ECONOMICS COLLECTION
Philip J. Romero and Jeffrey A. Edwards, *Editors*

Seeing the Future

How to Build Basic Forecasting Models

Tam Bang Vu

BEP BUSINESS EXPERT PRESS

Seeing the Future

Seeing the Future

How to Build Basic Forecasting Models

Tam Bang Vu

BUSINESS EXPERT PRESS

Seeing the Future: How to Build Basic Forecasting Models

First published in 2015 by
Business Expert Press, LLC
222 East 46th Street, New York, NY 10017
www.businessexpertpress.com

ISBN-13: 978-1-60649-790-6 (paperback)
ISBN-13: 978-1-60649-791-3 (e-book)

Business Expert Press Economics Collection

Collection ISSN: 2163-761X (print)
Collection ISSN: 2163-7628 (electronic)

Cover and interior design by Exeter Premedia Services Private Ltd., Chennai, India

First edition: 2015

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America.

Abstract

If you are a business person, who desires to forecast demands and sales, or a researcher, who wishes to predict a future event, this textbook is for you. The book is also written for upper-division undergraduate students, first year MBA students, and other readers who wish to acquire fundamental knowledge of quantitative forecasting.

A team of fictional characters is introduced to enhance your learning experience. They comprise two instructors and 10 students. They will share their learning and working experiences with you.

The book emphasizes applied aspects of forecasting, so the only prerequisites for the course are high-school statistics and college algebra. The book discusses most of the forecasting methods frequently used in practice. All in-book analyses are accompanied by numerical examples that can easily be performed on a handheld calculator or Microsoft Excel, which is the only technical software required for all demonstrations and exercises in the book.

Two Excel folders are provided without charge on the BEP website. The first folder, Excel Demos, comprises all Excel files to reproduce the demonstrations in this book, including the data and commands in the cells corresponding to the ones in the text. The second folder, Exercise Data, consists of all data for the exercises in the book.

A folder including all solutions to the exercises is available to instructors once a book order is placed.

Keywords

associative analyses, business, economics, Excel, forecasting, time series

Contents

Preface

This textbook is written for business persons who desire to forecast consumer demands and sales. It is also written for economic, financial, governmental, and private practitioners who want to make informed decisions based on costs, benefits, economic trends, and growth. Additionally, any researcher who wishes to predict a future event will benefit from this book. Finally, the book is written for upper-division undergraduate students, first year MBA or MA students, and other readers who wish to acquire a fundamental knowledge of quantitative forecasting for the purpose of pursuing jobs related to data analyses or making personal plans for the future.

Fictional Characters in This Book

Quantitative topics are often very abstract, and students sometimes lose focus on the subjects. To enhance your learning experience, a team of characters is introduced in this textbook. These characters are global citizens who will share their working and learning experiences with you throughout the book. By observing the applications of theoretical concepts into the business lives of these characters, you will learn the practical aspects of forecasting. Now, let me introduce our professors and classmates to you.

Professors

Dr. App is from Africa. She is an applied researcher and will teach applied forecasting in this course. Dr. Theo is from South America. He is a basic researcher and will be responsible for explaining any theoretical concept in this course.

Students

i. Ex and I are from Asia. Ex works for an import–export company called Excellency. My name is Na, and I am the owner of a nail salon called Nailight in this city.

ii. Alte and Rea are from the Pacific Islands. Alte runs an alteration service called Alcorner, and Rea works as a real-estate agent for a company called Realmart.
iii. Arti and Fligh are from North America. Arti is the director of an arts school called Artistown, and Fligh works for an airline company called Flightime.
iv. Sol and Mo are from Europe. Sol works for a solar energy company called Solarists, and Mo currently works for a motorcycle dealer called Motorland.
v. Fin is from Africa and works as an independent financial advisor.
vi. Cita is from South America and works for the city government.

The 10 of us are attending a forecasting class to either manage our businesses more efficiently or help our family and friends do so. We will join you in making the quantitative topics in this course more enjoyable to learn.

A Practical Approach

This book emphasizes the applied aspects of forecasting, so the only prerequisites for the course are high-school statistics and college algebra. The book discusses most of the forecasting methods frequently used in practice. All in-book analyses are accompanied by numerical examples that can be easily performed on a handheld calculator or Microsoft Excel, which is the only technical software required for all demonstrations and exercises in the book.

Two Excel folders are provided without charge on the BEP website. The first folder, Excel Demos, contains all Excel files that accompany the Excel demonstrations in this book, including the data and commands corresponding to the Excel applications in the text. Data in this folder are organized into chapters, and each of them is titled corresponding to the chapter number, for example, the file Ch01.xls consists of all data and commands for Chapter 1, with Sheet 1 for Figure 1.1 in the text, Sheet 2 for Figure 1.2 in the text, and so on. The second folder, Exercise Data, consists of all data for the exercises in the text. The data in this second folder are not accompanied by Excel commands so that you can become an active participant in the process of writing your own commands.

Topics in This Textbook

This book discusses most of the forecasting methods frequently used in practice, such as time series analyses, demand and sales, investment, short-term planning, long-term growth, and regression analyses. As the regression-based forecasts in this book are designed to be self-explanatory, knowledge of econometrics or time series modeling is not a prerequisite although it might be helpful. To keep the book concise, only a brief introduction to the autoregressive integrated moving average (ARIMA) models and the Box–Jenkins procedure is provided. Readers who wish to gain in-depth knowledge of ARIMA models are encouraged to read a book written specifically for time series modeling.

A folder including all solutions to the exercises is available to instructors once a book order is placed.

Acknowledgments

I wish to express my deepest appreciation to Dr. Philip Romero of the Business Expert Press (BEP), without whose encouragements and guidance this book will never be materialized.

Many thanks go to two former students of the University of Hawaii at Hilo, who are now my colleagues. Eric Santos, chair of Social Science Department, Mahidol University International Demonstration School in Thailand, went over the details of my writings. Shaun McKim, currently a senior analyst at the CVS Pharmacy Headquarters in Rhode Island, United States, shared some of the Excel applications used in the book.

I also wish to recognize help, updated information, and comments from Stewart Mattson, Sheri Dean, Scott Isenberg, Sean Kaneski, Jeff Edwards, and David Parker. The support provided from their respective roles has been crucial.

PART I

Basics

This part contains three chapters:

- Chapter 1 Introduction
- Chapter 2 Elementary Time Series Techniques
- Chapter 3 Evaluations and Adjustments

CHAPTER 1

Introduction

As you know from the preface, Mo works at Motorland in this city to supplement his college expenses. Presently, his boss wants him to estimate the demand for Honda motorcycles from their store next quarter so that they can place an order with their supplier. Here is what his boss says, “Hey Mo, we cannot afford to go by ear any more. Two years ago, we under-ordered and lost at least $10,000 in sales to our local competitors. The following year, we compensated by doubling our order, which left us with a warehouse full of unsold motorcycles. I want to limit our error. I don’t want to under-order or over-order by more than 20 percent. Let’s try to work on it.”

Mo understands his boss’ frustration and collects data on the sales of motorcycles from his store for the past 12 quarters. However, the first day of class will not start until the following week, and Mo does not know how to start his forecasting. Thus, he contacts Dr. Theo, who advises him to read this chapter and says that once he completes it, he will be able to:

1. Describe the basic steps of forecasting.
2. Distinguish qualitative from quantitative forecasting and choose the right method for his estimations.
3. Explain basic concepts of statistics.
4. Apply Excel operations into simple calculations, and chart, and obtain descriptive statistics for his data.

Mo sails through the section on “Starting” and one half of the section on “Basic Concepts” with ease. Here is what he reads in these sections.

Starting

Forecasting is used whenever the future is uncertain. Although forecast values are often not what actually occur, no one can have a good plan without a reasonable approach to form an *educational guess on the future*, which is called *forecasting*.

There are four basic steps in forecasting.

Step 1: Determining the Problem

A forecaster has to define what future information is needed and the time frame of the forecasts. For example, a firm must forecast how much of each good it should produce in the next several months so that it can order inputs from the supply chain and make a delivery plan to the market. To make this decision, the forecaster for the firm needs to understand the basic principles in production planning and communicate with people who have access to the monthly output data. The firm's leaders, then, have to decide how far into the future the information is needed (e.g., a month or a quarter ahead). This future period is the forecast horizon. Finally, the firm needs to decide how often the forecasts have to be updated and revised (e.g., weekly or monthly). This is the adjustment interval or forecast frequency.

Step 2: Selecting the Forecasting Method

Depending on the problem and the availability of the information, an appropriate method of forecasting should be decided. For the preceding example, historical data on the firm's production, constraints in labor, capital, and raw materials can be obtained. Hence, a quantitative method of a data analysis is appropriate. For many other situations, when historical data are not comprehensive or experiences are more important, a qualitative method utilizing the expertise of the professionals can be employed.

Step 3: Collecting and Analyzing the Data

Data can be collected directly by the user (primary data) or by someone other than the user (secondary data). Data analyses consist of constructing time series plots, obtaining descriptive statistics, and calculating the forecast values. Data analyses are crucial in the process of selecting a reliable model. No single model is good for all situations. When a theoretical model is introduced, data analysis should be performed based on this model. When no theoretical model is developed, a forecaster should experiment with several models. For the example in Step 1, a theoretical model of profit maximization or cost minimization is available and should be used.

Step 4: Evaluations and Adjustments

Evaluations are performed based on historical data to select the best model with the smallest forecast errors. Based on the magnitude of the forecast errors, adjustments are made to the existing models. Monitoring is then carried out because forecasting is a long-term process. Business and economic conditions change over time as do the forecasts. A good forecaster needs to track the changes so that the causes are pinpointed. New forecast errors need to be calculated. Either a new model might be developed, or a combination of several models might be introduced.

Basic Concepts

Qualitative Forecasting

There are three main qualitative methods in addition to the naïve approach of taking the current-period value as the forecast of the next-period value.

Individual judgment is a forecast made by an expert in a field based on the person's experience, the past performance of the market, and the current status of the business and economy. The expert can employ: (i) analogy-analysis technique by comparing similar items, (ii) analyzing possible scenarios in the near future, or (iii) collecting qualitative data by sending out survey forms to the respondents and performing a qualitative analysis of these data.

Panel of experts is a forecast made by a group of professionals, whose opinions are combined, averaged, and adjusted based on discussions and evaluations by all members of the group. An executive officer can lead the discussions but consensus has to be obtained at the end. All the techniques used in the individual judgment can be utilized.

Delphi method is similar to the panel of expert method, but the experts are not allowed to discuss the problem with each other. Instead, they are given an initial set of questions, to which their answers are anonymous to guarantee the objectivity of each member. Their answers to the first set of questions are the basis for the next set, and the process is repeated with the hope that the answers gradually converge. (Sometimes they do not obtain any consensus, and the process breaks down.)

Quantitative Forecasting

There are two large groups of quantitative forecast methods. The first group comprises time series analyses, and the second consists of associative analyses. Both share a common denominator of performing data analyses to draw conclusions on possible future outcomes.

Time series analyses examine only historical data of the time series itself instead of adding any external factor. The method comes from the notion that past performance might dictate the future performance of a market. Techniques for the time series analyses comprise moving averages, exponential smoothing, decomposition, autoregressive (AR), and autoregressive moving average (ARMA/ARIMA) models.

Associative analyses are based on the investigation of various external factors that affect the movements of a market. The associative analysis contains regression analysis and nonregression analysis. The regression analysis is based on the econometric technique and always involves an equation with a dependent variable on the left-hand side and one or more explanatory variables on the right-hand side. For example, spending on motorcycles at Motorland depends on the income level of the city residents.

The nonregression analysis involves variables that are related to each other in a certain manner that is not appropriate for the regression technique. For example, to forecast a turning point in the economy, various measures called economic indicators are developed. Each indicator depends on several factors. A composite index, which involves the most important indicators, is then calculated for each month. A forecaster can predict a turning point based on the changing direction of this index over time. Thus, a nonregression technique is appropriate for this case.

The focus of this book is on quantitative forecasting. Hence, all the aforementioned techniques for quantitative forecasting will be discussed. Most of them will be followed by Excel demonstrations. Once you master the quantitative forecast methods, you can combine quantitative and qualitative methods to obtain the best forecasting results.

Statistics Overview

Probability

Mo learns that probability is the likelihood of an event occurring and is measured by the ratio of the favorable case to the total number of possible cases. As an example, we can conduct the following experiment.

Let variable X equal throwing a coin, and let getting a head side of the coin be the objective, which is the favorable case in our question, then

x_1 = getting a head side = 1
x_2 = getting a tail side = 0

If the coin is fair, then you have the probability of getting a head $P(1) = f(x_1) = 0.5 = ½$ and the probability of getting a tail $P(0) = f(x_2) = 0.5 = ½$.

Throw the coin *three times*, and we will have a table of all probabilities as follows:

Probability	Number of heads
$P(0,0,0) = 0.5 * 0.5 * 0.5 = 0.125$	0 (= probability of getting no head three times in a row)
$P(0,0,1) = 0.125$	1
$P(0,1,0) = 0.125$	1
$P(0,1,1) = 0.125$	2
$P(1,0,0) = 0.125$	1
$P(1,0,1) = 0.125$	2
$P(1,1,0) = 0.125$	2
$P(1,1,1) = 0.125$	3

The graph of this table is the probability distribution function (*pdf*).

A discrete variable has only a handful of values. The variable X in the preceding example is a discrete variable. A continuous variable has numerous values, so its distribution is the area under a smooth curve. For example, if we let variable Y equal throwing the coin 10,000 times, then the variable Y can be considered a continuous variable. Graphing the *pdf* of this variable yields a smooth, bell shaped curve, which is called a normal distribution.

Measures of Central Tendency

Mo also understands that descriptive statistics are usually obtained before performing forecasts, and the following concepts are important to remember.

If X and Y are any two random variables whereas a and b are any two constants, then the expectation of X is the weighted mean (the weighted average) of all xs:

$$E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots + x_n P(X = x_n)$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX) = aE(X)$$

At this point, Mo gets lost. The formulas scare him. Luckily, time flies fast, and the class starts today. The following is Dr. Theo's example of the *expectation* concept.

"Think of a student's GPA for this course. Suppose there are equal weights to the two midterm exams and the final exam. If the student makes a C (2 points) in Midterm 1, a B (3 points) in Midterm 2, and an A in the Final Exam (4 points), what is your expectation of the student's GPA for this course?"

We all can answer this question as GPA = (2 + 3 + 4)/3 = 3

"So the student makes a B," we say in unison. Dr. Theo agrees and continues.

