Design & Implementation of Time Series Analysis Using Neural Network

A Dissertation

Submitted in Partial Fulfillment For the Award of The Degree of

Master of Engineering in

Computer Technology and Applications

By:

KUMKUM BAGCHI
(24/CTA/03)

Under the guidance of

Prof. D. Roy Choudhury

DEPARTMENT OF COMPUTER ENGINEERING
DELHI COLLEGE OF ENGINEERING
DELHI UNIVERSITY, DELHI-110042
2003-2005

CERTIFICATE

This is to certify that the project entitled " **Design & Implementation Time Series Analysis using Neural Network** " submitted by **Kumkum Bagchi,** class Roll No. 24/CTA/03, **University Roll No. 3017** in the partial fulfillment of the requirement for the award of degree of Master of Engineering in Computer Technology and Application, Delhi College of Engineering is an account of his authentic work carried out under my guidance and supervision. He has not submitted this work for the award of any other degree.

Project Guide:

Dr. D. Roy Choudhury
Professor
Dept. of Computer Engg.
Delhi College of Engg., Delhi

Head of Department:

Dr. D. Roy Choudhury
Professor
Dept. of Computer Engg.
Delhi College of Engg., Delhi

ACKNOWLEDGEMENTS

I here by take this opportunity to thank all those who have been a source of inspiration, help and encouragement during my project work.

It would be impossible to express in words my gratefulness to my advisors, Dr.D.Roy Choudhury. **Working with him has been a true privilege. His constant encouragement and valuable suggestions have guided me throughout this project. His enthusiasm and devotion have always inspired me during my hard times. I am again indebted to** Dr.D.Roy Choudhury **for taking keen interest, useful discussion, resourceful co-operation during my project work.**

I would like to convey my sincere thanks to Mr. Pranit Kumar (Senior Engr. (IT), BHEL) for his invaluable suggestions from time to time. Also, I'd like to express thanks to all my friends and family members who with their kind gestures have always encouraged me during the rough phases.

(KUMKUM BAGCHI)
DCE 2003-2005

DECLARATION BY THE CANDIDATE

I, hereby declare that the dissertation work entitled "Design & Implementation of Time Series Analysis using Neural Network" **is an authentic work carried out by me under the guidance of** Prof. D.Roy Choudhury **for the partial fulfillment and award of the degree of M.E. in computer Technology and Application. I have not submitted this work anywhere else for the award of any other degree.**

(KUMKUM BAGCHI)
M.E. (CTA) 24/CTA/03
Department of Computer Engineering
DCE, Delhi-1100042

Abstract

In this dissertation a detailed study of Time Series Forecasting & Analysis has been carried out. The implementation has been done using a Neural Network that learns from examples or historic live data. The designing of a suitable network and its mathematical approach is an important area of research that has been carried out extensively in MATLAB's nntool toolkit.

The following three different new examples have been taken into consideration for analyzing the forecasting techniques and see its accuracy in detail:

1. SERIES B, IBM STOCK PRICES
2. SUNSPOTS DATA FOR 300 YEARS (approx.) ON EARTH'S SURFACE, 1699-1999
3. TOTAL ANNUAL RAINFALL, INCHES, LONDON, ENGLAND, 1813-1912

Contents

CHAPTER 1

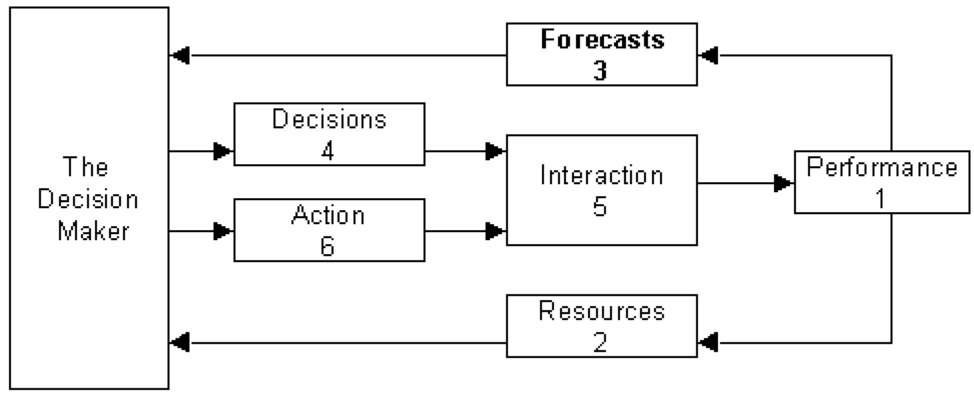
TIME-CRITICAL DECISION MODELING AND ANALYSIS

1.1 Introduction

The ability to model and perform decision modeling and analysis is an essential feature of many real-world applications ranging from emergency medical treatment in intensive care units to military command and control systems. Existing formalisms and methods of inference have not been effective in real-time applications where tradeoffs between decision quality and computational tractability are essential. In practice, an effective approach to time-critical dynamic decision modeling should provide explicit support for the modeling of temporal processes and for dealing with time-critical situations.

Almost all managerial decisions are based on forecasts. Every decision becomes operational at some point in the future, so it should be based on forecasts of future conditions.

Forecasts are needed throughout an organization -- and they should certainly not be produced by an isolated group of forecasters. Forecasting is never "finished". Forecasts are needed continually, and as time moves on, the impact of the forecasts on actual performance is measured; original forecasts are updated; and decisions are modified, and so on. This process is shown in the following figure:



Forecasting Within an Organization:
Forecasting and Managerial Decision Making

The decision-maker uses forecasting models to assist him or her in decision-making process. The decision-making often uses the modeling process to investigate the impact of different courses of action **retrospectively**; that is, "as if" the decision has already been made under a course of action. That is why the sequence of steps in the modeling process, in the above figure must be considered in reverse order. For example, the output (which is the result of the action) must be considered first.

It is helpful to break the components of decision making into three groups: Uncontrollable, Controllable, and Resources (that defines the problem situation). As indicated in the above activity chart, the decision-making process has the following components:

Performance measure (or indicator, or objective): Measuring business performance is the top priority for managers. The development of effective performance measures is seen as increasingly important in almost all organizations. However, the challenges of achieving this in the public and for non-profit sectors are arguably considerable. Performance measure provides the desirable level of outcome, i.e., objective of the decision. Objective is important in identifying the forecasting activity. The following table provides a few examples of performance measures for different levels of management:

Level	Performance Measure
Strategic	Return of Investment, Growth, and Innovations
Tactical	Cost, Quantity, and Customer satisfaction
Operational	Target setting, and Conformance with Standard

To improve a **system's performance**, an operational view is required. Such a view gets at how a forecasting system really works; for example, by what correlation its past output behaviors have generated. Forecasting activity is an iterative process. It starts with effective and efficient planning and ends in compensation of other forecasts for their performance.

What is a System?

Systems that are building blocks for other systems called *subsystems*

The Dynamics of a System: A system that does not change is a static system. Many of the business systems are dynamic systems, which mean their states change over time. The way a system changes over time as the system's behavior. And when the system's development follows a typical pattern, it is said that the system has a behavior pattern. Whether a system is static or dynamic depends on which time horizon you choose and on which variables you concentrate. The time horizon is the time period within which you study the system. The variables are changeable values on the system.

Resources: Resources are the constant elements that do not change during the time horizon of the forecast. Resources are the factors that define the decision problem.

Forecasts: Forecasts input come from the decision maker's environment. Uncontrollable inputs must be forecasted or predicted.

Decisions: Decisions inputs ate the known collection of all possible courses of action one might take.

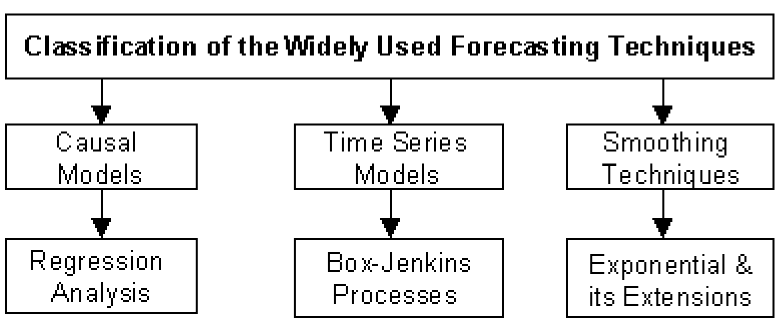
Interaction: Interactions among the above decision components are the logical, mathematical functions representing the cause-and-effect relationships among inputs, resources, forecasts, and the outcome.

Actions: Action is the ultimate decision and is the best course of strategy to achieve the desirable goal.

Decision-making involves the selection of a course of action (means) in pursue of the decision maker's objective (ends). The way that our course of action affects the outcome of a decision depends on how the forecasts and other inputs are interrelated and how they relate to the outcome.

Controlling the Decision Problem/Opportunity: Few problems in life, once solved, stay that way. Changing conditions tend to un-solve problems that were previously solved, and their solutions create new problems. One must identify and anticipate these new problems.

Forecasting is a prediction of what will occur in the future, and it is an uncertain process. Because of the uncertainty, the accuracy of a forecast is as important as the outcome predicted by the forecast. A general overview of business forecasting techniques are classified in the following figure:



Progressive Approach to Modeling: Modeling for decision-making involves two distinct parties, one is the decision-maker and the other is the model-builder known as the analyst. The analyst is to assist the decision-maker in his/her decision-making process. Therefore, the analyst must be equipped with more than a set of analytical methods.

Quantitative Decision Making: Schools of Business and Management are flourishing with more and more students taking up degree program at all level. In particular there is a growing market for conversion courses such as MSc in Business or Management and post experience courses such as MBAs. Specialists in model building are often tempted to study a problem, and then go off in isolation to develop an elaborate mathematical model for use by the manager (i.e., the decision-maker). Unfortunately the manager may not understand this model and may either use it blindly or reject it entirely. The specialist may believe that the manager is too ignorant and unsophisticated to appreciate the model, while the manager may believe that the specialist lives in a dream world of unrealistic assumptions and irrelevant mathematical language. Such **miscommunication** can be avoided if the manager works with the specialist to develop first a simple model that provides a crude but understandable analysis. After the manager has built up confidence in this model, additional detail and sophistication can be added, perhaps progressively only a bit at a time. This process requires an investment of time on the part of the manager and sincere interest on the part of the specialist in solving the manager's real problem, rather than in creating and trying to explain sophisticated models. This progressive model building is often referred to as **the bootstrapping approach** and is the most important factor in determining successful implementation of a decision model. Moreover the bootstrapping approach simplifies the otherwise difficult task of model validation and verification processes.

The time series analysis has three goals: **forecasting (also called predicting), modeling, and characterization.** Modeling is again the key, though out-of-sample forecasting may be used to test any model. Often modeling and forecasting proceed in an iterative way and there is no 'logical order' in the broadest sense. One may model to get forecasts, which enable better control, but iteration is again likely to be present and there are sometimes special approaches to control problems.

Outliers: Outliers can be one-time outliers or seasonal pulses or a sequential set of outliers with nearly the same magnitude and direction (level shift) or local time trends. A pulse is a difference of a step while a step is a difference of a time trend. In order to assess or declare "an unusual value" one must develop "the expected or usual value". Time series techniques extended for outlier detection, i.e. intervention variables like pulses, seasonal pulses, level shifts and local time trends can be useful in "data cleansing" or pre-filtering of observations.

1.2 Effective Modeling for Good Decision-Making

1.2.1 A Model

A Model is an external and explicit representation of a part of reality, as it is seen by individuals who wish to use this model to understand, change, manage and control that part of reality.

Descriptive and prescriptive models: A descriptive model is often a function of figuration, abstraction based on reality. However, a prescriptive model is moving from reality to a model a function of development plan, means of action, moving from model to the reality.

The distinction between descriptive and prescriptive models is in the perspective of a traditional analytical distinction between knowledge and action. The prescriptive models are in fact the furthest points in a chain cognitive, predictive, and decision making.

1.2.2 Need for modeling

The purpose of models is to aid in designing solutions. They are to assist understanding the problem and to aid deliberation and choice by allowing to evaluate the consequence of action before implementing them. The principle of bounded rationality assumes that the decision maker is able to optimize but only within the limits of his/her representation of the decision problem. Such a requirement is fully compatible with many results in the psychology of memory: an expert uses strategies compiled in the long-term memory and solves a decision problem with the help of his/her short-term working memory.

Problem solving is decision making that may involves heuristics such as satisfaction principle, and availability. Decision-making might be viewed as the achievement of a more or less complex information process and anchored in the search for a dominance structure: the Decision Maker updates his/her representation of the problem with the goal of finding a case where one alternative dominates all the others for example; in a mathematical approach based on dynamic systems under three principles:

1. Parsimony: the decision maker uses a small amount of information.
2. Reliability: the processed information is relevant enough to justify – personally or socially -- decision outcomes.
3. Decidability: the processed information may change from one decision to another.
4. Validation and Verification: As part of the calibration process of a model, the modeler must validate and verify the model. The term validation is applied to those processes, which seek to determine whether or not a model is correct with respect to the "real" system.

1.3 Balancing Success in Business

Without metrics, management can be a nebulous, if not impossible, exercise. A methodology for measuring success and setting goals from financial and operational viewpoints is required. With those measures, any business can manage its strategic vision and adjust it for any change. Setting a performance measure is a multi-perspective at least from financial, customer, innovation, learning, and internal business viewpoints processes.

- The **financial perspective** provides a view of how the shareholders see the company; i.e. the company's bottom-line.
- The **customer perspective** provides a view of how the customers see the company.
- While the financial perspective deals with the projected value of the company, the **innovation and learning perspective** sets measures that help the company compete in a changing business environment. The focus for this innovation is in the formation of new or the improvement of existing products and processes.
- The **internal business process perspective** provides a view of what the company must excel at to be competitive. The focus of this perspective then is the translation of customer-based measures into measures reflecting the company's internal operations.

Each of the above four perspectives must be considered with respect to four parameters:

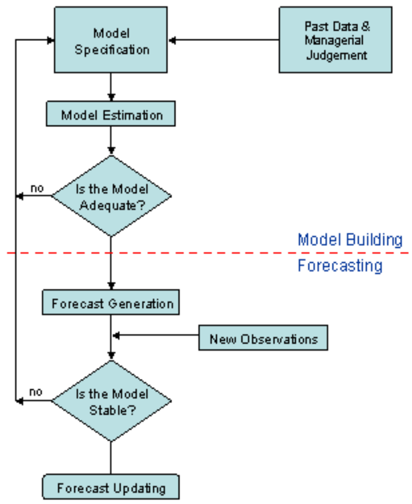
1. **Goals:** Targets needed to achieve to become successful.
2. **Measures:** Parameters required to be known for being successful.
3. **Targets:** Quantitative values used to determine success of the measure.
4. **Initiatives:** Requirements needed to meet goals.

1.4 Modeling for Forecasting:

1.4.1 Accuracy and Validation Assessments

Forecasting is a necessary input to planning whether in business or government. Often forecasts are generated subjectively and at great cost by group discussion even when relatively simple quantitative methods can perform just as well or, at least provide an informed input to such discussions.

The following flowchart highlights the systematic development of the modeling and forecasting phases:



Forecasting System:
The Model-Building and The Forecasting Phases

Modeling for Forecasting

The above modeling process is useful to:

- understand the underlying mechanism generating the time series. This includes

  describing and explaining any variations, seasonality, trend, etc.

- predict the future under "business as usual" condition.
- control the system, which is to perform the "what-if" scenarios.

1.4.2 Statistical Forecasting

The selection and implementation of the proper forecast methodology has always been an important planning and control issue for most firms and agencies. Often, the financial well-being of the entire operation rely on the accuracy of the forecast since such information will likely be used to make interrelated budgetary and operative decisions in areas of personnel management, purchasing, marketing and advertising, capital financing, etc. For example, any significant over-or-under sales forecast error may cause the firm to be overly burdened with excess inventory carrying costs or else create lost sales revenue through unanticipated item shortages. When demand is fairly stable unchanging or else growing or declining at a known constant rate, making an accurate forecast is less difficult. On the other hand if the firm has historically experienced an up-and-down sales pattern then the complexity of the forecasting task is compounded.

There are two main approaches to forecasting. Either the estimate of future value is based on an analysis of factors which are believed to influence future values i.e. the explanatory method, or else the prediction is based on an inferred study of past general data behavior over time, i.e. the extrapolation method. It is possible that both approaches will lead to the creation of accurate and useful forecasts but it must be remembered that even for a modest degree of desired accuracy the former method is often more difficult to implement and validate than the latter approach.

1.4.3 Autocorrelation

Autocorrelation is the serial correlation of equally spaced time series between its members one or more lags apart. Alternative terms are the lagged correlation, and persistence. Unlike the statistical data which are random samples allowing to perform statistical analysis the time series are strongly autocorrelated making it possible to predict and forecast. Three tools for assessing the autocorrelation of a time series are the time series plot, the lagged scatterplot, and at least the first and second order autocorrelation values.

1.4.4 Standard Error for a Stationary Time-Series

The sample mean for a time-series, has standard error not equal to $\mathbf{S / n^{1/2}}$, but $\mathbf{S[(1-r) / (n-nr)]^{1/2}}$, where S is the sample standard deviation, n is the length of the time-series, and r is its first order correlation.

1.4.5 Performance Measures and Control Chart for Examine Forecasting Errors

Beside the Standard Error there are other performance measures. The following are some of the widely used performance measures:

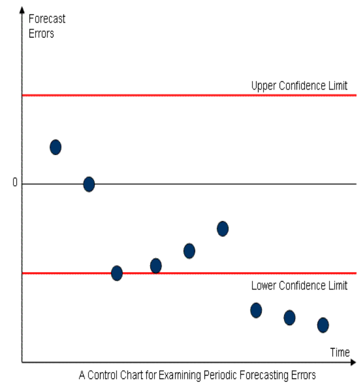
Forecast Error and Performance Measures

- Forecast error at time t = e(t) = x(t)-f(t)
- MAD Mean absolute deviation = $\sum |e(t)| / n$
- MSE Mean squared error = $\sum [e(t)]^2 / n$
- CFE Cumulative forecast error = $\sum e(t)$
- MAPE Mean absolute percentage error = $100 \sum \left[|e(t)| / x(t) \right] / n$
- Tracking signal = CFE / MAD

If the forecast error is stable then the distribution of it is approximately normal. If the forecast error is stable, then the distribution of it is approximately normal. With this in mind, we can plot and then analyze the on the control charts to see if they might be a need to revise the forecasting method being used. To do this, if we divide a normal distribution into zones, with each zone one standard deviation wide, then one obtains the approximate percentage we expect to find in each zone from a stable process.

1.4.6 Modeling for Forecasting with Accuracy and Validation Assessments

Control limits could be one-standard-error, or two-standard-error, and any point beyond these limits (i.e., outside of the error control limit) is an indication the need to revise the forecasting process, as shown below:

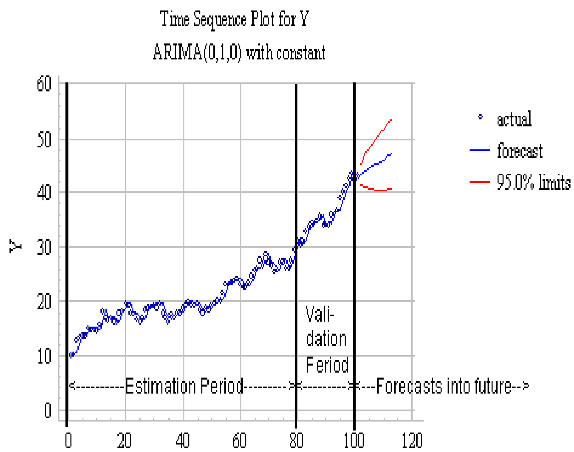


A Zone on a Control Chart for Controlling Forecasting Errors

The plotted forecast errors on this chart, not only should remain with the control limits, they should not show any obvious pattern, collectively.

Since validation is used for the purpose of establishing a model's credibility it is important that the method used for the validation is, itself, credible. Features of time series which might be revealed by examining **its graph**, with the forecasted values and the residuals behavior condition forecasting modeling.

An effective approach to modeling forecasting validation is to hold out a specific number of data points for estimation validation (i.e. estimation period) and a specific number of data points for forecasting accuracy (i.e. validation period). The data which are not held out are used to estimate the parameters of the model. The model is then tested on data in the validation period if the results are satisfactory and forecasts are then generated beyond the end of the estimation and validation periods. As an illustrative example the following graph depicts the above process on a set of data with trend component only:



Estimation Period, Validation Period, and the Forecasts

In general, the data in the estimation period are used to help select the model and to estimate its parameters. Forecasts into the future are "real" forecasts that are made for time periods beyond the end of the available data. The data in the validation period are held out during parameter estimation. These values may also be withhold during the forecasting analysis after model selection, and then one-step-ahead forecasts be made. A good model should have small error measures in both the estimation and validation periods compared to other models and its validation period statistics should be similar to its own estimation period statistics. Holding data out for validation purposes is probably the single most important diagnostic test of a model: it gives the best indication of the accuracy that can be expected when forecasting the future. **It is a rule-of-thumb that one should hold out at least 20% of data for validation purposes.**

1.5 Stationary Time Series

Stationarity has always played a major role in time series analysis. To perform forecasting, most techniques required stationarity conditions. Therefore, it is needed to establish some conditions, e.g. time series must be a first and second order stationary process.

