to 11 do
   begin
     y1[i]:=y[i+1]-y[i];
   end;
sumy:=0;
sumx:=0;
sumyy:=0;
sumxx:=0;
```

```pascal
    sumxy:=0;
   for i:=1 to 10 do
     begin
      v[i]:=y1[i];
      w[i]:=y1[i+1];
     end;
    for i:=1 to 10 do
      begin
       yy[i]:=w[i]*w[i];
       xx[i]:=v[i]*v[i];
       xy[i]:=v[i]*w[i];
       sumy:=sumy+w[i];
       sumx:=sumx+v[i];
       sumxx:=sumxx+xx[i];
       sumyy:=sumyy+yy[i];
       sumxy:=sumxy+xy[i];
      end;
  xaverage:=sumx/10;
  yaverage:=sumy/10;
  sxx:=sumxx-(sumx*sumx)/10;
  sxy:=sumxy-(sumx*sumy)/10;
  b1:=sxy/sxx;
  b2:=yaverage-b1*xaverage;
  mad:=0;
  f[1]:=y1[1];
   for i:=2 to 12 do
     begin
      f[i]:=b2+b1*y1[i-1];
      f1[i]:=y[i-1]+f[i];
      dev[i]:=Abs(f1[i]-y[i]);
      mad:=mad+dev[i];
    end;
  acc:=mad/11;
    for i:=2 to 12 do
     begin
      series1.AddXY(i,f1[i],'',clRed);
      series2.AddXY(i,y[i],'',clBlue);
      listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0.00',f1[i]));
      listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0.00',y[i]));
     end;
listbox3.Items.Add(formatfloat('0.00',acc));
end;


procedure TForm2.N1111Click(Sender: TObject);
var
v,w,vv,ww,yy,xx,xy,dev,devv:array[1..10]of real;
```

```
y1:array[1..11] of real;
f:array[1..12] of real;
f1,f2:array[2..12] of real;
i:integer;
xaverage,yaverage,sumx,sumy,sumxx,sumyy,sumxy,sxx,sxy,b1,b2,t1,acc,mad:real;
begin
  for i:=1 to 11 do
    begin
      y1[i]:=y[i+1]-y[i];
    end;
sumy:=0;
sumx:=0;
sumyy:=0;
sumxx:=0;
sumxy:=0;
for i:=1 to 10 do
  begin
    v[i]:=y1[i];
    w[i]:=y1[i+1];
  end;
  for i:=1 to 10 do
    begin
      yy[i]:=w[i]*w[i];
      xx[i]:=v[i]*v[i];
      xy[i]:=v[i]*w[i];
      sumy:=sumy+w[i];
      sumx:=sumx+v[i];
      sumxx:=sumxx+xx[i];
      sumyy:=sumyy+yy[i];
      sumxy:=sumxy+xy[i];
    end;
xaverage:=sumx/10;
yaverage:=sumy/10;
sxx:=sumxx-(sumx*sumx)/10;
sxy:=sumxy-(sumx*sumy)/10;
b1:=sxy/sxx;
b2:=yaverage-b1*xaverage;
mad:=0;
f[1]:=y1[1];
  for i:=2 to 12 do
    begin
      f[i]:=b2+b1*y1[i-1];
      f1[i]:=y[i-1]+f[i];
      dev[i]:=Abs(f1[i]-y[i]);
    end;
for i:=2 to 11 do
  begin
```

```
    vv[i]:=dev[i];
    ww[i]:=dev[i+1];
end;
for i:=2 to 11 do
   begin
     yy[i]:=0;
     xx[i]:=0;
     xy[i]:=0;
   end;
   sumx:=0; sumy:=0; sumxy:=0; sumxx:=0; sumyy:=0;
  for i:=2 to 11 do
   begin
     yy[i]:=ww[i]*ww[i];
     xx[i]:=vv[i]*vv[i];
     xy[i]:=v[i]*w[i];
     sumy:=sumy+ww[i];
     sumx:=sumx+vv[i];
     sumxx:=sumxx+xx[i];
     sumyy:=sumyy+yy[i];
     sumxy:=sumxy+xy[i];
   end;
sxx:=sumxx-(sumx*sumx)/10;
sxy:=sumxy-(sumx*sumy)/10;
t1:=sxy/sxx;
f[1]:=y1[1];

  for i:=2 to 12 do
   begin
     f[i]:=b1*y1[i-1]-t1*dev[i-1];
     f2[i]:=y[i-1]+f[i];
     devv[i]:=Abs(f2[i]-y[i]);
     mad:=mad+devv[i];
   end;
acc:=mad/11;
  for i:=4 to 11 do
   begin
     series1.AddXY(i,f2[i],'',clRed);
     series2.AddXY(i,y[i],'',clBlue);
     listbox1.Items.Add('f['+formatfloat('0',i)+']='+ formatfloat('0',f2[i]));
     listbox2.Items.Add('y['+formatfloat('0',i)+']='+ formatfloat('0',y[i]));
   end;
listbox3.Items.Add(formatfloat('0.00',acc));
Edit3.Text:=formatfloat('0.00', b1);
Edit4.Text:=formatfloat('0.00', t1);
end;
end.
```