"Suppose that Midterm 1 receives a 20 percent weight, Midterm 2 receives a 30 percent weight, and Final Exam receives a 50 percent weight in this class. What is your expectation of the student's GPA for this course?"

We apply the new weights: GPA = 0.20 * 2 + 0.30 * 3 + 0.5 * 4 = 3.3.

"Wow, the guy now makes a B+. That is cool! This is because he scored best (an A) on the highest-weighted exam (the final)," we exclaim. We now understand the concept of *expectation* and move on to the variance.

The variance of X is the average of the squared difference between X and $E(X)$.

The variance measures the dispersion of a distribution:

$$\mathrm{Var}(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2)[E(X)]^2$$

$$\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$$

Alte raises her hand and asks, "Why do we need to calculate the square of the difference?"

Dr. Theo replies with another example, "There is an avocado tree in my backyard with a lot of ripe avocados, but they are very high above the ground. I tried to hit a fruit with a long pole. At the first strike, I missed the fruit roughly five inches to the right. At the second strike, I missed it roughly five inches to the left. May I boast to you that *on average* I got the avocado because the average distance from my hits to the avocado is $D = (5 - 5)/2 = 0$?"

We all laugh. Of course, the value cannot be zero.

Hence, Dr. Theo concludes, "If we do not square the distance before taking the average of the difference, then the negative values cancel out the positive ones, and the average is zero, so we cannot measure the dispersion of X. The covariance is also easy to understand if you think of your relationship with your mom, who is much closer to you than to a person on the street."

We find that the formula for the covariance is similar to the one for the variance, except that we enter Y in place of another X. And thus, the covariance of X and Y measures the linear association between them:

$$\mathrm{Cov}(X, Y) = E\{[(X - E(X)][Y - E(Y)]\}$$

If the two are independent, they have zero covariance, but the reverse is not true. For example, Y and X in the function $Y^2 + X^2 = 1$ have zero covariance because the equation is not linear. However, these two variables belong to a nonlinear equation, so they are not independent of each other:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y)$$

$$\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + ab\,\mathrm{Cov}(X, Y)$$

If the two are independent or not correlated linearly, then:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$$

We all understand now. The next several concepts are easy and so we do not need further explanations.

The median is the value of the middle observation.
The standard deviation (SD = σ) is the square root of the variance.
The standard error (SE = s) is the sample version of the SD.
The skewness is the asymmetry of a distribution.

$\mathrm{SKEW}(X) = E[X - E(X)]^3$
A left-skewed (negatively skewed) distribution has a long left tail.
A right-skewed (positively skewed) distribution has a long right tail.
Also, a left-skewed sample has the mean smaller than the median.
A right-skewed sample has the mean greater than the median.
A symmetrical dataset has the mean equal to the median.
A perfectly normal distribution has skewness = 0.
Any distribution with |SKEW| > 1 is considered far from normal.

Next, we come to a little more abstract concept: the Kurtosis.

The Kurtosis is the peakedness (the height) of the distribution (Poirier 1996).
It can also be a measure of the thickness of the tail (Greene 2012):

$$\mathrm{KUR}(X) = E[X - E(X)]^4$$

Excess Kurtosis = EK = |KUR| – 3 (Davidson and MacKinnon 2004)

Dr. Theo reminds us that details of Kurtosis are explained in Brown (2014).

A perfectly normal distribution has KUR = 3 or an EK = 0.
Any distribution with EK > 1 is considered far from normal.

Rea raises a doubt, "Does that mean that if KUR = 5, then we do not have a normal distribution?" Sol volunteers to answer, "I think so, because in that case, EK = 5 − 3 = 2 > 1."

Dr. Theo commends Sol on her correct answer and asks if we have any question on the next two terms. We think these new terms are easy to understand:

The mode is the value that occurs most frequently.
The range is the difference between the highest and the lowest values.

Hence, nobody asks anymore questions. Dr. Theo concludes the theoretical section with a summary of the important concepts and reminds the class to read the next section, which will be taught by Dr. App.

Basic Excel Applications

Dr. App starts with easy mathematical operations in Excel.

Mathematical Operations

We learn that all Excel calculations start with an *equal* sign (=). For example, to add variable *X* in cell A2 to variable *Y* in cell B2, type

= A2 + B2 and press Enter.

To form a product (or quotient) of *X* and *Y*, type

= A2 * B2 (or A2/B2) and press Enter

For consistency, the notation * will denote a multiplicative operation throughout this book.

To calculate $X^{1/2}$, type

= A2^(1/2) and press Enter

To take the logarithm of variable X, type

= ln(A2) and press Enter

Excel uses only parentheses for all mathematical operations. For example, to enter this mathematical formula, $\{[(XY - Y) + (X + Y)^2] - Y^3\}/X$, type in Excel

= (((A2 * B2 – B2) + (A2 + B2)^2) – (B2)^3)/A2 and press Enter.

Dr. App reminds the class that once a formula is formed, you can copy and paste it into any cell using the *copy* and *paste* commands. She also assures us that more operations will come later. We then proceed to play around with the data. The following are the topics that we are learning.

Data Operations

Three Types of Data

There are cross-sectional, time series, and longitudinal (panel) data. Cross-sectional data are for many identities in a single period. The identities could be persons, companies, industries, regions, or countries. Time series data are for a single identity over many periods. The periods could be days, weeks, months, quarters, years, or many years. Longitudinal and panel data are for many identities over many periods.

Figure 1.1 provided an example of the three datasets on the number of associate degrees awarded for three economic regions in the United States from 2002 to 2010. Dr. App reminds us that all demonstrations are available in the Excel Demos folder. Data for Figure 1.1 are in the file Ch01.xls (see sheet Fig.1.1). The cross-sectional dataset is for three economic regions in the United States in one period (2002–2004). The time series dataset is for one region over three periods (2002–2004, 2005–2007, and 2008–2010), and the panel dataset is for the three regions over the three periods.

	A	B	C	D	E	F	G	H	I
1	Cross Sectional	2002-2004		Time Series:	Mideast		Longitudinal		
2	**Region**	Degree		Period	Degree		Period	Region	Degree
3	New England	81084		2002-2004	303661		2002-2004	New England	81084
4	Mideast	303661		2005-2007	335245		2005-2007	New England	81125
5	Great Lakes	277937		2008-2010	375972		2008-2010	New England	89356
6							2002-2004	Mideast	303661
7							2005-2007	Mideast	335245
8							2008-2010	Mideast	375972
9							2002-2004	Great Lakes	277937
10							2005-2007	Great Lakes	331640
11							2008-2010	Great Lakes	367297

Sheet1 **Sheet2** Sheet3

Figure 1.1 Three types of data

Data Source: Adapted into three-year intervals from the yearly data from National Center for Education Statistics, United States.

Charting Tools

These tools can be used to construct a time series plot and examine its pattern and trend.

Alte raises her hand and says that she has a dataset from her Alcorner with which she wishes to construct a plot. We are happy to accommodate her. Its plot is displayed in Figure 1.2, where her weekly sale values are rounded off to hundreds of dollars. Dr. App tells us to open the file Ch01.xls, Fig.1.2. She says that we should first obtain a simple image of the time series and then perform the following steps:

Remove the label in cell B1 if there is any.
Highlight cell B1 through C10.
Click on the Insert tab from the Ribbon.
Click on Line icon under the Chart section.
Select any 2D line you wish to use.
For example, clicking on the first choice beneath the Line icon gives us the plot in Figure 1.2.

We now have a simple plot sketched and decide to retype the label in cell B1 (Time in Figure 1.3). Dr. App says that to label the axes, we have to open the file Ch01.xls, Fig.1.3, and perform the following steps:

Click on the graph and go to Layout under the Chart Tools section.
In the Layout click on Axis Titles.

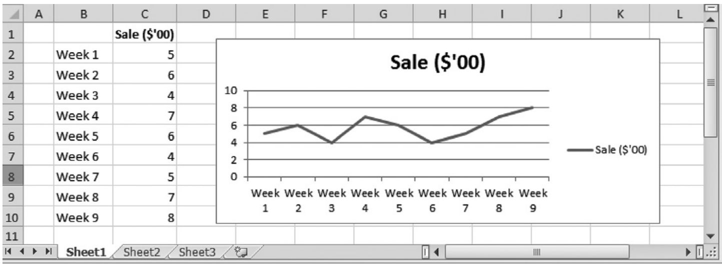


Figure 1.2 Obtaining a simple image of a time series

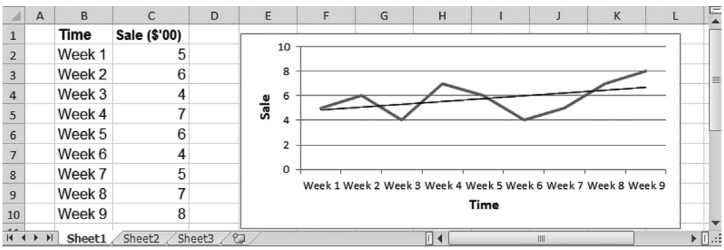


Figure 1.3 Obtaining a detailed time series plot

In Primary Horizontal Axis Title, choose Title below Axis.
Type a title in the box below the axis.
Repeat the same for the vertical axis if you wish to add a title to it.
To add a trend line, right click on the chart line.
A list of commands will appear. Choose Add Trendline.
A new dialog box will appear. Choose Linear (nonlinear is possible) and click Close.

We can also change the line style and color style in the file Ch01.xls, Fig.1.4, by following these steps:

Right click on the chart line.
Choose Format Data Series from the dropdown menu.
A dialog box will appear.
Click on Line Style from the menu on the left-hand side column.
To experiment, click on the arrow in the Dash Type box to open a drop down menu.

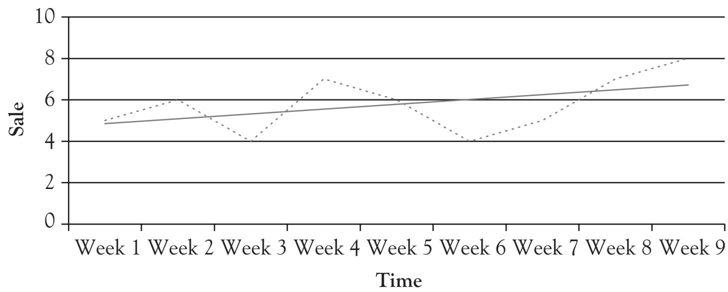


Figure 1.4 Changing line style

Click on the Round Dot option and then click Close to obtain the result shown in Figure 1.4.

We learn that we can follow the same procedure to change the color.

Add-in Tools

Next we need to install a tool to perform data analysis, so we work with Dr. App, who gives the following instruction.

For Microsoft Office (MO) 1997 to 2003:

Go to Data Tools, click on Add-Ins from the drop down menu.
Click on Analysis Toolpak from the new drop down menu then click OK.
Whenever you need this tool, click on Data Tools and then Analysis ToolPak.

For MO 2007:

Click on the office logo at the top left that you have to click to open any file.
Click on Excel Options at the bottom center.
Click on Add-Ins from the menu at the bottom of the left column in the Excel Options.
The View and Manage Microsoft Office Add-Ins window will appear.

In this window, click on Go at the bottom center.

A new dialog box will appear, check the Analysis ToolPak box and then click OK.

Whenever you need this tool, click on Data and then Data Analysis on the Ribbon.

For MO 2010:

Click on File and then Options at the bottom left column.

Click on Add-Ins from the menu at the bottom of the left column in the Excel Options.

The View and Manage Microsoft Office Add-Ins window will appear.

In this window, click on Go at the bottom center.

A new dialog box will appear, check the Analysis ToolPak box and then click OK.

Whenever you need this tool, click on Data and then Data Analysis on the Ribbon.

Data Manipulations

We learn that if a column is not wide enough, Excel displays ##, as shown in the left-hand cells of Figure 1.5, instead of values. To change the display, we need to open the file Ch01.xls, Fig.1.5, and follow these steps:

Highlight columns B, C, and D.

Click on Format on the Ribbon then choose Column Width.

A dialog box will appear.

Enter a value larger than the default value.

	A	B	C	D	E	F	G	H
1	Population	(Persons)						
2		2006	2007	2008		2006	2007	2008
3	Canada	########	########	########		32,656,679	32,935,961	33,212,696
4	United States	########	########	########		298,818,000	301,696,000	304,543,000
5	Mexico	########	########	########		107,449,525	108,700,891	109,955,400
6								

Sheet1 Sheet2 Sheet3

Figure 1.5 Changing column width

Data Source: U.S. Department of Agriculture.

	A	B	C	D	E	F	G	H	I
1	Population	(Persons)							
2		2006	2007	2008			Canada	United States	Mexico
3	Canada	32,656,679	32,935,961	33,212,696		2006	32,656,679	298,818,000	107,449,525
4	United States	298,818,000	301,696,000	304,543,000		2007	32,935,961	301,696,000	108,700,891
5	Mexico	107,449,525	108,700,891	109,955,400		2008	33,212,696	304,543,000	109,955,400
6									

Sheet1 Sheet2 Sheet3

Figure 1.6 Transposing the data

Data Source: U.S. Department of Agriculture.

In this case, enter 12 and then click OK.

The results are shown in the right-hand cells of Figure 1.5.

Fligh discovers that we can follow a similar procedure to change the row height, and we are delighted to follow him.

Dr. App points out that data are always analyzed with time displayed vertically. However, many data sources have time displayed horizontally to ulilize the available space as shown in the left-hand cells of Figure 1.6. She says that to change from a horizontal to a vertical time arrangement, we need to open the file Ch01.xls, Fig.1.6, and perform the following steps:

Copy cells A2 through D5 then right click cell F2.

Under Paste Options choose Paste Special.

A dialog box will appear. Choose Transpose and then click OK.

We now see the new vertical time arrangement displayed in the right-hand cells of Figure 1.6.

Formating Cells

Arti, the director of Artistown, raises her hand and says that her school usually keeps a spreadsheet of the books demanded from her students with reserve quantity equaling 50 percent of the demand. She shows us Figure 1.7, which displays the reserved books with decimal places in column D. This is incorrect because the numbers of books are always in integers. So she asks, “How can I change the quantity of books to integers?” Dr. App commends her on the question and says that we need to open the file Ch01.xls, Fig.1.7, and perform the following steps:

	A	B	C	D	E	F	G	H
1		Month	Book Demanded	Reserve		Month	Book Demanded	Reserve
2		January	31	15.5		January	31	16
3		February	42	21		February	42	21
4		March	50	25		March	50	25
5		April	70	35		April	70	35
6		May	46	23		May	46	23
7		June	43	21.5		June	43	22

Figure 1.7 Removing decimal places in column D to obtain integers in column H

Highlight colum D then right click on this column.
Click on Format Cells from the dropdown menu.
A dialog box will appear.
Click on Number that is on the left-hand side of the box.
Use the arrow in the Decimal Places box to select 0 then click OK.
We now see that the correct quantities are in column H of Figure 1.7.