1.5.1 First Order Stationary:

A time series is a first order stationary if expected value of X(t) remains the same for all t.

For example in economic time series, a process is first order stationary when one removes any kinds of trend by some mechanisms such as differencing.

1.5.2 Second Order Stationary

A time series is a second order stationary if it is first order stationary and covariance between X(t) and X(s) is function of length (t-s) only.

Again in economic time series, a process is second order stationary when its variance is also stabilized by some kind of transformations such as taking square root.

CHAPTER –2

CAUSAL MODELING AND FORECASTING

2.1 A Summary of Forecasting Methods

Organizations usually assign crucial forecast responsibilities to those departments and/or individuals that are best qualified and have the necessary resources at hand to make such forecast estimations under complicated demand patterns. A firm with a large ongoing operation and a technical staff comprised of statisticians, management scientists, computer analysts, etc. is in a much better position to select and make proper use of sophisticated forecast techniques than is a company with more limited resources. Notably, the bigger firm through its larger resources has a competitive edge over an unwary smaller firm and can be expected to be very diligent and detailed in estimating forecast Multi-predictor regression methods include logistic models for binary outcomes, the Cox model for right-censored survival times, repeated-measures models for longitudinal and hierarchical outcomes, and generalized linear models for counts and other outcomes. Below are outline some effective forecasting approaches, especially for short to intermediate term analysis and forecasting:

2.2 Modeling the Causal Time Series

With multiple regressions, one can use more than one predictor. It is always best however, to be parsimonious that is to use as few variables as predictors as necessary to get a reasonably accurate forecast. Multiple regressions are best modeled with commercial package such as SAS or SPSS. The forecast takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n,$$

where β_0 is the intercept, β_1, β_2, . . . β_n are coefficients representing the contribution of the independent variables X_1, X_2,..., X_n.

2.3 How to do forecasting using regression analysis

Forecasting is a prediction of what will occur in the future, and it is an uncertain process. Because of the uncertainty, the accuracy of a forecast is as important as the outcome predicted by forecasting the independent variables X_1, X_2,..., X_n. A forecast control must be used to determine if the accuracy of the forecast is within acceptable limits. Two widely used methods of forecast control are a tracking signal, and statistical control limits.

Tracking signal is computed by dividing the total residuals by their mean absolute deviation (MAD). To stay within 3 standard deviations, the tracking signal that is within 3.75 MAD is often considered to be good enough.

Statistical control limits are calculated in a manner similar to other quality control limit charts, however the residual standard deviation are used.

Multiple regressions are used when two or more independent factors are involved and it is widely used for short to intermediate term forecasting. They are used to assess which factors to include and which to exclude. They can be used to develop alternate models with different factors.

2.4 Predictions by Regression

The regression analysis has three goals: **predicting, modeling, and characterization.** What would be the logical order in which to tackle these three goals such that one task leads to and /or and justifies the other tasks? Clearly, it depends on what the prime objective is. Sometimes you wish to model in order to get better prediction. Then the order is obvious. Sometimes, you just want to understand and explain what is going on. Then modeling is again the key, though out-of-sample predicting may be used to test any model. Often modeling and predicting proceed in an iterative way and there is no 'logical order' in the broadest sense. One may model to get predictions which enable better control but iteration is again likely to be present and there are sometimes special approaches to control problems.

The following contains the main essential steps during modeling and analysis of regression model building.

Formulas and Notations:

$\bar{x} = \Sigma x / n$, This is just the mean of the x values.

$\bar{y} = \Sigma y / n$, This is just the mean of the y values.

$S_{xx} = SS_{xx} = \Sigma(x(i) - \bar{x})^2 = \Sigma x^2 - (\Sigma x)^2 / n$

$S_{yy} = SS_{yy} = \Sigma(y(i) - \bar{y})^2 = \Sigma y^2 - (\Sigma y)^2 / n$

$S_{xy} = SS_{xy} = \Sigma(x(i) - \bar{x})(y(i) - \bar{y}) = \Sigma x \cdot y - (\Sigma x) \cdot (\Sigma y) / n$

Slope $m = SS_{xy} / SS_{xx}$

Intercept, $b = \bar{y} - m . \bar{x}$

y-predicted = yhat(i) = m·x(i) + b.

Residual(i) = Error(i) = y – yhat(i).

$SSE = S_{res} = SS_{res} = SS_{errors} = \Sigma[y(i) - yhat(i)]^2$.

Standard deviation of residuals = $s = S_{res} = S_{errors} = [SS_{res} / (n-2)]^{1/2}$.

Standard error of the slope (m) = $S_{res} / SS_{xx}^{1/2}$.

Standard error of the intercept (b) = $S_{res}[(SS_{xx} + n. \bar{x}^2) /(n \cdot SS_{xx}]^{1/2}$.

2.5 Planning, Development, and Maintenance of a Linear Model

2.5.1 Planning:

1. Define the problem; select response; suggest variables.
2. Are the proposed variables fundamental to the problem, and are they variables? Are they measurable/countable? Can one get a complete set of observations at the

same time? Ordinary regression analysis does not assume that the independent variables are measured without error. However, they are conditioned on whatever errors happened to be present in the independent data set.

3. Is the problem potentially solvable?

4. Find correlation matrix and first regression runs (for a subset of data).
Find the basic statistics, correlation matrix.
How difficult is the problem? Compute the Variance Inflation Factor:

$$VIF = 1/(1 - r_{ij}), \quad \text{for all } i, j.$$

5. Establish goal; prepare budget and time table.

a. The final equation should have **Adjusted** $R^2 = 0.8$ (say).

b. Coefficient of Variation of say; less than 0.10

c. Number of predictors should not exceed p (say, 3), (for example for $p = 3$, one needs at least 30 points). Even if all the usual assumptions for a regression model are satisfied, over-fitting can ruin a model's usefulness. The widely used approach is the data reduction method to deal with the cases where the number of potential predictors is large in comparison with the number of observations.

d. All estimated coefficients must be significant at $\mu = 0.05$ (say).

e. No pattern in the residuals

6. Are goals and budget acceptable?

2.5.2 Development of the Model

(i) Collect date; check the quality of date; plot; try models; check the regression conditions.

(ii) Consult experts for criticism.
Plot new variable and examine same fitted model. Also transformed Predictor Variable may be used.

(iii) Are goals met?

Have you found "the best" model

2.5.3. Validation and Maintenance of the Model

(i) Are parameters stable over the sample space?

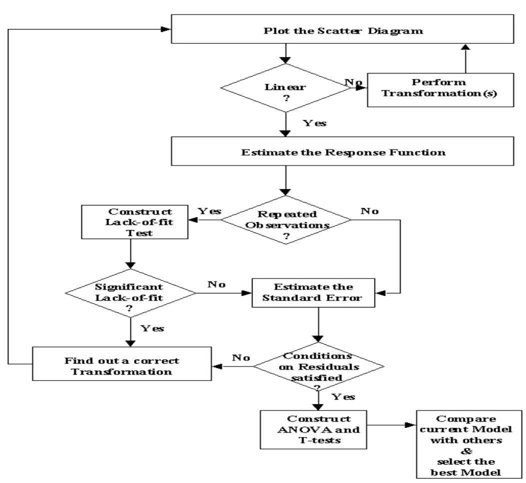
(ii) Is there a lack of fit?

(iii) Are the coefficients reasonable? Are any obvious variables missing? Is the equation usable for control or for prediction?

(iv) Maintenance of the Model.

Need to have control chart to check the model periodically by statistical techniques.

Steps in Regression Analysis Procedure



Regression Analysis Process

2.6 Trend Analysis

This uses linear and nonlinear regression with time as the explanatory variable. It is used where pattern over time have a long-term trend. Unlike most time-series forecasting techniques, the Trend Analysis does not assume the condition of equally spaced time series. Nonlinear regression does not assume a linear relationship between variables. It is frequently used when time is the independent variable.

2.7Modeling Seasonality and Trend

Seasonality is a pattern that repeats for each period. For example annual seasonal pattern has a cycle that is 12 periods long, if the periods are months, or 4 periods long if the periods are quarters. We need to get an estimate of the seasonal index for each month, or other periods, such as quarter, week, etc, depending on the data availability.

2.7.1. Seasonal Index

Seasonal index represents the extent of seasonal influence for a particular segment of the year. The calculation involves a comparison of the expected values of that period to the grand mean.

A seasonal index is how much the average for that particular period tends to be above (or below) the grand average. Therefore, to get an accurate estimate for the seasonal index, we compute the average of the first period of the cycle, and the second period, etc, and divide each by the overall average. The formula for computing seasonal factors is:

$$S_i = D_i/D,$$

where:

S_i = the seasonal index for i^{th} period,

D_i = the average values of i^{th} period,

D = grand average,

i = the i^{th} seasonal period of the cycle.

A seasonal index of 1.00 for a particular month indicates that the expected value of that month is 1/12 of the overall average. A seasonal index of 1.25 indicates that the expected value for that month is 25% greater than 1/12 of the overall average. A seasonal index of 80 indicates that the expected value for that month is 20% less than 1/12 of the overall average.

2.7.2. Deseasonalizing Process

Deseasonalizing the data, also called Seasonal Adjustment is the process of removing recurrent and periodic variations over a short time frame, e.g., weeks, quarters, months. Therefore, seasonal variations are regularly repeating movements in series values that can be tied to recurring events. The Deseasonalized data is obtained by simply dividing each time series observation by the corresponding seasonal index.

Almost all time series published by the US government are already deseasonalized using the seasonal index to unmasking the underlying trends in the data, which could have been caused by the seasonality factor.

2.7.3. Forecasting

Incorporating seasonality in a forecast is useful when the time series has both trend and seasonal components. The final step in the forecast is to use the seasonal index to adjust the trend projection. One simple way to forecast using a seasonal adjustment is to use a seasonal factor in combination with an appropriate underlying trend of total value of cycles.

2.7.4. A Numerical Application

The following table provides monthly sales ($1000) at a college bookstore. The sales show a seasonal pattern, with the greatest number when the college is in session and decrease during the summer months.

M T	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1	196	188	192	164	140	120	112	140	160	168	192	200	1972
2	200	188	192	164	140	122	132	144	176	168	196	194	2016
3	196	212	202	180	150	140	156	144	164	186	200	230	2160
4	242	240	196	220	200	192	176	184	204	228	250	260	2592
Mean:	208.6	207.0	192.6	182.0	157.6	143.6	144.0	153.0	177.6	187.6	209.6	221.0	2185
Index:	1.14	1.14	1.06	1.00	0.87	0.79	0.79	0.84	0.97	1.03	1.15	1.22	12

Suppose we wish to calculate seasonal factors and a trend, then calculate the forecasted sales for July in year 5.

The first step in the seasonal forecast will be to compute monthly indices using the past four-year sales. For example, for January the index is:

S(Jan) = D(Jan)/D = 208.6/181.84 = 1.14,

where D(Jan) is the mean of all four January months, and D is the grand mean of all past four-year sales.

Similar calculations are made for all other months. Indices are summarized in the last row of the above table. The mean (average value) for the monthly indices adds up to 12, which is the number of periods in a year for the monthly data.

Next, a **linear trend** often is calculated using the annual sales:

Y = 1684 + 200.4T,

The main question is whether this equation represents the trend.

Determination of the Annual Trend for the Numerical Example			
Year No:	**Actual Sales**	**Linear Regression**	**Quadratic Regression**
1	1972	1884	1981
2	2016	2085	1988
3	2160	2285	2188
4	2592	2486	2583

Often fitting a straight line to the seasonal data is misleading. By constructing the scatter-diagram, we notice that a Parabola might be a better fit. Using the **Polynomial Regression**, the estimated quadratic trend is:

$$\mathbf{Y = 2169 - 284.6T + 97T^2}$$

Predicted values using both the linear and the quadratic trends are presented in the above tables. Comparing the predicted values of the two models with the actual data indicates that the quadratic trend is a much superior fit than the linear one, as often expected.

We can now forecast the next annual sales; which, corresponds to year 5, or T = 5 in the above quadratic equation:

$$\mathbf{Y = 2169 - 284.6(5) + 97(5)^2 = 3171}$$

sales for the following year. The average monthly sales during next year is, therefore: 3171/12 = 264.25.

Finally, the forecast for month of July is calculated by multiplying the average monthly sales forecast by the July seasonal index, which is 0.79; i.e., (264.25).(0.79) or 209.

2.8 Trend Removal and Cyclical Analysis

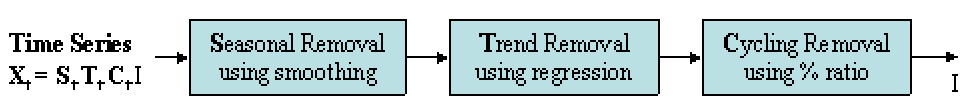
The cycles can be easily studied if the trend itself is removed. This is done by expressing each actual value in the time series as a percentage of the calculated trend for the same date. The resulting time series has no trend, but oscillates around a central value of 100.

2.9 Decomposition Analysis

It is the pattern generated by the time series and not necessarily the individual data values that offers to the manager who is an observer, a planner, or a controller of the system. Therefore, the Decomposition Analysis is used to identify several patterns that appear simultaneously in a time series. A variety of factors are likely influencing data. It is very important in the study that these different influences or components be separated or decomposed out of the 'raw' data levels. In general, there are four types of components in time series analysis: **S**easonality, **T**rend, **C**ycling and **I**rregularity.

$$X_t = S_t \cdot T_t \cdot C_t \cdot I$$

The first three components are deterministic which are called "Signals", while the last component is a random variable, which is called "Noise". To be able to make a proper forecast, we must know to what extent each component is present in the data. Hence, to understand and measure these components, the forecast procedure involves initially removing the component effects from the data (decomposition). After the effects are measured, making a forecast involves putting back the components on forecast estimates (recomposition). The time series decomposition process is depicted by the following flowchart:



The Three Signals Decomposition and Its Reversal Processes For Forecasting

Definitions of the major components in the above flowchart:

Seasonal variation: When a repetitive pattern is observed over some time horizon, the series is said to have seasonal behavior. Seasonal effects are usually associated with calendar or climatic changes. Seasonal variation is frequently tied to yearly cycles.

Trend: A time series may be stationary or exhibit trend over time. Long-term trend is typically modeled as a linear, quadratic or exponential function.

Cyclical variation: An upturn or downturn not tied to seasonal variation. Usually results from changes in economic conditions.

(i) Seasonalities are regular fluctuations which are repeated from year to year with about the same timing and level of intensity. The first step of a times series decomposition is to remove seasonal effects in the data. Without deseasonalizing the data, we may, for example, incorrectly infer that recent increase patterns will continue indefinitely; i.e., a growth trend is present, when actually the increase is 'just because it is that time of the year'; i.e., due to regular seasonal peaks. To measure seasonal effects, we calculate a series of seasonal indexes. A practical and widely used method to compute these indexes is the ratio-to-moving-average approach. From such indexes, we may quantitatively measure how far above or below a given period stands in comparison to the expected or 'business as usual' data period (the expected data are represented by a seasonal index of 100%, or 1.0).

(ii) Trend is growth or decay that is the tendencies for data to increase or decrease fairly steadily over time. Using the deseasonalized data, we now wish to consider the growth trend as noted in our initial inspection of the time series. Measurement of the trend component is done by fitting a line or any other function. This fitted function is calculated by the method of least squares and represents the overall trend of the data over time.

(iii) Cyclic oscillations are general up-and-down data changes; due to changes e.g., in the overall economic environment (not caused by seasonal effects) such as recession-and-expansion. To measure how the general cycle affects data levels, we calculate a series of cyclic indexes. Theoretically, the deseasonalized data still contains trend, cyclic, and irregular components. Also, we believe predicted data levels using the trend equation do represent pure trend effects. Thus, it stands to reason that the ratio of these respective data values should provide an index which reflects cyclic and irregular components only. As the business cycle is usually longer than the seasonal cycle, it should be understood that cyclic analysis is not expected to be as accurate as a seasonal analysis.

Due to the tremendous complexity of general economic factors on long term behavior, a general approximation of the cyclic factor is the more realistic aim. Thus, the specific sharp upturns and downturns are not so much the primary interest as the general tendency of the cyclic effect to gradually move in either direction. To study the general cyclic movement rather than precise cyclic changes (which may falsely indicate more accurately than is present under this situation), we 'smooth' out the cyclic plot by replacing each index calculation often with a centered 3-period moving average. The reader should note that as the number of periods in the moving average increases, the smoother or flatter the data become. The choice of 3 periods perhaps viewed as slightly subjective may be justified as an attempt to smooth out the many up-and-down minor actions of the cycle index plot so that only the major changes remain.

(iv) Irregularities (**I**) are any fluctuations not classified as one of the above. This component of the time series is unexplainable; therefore it is unpredictable. Estimation of **I** can be expected only when its variance is not too large. *Otherwise, it is not possible to decompose the series*. If the magnitude of variation is large, the projection for the future values will be inaccurate. The best one can do is to give a probabilistic interval for the future value given the probability of **I** is known.

(v) Making a Forecast: At this point of the analysis, after we have completed the study of the time series components, we now project the future values in making forecasts for the next few periods. The procedure is summarized below.

Step 1: Compute the future trend level using the trend equation.

Step 2: Multiply the trend level from Step 1 by the period seasonal index to include seasonal effects.

Step 3: Multiply the result of Step 2 by the projected cyclic index to include cyclic effects and get the final forecast result.

CHAPTER 3

SMOOTHING TECHNIQUES

3.1 Introduction

Smoothing Techniques: A time series is a sequence of observations, which are ordered in time. Inherent in the collection of data taken over time is some form of random variation. There exist methods for reducing of canceling the effect due to random variation. A widely used technique is "smoothing". This technique, when properly applied, reveals more clearly the underlying trend, seasonal and cyclic components.

Smoothing techniques are used to reduce irregularities (random fluctuations) in time series data. They provide a clearer view of the true underlying behavior of the series. Moving averages rank among the most popular techniques for the preprocessing of time series. They are used to filter random "white noise" from the data, to make the time series smoother or even to emphasize certain informational components contained in the time series.

Exponential smoothing is a very popular scheme to produce a smoothed time series. Whereas in moving averages the past observations are weighted equally, Exponential Smoothing assigns exponentially decreasing weights as the observation get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. Double exponential smoothing is better at handling trends. Triple Exponential Smoothing is better at handling parabola trends.

Exponential smoothing is a widely method used of forecasting based on the time series itself. Unlike regression models, exponential smoothing does not imposed any deterministic model to fit the series other than what is inherent in the time series itself.

3.2 Moving Averages and Weighted Moving Averages

3.2.1 Simple Moving Averages

The best-known forecasting methods is the moving averages or simply takes a certain number of past periods and add them together; then divide by the number of periods. Simple Moving Averages (MA) is effective and efficient approach provided the time series is stationary in both mean and variance. The following formula is used in finding the moving average of order n, MA(n) for a period t+1,

$MA_{t+1} = [D_t + D_{t-1} + ... + D_{t-n+1}] / n$

where n is the number of observations used in the calculation.

The forecast for time period t + 1 is the forecast for all future time periods. However, this forecast is revised only when new data becomes available.

3.2.2 Weighted Moving Average

Very powerful and economical. They are widely used where repeated forecasts required-uses methods like sum-of-the-digits and trend adjustment methods. As an example, a Weighted Moving Averages is:

Weighted $MA(3) = w_1.D_t + w_2.D_{t-1} + w_3.D_{t-2}$

where the weights are any positive numbers such that: w1 + w2 + w3 = 1. A typical weights for this example is, $w_1 = 3/(1 + 2 + 3) = 3/6$, $w_2 = 2/6$, and $w_3 = 1/6$.

3.2.3 An illustrative numerical example

The moving average and weighted moving average of order five are calculated in the following table.

Week	Sales ($1000)	MA(5)	WMA(5)
1	105	-	-
2	100	-	-
3	105	-	-
4	95	-	-
5	100	101	100
6	95	99	98
7	105	100	100
8	120	103	107
9	115	107	111
10	125	117	116
11	120	120	119
12	120	120	119

3.3 Exponential Smoothing Techniques

Exponential Smoothing Techniques: One of the most successful forecasting methods is the exponential smoothing (ES) techniques. Moreover, it can be modified efficiently to use effectively for time series with seasonal patterns. It is also easy to adjust for past errors-easy to prepare follow-on forecasts, ideal for situations where many forecasts must be prepared, several different forms are used depending on presence of trend or cyclical variations. In short, an ES is an averaging technique that uses unequal weights; however, the weights applied to past observations decline in an exponential manner.