We learn that we can change any cell's format by following similar steps in the Format Cells dialog box.

Descriptive Statistics

This is an abstract concept, so Dr. App reminds us that descriptive statistics provide information on a specific sample used in forecasting instead of information on a whole population. It gives an overall feeling on the data without concerns on the whole market it represents. She then tells us to open the file Ch01.xls, Fig.1.8 and Fig.1.9, and perform the following steps to obtain descriptive statistics:

Go to Data and click on Data Analysis on the Ribbon in Excel.
A dialogue box will appear; select Descriptive Statistics and click OK.
Another dialog box will appear as shown in Figure 1.8.
Enter C1:C10 in the Input Range box.
Choose Label in the First Row and Summary Statistics.
Check the Output Range button and enter E1.
Click OK. A new dialogue box will appear.
Click OK to overwrite data in range.
The Descriptive Statistics are shown in Figure 1.9.

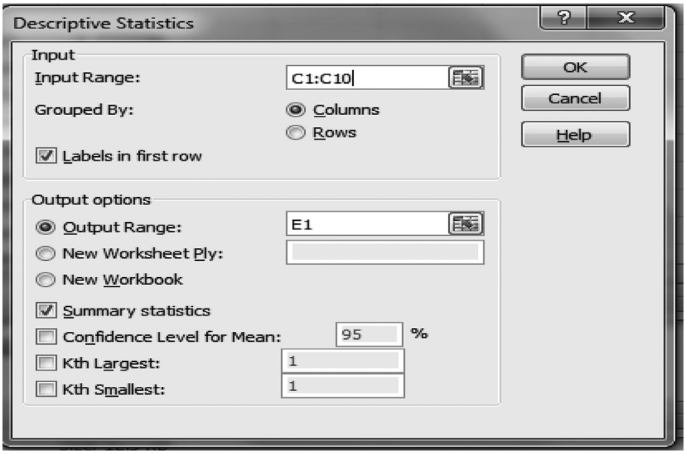


Figure 1.8 Dialog box for the descriptive statistics

	A	B	C	D	E	F
1		Time	Sale ($'00)	Sorted Data	*Sale ($Th)*	
2		Week 1	5	4	Mean	5.888888889
3		Week 2	6	5	Standard Error	0.423098506
4		Week 3	4	5	Median	6
5		Week 4	7	5	Mode	5
6		Week 5	6	6	Standard Deviation	1.269295518
7		Week 6	5	6	Sample Variance	1.611111111
8		Week 7	5	7	Kurtosis	-0.700186852
9		Week 8	7	7	Skewness	0.260025547
10		Week 9	8	8	Range	4
11					Minimum	4
12					Maximum	8
13					Sum	53
14					Count	9

Sheet4 | Sheet1 | Sheet2 | Sheet3

Figure 1.9 The descriptive statistics and sorted data

Ex points out that it is also helpful to sort the data so that we can have a visual image of the series. He shares his experience in sorting the data with the class.

To sort the data from the smallest value to the largest one in Figure 1.9:

Copy the series in cells C2 to C10 and paste it into cells D2 to D10; name the column as Sorted Data.

Highlight D1:D10 then go to Data on the Ribbon and click on Sort.

Choose Continue with the Current Selection and then click Sort.

A new dialog box will appear as shown in Figure 1.10.

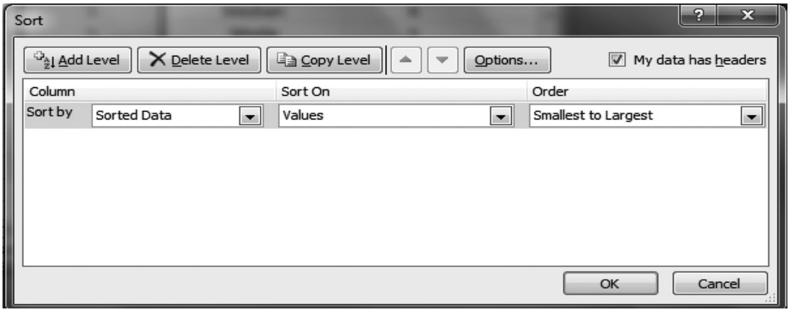


Figure 1.10 Sorting a time series

Choose Sorted Data in the 1st box and Values in the 2nd box.

Choose Smallest to Largest in the 3rd box and click OK to obtain the results in column D of Figure 1.9.

Dr. App says that the descriptive statistics provide a hint of what the sample looks like and what you should do with it in order to obtain reliable forecast values. For example, if the mean and the median are very different from each other, you might have an outlier that needs to be eliminated before data analyses can be performed. In Figure 1.9, since the mean and the median are close to each other, there is no need to eliminate any observation. A glance through the sorted data reveals that the median and the mode are indeed \$6,000 and \$5,000, respectively.

We finish with the applied section of the chapter and look forward to the next chapter, where we will learn about the simplest techniques of forecasting.

Exercises

1. Given a discrete variable X that is a single roll of a fair die,
   a. What is the probability that $X = 2$?
   b. What is the probability that $X = 2$ or 3?
2. Let X be a discrete random variable with the values $x = 0, 1, 2$ and let the probabilities be

$$P(X = 0) = 0.25$$
$$P(X = 1) = 0.50$$
$$P(X = 2) = 0.25.$$

a. Calculate $E(X)$
b. Find $E(X^2)$
c. Find $\text{Var}(X)$
d. Given a new function

$$g(X) = 3X + 2$$

Find the expectation and variance of this function.

3. Data for the Alcorner are provided in the file Alter.xls of the folder Exercise Data. Provide a time series plot of this sample, including the axis labels and a linear trend line. Give comments on the trend.
4. Data on sale values from Solarists are provided in the file Sales.xls of the folder Exercise Data. The sale values are in hundreds of dollars. Sort the data and then obtain descriptive statistics for this sample. Compare the mean to the median. Are they very close? What is the implication?

CHAPTER 2

Elementary Time Series Techniques

Having finished reading Chapter 1, Mo decides that he has collected enough historical data to employ a quantitative forecasting method. However, he worries that he does not know any quantitative techniques. Dr. Theo assures him that this chapter will introduce the two easiest ones: simple moving averages (MA) and exponential smoothing (ES). Both of them are used to obtain one-period forecasts. As Mo has to forecast how many Honda motorcycles to order next quarter, these techniques are good starting steps for his task. Dr. Theo says that once we finish with this chapter, we will be able to:

1. Describe the four components of a time series.
2. Distinguish the MAs from the weighted moving averages (WMs).
3. Explain the ES technique.
4. Apply the Excel commands into calculating one-period forecasts.
5. Analyze and demonstrate the technique of converting nominal values to real values.

Mo looks through the chapter and sees that the section on "Components of Time Series" is easy and fun to read.

Components of Time Series

Time series data comprise trend, seasonal, cyclical, and random components.

The trend component measures the overall direction of a time series, which could have an upward, a downward, or an ambiguous moving direction (not having a trend). Figure 2.1 shows a plot of the monthly visitor arrivals by air at Honolulu.

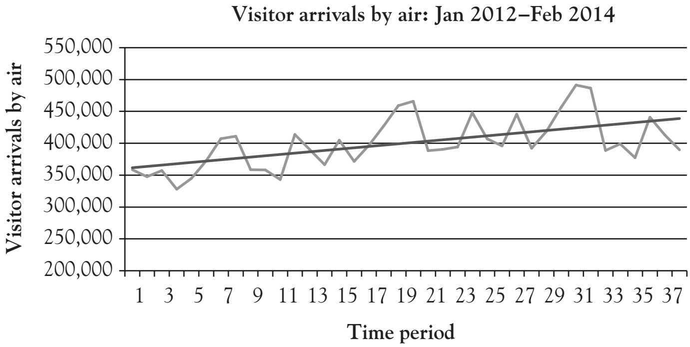


Figure 2.1 Visitor arrivals by air at Honolulu: trend component

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

Dr. Theo reminds us that the data and Excel commands for this chapter are in the file Ch02.xls. The monthly dataset comes from Fligh, an employee of Flightime Airlines. He has carried out a research on the visitor arrivals to U.S. cities and is eager to share the information with us. This monthly dataset is available in the file Ch02.xls, Fig.2.1, for the period from January 2011 to February 2014. The trend line reveals an upward direction in this case.

The seasonal component is the short-term movement that occurs periodically in a single year. To see this pattern more clearly, we chart the monthly data from the file Ch02.xls, Fig.2.2, for the period from February 2012 to January 2013 and display the plot in Figure 2.2. From this figure, the peak seasons of tourists coming to Honolulu are clear. Visitor arrivals are high in December followed by the month of August. And visitor arrivals are low in April followed by the month of November.

The cyclical component is the long-term fluctuation that also occurs periodically, for several years and sometimes for several decades. For example, Hawaiian tourism experienced a long declining period from 1992 to 2002 without any substantial interruption. Hence, we chart yearly data from the file Ch02.xls, Fig.2.3, for the period 1992–2013 to see a full cyclical pattern from a lowest point (the trough) to a highest point (the peak). Figure 2.3 reports the results, which reveal a full cycle of peak-to-peak from 1992 to 2006, or trough-to-trough from 2002 to 2009.

The random component represents the unpredictable fluctuations of any time series. They are the irregular movements resulting from shocks

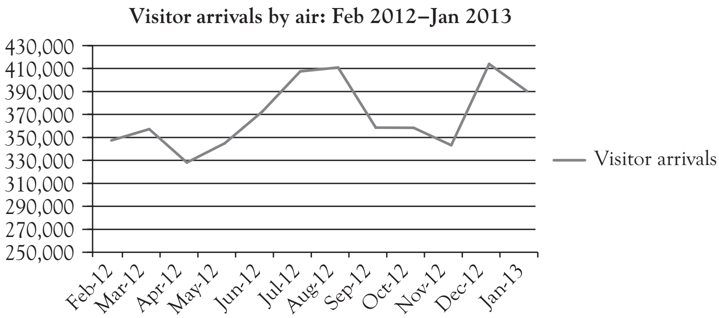


Figure 2.2 Visitor arrivals by air at Honolulu: seasonal component

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

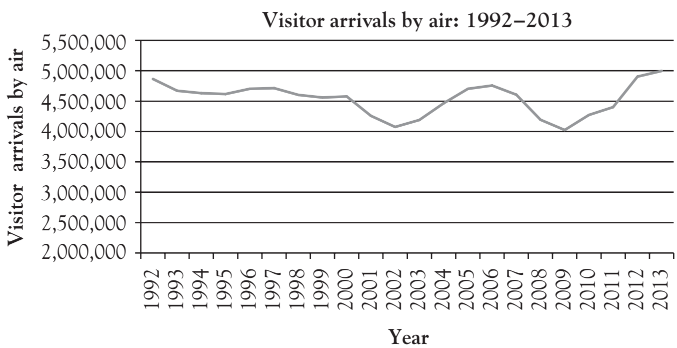


Figure 2.3 Visitor arrivals by air at Honolulu: cyclical component

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

that affect supplies or demands in the economy. Examples of these shocks are hurricanes, earthquakes, wildfires, wars, and so on. This component is the remaining movement once the other three components, trend, seasonal, and cyclical, are factored out. Not all random movements are identifiable. However, when identification is possible, we can incorporate the randomness into our forecasts.

Dr. Theo emphasizes, "In this chapter, we assume that the seasonal and cyclical components of a dataset are not clear, that is, the time series comprises only a deterministic (with clear pattern) component and a stochastic (random movement) component. Under this assumption, forecast values can be obtained using the MA and ES techniques to smooth out any randomness in the data."

At this point Dr. Theo reminds the class that the section on "Moving Averages" is very important because it is the basis for the later chapters. The class starts smoothly with the following section.

Moving Averages

Concept

We learn that the MA technique is the simplest in time series analyses. This technique can be used for short-term forecasts; for example, Mo has to forecast how many Honda motorcycles to order next quarter based on the historical data of the sales at his Motorland store. The technique assumes that the forecast value for the next period is an average of the previous and current-period values.

Simple MAs

The MA model assigns equal weights to past values and a current value with regard to their influences on the future values. Mathematically, a simple MA of order k can be written as:

$$F_{t+1} = \text{MA}(k)_t = \frac{\sum_{i=t-k+1}^{t} A_i}{k} \tag{2.1}$$

where
F_{t+1} = the forecast value for period $(t + 1)$
A_t = the actual value for period t
k = the order of the MAs
t = the time period.

At this point, Ex raises his hand and says that he does not understand why the formula has two parameters k and t in it. Dr. Theo explains that k and t are separate because k is fixed while t is moving in this formula: Each average value is formed by removing the oldest observation while adding the newest.

He then asks if any student can provide some examples. Cita says she can. Dr. Theo encourages her to go to the board, and she enthusiastically stands up. Cita offers the following discussion.

As an example, if you wish to experiment with an MA of order 3 ($k = 3$), and you have data for four periods, then the forecasting value for period 4 ($F_{t+1} = F_4$) is the MA of the first three periods, $\text{MA}(3)_3$, at time period $t = 3$:

$$F_4 = \text{MA}(3)_3 = \frac{A_1 + A_2 + A_3}{3} \tag{2.2}$$

Thus, the sum in Equation 2.1 starts with:

$$i = t - k + 1 = 3 - 3 + 1 = 1$$

and ends with $i = t = 3$.

Cita cheerfully fills the white board with her calculations. We are all awe-struck by her mathematical skills. Cita continues:

Next, the forecasting value for period 5 ($F_{t+1} = F_5$) is the MA of the second through the fourth periods, $\text{MA}(3)_4$, at time period $t = 4$:

$$F_5 = \text{MA}(3)_4 = \frac{A_2 + A_3 + A_4}{3} \tag{2.3}$$

Hence, the sum in Equation 2.1 starts with

$$i = t - k + 1 = 4 - 3 + 1 = 2$$

and ends with $i = t = 4$.

Ex understands now and even volunteers to go to the board to provide the class with an example from his import–export company: The actual sale values from his company in August, September, October, and November are $A_1 = 20$, $A_2 = 30$, $A_3 = 40$, and $A_4 = 35$, respectively, all in thousands of dollars. Thus, the forecasts for his company's sales in November and December will be:

$$F_{\text{November}} = F_4 = \text{MA}(3)_3 = (20 + 30 + 40)/3 = 30 \text{ (\$ thousands)} = \$30{,}000$$

$$F_{\text{December}} = F_5 = \text{MA}(3)_4 = (30 + 40 + 35)/3 = 35 \text{ (\$ thousands)} = \$35{,}000$$

The class is very impressed with Ex's demonstration. Dr. Theo then reminds us that a forecaster will have to experiment with several orders and perform evaluations to choose the best model with the lowest forecast errors, which will be discussed in Chapter 3. With that, we feel ready to move on, and Dr. Theo leads us to the next section.

Weighted Moving Averages

The WM technique is a little more advanced than the MA technique. A forecaster often assigns a higher weight to the more current values. The weights in the WM technique are often integers, which are assigned arbitrarily, that is, there is no limitation on the weights. Thus, Equations 2.2 and 2.3 will be transformed to the following equations:

$$F_4 = \text{WM}(3)_3 = \frac{w_1 A_1 + w_2 A_2 + w_3 A_3}{w_1 + w_2 + w_3};\ w_3 > w_2 > w_1$$

$$F_5 = \text{WM}(3)_4 = \frac{w_1 A_2 + w_2 A_3 + w_3 A_4}{w_1 + w_2 + w_3};\ w_3 > w_2 > w_1 \quad (2.4)$$

Sol remarks that in the previous example given by Ex, if we assign the weights of 1, 2, and 3 to August, September, and October, respectively, then the forecasts for his company's sales in November and December will be:

$$F_4 = \text{WM}(3)_3 = (20 * 1 + 30 * 2 + 40 * 3)/6 = (20 + 60 + 120)/6$$
$$= 33.33 \text{ (\$ thousands)}$$

$$F_5 = \text{WM}(3)_4 = (30 * 1 + 40 * 2 + 35 * 3)/6 = (30 + 80 + 105)/6$$
$$= 35.83 \text{ (\$ thousands)}$$

Dr. Theo is very pleased, saying that her calculations are correct. He also points out that the simple MA and WM techniques in this chapter only provide one-period forecasts, and that Chapter 4 will discuss forecasting techniques that provide forecasts for two or more periods ahead. We are very excited to hear this.

Excel Applications

Dr. App starts our Excel application section with the simplest model.

Simple MA

We see that Mo is very generous to share his data with us. He also reminds us that the data and commands are in the file Ch02.xls, Fig.2.4. Dr. App points out that Figure 2.4 displays the quarterly data on demand for Honda motorcycles from the Motorland for the period September 2011 through June 2014, the MA(3) and MA(4) values, and the forecast values using the simple MA technique. We learn that we should follow these steps to obtain the MA(3) forecasts:

In cell E5, type = (D3 + D4 + D5)/3 and press Enter
Copy and paste this formula into cells E6 through E14
To obtain forecasting values $F_{t+1}(3)$, copy cells E5 through E14
Right click cell F6 and select Paste Special
When the dialog box appears, select Values, then click OK
Excel will paste the values (instead of the formula) from column E into column F
The forecast value for the third quarter of 2014 using MA(3) is shown in cell F15

	A	B	C	D	E	F	G	H	I
1		Quarter	Period	Actual Demand	k = 3	Forecast	k=4	Forecast	
2		Q	t	A_t	$MA(3)_t$	$F_{t+1}(3)$	$MA(4)_t$	Ft+1(4)	
3		9/30/2011	1	7812					
4		12/30/2011	2	8654					
5		3/30/2012	3	9407	8624				
6		6/30/2012	4	9625	9228	8624	8874		
7		9/30/2012	5	11114	10048	9228	9700	8874	
8		12/30/2012	6	12080	10939	10048	10556	9700	
9		3/30/2013	7	13131	12108	10939	11487	10556	
10		6/30/2013	8	13275	12829	12108	12400	11487	
11		9/30/2013	9	14588	13665	12829	13268	12400	
12		12/30/2013	10	15852	14572	13665	14211	13268	
13		3/30/2014	11	15782	15407	14572	14874	14211	
14		6/30/2014	12	16044	15893	15407	15567	14874	
15		9/30/2014				15893		15567	

Sheet1 Sheet2 Sheet3 Sheet4 Sheet5

Figure 2.4 Simple MA: obtaining forecast values

Repeat the same procedure for the MA(4) forecasts:

In cell G6, type = (D3 + D4 + D5 + D6)/4 and press Enter
Copy and paste this formula into cells G7 through G14
To obtain forecasting values $F_{t+1}(4)$, copy cells G6 through G14
Right click cell H7 and select Paste Special
When the dialog box appears, select Values, then click OK
The forecast value for the third quarter of 2014 using MA(4) is shown in cell H15

Weighted Moving Averages

For comparative purpose, we use the same quarterly data on demand as in Figure 2.4. We find that Figure 2.5 displays the data, WM(3) values, WM(4) values, and the forecast values. The data are available in the file Ch02.xls, Fig.2.5. To proceed with WM(3), we do the following steps:

In cell E5, type = (1 * D3 + 2 * D4 + 3 * D5)/6 and press Enter
Copy and paste this formula into cells E6 through E14
To obtain forecasting values $F_{t+1}(3)$, copy cells E5 through E14
Right click cell F6 and select Paste Special
When the dialog box appears, select Values and click OK

	A	B	C	D	E	F	G	H
1		Quarter	Period	Actual Demand	k = 3	Forecast	k = 4	Forecast
2		Q	t	A_t	$WM(3)_t$	$F_{t+1}(3)$	$WM(4)_t$	$F_{t+1}(4)$
3		9/30/2011	1	7812				
4		12/30/2011	2	8654				
5		3/30/2012	3	9407	8890			
6		6/30/2012	4	9625	9390	8890	9184	
7		9/30/2012	5	11114	10333	9390	10080	9184
8		12/30/2012	6	12080	11349	10333	11032	10080
9		3/30/2013	7	13131	12444	11349	12061	11032
10		6/30/2013	8	13275	13028	12444	12777	12061
11		9/30/2013	9	14588	13908	13028	13652	12777
12		12/30/2013	10	15852	15001	13908	14685	13652
13		3/30/2014	11	15782	15606	15001	15314	14685
14		6/30/2014	12	16044	15925	15606	15781	15314
15		9/30/2014				15925		15781

Fig.2.1 | Fig2.2 | Fig.2.3 | Fig.2.4 | Fig.2.5 | Fig.2

Figure 2.5 WM: obtaining forecast values

Excel will paste the values from column E into column F
The forecast value for the third quarter of 2014 using WM(3) is shown in cell F15

For WM(4), repeat the same procedure:

In cell G6, type = (1 * D3 + 2 * D4 + 3 * D5 + 4 * D6)/10 and press Enter
Copy and paste this formula into cells G7 through G14
To obtain forecasting values $F_{t+1}(4)$, copy cells G6 through G14
Right click cell H7 and select Paste Special
When the dialog box appears, select Values and click OK
The forecast value for the third quarter of 2014 using WM(4) is shown in cell H15

Dr. App concludes this section by reminding us that the Data Analysis tools in Excel has commands for MA and ES. Additionally, we can use the Math Function in Excel to calculate the average. However, she points out that one saves more time doing the techniques manually rather than using the Math Function, and a manual calculation helps us understand the concepts more clearly. We are happy to follow her advice.

Exponential Smoothing

Dr. Theo takes over the class with the next theoretical section on ES.

Concept

We learn that similar to the WM technique, the ES technique assumes that the forecast value for the next period is a weighted average of the previous and current-period values. Different from the WM, only one weight is used in ES models, and the weight is confined between zero and one. This weight is also changeable and depends on the changes in market conditions. Mathematically, the equation of forecasts using the ES model can be written as:

$$F_{t+1} = aA_t + (1-a)F_t \tag{2.5}$$

where F_{t+1} and A_t are the same as in Equation 2.1
F_t is the forecast value for period t
a is the smoothing factor, also called smoothing constant ($0 < a < 1$).

This equation states that the forecast value of period ($t+1$) is equal to the weighted actual value of period t plus the weighted forecast value of period t. There are two common ways to obtain the first forecast value: calculate an average of the first several values in the actual data or take the first actual value. Montgomery, Jennings, and Kulahci (2008) also recommend a third approach of calculating an average of all values in the available data. However, this might result in a very large first value if the series moves upward sharply over time. For this reason, most textbooks take either the first value or the average of the first several values in the actual data.

Our textbook follows Hyndman and Athanasopoulos (2014), Krueger (2010), and Lawrence, Klimberg, and Lawrence (2009) in taking the first value in the actual data, that is, $F_1 = A_1$, so

$$\begin{aligned}
F_1 &= A_1 \\
F_2 &= aA_1 + (1-a)F_1 \\
&= aA_1 + F_1 - aF_1 \\
&= aA_1 + A_1 - aA_1 \\
&= A_1 \\
F_3 &= aA_2 + (1-a)F_2 \\
&= aA_2 + (1-a)A_1
\end{aligned} \tag{2.6}$$

Because of this gradual approach, the series does not settle until reaching period 4 and up. Also, since there is only one weight varying between zero and one, the weights assigned to past periods decrease so that the current period is given more weight. We notice the pattern of changes as follows:

$$F_{t+1} = aA_t + (1-a)F_t$$

$$F_{t+2} = aA_{t+1} + (1-a)F_{t+1} = aA_{t+1} + (1-a)[aA_t + (1-a)F_t]$$

$$= aA_{t+1} + a(1-a)A_t + (1-a)(1-a)F_t$$

Since $0 < a < 1$, the weight decreases over time.

To this point, Fligh raises his hand and asks for an example. Fin volunteers to go to the board. The following is his discussion.

For example, if $a = 0.2$, then the weight assigned to the actual value in each period is as in the following table and continues with higher exponential orders:

Time period	Cumulative value	Weight
Current period	0.2	0.2
One period apart	0.2 * (1 – 0.2)	0.16
Two periods apart	0.2 * (1 – 0.2) * (1 – 0.2) = $0.2 * (1 - 0.2)^2$	0.128

"Smart guy!" We exclaim. Now we know why the technique bears the name exponential smoothing.

Dr. Theo also emphasizes the advantage of using ES over WM: there is only one weight, which can be adjusted easily from zero to one, with a larger smoothing factor resulting in more weight given to the current period. For example, if $a = 0.9$, then the weight for the current period is 0.9, and the following weight drops sharply to 0.09.

Alte asks, "Does that means that if $a = 0.1$, then the weights are almost equal?" Dr. Theo commends her on the remark because in this case, the weight for the current period is 0.1, and the following weight is still 0.09, which is very close to 0.1.

Dr. Theo also reminds us that the sum of all weights over time equals to one.

At this point, Arti mentions that yesterday she read a book, in which Lawrence, Klimberg, and Lawrence (2009) offer a different way to look at the ES model by manipulating the original equation further:

$$F_{t+1} = aA_t + (1-a)F_t = aA_t + F_t - aF_t \rightarrow F_{t+1} = F_t + a(A_t - F_t) \quad (2.7)$$

The last expression $(A_t - F_t)$ is the forecast error, which measures the difference between the actual value and the forecast value. In view of Equation 2.7, the ES technique provides the next-period forecast, which equals the sum of the current-period forecast and the weighted adjustment of the forecast error in the current-period forecast.

Dr. Theo is very pleased, and we find that this alternative way to look at the ES model is interesting because it shows that the ES process adjusts itself based on the forecast errors.

Rea raises his hand and suggests that we calculate the sale values from Ex's company in the section on "Moving Averages," again using Equation 2.5 and a calculator with the actual sales for the first three months $A_1 = 20$, $A_2 = 30$, and $A_3 = 40$.

Sol asks, "So what smoothing factor should we use?" Ex says, "Let's use the smoothing factor $a = 0.5$, because our company doesn't emphasize too much on the current period but does not ignore it completely either." Mo then adds that we should let $F_1 = A_1$ because the sales at Ex's company increase quite rapidly over time; if we calculate the average of the three values, the initial forecast might be too large. We all think this is a good idea and work on the calculations shown in the following text.

August: $F_1 = A_1 = 20$
September: $F_2 = 0.5 * 20 + (1 - 0.5) * 20 = 20$
October: $F_3 = 0.5 * 30 + (1 - 0.5) * 20 = 15 + 10 = 25$
November: $F_4 = 0.5 * 40 + (1 - 0.5) * 25 = 20 + 12.5 = 32.5$

Dr. Theo reminds us that we can verify our calculations using Equation 2.7 when we get home and to read the Excel application in the following section.

Excel Application

Dr. App starts this section by showing us Figure 2.6, which displays the data and the forecast values using the same dataset in Figures 2.4 and 2.5, setting $F_1 = A_1$, with three different smoothing factors, $a = 0.3$, $a = 0.5$, and $a = 0.7$. The data are in the file Ch02.xls, Fig.2.6. We learn that we should perform the following steps to obtain the ES forecasts:

In cell E3, type = 0.3 * D2 + (1 – 0.3) * E2 and press Enter
In cell F3, type = 0.5 * D2 + (1 – 0.5) * F2 and press Enter
In cell G3, type = 0.7 * D2 + (1 – 0.7) * G2 and press Enter
Copy and paste the formulas in cells E3, F3, and G3 into cells E4 through E14, F4 through F14, and G4 through G14, respectively

Dr. App reminds us that we can experiment with Equation 2.7 using our knowledge of Excel mathematical operations learned in

	A	B	C	D	E	F	G
1		Quarter (Q)	Period (t)	Actual Demand (A_t)	$F(0.3)_{t+1}$	$F(0.5)_{t+1}$	$F(0.7)_{t+1}$
2		9/30/2011	1	7812	7812	7812	7812
3		12/30/2011	2	8654	7812	7812	7812
4		3/30/2012	3	9407	8064	8233	8401
5		6/30/2012	4	9625	8467	8820	9105
6		9/30/2012	5	11114	8814	9222	9469
7		12/30/2012	6	12080	9504	10168	10620
8		3/30/2013	7	13131	10277	11124	11642
9		6/30/2013	8	13275	11133	12127	12684
10		9/30/2013	9	14588	11776	12701	13098
11		12/30/2013	10	15852	12619	13645	14141
12		3/30/2014	11	15782	13589	14748	15339
13		6/30/2014	12	16044	14247	15265	15649
14		9/30/2014			14786	15655	15926

Fig.2.1 Fig2.2 Fig.2.3 Fig.2.4 Fig.2.5 **Fig.2.6**

Figure 2.6 ES: obtaining forecast values

Chapter 1 and type the formula into corresponding cells in the dataset for Figure 2.6, and that we should be able to obtain the same results using either equation.