Single Exponential Smoothing: It calculates the smoothed series as a damping coefficient times the actual series plus 1 minus the damping coefficient times the lagged value of the smoothed series. The extrapolated smoothed series is a constant, equal to the last value of the smoothed series during the period when actual data on the underlying series are available. While the simple Moving Average method is a special case of the ES, the ES is more parsimonious in its data usage.

$$F_{t+1} = \alpha D_t + (1 - \alpha) F_t$$

where:

- D_t is the actual value
- F_t is the forecasted value
- α is the weighting factor, which ranges from 0 to 1
- t is the current time period.

It can be noticed that the smoothed value becomes the forecast for period t + 1.

A small α provides a detectable and visible smoothing. While a large α provides a fast response to the recent changes in the time series but provides a smaller amount of smoothing. Notice that the exponential smoothing and simple moving average techniques will generate forecasts having the same average age of information if moving average of order n is the integer part of $(2-\alpha)/\alpha$.

An exponential smoothing over an already smoothed time series is called **double-exponential smoothing**. In some cases, it might be necessary to extend it even to a **triple-exponential smoothing**. While simple exponential smoothing requires stationary condition, the double-exponential smoothing can capture linear trends, and triple-exponential smoothing can handle almost all other business time series.

Double Exponential Smoothing: It applies the process described above three to account for linear trend. The extrapolated series has a constant growth rate, equal to the growth of the smoothed series at the end of the data period.

Triple Double Exponential Smoothing: It applies the process described above three to account for nonlinear trend.

3.4 Exponenentially Weighted Moving Average

Exponenentially Weighted Moving Average: Suppose each day's forecast value is based on the previous day's value so that the weight of each observation drops exponentially the further back (k) in time it is. The weight of any individual is

$\alpha(1 - \alpha)^k$,   where a is the smoothing constant.

An exponenentially weighted moving average with a smoothing constant α, corresponds roughly to a simple moving average of length n, where α and n are related by

$\alpha = 2/(n+1)$ OR $n = (2 - \alpha)/\alpha$.

Thus, for example, an exponenentially weighted moving average with a smoothing constant equal to 0.1 would correspond roughly to a 19 day moving average. And a 40-day simple moving average would correspond roughly to an exponentially weighted moving average with a smoothing constant equal to 0.04878.

This approximation is helpful, however, it is harder to update, and may not correspond to an optimal forecast.

3.5 Holt's Linear Exponential Smoothing Technique

Holt's Linear Exponential Smoothing Technique: Let the series $\{ y_t \}$ is non-seasonal but does display trend. Both the current level and the current trend are to be estimated. Here the trend T_t at time t can be defined as the difference between the current and previous level.

The updating equations express ideas similar to those for exponential smoothing. The equations are:

$L_t = \alpha\, y_t + (1 - \alpha)\, F_t$

for the level and

$$T_t = \beta\, (L_t - L_{t-1}) + (1 - \beta)\, T_{t-1}$$

for the trend. There are two smoothing parameters α and β; both must be positive and less than one. Then the forecasting for k periods into the future is:

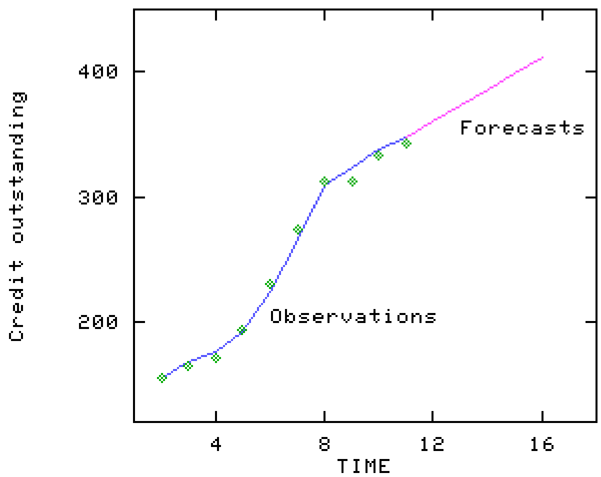
$F_{n+k} = L_n + k.\, T_n$

Given that the level and trend remain unchanged, the initial (starting) values are

$$T_2 = y_2 - y_1, \qquad L_2 = y_2, \quad \text{and} \quad F_3 = L_2 + T_2$$

An Application: A company's credit outstanding has been increasing at a relatively constant rate over time:

Applying the Holt's techniques with smoothing with parameters $\alpha = 0.7$ and $\beta = 0.6$, a graphical representation of the time series, its forecasts, together wit a few-step ahead forecasts, are depicted below:



Year-end Past credit	
Year	**credit (in millions)**
1	133
2	155
3	165
4	171
5	194
6	231
7	274
8	312
9	313
10	333
11	343

K-Period Ahead Forecast	
K	**Forecast (in millions)**
1	359.7
2	372.6
3	385.4
4	398.3

Demonstration of the calculation procedure, with $\alpha = 0.7$ and $\beta = 0.6$

$\mathbf{L_2 = y_2 = 155}$, $\mathbf{T_2 = y_2 - y_1 = 155 - 133 = 22}$

$\mathbf{L_3 = .7\, y_3 + (1 - .7)\, F_3}$, $\mathbf{T_3 = .6\, (L_3 - L_2) + (1 - .6)\, T_2}$

$F_4 = L_3 + T_3$, $F_3 = L_2 + T_2$, $L_3 = .7\, y_3 + (1 - .7)\, F_3$, $T_3 = .6\, (L_3 - L_2) + (1 - .6)\, T_2$, $F_4 = L_3 + T_3$

3.6 The Holt-Winters' Forecasting Technique:

Now in addition to Holt parameters, suppose that the series exhibits multiplicative seasonality and let S_t be the multiplicative seasonal factor at time t. Suppose also that there are s periods in a year, so s=4 for quarterly data and s=12 for monthly data. S_{t-s} is the seasonal factor in the same period last year.

In some time series, seasonal variation is so strong it obscures any trends or cycles, which are very important for the understanding of the process being observed. Winters' smoothing method can remove seasonality and makes long term fluctuations in the series stand out more clearly. A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time it can be said that there is evidence of a trend in the series. The updating equations are:

$$\mathbf{L_t = \alpha\, (L_{t-1} + T_{t-1}) + (1 - \alpha)\, y_t / S_{t-s}}$$

for the level,

$$\mathbf{T_t = \beta\, (L_t - L_{t-1}) + (1 - \beta)\, T_{t-1}}$$

for the trend, and

$$\mathbf{S_t = \gamma\, S_{t-s} + (1- \gamma)\, y_t / L_t}$$

for the seasonal factor.

There are thus three smoothing parameters α , β, and γ all must be positive and less than one.

To obtain starting values, one may use the first a few year data. For example for quarterly data, to estimate the level, one may use a centered 4-point moving average:

$$\mathbf{L_{10} = (y_8 + 2y_9 + 2y_{10} + 2y_{11} + y_{12}) / 8}$$

as the level estimate in period 10. This will extract the seasonal component from a series with 4 measurements over each year.

$$T_{10} = L_{10} - L_9$$

as the trend estimate for period 10.

$$S_7 = (y_7 / L_7 + y_3 / L_3) / 2$$

as the seasonal factor in period 7. Similarly,

$S_8 = (y_8 / L_8 + y_4 / L_4) / 2$, $S_9 = (y_9 / L_9 + y_5 / L_5) / 2$, $S_{10} = (y_{10} / L_{10} + y_6 / L_6) / 2$

For Monthly Data, then correspondingly one uses a centered 12-point moving average:

$$L_{30} = (y_{24} + 2y_{25} + 2y_{26} +.....+ 2y_{35} + y_{36}) / 24$$

as the level estimate in period 30.

$$T_{30} = L_{30} - L_{29}$$

as the trend estimate for period 30.

$$S_{19} = (y_{19} / L_{19} + y_7 / L_7) / 2$$

as the estimate of the seasonal factor in period 19, and so on, up to 30:

$$S_{30} = (y_{30} / L_{30} + y_{18} / L_{18}) / 2$$

Then the forecasting k periods into the future is:

$$F_{n+k} = (L_n + k. T_n) S_{t+k-s}, \quad \text{for } k = 1, 2, ...,s$$

3.7 Conclusion

A time series is a sequence of observations which are ordered in time. Inherent in the collection of data taken over time is some form of random variation. There exist methods for reducing of canceling the effect due to random variation. Widely used techniques are "smoothing". These techniques, when properly applied, reveals more clearly the underlying trends. In other words, smoothing techniques are used to reduce irregularities (random fluctuations) in time series data. They provide a clearer view of the true underlying behavior of the series.

Exponential smoothing has proven through the years to be very useful in many forecasting situations. Holt first suggested it for non-seasonal time series with or without trends. Winters generalized the method to include seasonality, hence the name: Holt-Winters Method. Holt-Winters method has 3 updating equations, each with a constant that ranges from (0 to 1). The equations are intended to give more weight to recent observations and less weight to observations further in the past. This form of exponential smoothing can be used for less-than-annual periods (e.g., for monthly series). It uses smoothing parameters to estimate the level, trend, and seasonality. Moreover, there are two different procedures, depending on whether seasonality is modeled in an additive or multiplicative way. Its multiplicative version is presented; the additive can be applied on an ant-logarithmic function of the data.

The single exponential smoothing emphasizes the short-range perspective; it sets the level to the last observation and is based on the condition that there is no trend. The linear regression, which fits a least squares line to the historical data (or transformed historical data), represents the long range, which is conditioned on the basic trend. Holt’s linear exponential smoothing captures information about recent trend. The parameters in Holt’s model are the levels-parameter which should be decreased when the amount of data variation is large, and trends-parameter should be increased if the recent trend direction is supported by the causal some factors.

Since finding three optimal, or even near optimal, parameters for updating equations is not an easy task, an alternative approach to Holt-Winters methods is to deseasonalize the data and then use exponential smoothing. Moreover, in some time series, seasonal variation is so strong it obscures any trends or cycles, which are very important for the understanding of the process being observed. Smoothing can remove seasonality and makes long term fluctuations in the series stand out more clearly. A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time one can say that there is evidence of a trend in the series.

How to compare several smoothing methods: Although there are numerical indicators for assessing the accuracy of the forecasting technique, the most widely approach is in using visual comparison of several forecasts to assess their accuracy and choose among the various forecasting methods. In this approach, one must plot (using, e.g., Excel) on the same graph the original values of a time series variable and the predicted values from several different forecasting methods, thus facilitating a visual comparison.

CHAPTER 4

BOX-JENKINS METHODOLOGY

4.1 Introduction

Forecasting Basics: The basic idea behind self-projecting time series forecasting models is to find a mathematical formula that will approximately generate the historical patterns in a time series.

Time Series: A time series is a set of numbers that measures the status of some activity over time. It is the historical record of some activity, with measurements taken at equally spaced intervals (exception: monthly) with a consistency in the activity and the method of measurement.

Approaches to time Series Forecasting: There are two basic approaches to forecasting time series: the self-projecting time series and the cause-and-effect approach. Cause-and-effect methods attempt to forecast based on underlying series that are believed to cause the behavior of the original series. The self-projecting time series uses only the time series data of the activity to be forecast to generate forecasts. This latter approach is typically less expensive to apply and requires far less data and is useful for short, to medium-term forecasting.

Box-Jenkins Forecasting Method: The univariate version of this methodology is a self- projecting time series forecasting method. The underlying goal is to find an appropriate formula so that the residuals are as small as possible and exhibit no pattern. The model- building process involves a few steps, repeated as necessary, to end up with a specific formula that replicates the patterns in the series as closely as possible and also produces accurate forecasts.

Box-Jenkins Methodology

Box-Jenkins forecasting models are based on statistical concepts and principles and are able to model a wide spectrum of time series behavior. It has a large class of models to choose from and a systematic approach for identifying the correct model form. There are both statistical tests for verifying model validity and statistical measures of forecast uncertainty. In contrast, traditional forecasting models offer a limited number of models relative to the complex behavior of many time series, with little in the way of guidelines and statistical tests for verifying the validity of the selected model.

Data: The misuse, misunderstanding, and inaccuracy of forecasts are often the result of not appreciating the nature of the data in hand. The consistency of the data must be insured, and it must be clear what the data represents and how it was gathered or calculated. As a rule of thumb, Box-Jenkins requires at least 40 or 50 equally-spaced periods of data. The data must also be edited to deal with extreme or missing values or other distortions through the use of functions such as log or inverse to achieve stabilization.

Preliminary Model Identification Procedure: A preliminary Box-Jenkins analysis with a plot of the initial data should be run as the starting point in determining an appropriate model. The input data must be adjusted to form a stationary series, one whose values vary more or less uniformly about a fixed level over time. Apparent trends can be adjusted by having the model apply a technique of "regular differencing," a process of computing the difference between every two successive values, computing a differenced series which has overall trend behavior removed. If a single differencing does not achieve stationarity, it may be repeated, although rarely, if ever, are more than two regular differencing required. Where irregularities in the differenced series continue to be displayed, log or inverse functions can be specified to stabilize the series, such that the remaining residual plot displays values approaching zero and without any pattern. This is the error term, equivalent to pure, white noise.

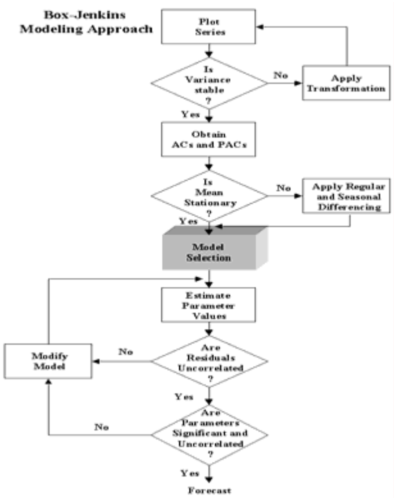
Pure Random Series: On the other hand, if the initial data series displays neither trend nor seasonality, and the residual plot shows essentially zero values within a 95% confidence level and these residual values display no pattern, then there is no real-world statistical problem to solve and one goes on to other things.

Model Identification Background

Basic Model: With a stationary series in place, a basic model can now be identified. Three basic models exist, AR (autoregressive), MA (moving average) and a combined ARMA in addition to the previously specified RD (regular differencing): These comprise the available tools. When regular differencing is applied, together with AR and MA, they are referred to as ARIMA, with the I indicating "integrated" and referencing the differencing procedure.

Seasonality: In addition to trend, which has now been provided for, stationary series quite commonly display seasonal behavior where a certain basic pattern tends to be repeated at regular seasonal intervals. The seasonal pattern may additionally frequently display constant change over time as well. Just as regular differencing was applied to the overall trending series, seasonal differencing (SD) is applied to seasonal non-stationarity as well. And as autoregressive and moving average tools are available with the overall series, so too, are they available for seasonal phenomena using seasonal autoregressive parameters (SAR) and seasonal moving average parameters (SMA).

Establishing Seasonality: The need for seasonal autoregression (SAR) and seasonal moving average (SMA) parameters is established by examining the autocorrelation and partial autocorrelation patterns of a stationary series at lags that are multiples of the number of periods per season. These parameters are required if the values at lags s, 2s, etc. are nonzero and display patterns associated with the theoretical patterns for such models. Seasonal differencing is indicated if the autocorrelations at the seasonal lags do not decrease rapidly.



B-J Modeling Approach to Forecasting

Referring to the above flow chart know that, the variance of the errors of the underlying model must be invariant, i.e., constant. This means that the variance for each subgroup of data is the same and does not depend on the level or the point in time. If this is violated then one can remedy this by stabilizing the variance. Make sure that there are no deterministic patterns in the data. Also, one must not have any pulses or one-time unusual values. Additionally, there should be no level or step shifts. Also, no seasonal pulses should be present.

The reason for all of this is that if they do exist, then the sample autocorrelation and partial autocorrelation will seem to imply ARIMA structure. Also, the presence of these kinds of model components can obfuscate or hide structure. For example, a single outlier or pulse can create an effect where the structure is masked by the outlier.

4.2 Autoregressive Models

The autoregressive model is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous outputs and inputs, such as:

$Y(t) = \beta_1 + \beta_2 Y(t\text{-}1) + \beta_3 X(t\text{-}1) + \varepsilon_t$, where X(t-1) and Y(t-1) are the actual value (inputs) and the forecast (outputs), respectively. These types of regressions are often referred to as *Distributed Lag Autoregressive Models*, *Geometric Distributed Lags,* and *Adaptive Models in Expectation* , among others.

A model which depends only on the previous outputs of the system is called an autoregressive model (AR), while a model which depends only on the inputs to the system is called a moving average model (MA), and of course a model based on both inputs and outputs is an autoregressive-moving-average model (ARMA). Note that by definition, the AR model has only poles while the MA model has only zeros. Deriving the autoregressive model (AR) involves estimating the coefficients of the model using the method of least squared error.

Autoregressive processes as their name implies, regress on themselves. If an observation made at time (t), then, p-order, [AR(p)], autoregressive model satisfies the equation:

$$\mathbf{X(t) = \Phi_0 + \Phi_1 X(t\text{-}1) + \Phi_2 X(t\text{-}2) + \Phi_2 X(t\text{-}3) + \dots + \Phi_p X(t\text{-}p) + \varepsilon_t,}$$

where ε_t is a White-Noise series.

The current value of the series is a linear combination of the p most recent past values of itself plus an error term, which incorporates everything new in the series at time t that is not explained by the past values. This is like a multiple regressions model but is regressed not on independent variables, but on past values; hence the term "Autoregressive" is used.

Autocorrelation: An important guide to the properties of a time series is provided by a series of quantities called sample autocorrelation coefficients or serial correlation coefficient, which measures the correlation between observations at different distances apart. These coefficients often provide insight into the probability model which generated the data. The sample autocorrelation coefficient is similar to the ordinary correlation coefficient between two variables

(x) and (y), except that it is applied to a single time series to see if successive observations are correlated.

Given (N) observations on discrete time series one can form (N - 1) pairs of observations. Regarding the first observation in each pair as one variable, and the second observation as a second variable, the correlation coefficient is called autocorrelation coefficient of order one.

Correlogram: A useful aid in interpreting a set of autocorrelation coefficients is a graph called a correlogram, and it is plotted against the lag(k); where is the autocorrelation coefficient at lag(k). A correlogram can be used to get a general understanding on the following aspects of our time series:

1.A random series: if a time series is completely random then for Large (N), will be approximately zero for all non-zero values of (k).
2.Short-term correlation: stationary series often exhibit short-term correlation characterized by a fairly large value of 2 or 3 more correlation coefficients which, while significantly greater than zero, tend to get successively smaller.
3.Non-stationary series: If a time series contains a trend, then the values of will not come to zero except for very large values of the lag.
4.Seasonal fluctuations: Common autoregressive models with seasonal fluctuations, of period s are:

$$\mathbf{X(t) = a + b\ X(t\text{-}s) + \varepsilon_t}$$

and

$$\mathbf{X(t) = a + b\ X(t\text{-}s) + c\ X(t\text{-}2s) + \varepsilon_t}$$

where ε_t is a White-Noise series.

Partial Autocorrelation: A partial autocorrelation coefficient for order k measures the strength of correlation among pairs of entries in the time series while accounting for (i.e., removing the effects of) all autocorrelations below order k. For example, the partial autocorrelation coefficient for order k=5 is computed in such a manner that the effects of the k=1, 2, 3, and 4 partial autocorrelations have been excluded. The partial autocorrelation coefficient of any particular order is the same as the autoregression coefficient of the same order.

Fitting an Autoregressive Model: If an autoregressive model is thought to be appropriate for modeling a given time series then there are two related questions to be answered: (1) What is the order of the model? and (2) How can one estimate the parameters of the model?

The parameters of an autoregressive model can be estimated by minimizing the sum of squares residual with respect to each parameter, but to determine the order of the autoregressive model is not easy particularly when the system being modeled has a biological interpretation.

One approach is, to fit AR models of progressively higher order, to calculate the residual sum of squares for each value of p; and to plot this against p. It may then be possible to see the value of p where the curve "flattens out" and the addition of extra parameters gives little improvement in fit.

Selection Criteria: Several criteria may be specified for choosing a model format, given the simple and partial autocorrelation correlogram for a series:

1. If none of the simple autocorrelations is significantly different from zero, the series is essentially a random number or white-noise series, which is not amenable to autoregressive modeling.
2. If the simple autocorrelations decrease linearly, passing through zero to become negative, or if the simple autocorrelations exhibit a wave-like cyclical pattern, passing through zero several times, the series is not stationary; it must be differenced one or more times before it may be modeled with an autoregressive process.

3. If the simple autocorrelations exhibit seasonality; i.e., there are autocorrelation peaks every dozen or so (in monthly data) lags, the series is not stationary; it must be differenced with a gap approximately equal to the seasonal interval before further modeling.

4. If the simple autocorrelations decrease exponentially but approach zero gradually, while the partial autocorrelations are significantly non-zero through some small number of lags beyond which they are not significantly different from zero, the series should be modeled with an autoregressive process.