We then move to the next topic of transforming the data to obtain real values, which are adjusted against inflation, so that we can produce more accurate forecasts.

Nominal versus Real Values

Concept

We learn that in real life situations, some data are in quantity values that are not distorted by inflation or deflation, for example, data on building permits, quantity of books demanded, and so on. However, some data are expressed in currency values, for example, data on sales, gross domestic product (GDP), or house prices are in domestic currencies. If the values in the data are reported using different price levels for different periods (called current prices), then these values are called nominal values. The nominal values have to be converted to the *real values*, which are values calculated using the price level in a base period (called the constant price).

To this point, Rea asks, “Why is using real values important?” Arti offers her own experience, “Two years ago, my school made a profit of $250,000. Last year, we made $256,000. The board members and I were happy that we were growing. My accountant then pointed out to us that we were worse off because our growth rate was:

$$G = (256{,}000 - 250{,}000)/250{,}000 = 0.024 = 2.4\%$$

However, the inflation rate during that period was 3.4 percent, so our real growth rate was 1 percent. It was a wake-up call to us. Since then, my school has been very determined to use real value in our data so that we can compensate for the inflation rate."

"Wow, that is a good example!" We say in unison and now understand why we have to master this section.

The data conversion helps a forecaster avoid inflation and deflation distortions before applying any forecast technique. The nominal values are converted to the real values using price indexes. The two frequently used price indexes are the consumer price index (CPI) and the GDP deflator (GDPD).

Consumer Price Index

The CPI is the cost of a fixed basket of goods and services purchased by a consumer representative in the urban areas in a current year relative to the cost of the same basket in a base year. We are surprised to learn that the measurement was introduced over a century ago by the German economist and statistician, Etienne Laspeyres (1834–1913). Dr. Theo points out that the CPI measures the cost of living and so it is often used when consumer goods are examined. The equation for calculating the CPI for a current year using a base year is:

$$\mathrm{CPI}_C = \frac{\sum_{i=1}^{n} Q_{i,b} P_{i,c}}{\sum_{i=1}^{n} Q_{i,b} P_{i,b}} * 100 = \frac{\text{Nominal Value}}{\text{Real Value}} * 100 \qquad (2.8)$$

where

CPI_C = the consumer price index for the current year

$Q_{i,b}$ = the fixed quantity of good i in the base year

$P_{i,c}$ = the price of good i in the current year
$P_{i,b}$ = the price of good i in the base year.

We then break into small groups to work on an example. Alte, Fligh, and I are in the same group. Here is the exercise.

Suppose the representative consumer buys a nominal value of $11,000 in consumer goods in 2014 and the real value of this basket in 2010 was $10,000, then the CPI for this basket in 2014 using 2010 as the base is:

$$\text{CPI}_{2014} = (\$11{,}000/\$10{,}000) * 100 = 110$$

Rea raises a question, "Does the result reveal that the inflation for the period 2010–2014 is 10 percent, implying an inflation rate of 2.5 percent per year on average?" Dr. Theo commends him on the correct interpretation and leads us to the next calculation: When the nominal value of $11,000 and the CPI of 110 are provided, the real value is:

$$\text{Real value} = (\$11{,}000/110) * 100 = \$10{,}000$$

We now know that we should use this real value to perform forecasting instead of the nominal value.

GDP Deflator

The GDPD was introduced by the German economist and statistician Hermann Paasche (1851–1925) and has been often used in macroeconomics. The index is also called the implicit price level, or the implicit GDPD. The equation for calculating the GDPD for a current year using a base year is:

$$\text{GDPD}_C = \frac{\sum_{i=1}^{n} Q_{i,c} P_{i,c}}{\sum_{i=1}^{n} Q_{i,c} P_{i,b}} * 100 = \frac{\text{NGDP}_C}{\text{RGDP}_C} * 100 \tag{2.9}$$

where

GDPD_C = the GDPD for the current year
$Q_{i,c}$ = the quantity of good i in the current year

$P_{i,c}$ = the price of good i in the current year

$P_{i,b}$ = the price of good i in the base year

$NGDP_C$ = the nominal GDP for the current year

$RGDP_C$ = the real GDP for the current year.

For example, when the NGDP = \$10.8 million and the $GDPD_C$ = 108, then

$$RGDP = (\$10.8\ million/108) * 100 = \$10\ million.$$

We find it is interesting to know that the GDPD has an advantage of not using a fixed basket of goods and services the way the CPI does. Hence, the disappearance of a good or appearance of a new one in a current year will be reflected in GDPD. However, the GDPD has a disadvantage of not including imported goods and imported services as the CPI does. For this reason, the GDPD usually understates the price level whereas the CPI overstates it.

At this moment, Dr. Theo asks us to give an example, and Fligh raises his hand to provide one: A hurricane destroys all oranges in Florida, so the overall price level of a few imported oranges increases sharply. Since the quantity of oranges in the CPI basket does not change, the rise in the orange price causes the CPI in Florida to go up a great deal and overstates the inflation in the state. In the meantime, the quantity of oranges is dropped from the GDPD calculation for Florida, so the total value of oranges in the GDPD is zero, and the GDPD understates the inflation in Florida.

We are truly impressed with Fligh's intelligence. Dr. Theo is very pleased with the class and lets Dr. App work with us on the next section.

Excel Application

Ex has studied the market for exports in order to help his company and is able to provide us with the data on the exports from China to Canada. The data are in the file Ch02.xls, Fig.2.7. We see that Figure 2.7 displays data from 2000 to 2011 in current U.S. dollars and the GDPD with 2004 as the base.

	A	B	C	D	E	F
1	Exprots from	To	Year	Exports	Deflator	Real Exports
2	China, P.R.: Mainland	Canada	2000	3158045000	93.859	3.36E+09
3	China, P.R.: Mainland	Canada	2001	3349754000	95.415	3.51E+09
4	China, P.R.: Mainland	Canada	2002	4304719000	96.475	4.46E+09
5	China, P.R.: Mainland	Canada	2003	5633218000	97.868	5.76E+09
6	China, P.R.: Mainland	Canada	2004	8164933000	100	8.16E+09
7	China, P.R.: Mainland	Canada	2005	11657641000	102.402	1.14E+10
8	China, P.R.: Mainland	Canada	2006	15519802000	104.193	1.49E+10
9	China, P.R.: Mainland	Canada	2007	19362909000	106.409	1.82E+10
10	China, P.R.: Mainland	Canada	2008	21789889000	109.429	1.99E+10
11	China, P.R.: Mainland	Canada	2009	17673260000	112.744	1.57E+10
12	China, P.R.: Mainland	Canada	2010	22208071511	114.232	1.94E+10
13	China, P.R.: Mainland	Canada	2011	25249130127	116.603	2.17E+10

Sheet1 Sheet2 Sheet3

Figure 2.7 Converting nominal values to real values

Data Source: IMF.com: Direction of Trade Statistics (2014).

Since this is a macroeconomic dataset, Dr. App says that using GDPD is appropriate. We learn that to convert the nominal values to the real ones, we have to perform the following steps:

In cell F2, type = (D2/E2) * 100 and press Enter
Copy and paste the formula into cells F3 through F13
The real values are displayed in column F

Dr. App explains that a similar process can be followed to obtain real values using CPI. She also reminds us to read Chapter 3 of the textbook so that we can learn how to evaluate and adjust our forecasts in the following class.

Exercises

1. Data on the actual number of permits for residential buildings in the city are provided by Rea from Realmart and can be found in the file Real.xls. The data are from December 2012 to February 2014. Perform the WM(4) procedure on this dataset to obtain one-period forecasts with the weights of 1, 2, 3, and 4 for periods 1, 2, 3, and 4, respectively. Construct columns in Excel similar to the ones in Figure 2.5.

2. Perform the ES procedures with $a = 0.4$ and $a = 0.6$ on the dataset Real.xls to obtain one-period forecasts and construct columns in Excel similar to the ones in Figure 2.6.
3. Use the dataset Sales.xls, a handheld calculator, and the ES technique with $a = 0.8$ to calculate forecasts for weeks 2 through 5 of this series. Show how the steps of your calculations are similar to those in the section "Concept" under "Exponential Smoothing."
4. The file Revenue.xls contains data on monthly revenue in current dollars for Artistown and the monthly CPI index for the city. Convert the nominal revenue values into real values.

CHAPTER 3

Evaluations and Adjustments

At the end of our last class, we obtained the one-period forecasts for Motorland store. Alte raises a question, "In Chapter 2, a long series of data is analyzed only to obtain one forecast value for the 13th period. All other 'forecasts' are current values that are known to us. Why do we need a long time series then?"

Dr. Theo commends Alte on asking a good question and explains, "We need a long series because we have to compare the forecast values with the actual values in order to evaluate the accuracy of our forecast models." Mo exclaims, "Oh yes, my boss did not want to under-order or over-order by more than 20 percent, so we had better check on our model." Dr. Theo says, "Right, so the error interval in this case is (–20%; +20%). Hence, we want to know if this interval condition is met. If not, adjustments can be made to the model to improve its predictive power. Monitoring then can be carried out to obtain the best forecast model. These are the topics of this chapter, so once we finish with the chapter, you will be able to:

1. Explain the concept of error measurements.
2. Apply the concept into calculating each error measurement using Excel.
3. Adjust our forecast models based on these error measurements.
4. Explain other methods of model evaluations.
5. Apply these other methods into model evaluations and adjustments."

We are excited to hear that we can evaluate our models. However, when we look through the chapter, we see that the section on "Error Evaluations" is quite abstract and we freak out. Dr. Theo then promises to go slow and to provide the class with numerous examples. With that, we start our next theoretical section.

Error Evaluations

We learn that error measurements are crucial in evaluating and adjusting a forecast model. As the future is not certain, there are always forecast errors.

Concept

The forecast error is the deviation of the forecast value from its actual counterpart:

$$f_t = A_t - F_t \tag{3.1}$$

where
f_t = the forecast error for period t
A_t = the actual value in period t
F_t = the forecast value for period t.

The simplest way to see the difference is to plot the two series. We examine the plots in Figure 3.1, which displays the actual demand and forecast values from Figures 2.4 and 2.5 using the moving average of order 3 (MA(3)) and the weighted moving average of order 3 (WM (3)). We are able to find the data in the file Ch03.xls, Fig.3.1, and gain a hands-on experience by charting the plots ourselves.

Observing this figure, we see that forecast values obtained by the WM technique (the solid line) appear to mimic the actual data more closely than the simple MA (the dashed line). Overall, both techniques underestimated the sale values. Dr. Theo says that this is a common characteristic

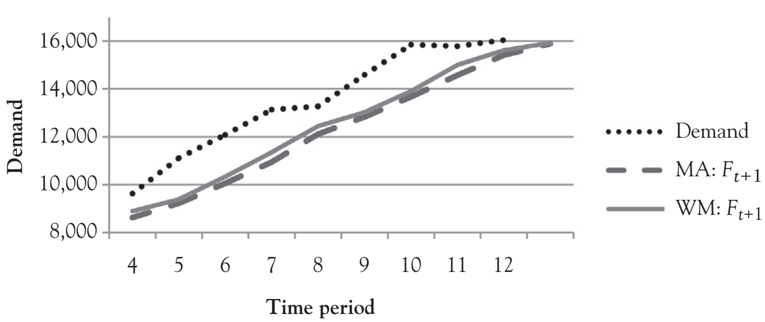


Figure 3.1 Comparing MA(3) and WM(3) forecasts by charting

of elementary forecast techniques, including the exponential smoothing (ES), partly because they ignore the trend in their data analyses. Chapter 4 will add this trend and introduce more advanced techniques.

Dr. Theo then asks us whether we can say for sure that the WM is better than the MA. We all agree that we cannot be sure because no quantitative value is provided, so more sophisticated measures have to be learned. Dr. Theo then guides us through the quantitative measurements in the following section.

Standard Deviation of the Forecast

This measure is similar to that in statistics except that the standard deviation of the forecast (SDF) measures the deviation from the mean of the forecast errors instead of the deviation from the mean of a population. Some textbooks use the notation σ (sigma) for this measure:

$$\text{SDF} = \sqrt{E\{|f| - E|f|\}^2} \tag{3.2}$$

Since this is a complicated formula, we study Table 3.1, which provides an example of the SDF for a dataset with A as the actual value and F as the forecast value. We are divided into groups, and we discuss how to calculate the SDF following step by step instruction in Table 3.1.

Sol and Rea and I are in the same group. Rea points out that the results in column (5) come from the calculation of $E|f| = (1 + 2 + 3)/3 = 2$. I am very grateful for his remark because I have not noticed so.

We are continue to look through Table 3.1 when Fin from the adjacent group calls out, “I think that taking the average of the results in column (6) gives us the variance of the forecast error.” I look over Fin’s shoulders to see that he is writing: $E\{|f| - E|f|\}^2 = (1 + 0 + 1)/3 = 2/3$.

Dr. Theo praises Fin for his correct observation, so all we have to do is to take the square root of the variance to obtain SDF:

$$\text{SDF} = \sqrt{2/3} = 0.8165$$

Dr. Theo guides us through every step of the calculation, and we feel a little better now.

Table 3.1 Calculating the expression $\{|f| - E|f|\}^2$

(1)	(2)	(3)	(4)	(5)	(6)
Period	A	F	$\lvert f\rvert = \lvert A - F\rvert$	$E\lvert f\rvert$	$\{\lvert f\rvert - E\lvert f\rvert\}^2$
4	9	8	$\lvert 9 - 8\rvert = 1$	2	$\{1 - 2\}^2 = 1$
5	10	8	$\lvert 10 - 8\rvert = 2$	2	$\{2 - 2\}^2 = 0$
6	8	11	$\lvert 8 - 11\rvert = 3$	2	$\{3 - 2\}^2 = 1$

Mean Squared Error

The mean squared error (MSE) is the average of the squared errors of the forecasts. This concept is easier to learn. Dr. Theo reminds us that since we use the squared errors, MSE gives greater weight to the larger errors and can be a good measure if the objective is to minimize the larger errors:

$$\text{MSE} = E(f)^2 = \frac{\sum_{t=1}^{T}(A_t - F_t)^2}{T} \quad (3.3)$$

We then work on an MSE example, using the information on $(A - F)$ in column (4) of Table 3.1. Because all *f*s will be squared, the sign does not matter:

$$\text{MSE} = (1^2 + 2^2 + 3^2)/3 = (1 + 4 + 9)/3 = 4.67$$

Alte points out that she often sees the square root of MSE, which is calculated as:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{4.67} = 2.16$$

Dr. Theo commends her on the remark and says that this measure is called root mean squared error.