5. If the partial autocorrelations decrease exponentially but approach zero gradually, while the simple autocorrelations are significantly non-zero through some small number of lags beyond which they are not significantly different from zero, the series should be modeled with a moving average process.

6. If the partial and simple autocorrelations both converge upon zero for successively longer lags, but neither actually reaches zero after any particular lag, the series may be modeled by a combination of autoregressive and moving average process.

Forecasting: The estimates of the parameters are used in Forecasting to calculate new values of the series, beyond those included in the input data set and confidence intervals for those predicted values.

An Illustrative Numerical Example: The analyst at Aron Company has a time series of readings for the monthly sales to be forecasted. The data are shown in the following table:

Aron Company Monthly Sales ($1000)									
t	X(t)	t	X(t)	t	X(t)	t	X(t)	t	X(t)
1	50.8	6	48.1	11	50.8	16	53.1	21	49.7
2	50.3	7	50.1	12	52.8	17	51.6	22	50.3
3	50.2	8	48.7	13	53.0	18	50.8	23	49.9
4	48.7	9	49.2	14	51.8	19	50.6	24	51.8
5	48.5	10	51.1	15	53.6	20	49.7	25	51.0

By constructing and studying the plot of the data one notices that the series drifts above and below the mean of about 50.6.

$$\mathbf{X(t) = \Phi_0 + \Phi_1 X(t\text{-}1) + \varepsilon_t,}$$

where ε_t is a White-Noise series.

Stationary Condition: The AR(1) is stable if the slope is within the open interval (-1, 1), that is:

$$\mathbf{| \Phi_1| < 1}$$

is expressed as a null hypothesis H_0 that must be tested before forecasting stage. To test this hypothesis, one must replace the t-test used in the regression analysis for testing the slope with the τ-test introduced by the two economists, Dickey and Fuller. The estimated AR (1) model is:

$$\mathbf{X(t) = 14.44 + 0.715\, X(t\text{-}1)}$$

The 3-step ahead forecasts are:

$$\mathbf{X(26) = 14.44 + 0.715\, X(25) = 14.44 + 0.715\, (51.0) = 50.91}$$
$$\mathbf{X(27) = 14.44 + 0.715\, X(26) = 14.44 + 0.715\, (50.91) = 50.84}$$
$$\mathbf{X(28) = 14.44 + 0.715\, X(27) = 14.44 + 0.715\, (50.84) = 50.79}$$

- As always, it is necessary to construct the graph and compute statistics and check for stationary both in mean and variance, as well as the seasonality test.

CHAPTER 5

FILTERING TECHNIQUES

5.1 Adaptive Filtering

Filtering Techniques: Often on must filters an entire, e.g., financial time series with certain filter specifications to extract useful information by a transfer function expression. The aim of a filter function is to filter a time series in order to extract useful information hidden in the data, such as cyclic component. The filter is a direct implementation of and input-output function.

Data filtering is widely used as an effective and efficient time series modeling tool by applying an appropriate transformation technique. Most time series analysis techniques involve some form of filtering out noise in order to make the pattern more salient.

Differencing: A special type of filtering which is particularly useful for removing a trend, is simply to difference a given time series until it becomes stationary. This method is useful in Box-Jenkins modeling. For non-seasonal data, first order differencing is usually sufficient to attain apparent stationarity, so that the new series is formed from the original series.

Adaptive Filtering Any smoothing techniques such as moving average which includes a method of learning from past errors can respond to changes in the relative importance of trend, seasonal, and random factors. In the adaptive exponential smoothing method, one may adjust α to allow for shifting patterns.

5.2 Hodrick-Prescott Filter

Hodrick-Prescott Filter: The Hodrick-Prescott filter or H-P filter is an algorithm for choosing smoothed values for a time series. The H-P filter chooses smooth values $\{s_t\}$ for the series $\{x_t\}$ of T elements (t = 1 to T) that solve the following minimization problem:

min { $\{(x_t-s_t)^2$ ... etc. }

the positive parameter λ is the penalty on variation, where variation is measured by the average squared second difference. A larger value of λ makes the resulting $\{s_t\}$ series smoother; less high-frequency noise. The commonly applied value of l is 1600.

For the study of business cycles one uses not the smoothed series, but the jagged series of residuals from it. H-P filtered data shows less fluctuation than first-differenced data, since the H-P filter pays less attention to high frequency movements. H-P filtered data also shows more serial correlation than first-differenced data.

This is a smoothing mechanism used to obtain a long term trend component in a time series. It is a way to decompose a given series into stationary and non-stationary components in such a way that their sum of squares of the series from the non-stationary component is minimum with a penalty on changes to the derivatives of the non-stationary component.

5.3 Kalman Filter

Kalman Filter: The Kalman filter is an algorithm for sequentially updating a linear projection for a dynamic system that is in state-space representation. Application of the Kalman filter transforms a system of the following two-equation kind into a more solvable form:

$x_{t+1}=Ax_t+Cw_{t+1}$, and $y_t=Gx_t+v_t$ in which: A, C, and G are matrices known as functions of a parameter q about which inference is desired where: t is a whole number, usually indexing time; x_t is a true state variable, hidden from the econometrician; y_t is a measurement of x with scaling factor G, and measurement errors v_t, w_t are innovations to the hidden x_t process, $E(w_{t+1}w_t')=1$ by normalization (where, ' means the transpose), $E(v_tv_t)=R$, an unknown matrix, estimation of which is necessary but ancillary to the problem of interest, which is to get an estimate of q. The Kalman filter defines two matrices S_t and K_t such that the system described above can be transformed into the one below, in which estimation and inference about q and R is more straightforward; e.g., by regression analysis:

$z_{t+1}=Az_t+Ka_t$, and $y_t=Gz_t+a_t$ where z_t is defined to be $E_{t-1}x_t$, a_t is defined to be y_t-$E(y_{t-1}y_t$, K is defined to be limit K_t as t approaches infinity.

The definition of those two matrices S_t and K_t is itself most of the definition of the Kalman filters: $K_t=AS_tG'(GS_tG'+R)^{-1}$, and $S_{t-1}=(A-K_tG)S_t\,(A-K_tG)'+CC'+K_t\,RK_t'$, K_t is often called the Kalman gain.

CHAPTER 6

A SUMMARY OF SPECIAL MODELING TECHNIQUES

6.1 Neural Network

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

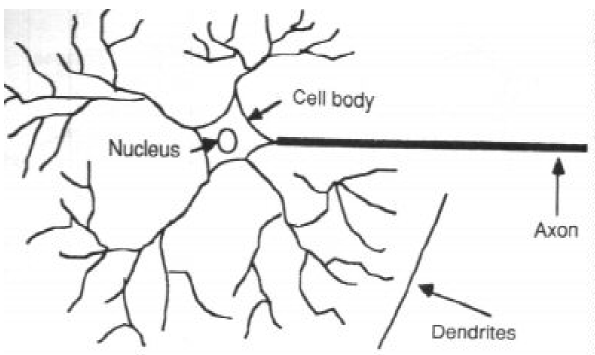
Neural network simulations appear to be a recent development. However, this field was established before the advent of computers, and has survived at least one major setback and several eras. Many important advances have been boosted by the use of inexpensive computer emulations. Following an initial period of enthusiasm, the field survived a period of frustration and disrepute. During this period when funding and professional support was minimal, important advances were made by relatively few researchers. These pioneers were able to develop convincing technology that surpassed everything that had evolved before. Currently, the neural network field enjoys a resurgence of interest and a corresponding increase in funding.

6.2 How the Human Brain Learns?

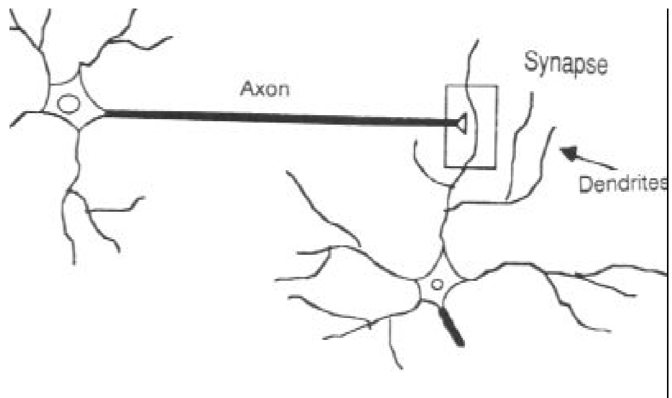
Much is still unknown about how the brain trains itself to process information, so theories abound. In the human brain, a typical neuron collects signals from others through a host of fine structures called *dendrites*. The neuron sends out spikes of electrical activity through a long, thin stand known as an *axon*, which splits into thousands of branches. At the end of each branch, a structure called a *synapse* converts the activity from the axon into electrical effects that inhibit or excite activity from the axon into electrical effects that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning occurs by changing the effectiveness of the synapses so that the influence of one neuron on another changes.

These neural networks are conducted by first trying to deduce the essential features of neurons and their interconnections. We then typically program a computer to simulate these features. However because our knowledge of neurons is incomplete and our computing power is limited, our models are necessarily gross idealizations of real networks of neurons.

Working of Brain Cells



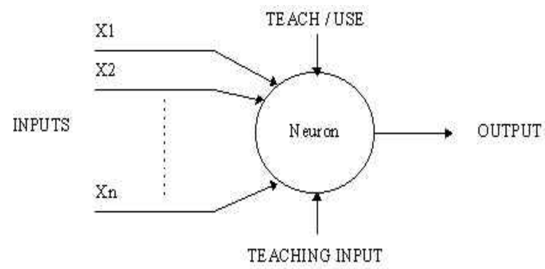
Components of a neuron



The synapse

THE NEURON MODEL

A Simple Neuron



An artificial neuron is a device with many inputs and one output.

6.3 Definition of Neural Network

A Neural Network is a system composed of many simple processing elements operating in parallel which can acquire, store, and utilize experiential knowledge.

6.3.1 Basic (Intuitive) Idea-

- ► Start with desired network architecture (number of neurons/nodes, layers, connections).
- ► Initialize randomly the weights with small values Iterate over each training example (stochastic gradient descent)
- ► Present each training example, find the error term at each output node, calculate E (W).
- ► Compute the gradient of E (W) for this example.
- ► Update ALL weights in the network (BP algorithm shows how)
- ► Repeat iterations (thousands of times maybe), with multiple use of the training examples, until reasonable performance, or a preset number of iterations, or the error rate with a test set of examples is below a preset threshold value. Too little iteration will lead to undertraining, too many leads to overfitting.

6.3.2 Two Modes Of Operation Of A Neuron

- TRAINING MODE: In the training mode, the neuron can be trained to fire (or not), for particular input patterns.

- USING MODE : In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not.

6.3.3 Basic Artificial Model

- Consists of simple processing elements called neurons, units or nodes
- Each neuron is connected to other nodes with an associated weight(strength) which typically multiplies the signal transmitted
- Each neuron has a single threshold value
- Weighted sum of all the inputs coming into the neuron is formed and the threshold is subtracted from this value = activation
- Activation signal is passed through an activation function(i.e. transfer function) to produce the output of the neuron.

6.3.4 Characterization

- Architecture: the pattern of nodes and connections between them
- Learning algorithm or training method: method for determining weights of the connections
- Activation function: function that produces an output based on the input values received by node.

6.4 CHOOSING THE NETWORK STRUCTURE

There are many ways that feed forward neural networks can be constructed. We must decide how many neurons will be inside the input and output layers. We must also decide how many hidden layers we're going to have, as well as how many neurons will be in each of these hidden layers. There are many techniques for choosing these parameters.

6.4.1 The Input Layer

The input layer to the neural network is the conduit through which the external environment presents a pattern to the neural network. Once a pattern is presented to the input later of the neural network the output layer will produce another pattern. In essence this is all the neural network does. The input layer should represent the condition for which we are training the neural network. Every input neuron should represent some independent variable that has an influence over the output of the neural network.

It is important to remember that the inputs to the neural network are floating point numbers. These values are expressed as the primitive Java data type "double". This is not to say that only numeric data can be processed with the neural network. If we wish to process a form of data that is non-numeric we must develop a process that normalizes this data to a numeric representation.

6.4.2 The Output Layer

The output layer of the neural network is what actually presents a pattern to the external environment. Whatever patter is presented by the output layer can be directly traced back to the input layer. The number of output neurons should directly relate to the type of work that the neural network is to perform.

To consider the number of neurons to use in output layer we must consider the intended use of the neural network. If the neural network is to be used to classify items into groups, then it is often preferable to have one output neurons for each group that the item is to be assigned into. If the neural network is to perform noise reduction on a signal then it is likely that the number of input neurons will match the number of output neurons. In this sort of neural network we would one day want the patterns to leave the neural network in the same format as they entered.

6.5 Neural Network for Time Series

Neural Network: For time series forecasting, the prediction model of order p, has the general form:

$D_t = f(D_{t-1}, D_{t-1}, ..., D_{t-p}) + e_t$

Neural network architectures can be trained to predict the future values of the dependent variables. What is required are design of the network paradigm and its parameters. The *multi-layer feed-forward neural network* approach consists of an input layer, one or several hidden layers and an output layer. Another approach is known as the partially *recurrent neural network* that can learn sequences as time evolves and responds to the same input pattern differently at different times, depending on the previous input patterns as well. None of these approaches is superior to the other in all cases; however, an additional dampened feedback, that possesses the characteristics of a dynamic memory, will improve the performance of both approaches.

Outlier Considerations: Outliers are a few observations that are not well fitted by the "best" available model. In practice, any observation with standardized residual greater than 2.5 in absolute value is a candidate for being an outlier. In such case, one must first investigate the source of data. If there is no doubt about the accuracy or veracity of the observation, then it should be removed, and the model should be refitted.

Whenever data levels are thought to be too high or too low for "business as usual", such points are called the outliers. A mathematical reason to adjust for such occurrences is that the majority of forecast techniques are based on averaging. It is well known that arithmetic averages are very sensitive to outlier values; therefore, some alteration should be made in the data before continuing.

One approach is to replace the outlier by the average of the two sales levels for the periods, which immediately come before and after the period in question and put this number in place of the outlier. This idea is useful if outliers occur in the middle or recent part of the data. However, if outliers appear in the oldest part of the data, a second alternative may be followed, which is to simply throw away the data up to and including the outlier.

In light of the relative complexity of some inclusive but sophisticated forecasting techniques, it is recommended that management go through an evolutionary progression in adopting new forecast techniques. That is to say, a simple forecast method well understood is better implemented than one with all inclusive features but unclear in certain facets.

6.6 Modeling and Simulation

Modeling and Simulation: Dynamic modeling and simulation is the collective ability to understand the system and implications of its changes over time including forecasting. System Simulation is the mimicking of the operation of a real system, such as the day-to-day operation of a bank, or the value of a stock portfolio over a time period. *By advancing the simulation run into the future*, managers can quickly find out how the system might behave in the future, therefore making decisions as they deem appropriate.

In the field of simulation, the concept of "principle of computational equivalence" has beneficial implications for the decision-maker. Simulated experimentation accelerates and replaces effectively the "wait and see" anxieties in discovering new insight and explanations of future behavior of the real system.

6.7 Probabilistic Models

Probabilistic Models: Uses probabilistic techniques, such as Marketing Research Methods, to deal with uncertainty, gives a range of possible outcomes for each set of events. For example, it is desired to identify the prospective buyers of a new product within a community of size N. From a survey result, one may estimate the probability of selling p, and then estimate the size of sales as Np with some confidence level.

An Application: Suppose it is desired to forecast the sales of new toothpaste in a community of 50,000 housewives. A free sample is given to 3,000 selected randomly, and then 1,800 indicated that they would buy the product.

Using the binomial distribution with parameters (3000, 1800/3000), the standard error is 27, and the expected sale is 50000(1800/3000) = 30000. The 99.7% confidence interval is within 3 times standard error 3(27) = 81 times the total population ratio 50000/3000; i.e., 1350. In other words, the range (28650, 31350) contains the expected sales.

6.8 Event History Analysis

Event History Analysis: Sometimes data on the exact time of a particular event (or events) are available, for example on a group of patients. Examples of events could include asthma attack; epilepsy attack; myocardial infections; hospital admissions. Often, occurrence (and non-occurrence) of an event is available on a regular basis, e.g., daily and the data can then be thought of as having a repeated measurements structure. An objective may be to determine whether any concurrent events or measurements have influenced the occurrence of the event of interest. For example, daily pollen counts may influence the risk of asthma attacks; high blood pressure might precede a myocardial infarction. PROC GENMOD available in SAS can be used for the event history analysis.

6.9 Predicting Market Response

Predicting Market Response: As applied researchers in business and economics, faced with the task of predicting market response, the functional form of the response is seldom known. Perhaps market response is a nonlinear monotonic, or even a non-monotonic function of explanatory variables. Perhaps it is determined by interactions of explanatory variable. Interaction is logically independent of its components.

When it is desired to represent complex market relationships within the context of a linear model, using appropriate transformations of explanatory and response variables, it is understood how hard the work of statistics can be. Finding reasonable models is a challenge, and justifying our choice of models to our peers can be even more of a challenge. Alternative specifications abound.

Modern regression methods, such as generalized additive models, multivariate adaptive regression splines, and regression trees, have one clear advantage: They can be used without specifying a functional form in advance. These data-adaptive, computer- intensive methods offer a more flexible approach to modeling than traditional statistical methods. Some modern regression methods perform quite well based on the results of simulation studies.

6.10 Prediction Interval for a Random Variable

Prediction Interval for a Random Variable: In many applied business statistics, such as forecasting, one is interested in construction of statistical interval for random variable rather than a parameter of a population distribution. For example, let X be a random variable distributed normally with estimated mean $\bar{X}$ and standard deviation S, then a prediction interval for the sample mean $\bar{X}$ with 100(1-α)% confidence level is:

$\bar{X}$ - t . S (1 + 1/n) , $\bar{X}$ + t . S (1 + 1/n)

This is the range of a random variable $\bar{x}$ with 100(1- α)% confidence, using t-table. Relaxing the normality condition for sample mean prediction interval requires a large sample size, say n over 30.

6.11 Census II Method of Seasonal Analysis

Census II Method of Seasonal Analysis:

Census-II is a variant of X-11. The X11 procedure provides seasonal adjustment of time series using the Census X-11 or X-11 ARIMA method. The X11 procedure is based on the US Bureau of the Census X-11 seasonal adjustment program, and it also supports the X-11 ARIMA method developed by Statistics Canada.

6.12 Delphi Analysis

Delphi Analysis: Delphi Analysis is used in the decision making process, in particular in forecasting. Several "experts" sit together and try to compromise on something upon which they cannot agree.

6.13 System Dynamics Modeling

System Dynamics Modeling: System dynamics (SD) is a tool for scenario analysis. Its main modeling tools are mainly the dynamic systems of differential equations and simulation. The SD approach to modeling is an important one for the following, not the least of which is that e.g., econometrics is the established methodology of system dynamics. However, from a philosophy of social science perspective, SD is deductive and econometrics is inductive. SD is less tightly bound to actuarial data and thus is free to expand out and examine more complex, theoretically informed, and postulated relationships. Econometrics is more tightly bound to the data and the models it explores, by comparison, are simpler. This is not to say the one is better than the other: properly understood and combined, they are complementary. Econometrics examines historical relationships through correlation and least squares regression model to compute the fit. In contrast, consider a simple growth scenario analysis; the initial growth portion of say, population is driven by the amount of food available. So there is a correlation between population level and food. However, the usual econometrics techniques are limited in their scope. For example, changes in the direction of the growth curve for a time population is hard for an econometrics model to capture.

6.14 Transfer Functions Methodology

It is possible to extend regression models to represent dynamic relationships between variables via appropriate transfer functions used in the construction of feedforward and feedback control schemes. The Transfer Function Analyzer module in SCA forecasting & modeling package is a frequency spectrum analysis package designed with the engineer in mind. It applies the concept of the Fourier integral transform to an input data set to provide a frequency domain representation of the function approximated by that input data. It also presents the results in conventional engineering terms.

6.15 Testing for and Estimation of Multiple Structural Changes

Testing for and Estimation of Multiple Structural Changes

The tests for structural breaks that I have seen are designed to detect only one break in a time series. This is true whether the break point is known or estimated using iterative methods. For example, for testing any change in level of the dependent series or model specification, one may use an iterative test for detecting points in time by incorporating level shift (0,0,0,0,...,1,1,1,1,1) variables to account for a change in intercept. Other causes are the change in variance and changes in parameters.

6.16 Combination of Forecasts

Combination of Forecasts: Combining forecasts merges several separate sets of forecasts to form a better composite forecast. The main question is "how to find the optimal combining weights?" The widely used approach is to change the weights from time to time for a better forecast rather than using a fixed set of weights on a regular basis or otherwise.

All forecasting models have either an implicit or explicit error structure, where error is defined as the difference between the model prediction and the "true" value. Additionally, many data snooping methodologies within the field of statistics need to be applied to data supplied to a forecasting model. Also, diagnostic checking, as defined within the field of statistics, is required for any model which uses data.

Using any method for forecasting one must use a performance measure to assess the quality of the method. Mean Absolute Deviation (MAD), and Variance are the most useful measures. However, MAD does not lend itself to making further inferences, but the standard error does. For error analysis purposes, variance is preferred since variances of independent (uncorrelated) errors are additive; however, MAD is not additive.

Regression and Moving Average: When a time series is not a straight line one may use the moving average (MA) and break-up the time series into several intervals with common straight line with positive trends to achieve linearity for the whole time series. The process involves transformation based on slope and then a moving average within that interval. For most business time series, one the following transformations might be effective:

- slope/MA,
- log (slope),
- log(slope/MA),
- log(slope) - 2 log(MA).

6.17 Measuring for Accuracy

Measuring for Accuracy

The most straightforward way of evaluating the accuracy of forecasts is to plot the observed values and the one-step-ahead forecasts in identifying the residual behavior over time.

The widely used statistical measures of error that can help you **to identify a method or the optimum value of the parameter within a method** are:

Mean absolute error: The mean absolute error (MAE) value is the average absolute error value. Closer this value is to zero the better the forecast is.

Mean squared error (MSE): Mean squared error is computed as the sum (or average) of the squared error values. This is the most commonly used lack-of-fit indicator in statistical fitting procedures. As compared to the mean absolute error value, this measure is very sensitive to any outlier; that is, unique or rare large error values will impact greatly MSE value.

Mean Relative Percentage Error (MRPE): The above measures rely on the error value without considering the magnitude of the observed values. The MRPE is computed as the average of the APE values:

Relative Absolute Percentage $\text{Error}_t = 100|(X_t - F_t)/X_t|\%$

Durbin-Watson statistic quantifies the serial correlation of serial correlation of the errors in time series analysis and forecasting. D-W statistic is defined by:

D-W statistic $= \Sigma_2^n (e_j - e_{j-1})^2 / \Sigma_1^n e_j^2$,
where e_j is the j^{th} error. D-W takes values within [0, 4]. For no serial correlation, a value close to 2 is expected. With positive serial correlation, adjacent deviates tend to have the same sign; therefore D-W becomes less than 2; whereas with negative serial correlation, alternating signs of error, D-W takes values larger than

2. For a forecasting where the value of D-W is significantly different from 2, the estimates of the variances and covariances of the model's parameters can be in error, being either too large or too small.

In measuring the forecast accuracy one should first determine a loss function and hence a suitable measure of accuracy. For example, quadratic loss function implies the use of MSE. Often one has a few models to compare and one tries to pick the "best". Therefore one must be careful to standardize the data and the results so that one model with large variance does not 'swamp' the other model.

CHAPTER 7

LEARNING AND THE LEARNING CURVE

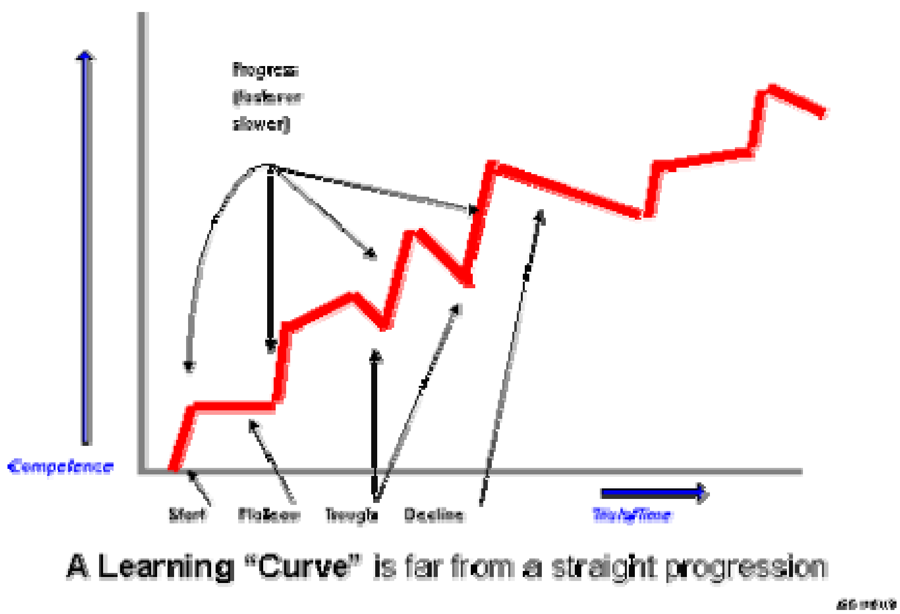
7.1 Introduction

The concept of the learning curve was introduced to the aircraft industry in 1936 when T. P. Wright published an article in the February 1936 Journal of the Aeronautical Science. Wright described a basic theory for obtaining cost estimates based on repetitive production of airplane assemblies. Since then, learning curves (also known as progress functions) have been applied to all types of work from simple tasks to **complex jobs** like manufacturing. The theory of learning recognizes that repetition of the same operation results in less time or effort expended on that operation. Its underlying concept is that, for example the direct labor man-hours necessary to complete a unit of production will decrease by a constant percentage each time the production quantity is doubled. If the rate of improvement is 20% between doubled quantities, then the learning percent would be 80% (100-20=80). While the learning curve emphasizes time, it can be easily extended to cost as well.

7.2 Psychology of Learning

Based on the theory of learning it is easier to learn things that are related to what you already know. The likelihood that new information will be retained is related to how much previous learning there is that provides "hooks" on which to hang the new information. In other words, to provide new connectivity in the learner's **neural mental network**. For example, it is a component of my teaching style to provide a preview of the course contents and review of necessary topics form prerequisites courses (if any) during the first couple of class meeting, before teaching them to course topics in detail. Clearly, change in one's mental model happens more readily when you have a mental model similar to the one you're trying to learn; and that it will be easier to change a mental model after you become more consciously aware.

A steep learning curve is often referred to indicate that something is difficult to learn. In practice, a curve of the amount learned against the number of trials (in experiments) or over time (in reality) is just the opposite: if something is difficult, the line rises slowly or shallowly. So the steep curve refers to the demands of the task rather than a description of the process. The following figure is of a fairly typical of a learning curve. It depicts the fact that the learning curve does not proceed smoothly: the plateaus and troughs are normal features of the process.



A Typical Learning Curve

The goal is to make the "valley of despair" as Shallow and as Narrow as possible. To make it narrow, you must give plenty of training, and follow it up with continuing floor support, help desk support, and other forms of just-in-time support so that people can quickly get back to the point of competence. If they stay in the valley of despair for too long, they will lose hope and hate the new software and the people who made them switch.

Valley of Despair Characteristics:

- Who's dumb idea was this?
- I hate this
- I could do better the old way
- I cannot get my work done

Success Characteristic:

- How did I get along without this?

To make it as shallow as possible, minimize the number of things you try to teach people at once. Build gradually, and only add more to learn once people have developed a level of competence with the basic things. In the acquisition of skills, a major issue is the reliability of the performance. Any novice can get it right occasionally, but it is consistency which counts, and the progress of learning is often assessed on this basis. Need to train workers in new method based on the facts that the longer a person performs a task, the quicker it takes him/her:

1. Learn-on-the-job approach:
    - learn wrong method
    - bother other operators, lower production
    - anxiety
2. Simple written instructions: only good for very simple jobs
3. Pictorial instructions: "good pictures worth 1000 words"
4. Videotapes: dynamic rather than static
5. Physical training:
    - real equipment or simulators, valid
    - does not interrupt production

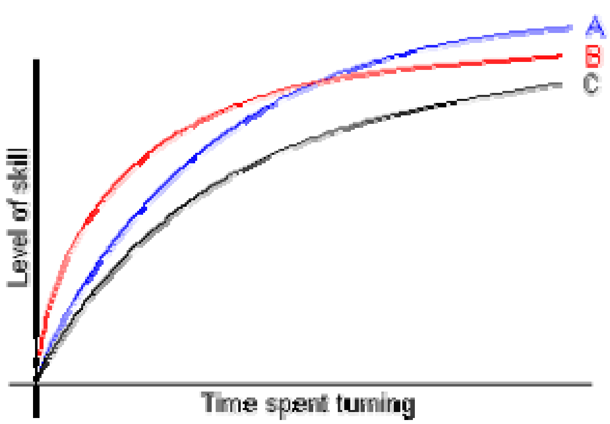
- monitor performance
- simulate emergencies

Factors that affect human learning:

1. Job complexity - long cycle length, more training, amount of uncertainty in movements, more C-type motions, simultaneous motions

2. Individual capabilities- age, rate of learning declines in older age, amount of prior training, physical capabilities, active, good circulation of oxygen to brain

Because of the differences between individuals, their innate ability, their age, or their previous useful experience then each turner will have his/her own distinctive learning curve. Some possible, contrasting, curves are shown in the following figure:



An Individuals Differences Classification

Individual C is a very slow learner but he improves little by little. Individual B is a quick learner and reaches his full capacity earlier than individuals A or C. But, although A is a slow learner, he eventually becomes more skilled than B.

Measuring and Explaining Learning Effects of Modeling: It is already accepted that modeling triggers learning, this is to say the modeler's mental model changes as effect of the activity "modeling". In "systems thinking" it also includes the way people approach decision situations by studying attitude changes model building.

7.3 Modeling the Learning Curve

Learning curves are all about ongoing improvement. Managers and researchers noticed, in field after field, from aerospace to mining to manufacturing to writing, that stable processes improve year after year rather than remain the same. Learning curves describe these patterns of long-term improvement. Learning curves help answer the following questions.

- How fast can you improve to a specific productivity level?
- What are the limitations to improvement?
- Are aggressive goals achievable?

The learning curve was adapted from the historical observation that individuals who perform repetitive tasks exhibit an improvement in performance as the task is repeated a number of times.

With proper instruction and repetition, workers learn to perform their jobs more efficiently and effectively and consequently, e.g., the direct labor hours per unit of a product are reduced. This learning effect could have resulted from better work methods, tools, product design, or supervision, as well as from an individual’s learning the task.

A Family of Learning Curves Funtions: Of the dozens of mathematic concepts of learning curves, the four most important equations are:

- Log-Linear: $y(t) = k\, t^b$
- Stanford-B: $y(t) = k\, (t + c)^b$

- DeJong: $y(t) = a + k\, t^b$
- S-Curve: $y(t) = a + k\, (t + c)^b$

The Log-Linear equation is the simplest and most common equation and it applies to a wide variety of processes. The Stanford-B equation is used to model processes where experience carries over from one production run to another, so workers start out more productively than the asymtote predicts. The Stanford-B equation has been used to model airframe production and mining. The DeJong equation is used to model processes where a portion of the process cannot improve. The DeJong equation is often used in factories where the assembly line ultimately limits improvement. The S-Curve equation combines the Stanford-B and DeJong equations to model processes where both experience carries over from one production run to the next and a portion of the process cannot improve.

7.4 An Application

The learning effect causes the time required to perform a task to reduce when the task is repeated. Applying this principle, the time required to perform a task will decrease at a declining rate as cumulative number of repetitions increase. This reduction in time follows the function: $y(t) = k\, t^{\,b}$, where b = log(r)/log (2), i.e., $2^b = r$, and r is the learning rate, a lower rate implies faster learning, a positive number less than 1, and k is a constant.

For example, industrial engineers have observed that the learning rate ranges from 70% to 95% in the manufacturing industry. An r = 80% learning curve denotes a 20% reduction in the time with each doubling of repetitions. An r = 100% curve would imply no improvement at all. For an r = 80% learning curve, b = log(0.8)/log(2) = -0.3219.

Numerical Example: Consider the first (number if cycles) and the third (their cycle times) columns for the following data set:

# Cycles	Log # Cycles	Cycle Time	Log Cycle Time
1	0	12.00	1.08
2	0.301	9.60	0.982
4	0.602	7.68	0.885
8	0.903	6.14	0.788
16	1.204	4.92	0.692
32	1.505	3.93	0.594

To estimate $y = k\, t^b$ one must use linear regression on the logarithmic scales, i.e., $\log y = \log(k) + b \log(t)$ using a data set, and then computing $r = 2^b$. Using the By By Regression Analysis **for the above data, we obtain:**

b = Slope = -0.32, y-Intercept = log(k) = 1.08

log y = log(k) + b log(t)

b = -0.32

$k = 10^{1.08} = 12$

$y(t) = 12\, t^{-0.32}$

$r = 2^b = 2^{-0.32} = 80\%$

Conclusions: As expected while number of cycles doubles, cycle time decreases by a constant %, that is, the result is a 20% decrease or 80% learning ratio or 80% learning curve with a mathematical model $y(t) = 12\ t^{-0.32}$

CHAPTER 8

NEURAL NETWORK AND TIME SERIES ANALYSIS

8.1 Time Series Prediction in ST Neural Networks

In time series problems, the objective is to predict ahead the value of a variable that varies in time using previous values of that and/or other variables. Typically the predicted variable is continuous so that time series prediction is usually a specialized form of regression. However, without this restriction time series can also do prediction of nominal variables (i.e. classification).

It is also usual to predict the next value in a series from a fixed number of previous values (looking ahead a single time step). When the next value in a series is generated, further values can be estimated by feeding the newly-estimated value back into the network together with other previous values: time series projection. Obviously, the reliability of projection drops the more steps ahead one tries to predict and if a particular distance ahead is required it is probably better to train a network specifically for that degree of lookahead.

Any type of network can be used for time series prediction (the network type must however, be appropriate for regression or classification depending on the problem type). The network can also have any number of input and output variables. However, most commonly there is a single variable that is both the input and (with the lookahead taken into account) the output. Configuring a network for time series usage alters the way that data is pre-processed (i.e., it is drawn from a number of sequential cases rather than a single case) but the network is executed and trained just as for any other problem.

The time series training data set therefore typically has a single variable, and this has type input/output (i.e. it is used both for network input and network output).

The most difficult concept in time series handling is the interpretation of training, selection, test and ignored cases. For standard data sets, each case is independent and these meanings are clear. However, with a time series network each pattern of inputs and outputs is actually drawn from a number of cases determined by the network's *Steps* and *Lookahead* parameters. There are two consequences of this:

The input pattern's type is taken from the type of the output case. For example, in a data set containing some cases, the first two ignored and the third test, with *Steps=2* and *Lookahead=1*, the first usable pattern has type Test, and draws its inputs from the first two cases, and its output from the third. Thus, the first two cases are used in the test set even though they are marked Ignore. Further, any given case may be used in three patterns and these may be any of training, selection and test patterns. In some sense data actually leaks between training, selection and test sets. To isolate the three sets entirely, contiguous blocks of train, verify or test cases would need to be constructed, separated by the appropriate number of ignore cases.

The first few cases can only be used as inputs for patterns. When selecting cases for time series use, the case number selected is always the output case. The first few clearly cannot be selected (as this would require further cases before the beginning of the data set), and are not available.

Thus here in our examples we use two types of Network Architecture.

8.2 The Back-Propagation Algorithm - A Mathematical Approach

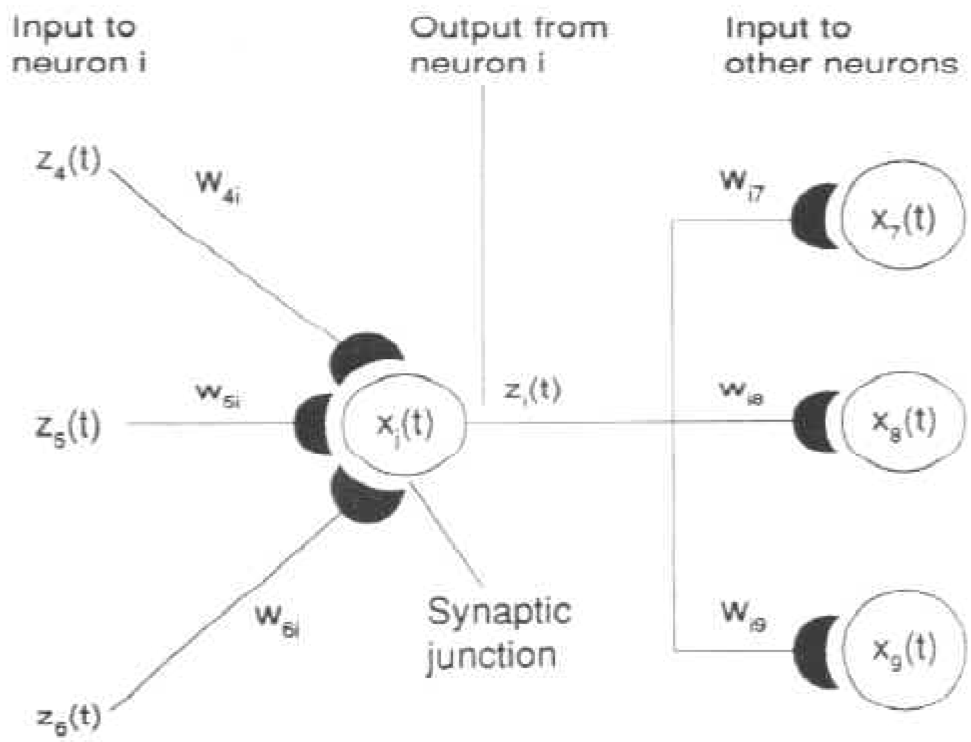
The Back-Propagation Algorithm

In order to train a neural network to perform some task, we must adjust the weights of each unit in such a way that the error between the desired output and the actual output is reduced. This process requires that the neural network compute the error derivative of the weights (**EW**). In other words, it must calculate how the error changes as each weight is increased or decreased slightly. The back propagation algorithm is the most widely used method for determining the **EW**.

The back-propagation algorithm is easiest to understand if all the units in the network are linear. The algorithm computes each **EW** by first computing the **EA**, the rate at which the error changes as the activity level of a unit is changed. For output units, the **EA** is simply the difference between the actual and the desired output. To compute the **EA** for a hidden unit in the layer just before the output layer, we first identify all the weights between that hidden unit and the output units to which it is connected. We then multiply those weights by the **EA**s of those output units and add the products. This sum equals the **EA** for the chosen hidden unit. After calculating all the **EA**s in the hidden layer just before the output layer, we can compute in like fashion the **EA**s for other layers, moving from layer to layer in a direction opposite to the way activities propagate through the network. This is what gives back propagation its name. Once the **EA** has been computed for a unit, it is straight forward to compute the **EW** for each incoming connection of the unit. The **EW** is the product of the EA and the activity through the incoming connection.

It is to be noticed that for non-linear units, the back-propagation algorithm includes an extra step. Before back-propagating, the **EA** must be converted into the **EI**, the rate at which the error changes as the total input received by a unit is changed.

Units are connected to one another. Connections correspond to the edges of the underlying directed graph. There is a real number associated with each connection, which is called the weight of the connection. We denote by W_{ij} the weight of the connection from unit u_i to unit u_j. It is then convenient to represent the pattern of connectivity in the network by a weight matrix W whose elements are the weights W_{ij}. Two types of connection are usually distinguished: excitatory and inhibitory. A positive weight represents an excitatory connection whereas a negative weight represents an inhibitory connection. The pattern of connectivity characterises the architecture of the network.



A unit in the output layer determines its activity by following a two step procedure.

- First, it computes the total weighted input x_j, using the formula:

$$X_j = \sum_i y_i W_{ij}$$

where y_i is the activity level of the jth unit in the previous layer and W_{ij} is the weight of the connection between the ith and the jth unit.

Next, the unit calculates the activity y_j using some function of the total weighted input. Typically we use the sigmoid function:

$$y_j = \frac{1}{1 + e^{-x_j}}$$

Once the activities of all output units have been determined, the network computes the error E, which is defined by the expression:

$$E = \frac{1}{2}\sum_i (y_i - d_i)^2$$

where y_j is the activity level of the jth unit in the top layer and d_j is the desired output of the jth unit.

The back-propagation algorithm consists of four steps:

1. Compute how fast the error changes as the activity of an output unit is changed. This error derivative (EA) is the difference between the actual and the desired activity.

$$\mathrm{EA}_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

2. Compute how fast the error changes as the total input received by an output unit is changed. This quantity (EI) is the answer from step 1 multiplied by the rate at which the output of a unit changes as its total input is changed.

$$\mathrm{EI}_j = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{dy_j}{dx_j} = \mathrm{EA}_j y_j (1 - y_j)$$

3. Compute how fast the error changes as a weight on the connection into an output unit is changed. This quantity (EW) is the answer from step 2 multiplied by the activity level of the unit from which the connection emanates.

$$\mathrm{E}W_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial W_{ij}} = \mathrm{EI}_j y_i$$

4. Compute how fast the error changes as the activity of a unit in the previous layer is changed. This crucial step allows back propagation to be applied to multilayer networks. When the activity of a unit in the previous layer changes, it affects the activites of all the output units to which it is connected. So to compute the overall effect on the error, we

add together all these seperate effects on output units. But each effect is simple to calculate. It is the answer in step 2 multiplied by the weight on the connection to that output unit.

$$\mathrm{EA}_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} \times \frac{\partial x_j}{\partial y_i} = \sum_j \mathrm{EI}_j W_{ij}$$

By using steps 2 and 4, we can convert the EAs of one layer of units into EAs for the previous layer. This procedure can be repeated to get the EAs for as many previous layers as desired. Once we know the EA of a unit, we can use steps 2 and 3 to compute the EWs on its incoming connections.