Mean Absolute Error

The mean absolute error (MAE) is also referred to as mean absolute deviation (MAD) in several textbooks. In our class, we use the terminology MAE stipulated by Greene (2012). In calculating MAE, the absolute value of each forecast error is used to find the average of the deviation from the actual values. This overcomes the problem of the negative and the positive errors cancelling each other out. It can be a good measure if the objective is to minimize all errors equally:

$$\text{MAE} = E|f| = \frac{\sum_{t=1}^{T} |A_t - F_t|}{T} \quad (3.4)$$

For example, from column (4) of Table 3.1, MAE $= E|f| = (1+2+3)/3 = 2$.

Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) measures the MAE in percentage. MAPE provides a more precise evaluation of error relative to the actual value. To see the difference, Dr. Theo asks the class, "Does anyone know why the percentage difference is important?"

I raise my hand and volunteer to give the following story: "Two months ago, the total profit from my Nailight Salon was $11,000 whereas my competitor across the street made $10,000 in total profit. Thus, we made $1,000 more than they did. Last month we decided to launch an aggressive campaign by giving out nail files and flyers to our customers and increased our profit to $15,100 while my competitor made $14,000 in profit. I later heard that they also launched massive promotions by giving out nail polish removers and posting Internet advertisements. Our profit was still $1,100 more than theirs last month. My partner was happy thinking that we improved because we were $1,100 instead of $1,000 ahead of them. I pointed out to her that they might surpass us one day if we progress at this rate. The reason is:

($1,000 / $10,000) * 100% = 10% (we were 10 percent ahead of them two months ago)

($1,100 / $14,000) * 100% = 7.9% (we were only 7.9 percent ahead of them last month)."

Dr. Theo commends me for the example. He then says that the same is true for the error measurement: An error of 20 units relative to the actual value of 100 units (a 20 percent error) is 10 times as large as an error of 20 units relative to an actual value of 1,000 units (a 2 percent error). For this reason, the MAPE measures the accuracy of a forecast model better than the preceding measurements.

$$\text{MAPE} = \frac{\sum_{t=1}^{T} | f | / A_t}{T} * 100 = \frac{\sum_{t=1}^{T} | A_t - F_t | / A_t}{T} * 100 \qquad (3.5)$$

We then work in small groups on an example using the information in column (4) of Table 3.1 again to calculate:

$$\text{MAPE} = \frac{\sum_{t=1}^{3} | A_t - F_t | / A_t}{3} * 100 = \frac{1/9 + 2/10 + 3/8}{3} * 100 = 22.87\ (\%)$$

Deviation in Percentage

The deviation in percentage (DPE) measures the percentage deviation from the mean of the errors:

$$\text{DPE} = \sqrt{\frac{\sum_{t=1}^{T} [(| f | / A_t) * 100 - \text{MAPE}]^2}{T}} \qquad (3.6)$$

Cita is quick to point out that Table 3.2 displays the calculation of DPE for the example in Table 3.1. Ex then notices that column (4) in Table 3.1 becomes column (2) in Table 3.2. Dr. Theo confirms their observations and asks us to break into small groups again and discuss the calculations with each other.

Table 3.2 Calculating the expression [{ |*f*|/A} **** 100 – MAPE***]²

(1)	(2)	(3)	(4)
Period	|*f*| = |A – *F*|	(|*f*|/A) * 100	[(|*f*|/A) * 100 – MAPE]²
4	|9 – 8| = 1	(1/9) * 100 = 11.11	{11.11 – 22.87}² = 138.30
5	|10 – 8| = 2	(2/10) * 100 = 20	{20 – 22.87}² = 8.24
6	|8 – 11| = 3	(3/8) * 100 = 37.5	{37.5 – 22.87}² = 214.04

From the results in Table 3.2, we are able to calculate the DPE:

$$\text{DPE} = \sqrt{\frac{138.30 + 8.24 + 214.04}{3}} = \sqrt{120.19} = 10.96\ (\%)$$

Forecast Bias

The forecast bias (FB) measures the direction of the errors to see whether the forecast model understates (downward bias) or overstates (upward bias) the actual values. Hence, the sign of the error becomes important in this measure:

$$\text{FB} = E(f) = \frac{\sum_{t=1}^{T}(A_t - F_t)}{T} \tag{3.7}$$

Dr. Theo reminds us to pay attention to the definition of the forecast error, which is $f = A - F$. Hence, if forecast values on average are lower than its actual values, then the value of FB is positive, and we have a downward bias, where the model underestimates the actual values. Therefore, we are advised to avoid using the term *positive bias* when FB > 0 because this term creates an impression of an overestimation, which is incorrect in this case. Similarly, we should also avoid the term *negative bias* when FB < 0.

To observe this point, we work on an example from columns (2) and (3) in Table 3.1, where $(A - F)$ values are: 9 − 8 = 1, 10 − 8 = 2, and 8 − 11 = −3. Hence, FB = (1 + 2 − 3)/3 = 0, that is, there is no bias in this forecast.

Next, we experiment with changing the last forecast error in Table 3.1 from −3 to −2 and find that the FB = (1 + 2 − 2)/3 = 1/3, which is a positive number and implies a downward bias (an underestimation of the actual values).

Finally, we also experiment with changing the last forecast value in Table 3.1 from −3 to −4 and find that the FB = (1 + 2 − 4)/3 = −1/3, which implies an upward bias (an overestimation of the actual values).

The theoretical section is then adjourned, and we work with Dr. App on the following section.

Excel Applications

Obtaining SDF

Dr. App points out that Figure 3.2 displays the actual demand and the MA(3) forecasts from Figure 2.4 for periods 4 through 12, as well as steps of calculations to obtain the SDF. The forecast value for period 13 is not displayed because actual demand for this period is not available. She also reminds us that the data are available in the file Ch03.xls, Fig.3.4. We learn that we need to perform the following steps:

In cell D2, type = ABS(B2 − C2) and press Enter
Copy and paste the formula into cells D3 through D10
In cell D11, type = average (D2:D10) and press Enter
Copy the value in cell D11
Right click cell E2 and select Paste Special
When the dialog box appears, select Values then click OK
Excel will paste the value (instead of the formula) from cell D11 into cell E2 (Henceforth, this command will read "paste-special the value(s) from cell … into cell ….")
Paste the value from cell E2 into cells E3 through E10

In cell F2, type = (D2 − E2)^2 and press Enter
Copy and paste the formula into cells F3 through F10
In cell F11, type = average (F2:F10) and press Enter
In cell E11, type = F11^(1/2) and press Enter

	A	B	C	D	E	F	G	H	I	J	K	L
1	Period	Demand	F_{t+1}	|f|	E |f|	$\{\|f\|-E\|f\|\}^2$		Demand	WM: F_{t+1}	|f|	E |f|	$\{\|f\|-E\|f\|\}^2$
2	4	9625	8624	1000	1563	316816		9625	8890	735	1282.39	300058
3	5	11114	9228	1885	1563	103655		11114	9390	1723	1282.39	194573
4	6	12080	10048	2032	1563	219491		12080	10333	1747	1282.39	216120
5	7	13131	10939	2191	1563	394113		13131	11349	1782	1282.39	249433
6	8	13275	12108	1167	1563	157023		13275	12444	831	1282.39	203856
7	9	14588	12829	1760	1563	38517		14588	13028	1560	1282.39	77274
8	10	15852	13665	2188	1563	389680		15852	13908	1945	1282.39	438520
9	11	15782	14572	1210	1563	124508		15782	15001	781	1282.39	251461
10	12	16044	15407	637	1563	858877		16044	15606	438	1282.39	713814
11				1563	538	289187				1282	542	293901

Fig.3.1 | Fig.3.4 | MSE-MAE-FB | Fig.3.5 | Fig.3.6 | Fig.3.8 | Fig.3.9 | Sh

Figure 3.2 Obtaining SDF for the MA(3) and WM(3) forecasts in Figures 2.4 and 2.5

The value in cell E11 is the SDF for the MA(3). Dr. App asks us to repeat the same process for the WM(3) forecasts in Figure 2.5 once we get home so that we can compare the two standard deviations. She also points out that the SDF for the WM(3) is displayed in cell K11.

Obtaining MSE, MAE, and FB

Since these are simple mathematical operations, we are provided with Excel commands without accompanied figures in the text. (For further reference, the commands are in the file Ch03.xls, on the sheet MSE-MAE-FB.) We follow her guidance to obtain the results for the MA(3) by performing the following steps.

RMSE: In cell D2, type = B2 − C2 and press Enter
In cell E2, type = D2^2 and press Enter
Copy and paste the formulas into cells D3 through D10 and E3 through E10
In cell E11, type = average (E2:E10) and press Enter to obtain the MSE
In cell E12, type =SQRT(E11) and press Enter to obtain the RMSE

MAE: In cell F2, type = ABS(B2 − C2) and press Enter
Copy and paste the formula into cells F3 through F10
Copy and paste the formula in cell E11 into cell F11 to obtain the MAE

FB: Copy and paste the formula in cell E11 into cell D11 to obtain the FB

Ex discovers that we can repeat the same procedure to obtain RMSE, MAE, and FB for the WM(3) model. We are very impressed with his discovery.

Obtaining MAPE

Sol notices that Figure 3.3 displays the actual demand, the MA(3) forecasts, and the WM(3) forecasts. The data are available in the file Ch03.xls, Fig.3.5.

	A	B	C	D	E	F	G	H	I	J
1	Period	Demand	MA: F_{t+1}	|f|	(|f|/A)*100		Demand	WM: F_{t+1}	|f|	(|f|/A)*100
2	4	9625	8624	1000	10.39		9625	8890	735	7.63
3	5	11114	9228	1885	16.96		11114	9390	1723	15.51
4	6	12080	10048	2032	16.82		12080	10333	1747	14.46
5	7	13131	10939	2191	16.69		13131	11349	1782	13.57
6	8	13275	12108	1167	8.79		13275	12444	831	6.26
7	9	14588	12829	1760	12.06		14588	13028	1560	10.70
8	10	15852	13665	2188	13.80		15852	13908	1945	12.27
9	11	15782	14572	1210	7.67		15782	15001	781	4.95
10	12	16044	15407	637	3.97		16044	15606	438	2.73
11				MAPE: MA(3	11.91				MAPE: WM(3	9.79

Sheet1 Sheet2 Sheet3 Sheet4 Sheet5 Sheet6 Sheet7 Sheet

Figure 3.3 Obtaining MAPE for the MA(3) and WM(3) forecasts in Figures 2.4 and 2.5

To find the MAPE for the MA(3) forecasts, we have to perform the following steps:

In cell D2, type = ABS(B2 – C2) and press Enter
Copy and paste the formula into cells D3 through D10
In cell E2, type = (D2/B2) * 100 and press Enter
Copy and paste the formula into cells E3 through E10
In cell E11, type = average (E2:E10) and press Enter to find the MAPE = 11.91 (%)
Repeat the same process for the WM(3) forecasts
As displayed in cell J11 for the WM(3), the MAPE = 9.79 (%)

Obtaining DPE

Cita is the first person to observe that Figure 3.4 displays data from the file Ch03.xls, Fig.3.6, and the steps to calculate the DPE. She also notices that column E in Figure 3.3 becomes column C in Figure 3.4. We learn that to find the DPE for the MA(3) forecast, the following steps are needed:

Copy and paste-special the value in cell C11 into cells D2 through D10
In cell E2, type = (C2 – D2)^2 and press Enter
Copy and paste the formula into cells E3 through E10
In cell E11, type = average (E2:E10) and press Enter to find the DPE = 19 (%)

	A	B	C	D	E	F	G	H	I
1	Period	Demand	|f|/A*100	MAPE	[(|f|/A)*100-MAPE]^2		|f|/A*100	MAPE	[(|f|/A)*100-MAPE]2
2	4	9625	10.39	11.91	2.28		7.63	9.79	4.64
3	5	11114	16.96	11.91	25.58		15.51	9.79	32.74
4	6	12080	16.82	11.91	24.14		14.46	9.79	21.89
5	7	13131	16.69	11.91	22.86		13.57	9.79	14.32
6	8	13275	8.79	11.91	9.70		6.26	9.79	12.44
7	9	14588	12.06	11.91	0.02		10.70	9.79	0.83
8	10	15852	13.80	11.91	3.59		12.27	9.79	6.16
9	11	15782	7.67	11.91	17.95		4.95	9.79	23.40
10	12	16044	3.97	11.91	63.02		2.73	9.79	49.83
11			11.91	DPE: MA(3)	19		9.79	DPE: WM(3)	18

Figure 3.4 Obtaining DPE for the MA(3) and WM(3) forecasts from Figures 2.4 and 2.5

Repeat the same process for the WM(3) forecasts

As displayed in cell I11, the DPE for the WM(3) is 18 (%)

We come to the class the next day to learn that Dr. Theo caught the flu, so Dr. App will teach both the theoretical and the applied sections.

Adjustments Based on Error Evaluations

Concept

To start the lecture, Dr. App emphasizes that interpretation and adjustments are important in order to improve a model's prediction power over time.

Interpretation

Once the error measurements are calculated, compare and contrast all measurements to obtain some meaningful interpretation. The model that has the greatest amounts of smaller error measurements is the best to use.

At this moment, Arti raises a question, "How can you know if a model with the greatest amounts of smaller error measurements is reliable enough to use in your forecasting?"

Dr. App responds that it all depends on the costs and benefits of your data analysis. If you just want to have an overall feel of the market, the simple techniques in Chapter 2 might be sufficient because they are easy to use and so the costs of developing the models and tracking for the changes are low. However, if you want to make a decision on how much stock is needed to appropriately fill your inventory for future sales, you might want to use more advanced forecast techniques introduced in the following chapters to reduce the errors.

Adjustments

We are relieved to learn that regardless of the forecast techniques, instant adjustments can be made to obtain a more reliable model.

Dr. App first introduces the method of central-tendency adjustment, which uses some of the information on the error measures such as FB and SDF. Theoretically, forecast errors are supposed to have a zero mean, implying a forecast bias FB = 0, and a constant SDF. If the FB is greater than zero, then the forecast series can be adjusted upward by adding the bias to the forecast values.