8.3 Linear Perceptron Network

The units each perform a biased weighted sum of their inputs and pass this activation level through a transfer function to produce their output, and the units are arranged in a layered feedforward topology. The network thus has a simple interpretation as a form of input-output model, with the weights and thresholds (biases) the free parameters of the model. Such networks can model functions of almost arbitrary complexity, with the number of layers, and the number of units in each layer, determining the function complexity. Important issues in Multilayer Perceptrons (MLP) design include specification of the number of hidden layers and the number of units in these layers.

Training Multilayer Perceptrons

Once the number of layers, and number of units in each layer, has been selected, the network's weights and thresholds must be set so as to minimize the prediction error made by the network. This is the role of the *training algorithms*. The number of input and output units is defined by the problem. The number of hidden units to use is far from clear. As good a starting point as any is to use one hidden layer, with the number of units equal to half the sum of the number of input and output units. The historical cases that have been gathered are used to automatically adjust the weights and thresholds in order to minimize this error. This process is equivalent to fitting the model represented by the network to the training data available. The error of a particular configuration of the network can be determined by running all the training cases through the network, comparing the actual output generated with the desired or target outputs. The differences are combined together by an *error function* to give the network error. The most common error functions are the *sum squared error* (used for regression problems), where the individual errors of output units on each case are squared and summed together, and the cross entropy functions.A helpful concept here is the error surface. Each of the *N* weights and thresholds of the network (i.e., the free parameters of the model) is taken to be a dimension in space. The *N+1*th dimension is the network error. For any possible configuration of weights the error can be plotted in the *N+1*th dimension, forming an *error surface*. The objective of network training is to find the lowest point in this many-dimensional surface.

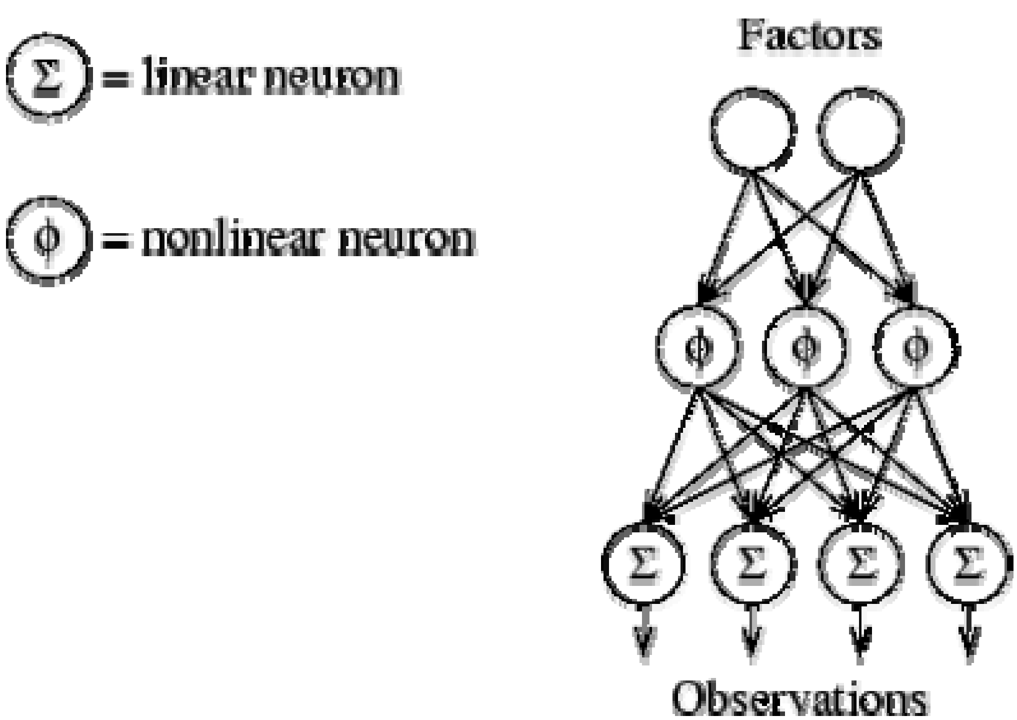
The choice of the set of functions **f** depends on the problem at hand, but a possible choice is the multi-layer perceptron (MLP) network. It has often been found to provide compact representations of mappings in real-world problems. The MLP network is composed of neurons which are very close to the ones represented in the case of the linear network. The linear neurons are modified so that a slight nonlinearity is added after the linear summation. The output *c* of each neuron is thus

$$c = \phi\left(\sum_i w_i a_i + b\right),$$

where a_i are the inputs of the neuron and w_i are the weights of the neuron. The nonlinear function ϕ is called the activation function as it determines the activation level of the neuron. This refers to interpreting the activation as the pulse rate of biological neurons. Due to the nonlinear activation function, a multi-layer network is not equivalent to any one-layer structure with the same activation function. In fact, it has been shown that one layer of suitable nonlinear neurons followed by a linear layer can approximate any nonlinear function with arbitrary accuracy, given enough nonlinear neurons. Thus, an MLP network is a universal function approximator.

The activation functions most widely used are the hyperbolic tangent **tanh(x)** and logistic sigmoid **1/(1+exp(-x)).** They are actually related as **(tanh(x)+1) /2 = 1/ (1+exp(-2x)).** These activation functions are used for their convenient mathematical properties and because they have a roughly linear behaviour around origin, which means that it is easy to represent close-to-linear mappings with the MLP network.

<u>A graphical representation of the computational structure of an MLP network with one hidden layer of nonlinear neurons</u>



The mapping of the network can be compactly described as:

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) = \mathbf{B}\phi(\mathbf{A}\mathbf{s}(t) + \mathbf{a}) + \mathbf{b} + \mathbf{n}(t).$$

According to the usual notation with MLP networks, the vector φ denotes a vector of functions which each operate on one of the components of the argument vector.

CHAPTER 9

IMPLEMENTATION OF TIME SERIES FORECASTING USING THREE DIFFERENT EXAMPLES.

EXAMPLE-1

SERIES B, IBM STOCK PRICES

- Closing price of common stock, daily, May 17 1961 to November 2 1962
- May 17 = 137th day

460 457 452 459 462 459 463 479 493 490 492 498 499 497
496 490 489 478 487 491 487 482 479 478 479 477 479 475
479 476 476 478 479 477 476 475 475 473 474 474 474 465
466 467 471 471 467 473 481 488 490 489 489 485 491 492
494 499 498 500 497 494 495 500 504 513 511 514 510 509
515 519 523 519 523 531 547 551 547 541 545 549 545 549
547 543 540 539 532 517 527 540 542 538 541 541 547 553
559 557 557 560 571 571 569 575 580 584 585 590 599 603
599 596 585 587 585 581 583 592 592 596 596 595 598 598
595 595 592 588 582 576 578 589 585 580 579 584 581 581
577 577 578 580 586 583 581 576 571 575 575 573 577 582
584 579 572 577 571 560 549 556 557 563 564 567 561 559
553 553 553 547 550 544 541 532 525 542 555 558 551 551
552 553 557 557 548 547 545 545 539 539 535 537 535 536
537 543 548 546 547 548 549 553 553 552 551 550 553 554
551 551 545 547 547 537 539 538 533 525 513 510 521 521
521 523 516 511 518 517 520 519 519 519 518 513 499 485
454 462 473 482 486 475 459 451 453 446 455 452 457 449

450 435 415 398 399 361 383 393 385 360 364 365 370 374
359 335 323 306 333 330 336 328 316 320 332 320 333 344
339 350 351 350 345 350 359 375 379 376 382 370 365 367
372 373 363 371 369 376 387 387 376 385 385 380 373 382
377 376 379 386 387 386 389 394 393 409 411 409 408 393
391 388 396 387 383 388 382 384 382 383 383 388 395 392
386 383 377 364 369 355 350 353 340 350 349 358 360 360
366 359 356 355 367 357 361 355 348 343 330 340 339 331
345 352 346 352 357

IBM STOCK IN MATLAB

INPUTDATA FOR TRAINING

[460 457 452 459 462 459 463 479 493 490 492 498 499 497 496 490 489 478 487 491
487 482 479 478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474
474 465 466 467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500
497 494 495 500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541
545 549 545 549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557
557 560 571 571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592
592 596 596 595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581
577 577 578 580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560
549 556 557 563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558
551 551 552 553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546
547 548 549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525
513 510 521 521 521 523 516 511 518 517 520 519 519 519 518 513 499 485 454 462
473 482 486 475 459 451 453 446;
457 452 459 462 459 463 479 493 490 492 498 499 497 496 490 489 478 487 491 487
482 479 478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474 474
465 466 467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500 497
494 495 500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541 545

549 545 549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557 557
560 571 571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592 592
596 596 595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581 577
577 578 580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560 549
556 557 563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558 551
551 552 553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546 547
548 549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525 513
510 521 521 521 523 516 511 518 517 520 519 519 519 518 513 499 485 454 462 473
482 486 475 459 451 453 446 455;

452 459 462 459 463 479 493 490 492 498 499 497 496 490 489 478 487 491 487 482
479 478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474 474 465
466 467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500 497 494
495 500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541 545 549
545 549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557 557 560
571 571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592 592 596
596 595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581 577 577
578 580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560 549 556
557 563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558 551 551
552 553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546 547 548
549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525 513 510
521 521 521 523 516 511 518 517 520 519 519 519 518 513 499 485 454 462 473 482
486 475 459 451 453 446 455 452;

459 462 459 463 479 493 490 492 498 499 497 496 490 489 478 487 491 487 482 479
478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474 474 465 466
467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500 497 494 495
500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541 545 549 545
549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557 557 560 571
571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592 592 596 596
595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581 577 577 578
580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560 549 556 557
563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558 551 551 552

553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546 547 548 549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525 513 510 521 521 521 523 516 511 518 517 520 519 519 519 518 513 499 485 454 462 473 482 486 475 459 451 453 446 455 452 457;
462 459 463 479 493 490 492 498 499 497 496 490 489 478 487 491 487 482 479 478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474 474 465 466 467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500 497 494 495 500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541 545 549 545 549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557 557 560 571 571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592 592 596 596 595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581 577 577 578 580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560 549 556 557 563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558 551 551 552 553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546 547 548 549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525 513 510 521 521 521 523 516 511 518 517 520 519 519 519 518 513 499 485 454 462 473 482 486 475 459 451 453 446 455 452 457 449]

TARGET

[459 463 479 493 490 492 498 499 497 496 490 489 478 487 491 487 482 479 478 479 477 479 475 479 476 476 478 479 477 476 475 475 473 474 474 474 465 466 467 471 471 467 473 481 488 490 489 489 485 491 492 494 499 498 500 497 494 495 500 504 513 511 514 510 509 515 519 523 519 523 531 547 551 547 541 545 549 545 549 547 543 540 539 532 517 527 540 542 538 541 541 547 553 559 557 557 560 571 571 569 575 580 584 585 590 599 603 599 596 585 587 585 581 583 592 592 596 596 595 598 598 595 595 592 588 582 576 578 589 585 580 579 584 581 581 577 577 578 580 586 583 581 576 571 575 575 573 577 582 584 579 572 577 571 560 549 556 557 563 564 567 561 559 553 553 553 547 550 544 541 532 525 542 555 558 551 551 552 553 557 557 548 547 545 545 539 539 535 537 535 536 537 543 548 546 547 548 549 553 553 552 551 550 553 554 551 551 545 547 547 537 539 538 533 525 513 510 521 521 521

523 516 511 518 517 520 519 519 519 518 513 499 485 454 462 473 482 486 475 459 451 453 446 455 452 457 449 450]

SIMULATED DATA

[450 435 415 398 399 361 383 393 385 360 364 365 370 374 359 335 323 306 333 330 336 328 316 320 332 320 333 344 339 350 351 350 345 350 359 375 379 376 382 370 365 367 372 373 363 371 369 376 387 387 376 385 385 380 373 382 377 376 379 386 387 386 389 394 393 409 411 409 408 393 391 388 396 387 383 388 382 384 382 383 383 388 395 392 386 383 377 364 369 355 350 353 340 350 349 358 360 360 366 359 356 355 367 357 361 355 348 343 330 340 339 331;
435 415 398 399 361 383 393 385 360 364 365 370 374 359 335 323 306 333 330 336 328 316 320 332 320 333 344 339 350 351 350 345 350 359 375 379 376 382 370 365 367 372 373 363 371 369 376 387 387 376 385 385 380 373 382 377 376 379 386 387 386 389 394 393 409 411 409 408 393 391 388 396 387 383 388 382 384 382 383 383 388 395 392 386 383 377 364 369 355 350 353 340 350 349 358 360 360 366 359 356 355 367 357 361 355 348 343 330 340 339 331 345;
415 398 399 361 383 393 385 360 364 365 370 374 359 335 323 306 333 330 336 328 316 320 332 320 333 344 339 350 351 350 345 350 359 375 379 376 382 370 365 367 372 373 363 371 369 376 387 387 376 385 385 380 373 382 377 376 379 386 387 386 389 394 393 409 411 409 408 393 391 388 396 387 383 388 382 384 382 383 383 388 395 392 386 383 377 364 369 355 350 353 340 350 349 358 360 360 366 359 356 355 367 357 361 355 348 343 330 340 339 331 345 352;
398 399 361 383 393 385 360 364 365 370 374 359 335 323 306 333 330 336 328 316 320 332 320 333 344 339 350 351 350 345 350 359 375 379 376 382 370 365 367 372 373 363 371 369 376 387 387 376 385 385 380 373 382 377 376 379 386 387 386 389 394 393 409 411 409 408 393 391 388 396 387 383 388 382 384 382 383 383 388 395 392 386 383 377 364 369 355 350 353 340 350 349 358 360 360 366 359 356 355 367 357 361 355 348 343 330 340 339 331 345 352 346;
399 361 383 393 385 360 364 365 370 374 359 335 323 306 333 330 336 328 316 320 332 320 333 344 339 350 351 350 345 350 359 375 379 376 382 370 365 367 372 373 363 371 369 376 387 387 376 385 385 380 373 382 377 376 379 386 387 386 389 394

393 409 411 409 408 393 391 388 396 387 383 388 382 384 382 383 383 388 395 392 386 383 377 364 369 355 350 353 340 350 349 358 360 360 366 359 356 355 367 357 361 355 348 343 330 340 339 331 345 352 346 352]

EXAMPLE-2

SUNSPOTS DATA FOR 300 YEARS ON EARTH'S SURFACE, 1699-1999

TRAINING DATA

[0.3138 0.4231 0.4362 0.2495 0.25 0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758 0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424 0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094 0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434;

0.4231 0.4362 0.2495 0.25 0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758 0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424

0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094 0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434 0.4215;

0.4362 0.2495 0.25 0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758 0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424 0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094 0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434 0.4215 0.3326;

0.2495 0.25 0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758 0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424 0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094

0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434 0.4215 0.3326 0.1966;

0.25 0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758 0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424 0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094 0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434 0.4215 0.3326 0.1966 0.1365]

<u>TARGET DATA</u>

[0.1606 0.0638 0.0502 0.0534 0.17 0.2489 0.2824 0.329 0.4493 0.3201 0.2359 0.1904 0.1093 0.0596 0.1977 0.3651 0.5549 0.5272 0.4268 0.3478 0.182 0.16 0.0366 0.1036 0.4838 0.8075 0.6585 0.4435 0.3562 0.2014 0.1192 0.0534 0.126 0.4336 0.6904 0.6846 0.6177 0.4702 0.3483 0.3138 0.2453 0.2144 0.1114 0.0837 0.0335 0.0214 0.0356 0.0758

0.1778 0.2354 0.2254 0.2484 0.2207 0.147 0.0528 0.0424 0.0131 0 0.0073 0.0262 0.0638 0.0727 0.1851 0.2395 0.215 0.1574 0.125 0.0816 0.0345 0.0209 0.0094 0.0445 0.0868 0.1898 0.2594 0.3358 0.3504 0.3708 0.25 0.1438 0.0445 0.069 0.2976 0.6354 0.7233 0.5397 0.4482 0.3379 0.1919 0.1266 0.056 0.0785 0.2097 0.3216 0.5152 0.6522 0.5036 0.3483 0.3373 0.2829 0.204 0.1077 0.035 0.0225 0.1187 0.2866 0.4906 0.501 0.4038 0.3091 0.2301 0.2458 0.1595 0.0853 0.0382 0.1966 0.387 0.727 0.5816 0.5314 0.3462 0.2338 0.0889 0.0591 0.0649 0.0178 0.0314 0.1689 0.284 0.3122 0.3332 0.3321 0.273 0.1328 0.0685 0.0356 0.033 0.0371 0.1862 0.3818 0.4451 0.4079 0.3347 0.2186 0.137 0.1396 0.0633 0.0497 0.0141 0.0262 0.1276 0.2197 0.3321 0.2814 0.3243 0.2537 0.2296 0.0973 0.0298 0.0188 0.0073 0.0502 0.2479 0.2986 0.5434 0.4215 0.3326 0.1966 0.1365 0.0743]

SIMULATED DATA

[0.0743 0.0303 0.0873 0.2317 0.3342 0.3609 0.4069 0.3394 0.1867 0.1109 0.0581 0.0298 0.0455 0.1888 0.4168 0.5983 0.5732 0.4644 0.3546 0.2484 0.16 0.0853 0.0502 0.1736 0.4843 0.7929 0.7128 0.7045 0.4388 0.363 0.1647 0.0727 0.023 0.1987 0.7411 0.9947 0.9665 0.8316 0.5873 0.2819 0.1961 0.1459 0.0534 0.079 0.2458 0.4906 0.5539 0.5518 0.5465 0.3483 0.3603 0.1987 0.1804 0.0811;
0.0303 0.0873 0.2317 0.3342 0.3609 0.4069 0.3394 0.1867 0.1109 0.0581 0.0298 0.0455 0.1888 0.4168 0.5983 0.5732 0.4644 0.3546 0.2484 0.16 0.0853 0.0502 0.1736 0.4843 0.7929 0.7128 0.7045 0.4388 0.363 0.1647 0.0727 0.023 0.1987 0.7411 0.9947 0.9665 0.8316 0.5873 0.2819 0.1961 0.1459 0.0534 0.079 0.2458 0.4906 0.5539 0.5518 0.5465 0.3483 0.3603 0.1987 0.1804 0.0811 0.0659;
0.0873 0.2317 0.3342 0.3609 0.4069 0.3394 0.1867 0.1109 0.0581 0.0298 0.0455 0.1888 0.4168 0.5983 0.5732 0.4644 0.3546 0.2484 0.16 0.0853 0.0502 0.1736 0.4843 0.7929 0.7128 0.7045 0.4388 0.363 0.1647 0.0727 0.023 0.1987 0.7411 0.9947 0.9665 0.8316 0.5873 0.2819 0.1961 0.1459 0.0534 0.079 0.2458 0.4906 0.5539 0.5518 0.5465 0.3483 0.3603 0.1987 0.1804 0.0811 0.0659 0.1428;

0.2317 0.3342 0.3609 0.4069 0.3394 0.1867 0.1109 0.0581 0.0298 0.0455 0.1888 0.4168 0.5983 0.5732 0.4644 0.3546 0.2484 0.16 0.0853 0.0502 0.1736 0.4843 0.7929 0.7128 0.7045 0.4388 0.363 0.1647 0.0727 0.023 0.1987 0.7411 0.9947 0.9665 0.8316 0.5873 0.2819 0.1961 0.1459 0.0534 0.079 0.2458 0.4906 0.5539 0.5518 0.5465 0.3483 0.3603 0.1987 0.1804 0.0811 0.0659 0.1428 0.4838;

0.3342 0.3609 0.4069 0.3394 0.1867 0.1109 0.0581 0.0298 0.0455 0.1888 0.4168 0.5983 0.5732 0.4644 0.3546 0.2484 0.16 0.0853 0.0502 0.1736 0.4843 0.7929 0.7128 0.7045 0.4388 0.363 0.1647 0.0727 0.023 0.1987 0.7411 0.9947 0.9665 0.8316 0.5873 0.2819 0.1961 0.1459 0.0534 0.079 0.2458 0.4906 0.5539 0.5518 0.5465 0.3483 0.3603 0.1987 0.1804 0.0811 0.0659 0.1428 0.4838 0.8127]

EXAMPLE 3

TOTAL ANNUAL RAINFALL, INCHES, LONDON, ENGLAND, 1813-1912

23.56 26.07 21.86 31.24 23.65 23.88
26.41 22.67 31.69 23.86 24.11 32.43
23.26 22.57 23.00 27.88 25.32 25.08
27.76 19.82 24.78 20.12 24.34 27.42
19.44 21.63 27.49 19.43 31.13 23.09
25.85 22.65 22.75 26.36 17.70 29.81
22.93 9.22 20.63 35.34 25.89 18.65
23.06 22.21 22.18 18.77 28.21 32.24
22.27 27.57 21.59 16.93 29.48 31.60
26.25 23.40 25.42 21.32 25.02 33.86
22.67 18.82 28.44 26.16 28.17 34.08
33.82 30.28 27.92 27.14 24.40 20.35
26.64 27.01 19.21 27.74 23.85 21.23
28.15 22.61 19.80 27.94 21.47 23.52
22.86 17.69 22.54 23.28 22.17 20.84
38.10 20.65 22.97 24.26 23.01 23.67 26.75 25.36 24.79 27.88

Conclusion and Comparison

Conclusion:

It has been observed that the Time series analysis Network designed in MATLAB gives approximately 97% accuracy in the above examples.

The following table shows the comparison chart for three examples used above:

S.No.	Desired Output	Actual Output	% Accuracy
EXAMPLE-1	459	463.4469	99.03
	463	467.3457	99.06
	479	464.7565	97.03
	493	484.0494	98.18
	490	495.3209	98.91
EXAMPLE-2	0.1606	0.15758	98.12
	0.0637	0.065957	96.46
	0.0502	0.051934	96.55
	0.0534	0.055105	96.81
	0.172	0.16936	98.47
EXAMPLE-3	23.88	23.2719	97.45
	26.41	26.6706	99.01
	22.67	23.0397	98.37
	24.11	23.3298	96.76
	23.26	24.0728	96.51

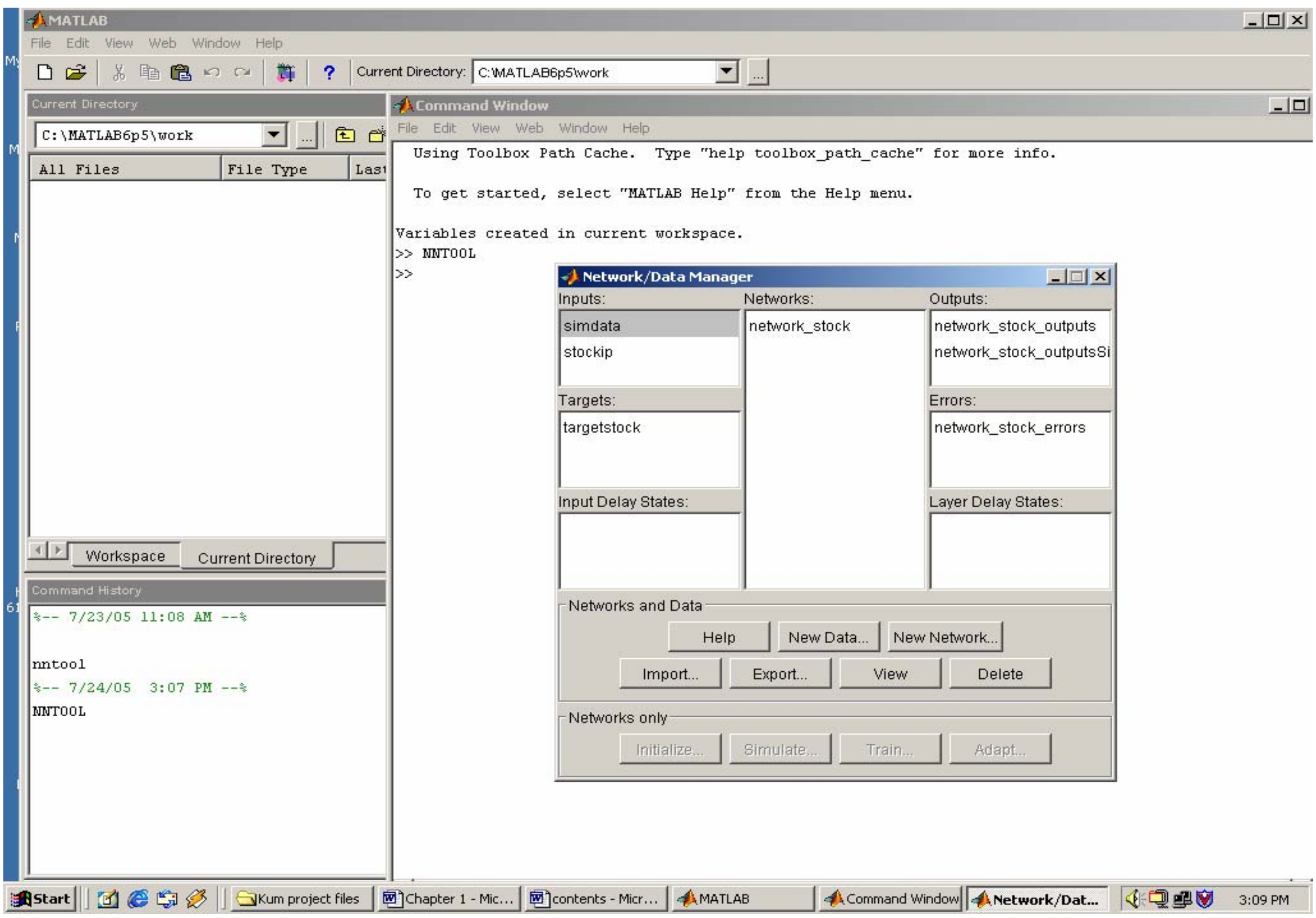
APPENDIX-A

SCREEN SNAPSHOTS OF TIME SERIES ANALYSIS IN MATLAB

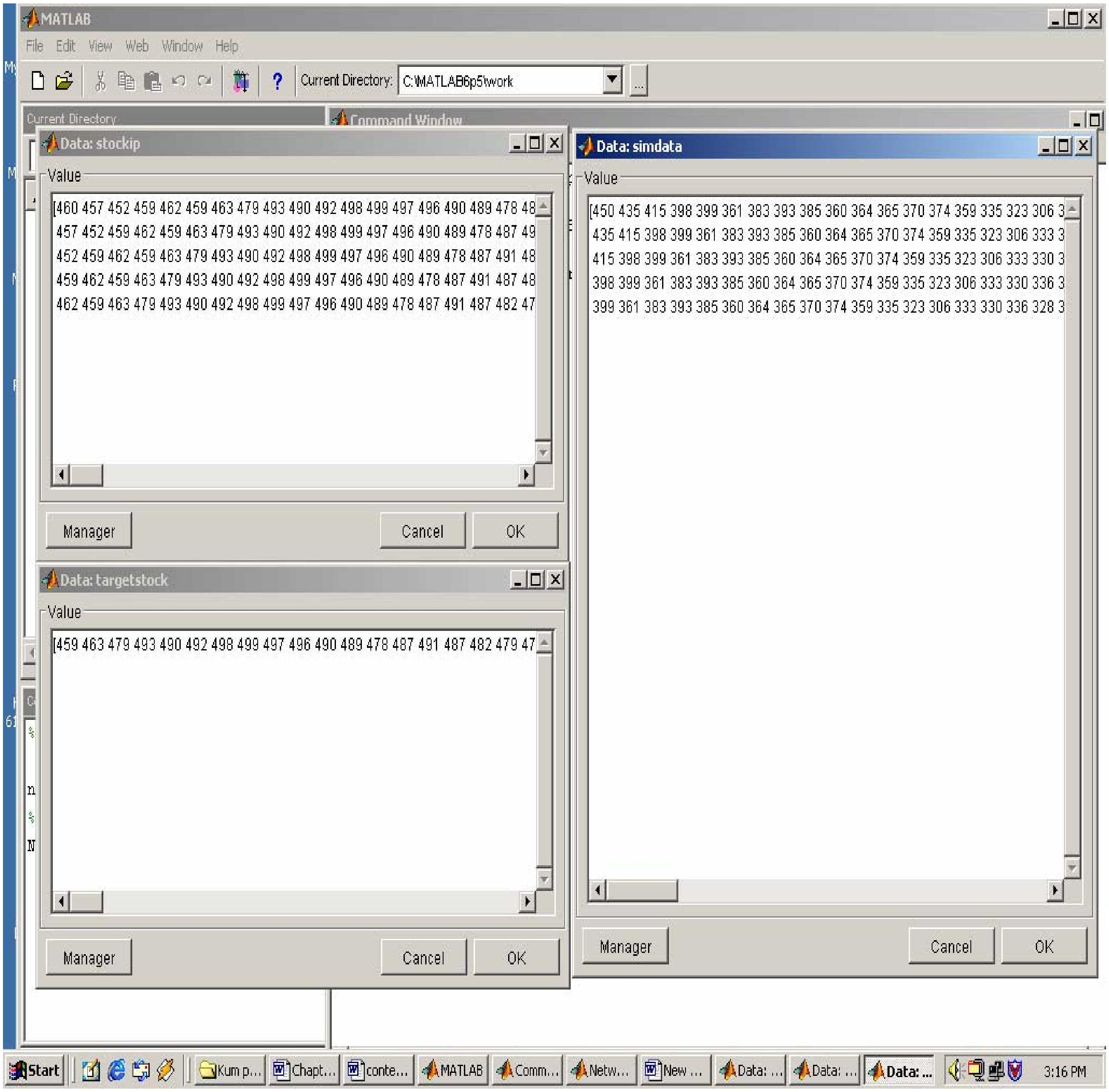
EXAMPLE-1

SERIES B, IBM STOCK PRICES

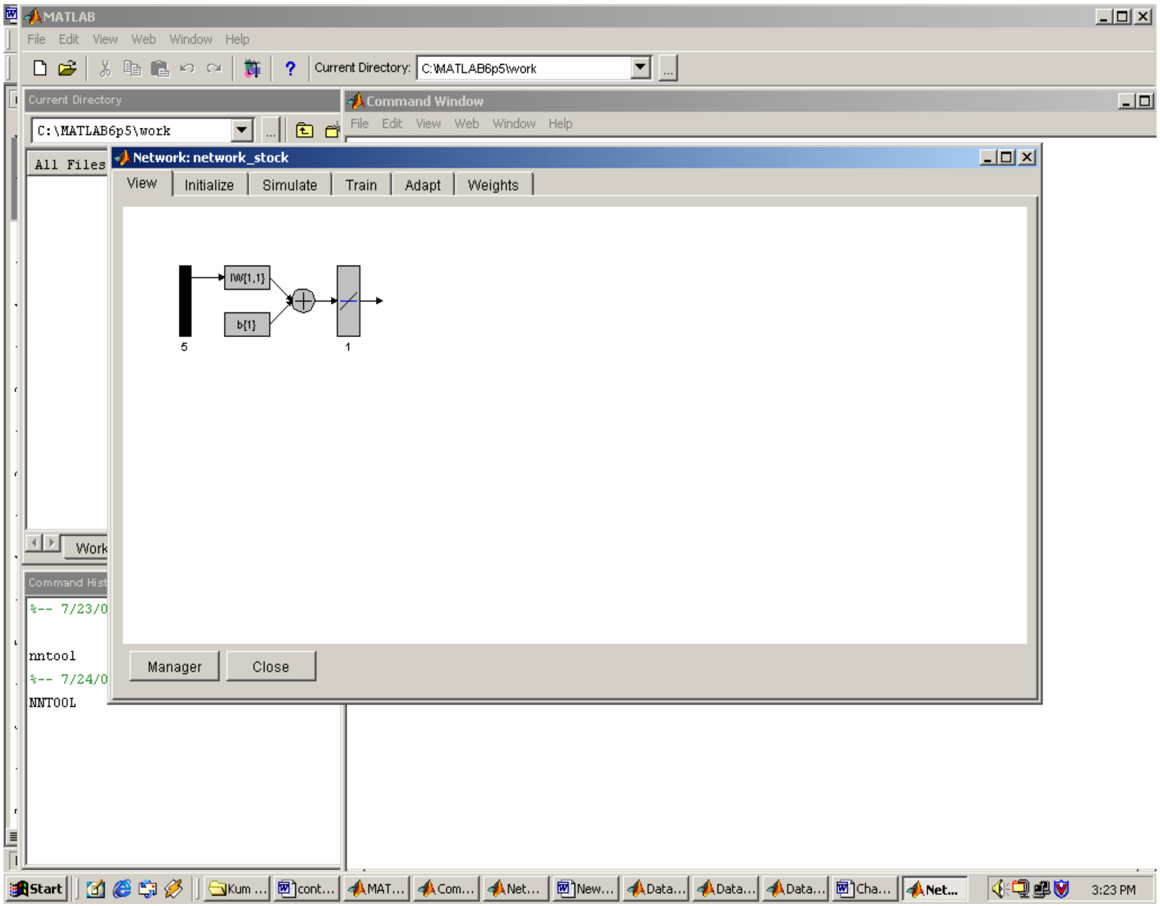
MAIN SCREEN



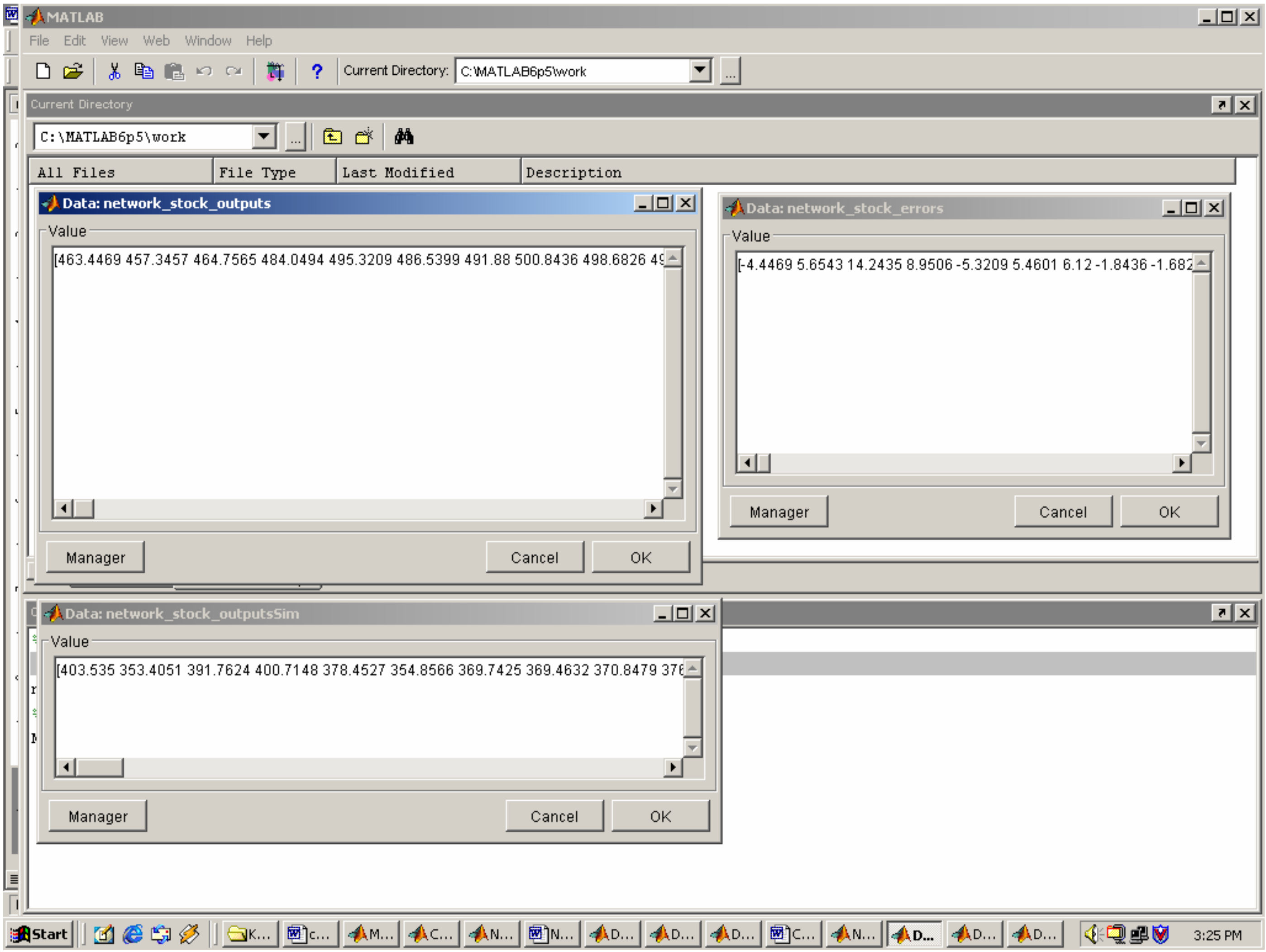
INPUT /TARGET/SIMULATED DATA FOR TRAINING



THE LINEAR LAYER DESIGN NETWORK STRUCTURE



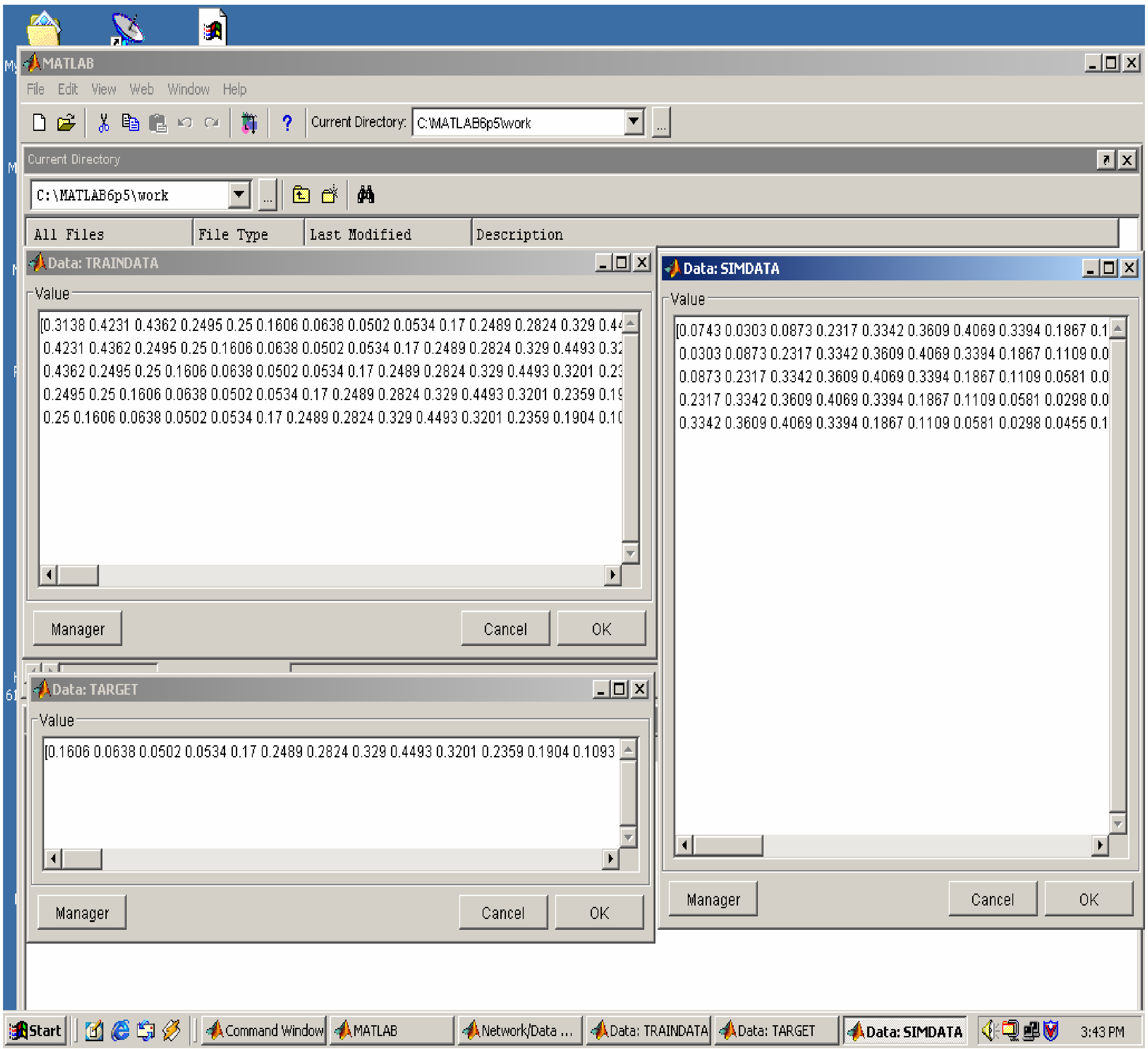
TRAINING OUTPUT AND SIMULATED OUTPUT DATA