Mo raises his hand to ask, “But markets change very frequently. How can we make just one instant adjustment and be sure that the model is appropriate?” Dr. App praises him for a good question and says that markets change frequently due to changes in consumer preferences, improvements in technology, company realignments, or any supply shocks such as gasoline price changes, wars, and so on. These changes necessitate the need to monitor the forecast errors and adjust the model over time once an instant adjustment is made.

We learn that the SDF should be monitored over time based on the common practice of keeping a band of the lower bound and upper bound between three standard deviations of the forecast ($\pm 3\,\text{SDF}$). If the forecast values fall out of this band for two consecutive periods, market research should be carried out to discover the causes and the patterns of the changes so that adjustments can be made to the model.

Next, we learn the method of error MA adjustments. The method calculates the MAs of the forecast errors and then adds these MAs to the original forecast values to adjust for the error differences.

Finally, we learn that once a model is adjusted, we need to set an adjustment interval to collect new data and continue the evaluation and adjustment process over time. The adjustment intervals can be weekly, monthly, quarterly, and so on depending on the sensitive level of the market.

Dr. App then instructs us to open the data for practicing evaluations and adjustments.

Excel Applications

To compare and contrast the two models in the section on "Error Evaluations," we study Table 3.3, which reports all performance measures for the MA(3) against the WM(3). Alte raises her hand and asks, "How come MAE and FB have the same values in this example?" For which Fin replies, "I think because all forecast values are below the actual values for both models, so the absolute values of the forecast errors are identical to the errors themselves."

Dr. App commends both of them for being such active learners. We also find that the results support our observation on the underestimation (downward bias) of the two models, as shown in Figure 3.1, because the FB values are positive for both models. The results also confirm our intuition from looking at the chart in Figure 3.1 that the WM(3) is a better forecast model than the MA(3): the former has smaller values than the latter in five out of six error measures.

Dr. App reminds us that the same process can be used to compare and contrast the three ES models with different smoothing factors presented in Chapter 2.

Rea then asks, "Since Table 3.3 reveals that the FB values are positive, implying a downward bias, can each forecast curve be adjusted by adding a constant equal to the FB?" We all say, "Yeah…let's try it" and work on the adjustment. We find the data in the file Ch03.xls, Fig.3.8, and display the results in Figure 3.5.

As you see, once the value FB = 1563 is added to the MA(3) curve and FB = 1282 is added to the WM(3) curve, the two models mimic the actual data more closely.

Table 3.3 Performance measures for the MA(3) and the WM(3)

Measures	Model	
	MA(3)	WM(3)
SDF	538	542
RMSE	1,653	1,392
MAE	1,563	1,282
FB	1,563	1,282
MAPE	11.91%	9.39%
DPE	19%	18%

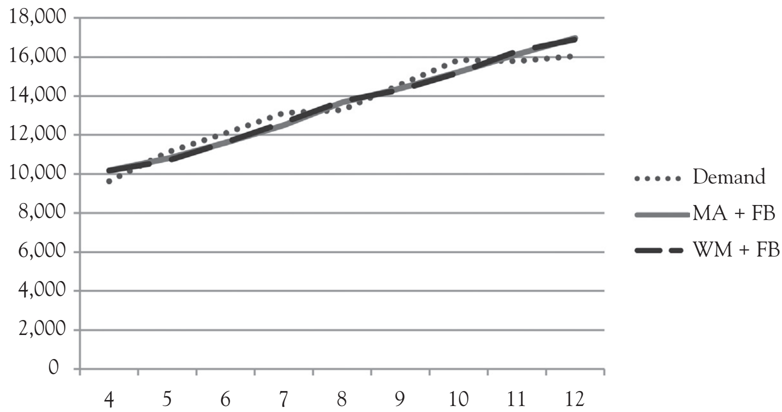


Figure 3.5 Adjusting the MA(3) and WM(3) by adding the FB values

	A	B	C	D	E	F	G	H	I	J	K
1	Period	Demand	MA: F_{t+1}	f_{MA}	MA(f)	Adjust (f)	New F_{t+1}	WM: Ft+1	f_{WM}	Adjust (f)	New Ft+1
2	4	9625	8624	1000				8890	735		
3	5	11114	9228	1885				9390	1723		
4	6	12080	10048	2032	1639			10333	1747		
5	7	13131	10939	2191	2036	1639	12579	11349	1782	1402	12750
6	8	13275	12108	1167	1797	2036	13747	12444	831	1751	14195
7	9	14588	12829	1760	1706	1797	14865	13028	1560	1453	14481
8	10	15852	13665	2188	1705	1706	15461	13908	1945	1391	15299
9	11	15782	14572	1210	1719	1705	16278	15001	781	1445	16447
10	12	16044	15407	637	1345	1719	17112	15606	438	1429	17035
11			15893			1345	17612	15925			15925

Fig.3.4 / MSE-MAE-FB / Fig.3.5 / Fig.3.6 / Fig.3.8 / Fig.3.9

Figure 3.6 Adjusting MA(3) and WM(3) forecasts by adding the forecast-error MA(3)

We then apply the method of error MA adjustments by calculating MAs of the forecast errors and then adding these MAs to the original forecast values. Figure 3.6 displays the data from the file Ch03.xls, Fig.3.9, and the results of our new calculations. Here is what you have to do to get along with us:

In cell E4, type = (D2 + D3 + D4)/3 and press Enter
Copy and paste this formula into cells E5 through E10
Copy and paste-special the values of cells E4 through E10 into cells F5 through F11
In cell G5, type = C5 + F5 and press Enter.
Copy and paste this formula into cells G6 through G11 to obtain new forecasts.

Repeat the same for WM(3) in columns H through K.
(We can skip one step for WM because the values in columns E and F are the same)

We are delighted to see that the new forecast values are closer to the actual values than the original forecasts.

Dr. App points out that there are several other methods of evaluations and adjustments that we might want to learn.

Other Evaluations and Adjustments

Holdout Samples

Holdout-sample evaluations and adjustments are also called out-of-sample forecasts, which consist of two techniques: (i) knife jacking for recalculations of the errors and (ii) cross validation for recalculations of the forecast values. First introduced by Michaelsen (1987), the strategies are discussed in detail by Mason (2004). Both techniques split the historical data on actual values into at least two subsets called in-sample and out-sample that are available for in-of-sample forecasts and out-of-sample evaluations and adjustments. Since past performances do not guarantee the same future performances, the idea is to create a forecast model of the future instead of making only an excellent description of the historical data.

In both techniques, a subset of several periods in the actual data is not used in the forecast model and is called the holdout subset, out-sample subset, or missing subset. The forecasting is performed using only the in-sample subset.

The knife-jacking strategy is used to recalculate the forecast errors employing all possible combination of the in-sample subset. The cross-validation strategy is used to recalculate the forecast values by employing the in-sample subset followed by the holdout subset to reduce bias for the forecasts.

For example, we can leave out the last four quarters of the actual series in Figure 2.4. Hence, the forecast model is first developed using data from September 2011 to June 2013 and provide forecast values up to June 2014. The forecast error measures are then calculated by comparing the forecast values with the actual values for the period September 2013 through June 2014. The forecast values are then recalculated based on these error measures.

Dr. App notifies us that the holdout sample has practical uses in multiperiod forecasting and so we will work on an Excel demonstration in Chapter 4 instead of doing it in this chapter.

At this point, Alte puts forth an issue at her Alcorner store: She and her partner were able to use the ES technique to obtain a series of forecast data for their next month's supplies, but they are not sure whether the forecast series requires further improvement.

Dr. App is quite pleased with her question because this is the very issue to be discussed next in the class.

Data Randomness

This approach investigates whether a model can be further improved by looking at the randomness of the forecast data. Time series forecasts are based on the notion that there are patterns in the past performance of a market that can be used to forecast its future performance. If a forecast series no longer exhibits any pattern, then it is random, and no further improvement can be made by adjusting the model.

Dr. App tells us Bradley (1968) is the researcher who offers a very helpful technique to check on the randomness of a forecast model. This technique is called the run test, which is summarized here for us. We first learn about a *run*.

Given a sample mean, a run is defined as a series of values above the mean or a series of values below the mean. A new run is formed each time the series moves from below to above the mean and vice versa. For example, the series 2, 5, 5, 6, 3, 2, 9, 8, 3, 1 has the sample mean:

$$\bar{x} = (2+5+5+6+3+2+9+8+3+1)/10 = 4.4$$

Hence, this series has 5 runs around its mean: (2) (5 5 6) (3 2) (9 8) and (3 1).

And so the number of runs in this series = $R = 5$.

The number of values above the mean = n_1 = 5 (consists of 5, 5, 6, 9, and 8).

The number of values below the mean = n_2 = 5 (consists of 2, 3, 2, 3, and 1).

The run test is performed in four steps:

i. The null and the alternative hypotheses
H_0: the series is random
H_a: the series is not random
ii. The test statistic

$$Z_{\text{STAT}} = \frac{R - \overline{R}}{s_R} \tag{3.8}$$

where

$$\overline{R} = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{3.9}$$

and

$$s_R = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}} \tag{3.10}$$

iii. Find the critical value of a normal distribution, Z-critical (Z_c), for a two-tail test.
iv. Decision: If $|Z_{\text{STAT}}| > Z_c$, reject the null, meaning the data is not random and implying improvements can be made by further adjusting the model.

As an example, the class breaks into groups and performs the test for the preceding series:

i. The hypotheses
H_0: the series is random
H_a: the series is not random
ii. The test statistic

$$\overline{R} = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \frac{2 * 5 * 5}{5 + 5} + 1 = 6$$

$$s_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2*5*5(2*5*5-5-5)}{(5+5)^2(5+5-1)}}$$

$$= \sqrt{2.2222} = 1.49$$

$$Z_{\text{STAT}} = \frac{R-\bar{R}}{s_R} = \frac{5-6}{1.49} = -0.67$$

iii. The critical value: we choose $\alpha = 0.05$ (a 5 percent significance level), so $\alpha/2 = 0.025$ and $1 - \alpha = 0.975$.

Fin has a table of the critical values for a normal distribution and is able to find that Z-critical = $Z_c = Z_{0.975} = 1.96$

It turns out that we do not need Fin's table because Excel provides Z-critical values. All we have to do is to type in any cell = NORMS-INV (0.975) and press Enter

Excel will report the $Z_c = 1.9599$, which we can round it off to 1.96.

iv. Decision: $|-0.67| = 0.67 < 1.96$, so we do not reject the null.
Meaning: The forecast series is random.
Implication: Improvements cannot be made, that is, no further adjustments of the model are needed.

Dr. App reminds us that if the forecast series is not random, then another round of forecasting should be performed using any technique discussed in this textbook and then the run test can be carried out again until adjustments are no longer needed.

Mo raises his hand and tells us another story: His boss now wants to raise the bar with the forecast results obtained by the class using Mo's data. He is wondering whether a model with a smaller error than the ones we have obtained can be developed. Dr. App assures us that the next section of the lecture will address this issue.

Combining Forecast Models

After trying many forecast models, with none of them providing highly desirable results, we learn that combining several of them might yield a better model. The simplest method is to calculate the average of the individual forecasts. For example, if the three forecast values are $F_1 = 4$, $F_2 = 6$, and $F_3 = 8$, then the simple average forecast is:

$$F_A = (4 + 6 + 8)/3 = 6$$

Suppose that the actual demand is $A = 5.5$, then the forecast error is:

$$f = 5.5 - 6 = -0.5.$$

Most of the time, this approach is too simple to produce a better model. Hence, a weighted adjustment technique is usually preferable. Any measures discussed in the section on "Error Evaluations" can be used for a combined model. For example, if MSE is used, the weight can be written as

$$w_i = \frac{1/\text{MSE}(F_i)}{\frac{1}{\text{MSE}(F_1)} + \frac{1}{\text{MSE}(F_2)} + \ldots + \frac{1}{\text{MSE}(F_i)} + \ldots + \frac{1}{\text{MSE}(F_n)}},$$

$$\text{where } \sum_{i=1}^{n} w_i = 1 \qquad (3.11)$$

The weighted forecast value is

$$F_W = w_1 F_1 + w_2 F_2 + \ldots + w_i F_i + \ldots + w_n F_n \qquad (3.12)$$

At this point, Dr. App asks us to go back to the preceding example and work on an exercise. She gives us the actual demand as $A = 5.5$. Additionally, she gives us the corresponding MSE of the three forecast values as follows:

$F_1 = 4$, and $\text{MSE}_{F1} = 1.2$
$F_2 = 6$, and $\text{MSE}_{F2} = 2.4$
$F_3 = 8$, and $\text{MSE}_{F3} = 2.8$
We proceed to calculate the weights as

$$w_1 = \frac{1/1.2}{1/1.2 + 1/2.4 + 1/2.8} = 0.5185; \quad w_2 = \frac{1/2.4}{1.6071} = 0.2593$$

$$w_3 = \frac{1/2.8}{1.6071} = 0.2222$$

We then calculate the weighted forecast value:

$$F_w = 0.5185 * 4 + 0.2593 * 6 + 0.2222 * 8$$
$$= 2.074 + 1.5558 + 1.7776 = 5.4074$$

Hence, the forecast error is:

$$f = 5.5 - 5.4074 = 0.0926$$

Comparing to the simple average forecast, which has the forecast error of –0.5, we conclude that the weighted forecast has a smaller forecast error and so is a better technique to follow.

Finally, Dr. App tells us that model adjustments be done by computer programs. The *adaptive smoothing* is a computer program that adjusts the smoothing factors whenever the computer notices a change in the pattern of the error terms. The *focus forecasting* allows you to try a variety of forecast models (Heizer and Render 2011). She says that if we want to become professional forecasters, we might want to invest or ask our company to invest in any professional software developed specifically for forecasting to improve our models and the subsequent results. However, she also warns us that no machine can replace a human being and so we still have to periodically keep an eye on the computer.

Exercises

1. The file Demand.xls contains data for the two forecast models with $a = 0.3$ and $a = 0.7$. Use the Excel commands learned in this chapter to calculate the SDF, MAPE, and DPE for each of the two models. Construct a table similar to the one shown in Table 3.3. Which model is better?
2. Use a handheld calculator to calculate the MAE and FB for the forecasts in Exercise 3 of Chapter 2.
3. Use a handheld calculator to perform the run test at 5 percent significance level for the series in the dataset Sales.xls, and show all four steps of the test, including the meaning and implication of your decision.

4. Use the method of MA adjustments to calculate the MA(3) of the forecast errors for the forecasts in Figure 2.6 with $a = 0.3$ then add this MA(3) to the original forecasts to obtain new forecast values. Construct an Excel spreadsheet similar to the one in Figure 3.6.

PART II

Intermediate Forecast Techniques

This part contains two chapters:

- Chapter 4 Intermediate Time Series Techniques
- Chapter 5 Simple Linear Regressions

CHAPTER 4

Intermediate Time Series Techniques

Mo comes to the class today with exciting news: His boss is very happy with the results from our combined forecasting. Additionally, his company is going to open a new branch in the south side of the city. For that reason, his boss asks him to extend the model to allow for forecasts for several periods so that they can stock up the inventory for the new branch. Ex tells us what one of his customers had asked him. "I heard that you are taking a forecasting class. Are you able to forecast your sales correctly?" For which Ex answered, "Sometimes yes, sometimes no, I guess that my forecasts are not always correct because there is a lot of uncertainty in the future."

Dr. App assures us that this week we will discuss multiperiod forecasting and interval forecasts, which address the future uncertainty. She says that after studying this chapter, we will be able to:

1. Explain the concept of double moving averages (DMs).
2. Explain the concept of double exponential smoothing (DE).
3. Apply the concepts in (1) and (2) while calculating multiperiod forecasts.
4. Obtain interval forecasts to account for the future uncertainty.

Dr. App informs the class that Dr. Theo is still on a sick leave, so she will start our new chapter.

Double Moving Average

As the name suggests, the DM technique calculates the MAs a second time. Dr. App notifies us that for notational simplicity, the first MA henceforth is notated as M' while the second MA is notated as M''.

Concept

In the DM technique, we have to calculate second MAs from the simple MAs and then combine with a trend equation to calculate the forecast values. Since we call the first MA M', Equation 2.1 becomes:

$$M'_t = \frac{\sum_{i=t-k+1}^{t} A_i}{k} \tag{4.1}$$

and the equation for second MA, M'', is

$$M''_t = \frac{\sum_{i=t-l+1}^{t} M'_i}{k} \tag{4.2}$$

where the variable definitions are the same as those in Chapter 2 except for M' and M''.

Theoretically, you can choose a different order for M''. Empirically, the two MAs are almost always selected to have the same orders. Therefore, k is used solely to denote the order of both MAs in this textbook. For comparison, Dr. App extends the simple example in Chapter 2 with three additional periods. The new dataset is displayed in Table 4.1.

Table 4.1 Calculating M′ and M″

Period	Actual value	M′	M″
1	20		
2	30		
3	40	(20 + 30 + 40)/3 = 30	
4	35	(30 + 40 + 35)/3 = 35	
5	40	(40 + 35 + 40)/3 = 38.33	(30 + 35 + 38.33)/3 = 34.44
6	50	(35 + 40 + 50)/3 = 41.67	(35 + 38.33 + 41.67)/3 = 38.33
7	60	(40+50+60)/3 = 50	(38.33 + 41.67 + 50)/3 = 43.33

Dr. App asks Cita to go over the simple MA technique using the data in Table 4.1. She is able to provide the forecast value for period 6 as:

$$F_{6,\,\mathrm{MA}} = M'_5 = 38.33$$

Ex then volunteers to calculate the forecast error for this simple MA as follows:

$$\text{The error } f_{6,\,\mathrm{MA}} = A_6 - F_6 = 50 - 38.33 = 11.67$$

We learn that the DM technique includes an equation for the trend T_t, which is written as:

$$T_t = \frac{2}{k-1}(M'_t - M''_t) \qquad (4.3)$$

$$\text{so } T_5 = [2/(3-1)] * (M_5' - M_5'') = 1 * (38.33 - 34.44) = 3.89$$

At this point, Alte asks, "Dr. App, can you explain Equation 4.3 to us?"

Dr. App says, "Yes, I will explain it very soon, but I want you to have fun calculating the forecast value for period 6 with the DM technique first so that you can compare the forecast error of the $F_{6,\,\mathrm{MA}}$ with that of the $F_{6,\,\mathrm{DM}}$. The equation for one-period forecasts using DM technique is:

$$F_{t+1} = M'_t + (M'_t - M''_t) + T_t \qquad (4.4)$$

where the term $(M'_t - M''_t)$ in (4.4) is the adjustment made by a forecaster upon learning of the error between the two MAs."

Fligh volunteers to calculate the forecast for period 6 using Equation 4.4:

$$F_{6,\,\mathrm{DM}} = 38.33 + (38.33 - 34.44) + 3.89 * 1 = 46.11$$

Arti then calculates the forecast error as follows:

$$\text{The error } f_{6,\,\mathrm{DM}} = 50 - 46.11 = 3.89$$

We all say "Wow, this forecast error is much smaller than that of the MA." Clearly, the DM technique can produce a better forecast model than the MA does.

Dr. App emphasizes that multiperiod forecasts are made available by using a trend line. Hence, we will calculate forecasts for period 7 and beyond after she explains the trend equation.

Trend Line in DMs

The trend is the slope of a line by definition. In the DM technique, it is the change in MAs, ΔM, over one unit change in time, Δt:

$$Trend = T_t = \Delta M \,/\, \Delta t \tag{4.5}$$

The problem is that when MAs are taken, the changes in time are often not one period, so an adjusting factor is needed. Figure 4.1 displays three groups: M′(3) and M″(3) in columns A through D, M′(4) and M″(4) in columns F through I, and M′(5) and M″(5) in columns K through N. We learn that all the data and commands for this chapter are available in the file Ch04.xls.

For heuristic purpose, Dr. App uses a dataset with small values so that we can practice using a handheld calculator. Moreover, the series of the actual values has a constant slope of 2 for easy comparison. For all MAs the change in M is shown as follows:

$$\Delta M = M' - M''$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Time	A	M'(3)	M"(3)		Time	A	M'(4)	M"(4)		Time	A	M'(5)	M"(5)
2	1	1				1	1				1	1		
3	2	3				2	3				2	3		
4	3	5	3			3	5				3	5		
5	4	7	**5**			4	7	4			4	7		
6	5	9	7	**5**		5	9	6			5	9	5	
7	6	11	9	7		6	11	8			6	11	7	
8						7	13	10	7		7	13	**9**	
9						8	15	12	9		8	15	11	
10											9	17	13	**9**

Sheet1 Sheet2 Sheet3

Figure 4.1 Changes in time when MAs are taken

The change in time when $k = 3$ is exactly one period as shown in cells C5 and D6 with boldfaced numbers for easy recognition of the change. Hence, for the denominator of Equation 4.5, the time change is from period 4 to period 5, whereas the numerator is $\Delta M = 7 - 5 = 2$. However,

$$k - 1 = 3 - 1 = 2$$

Thus, $\Delta t = 1 = (k - 1)/2 = (3 - 1)/2$.

Suppose you wish to use $k = 5$, then you are taking averages of a five-period subset and the change in time is two periods as shown in cells M8 and N10, again with boldfaced numbers to emphasize the time change from period 7 to period 9 while the numerator is:

$$\Delta M = M' - M'' = 13 - 9 = 4$$

and the denominator is:

$$\Delta t = 2 = (k - 1)/2 = (5 - 1)/2$$

When $k = 4$, as in columns F through I, it is a little more difficult to see because the interval falls between periods 5 and 6 where

$$(6 + 8)/2 = 7 = \text{the value in cell I8.}$$

This implies that the change in time is 1.5.

Since $k = 4$, we now have the denominator as

$$\Delta t = (k - 1)/2 = (4 - 1)/2 = 1.5$$

and the numerator is:

$$\Delta M = M' - M'' = 10 - 7 = 3.$$

Dr. App then says, "You should convince yourself when you get home that in any case, you always have $\Delta t = (k - 1)/2$ and that the trend equation is

$$T_t = \frac{\Delta M}{\Delta t} = \frac{M' - M''}{(k-1)/2} = \frac{2}{k-1}(M_t' - M_t'')$$

Therefore, Equation 4.3 holds and implies that the trend changes from one period to the next, as long as M' and M'' can be calculated."

We all understand the trend equation now and start to calculate forecasts for the next periods.

Multiperiod Forecasts

We realize that we have already calculated the value T_5:

$$T_5 = [2/(3-1)] * (M'_5 - M''_5) = 1 * (38.33 - 34.44) = 3.89$$

So we continue by calculating the next period trends:

$$T_6 = 1 * (41.67 - 38.33) = 3.34$$

$$T_7 = 1 * (50 - 43.33) = 6.67$$

Hence, forecast values for periods 7 and 8 are

$$F_{7,\,\mathrm{DM}} = M'_6 + (M'_6 - M''_6) + T_6 = 41.67 + (41.67 - 38.33) + 3.34 = 48.35$$

$$F_{8,\,\mathrm{DM}} = M'_7 + (M'_7 - M''_7) + T_7 = 50 + (50 - 43.33) + 6.67 = 63.34$$

Fin exclaims, "We have run out of trends already, T_7 is the last one!"

Dr. App smiles, "Yes, that is true. However, the trend itself can be projected into the future. For any period beyond period 8, some researchers use the last trend value ($T_7 = 6.67$ in the preceding example) multiplied by the number of periods ahead to be forecasted:

$$F_{t+m} = g_t + T_t\, m, \quad g_t = M'_t + (M'_t - M''_t), \tag{4.6}$$

where m is the number of periods ahead that have to be forecasted. Since T_7 is the last trend available, we have to use T_7 and g_7 (= 50 + 50 − 43.33 = 56.67) to calculate the forecast value for all later periods. For example,

$$F_{9,\,DM} = g_7 + 2 * T_7 = 56.67 + 2 * 6.67 = 70.01$$

$$F_{10,\,DM} = g_7 + 3 * T_7 = 56.67 + 3 * 6.67 = 76.68$$

$$F_{11,\,DM} = g_7 + 4 * T_7 = 56.67 + 4 * 6.67 = 83.35\text{, and so on.} \quad (4.7)$$

However, the fact that you have many trend values also implies that you can utilize them to obtain a forecast series that mimics the market movement more closely. Here are two examples: (1) you can calculate the averages of all trend values and (2) you can calculate averages of the trend subgroups and then use them alternatively. In the preceding sample, if you calculate the average of all trend values, T_a, then

$$T_a = (3.89 + 3.34 + 6.67)/3 = 4.63$$

$$F_{9,\,DM} = g_7 + 2 * T_a = 56.67 + 2 * 4.63 = 65.93$$

$$F_{10,\,DM} = g_7 + 3 * T_a = 56.67 + 3 * 4.63 = 70.56$$

$$F_{11,\,DM} = g_7 + 4 * T_a = 56.67 + 4 * 4.63 = 75.19\text{, and so on.} \quad (4.8)$$

If you calculate averages of the trend subgroups with a priority given to the current trend and use the results alternatively, then the trends could be

$$T_{a1} = (3.89 + 6.67)/2 = 5.28$$

$$T_{a2} = (3.34 + 6.67)/2 = 5.01$$

From Table 4.1, $g_6 = 41.67 + (41.67 - 38.33) = 45.01$.
Hence, the forecasts will be

$$F_{9,\,DM} = g_6 + 3 * T_{a1} = 45.01 + 3 * 5.28 = 60.85$$

$$F_{10,\,DM} = g_7 + 3 * T_{a2} = 56.67 + 3 * 5.01 = 71.70$$

$$F_{11,\,DM} = g_6 + 5 * T_{a1} = 45.01 + 5 * 5.28 = 71.41\text{, and so on.} \quad (4.9)$$

In brief, adding the trends enables you to forecast multiple periods ahead."

We are relieved that we are not running out of trends. We then turn to applied exercises.

Excel Applications

Figure 4.2 displays the data from the file Ch04.xls, Fig.4.3. We proceed with the following steps:

- Copy and paste the formula for simple MA in cell E4 into cells F6 through F13
- In cell G6, type = 2/(3 – 1) * (E6 – F6) and press Enter
- Copy and paste the formula in cell G6 into cells G7 through G13
- In cell H7, type = E6 + (E6 – F6) + G6 and press Enter
- Copy and paste the formula in cell H7 into cells H8 through H14 to obtain F_{t+1}

Regarding forecast values for the next three periods, use the last trend value:

- In cell H15 type = $E$13 + ($E$13 – $F$13) + C3 * $G$13 and press Enter
- (the $ signs are used to keep the same values throughout the cells. Additionally, Cell C3 is used because it has number 2 that will change into 3 and 4 in lower cells)
- Copy and paste the formula in cell H15 into cells H16 and H17

	A	B	C	D	E	F	G	H	I	J
1		Quarter	t	Demand A_t	M'(3)	M"(3)	T_t	$F_{t+m}(1)$	$F_{t+m}(2)$	$F_{t+m}(3)$
2		9/30/2011	1	7812						
3		12/30/2011	2	8654						
4		3/30/2012	3	9407	8624					
5		6/30/2012	4	9625	9228					
6		9/30/2012	5	11114	10048	9300	748			
7		12/30/2012	6	12080	10939	10072	867	11544		
8		3/30/2013	7	13131	12108	11032	1076	12674		
9		6/30/2013	8	13275	12829	11959	870	14260		
10		9/30/2013	9	14588	13665	12867	798	14568		
11		12/30/2013	10	15852	14572	13688	883	15260		
12		3/30/2014	11	15782	15407	14548	860	16339		743
13		6/30/2014	12	16044	15893	15291	602	17127		731
14		9/30/2014	13				872	17367		
15		12/30/2014	14					17699	18238	18495
16		3/30/2015	15					18301	19110	18687
17		6/30/2015	16					18903	19982	19981

Sheet1 Sheet2 Sheet3

Figure 4.2 DM with alternative trends for multiple-period forecasts

To use the average of all trend values:

In cell G14, calculate the average T_a using the average command learned in Chapter 2
In cell I15, type = $E$13 + ($E$13 – $F$13) + C3 * $G$14 and press Enter
Copy and paste the formula in cell I15 into cells I16 and I17

To use the two average trends alternatively:

In cell J12, type = (G11 + G13)/2 and press Enter
In cell J13, type = (G12 + G13)/2 and press Enter
In cell J15, type = E12 + (E12 – F12) + 3 * J12 and press Enter
Copy and paste the formula in cell J15 into cell J16
In cell J17, type = E12 + (E12 – F12) + 5 * J12 and press Enter

We then construct simple plots of these series against each other. Figure 4.3 displays these plots.

We notice that $F_{t+m}(1)$ and $F_{t+m}(2)$ only provide straight-line trends while $F_{t+m}(3)$ mimics the small fluctuations of the actual series. All three approaches yield better results than those of simple MAs.

Ex says, "I think it is hard to tell which DM model is the best." Dr. App replies, "That is true, so evaluations are needed. For this purpose, using hold-out sample is the best strategy if the sample for actual