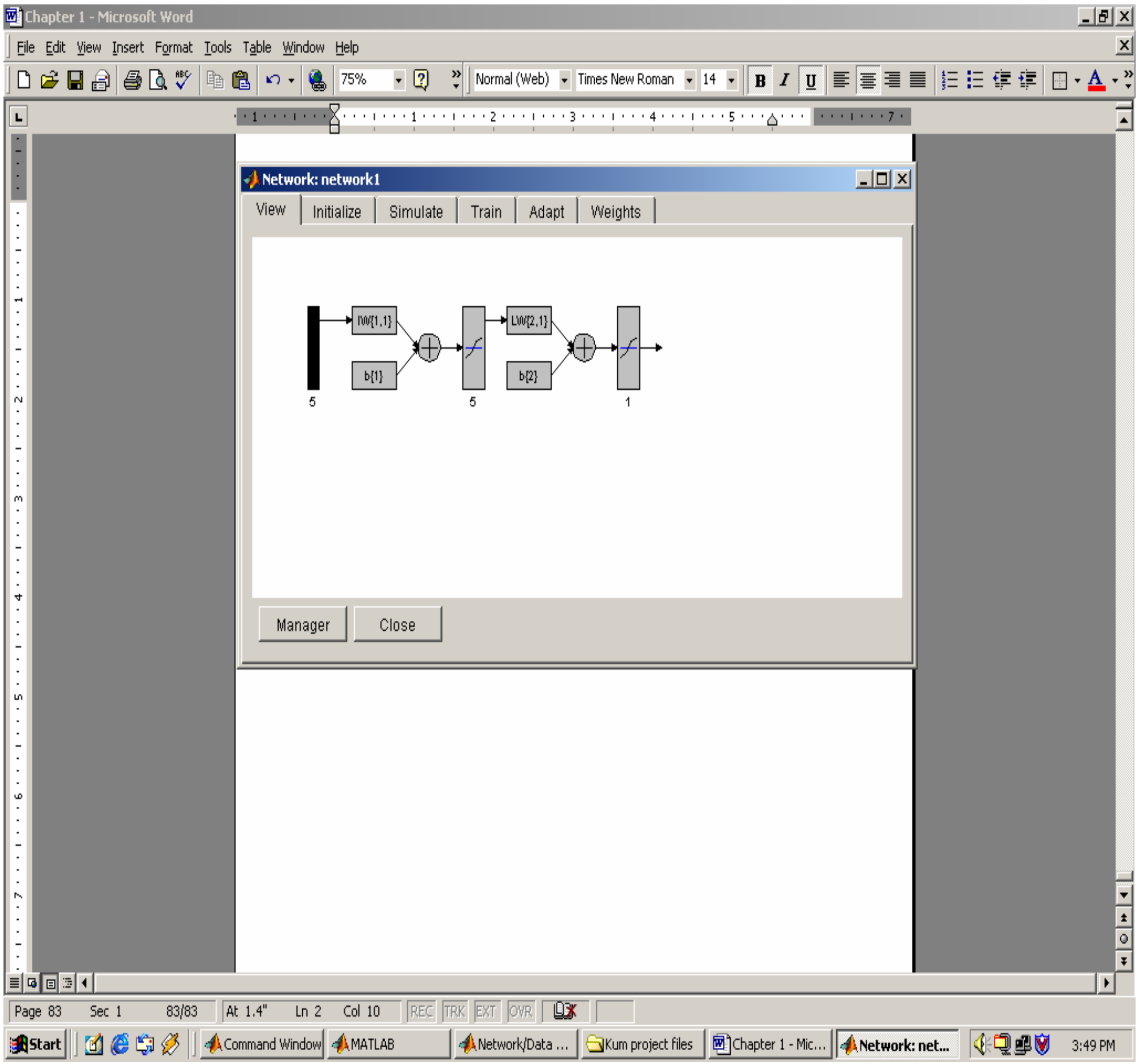
EXAMPLE-2

SUNSPOTS DATA FOR 300 YEARS ON EARTH'S SURFACE, 1699-1999

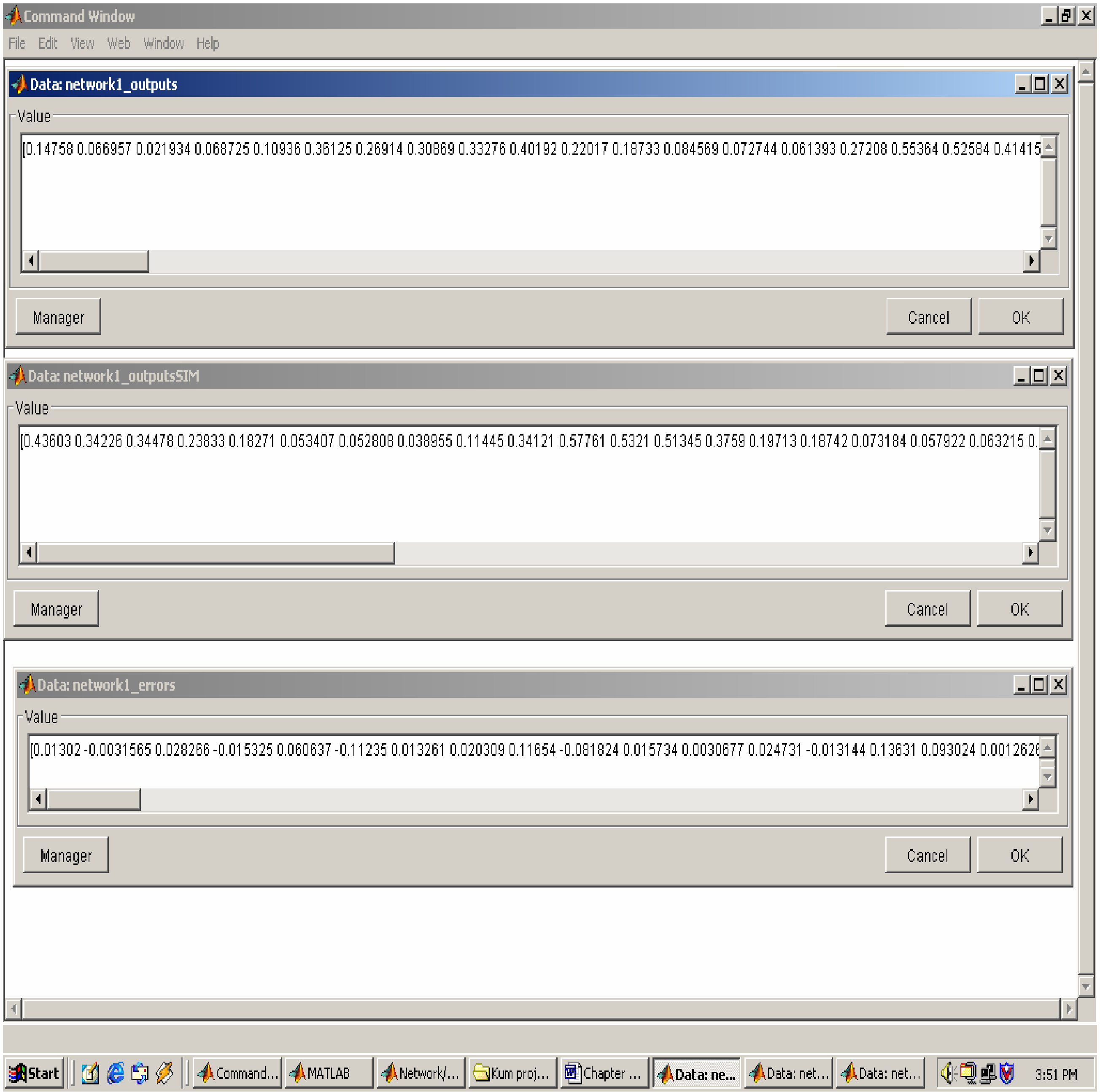
INPUT /TARGET/SIMULATED DATA FOR TRAINING



THE FEEDFORWARD BACKPROPAGATION NETWORK STRUCTURE



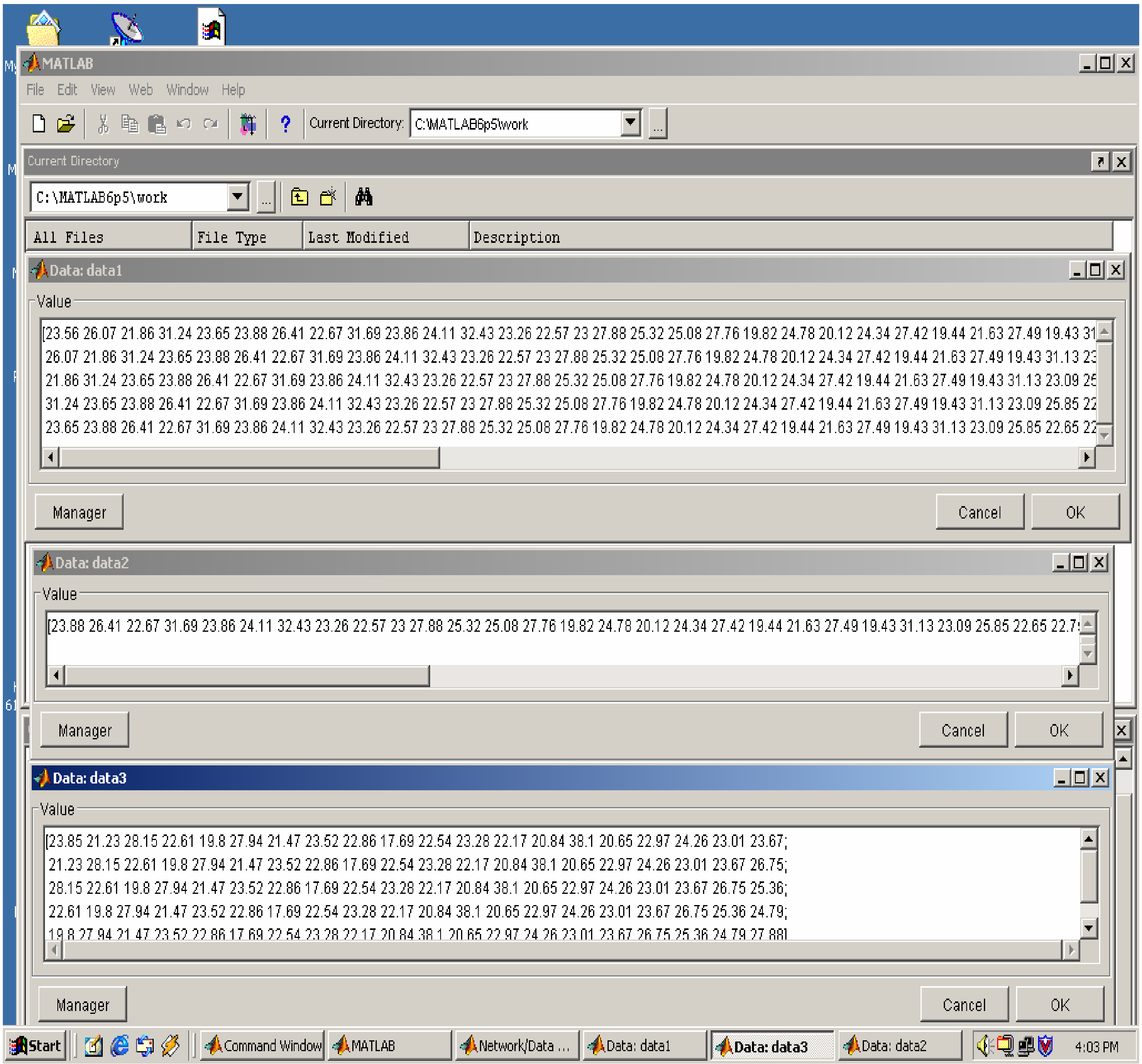
TRAINING OUTPUT AND SIMULATED OUTPUT DATA



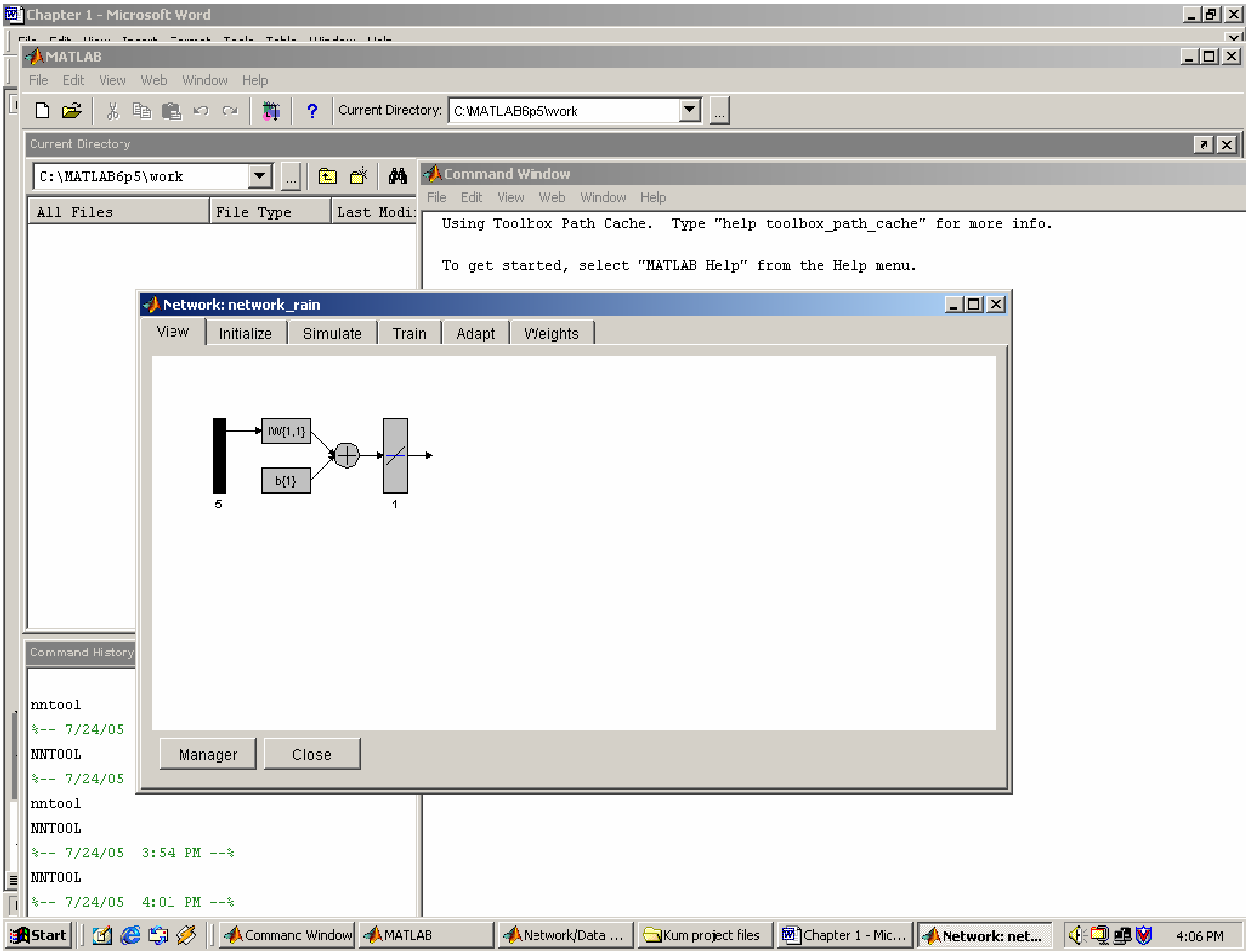
EXAMPLE 3

TOTAL ANNUAL RAINFALL, INCHES, LONDON, ENGLAND, 1813-1912

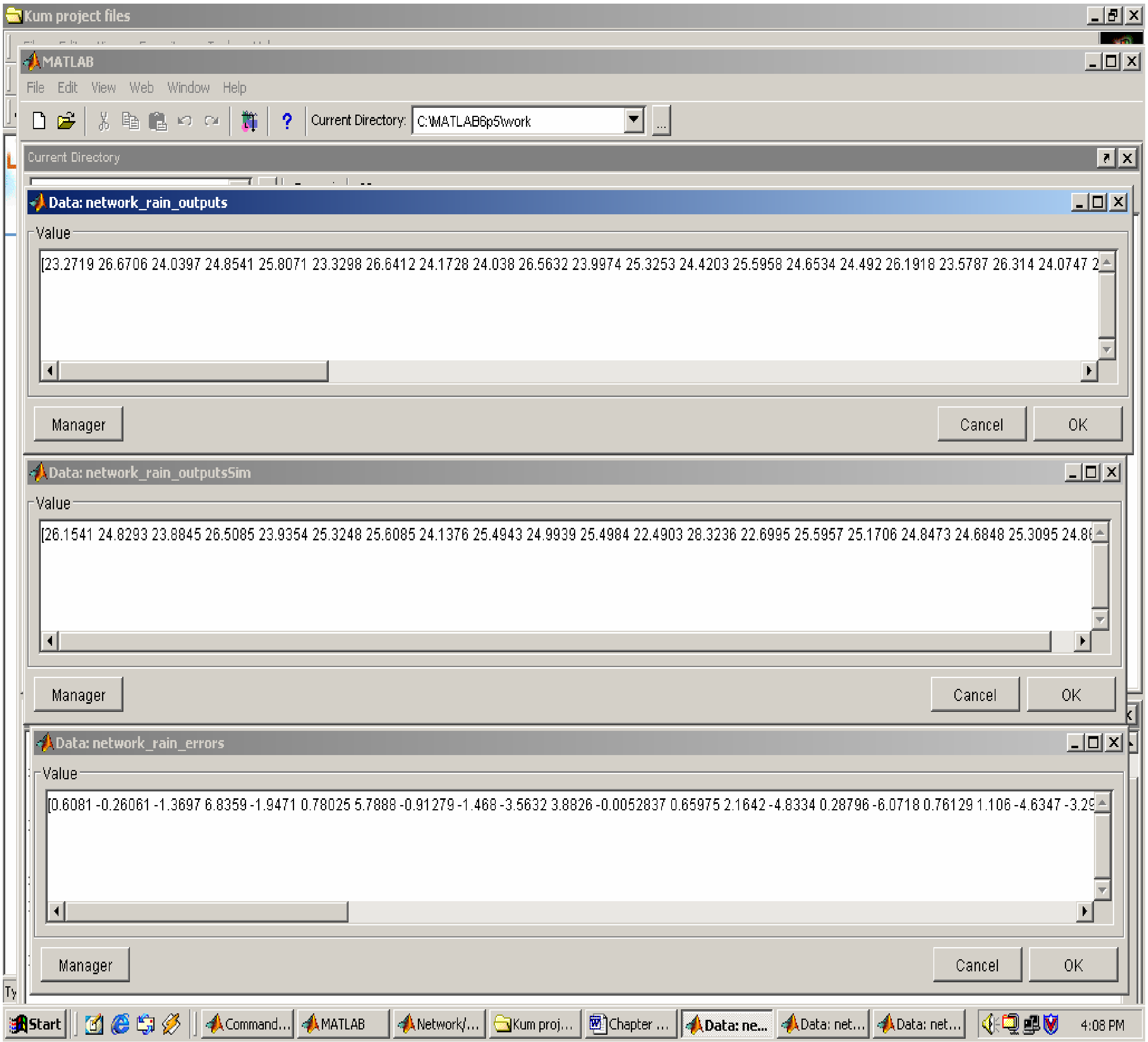
INPUT/ TARGET/ SIMULATED DATA FOR TRAINING



LINEAR LAYER DESIGN NETWORK STRUCTURE



TRAINING OUTPUT AND SIMULATED OUTPUT DATA



References and Bibliography

1. Craven B., and S. Islam, *Optimization in Economics and Finance*, Springer , 2005.
2. Kaplan R., and D. Norton, *The balanced scorecard: Measures that drive performance*, Harvard Business Review, 71, 1992.
3. Yar, M and C. Chatfield (1990), Prediction intervals for the Holt-Winters forecasting procedure, International Journal of Forecasting 6, 127-137.
4. Hamilton J, *Time Series Analysis*, Princeton University Press, 1994.
5. Harvey A., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, 1991.
6. Mills T., *The Econometric Modelling of Financial Time Series*, Cambridge University Press, 1995.
7. Cromwell J., W. Labys, and M. Terraza, *Univariate Tests for Time Series Models*, Sage Pub., 1994.
8. Ho S., M. Xie, and T. Goh, A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction, *Computers & Industrial Engineering*, 42, 371-375, 2002.
9. Kedem B., K. Fokianos, *Regression Models for Time Series Analysis*, Wiley, 2002
10. Kohzadi N., M. Boyd, B. Kermanshahi, and I. Kaastra , A comparison of artificial neural network and time series models for forecasting commodity prices, *Neurocomputing*, 10, 169-181, 1996.
11. Krishnamoorthy K., and B. Moore, Combining information for prediction in linear regression, *Metrika*, 56, 73-81, 2002.
12. Clements M., and D. Hendry, *Forecasting Non-Stationary Economic Time Series*, MIT Press, 1999.
13. Hamouda O., and J. Rowley, (Eds.), Time Series Models, Causality and Exogeneity, Edward Elgar Pub., 1999.
14. www-personal.buseco.monash.edu.-auhyndman-TSDL\Forecasting Methods and Applications.htm

15. www-personal.buseco.monash.edu.-auhyndman-TSDL\Time Series Data Library 1.htm
16. http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm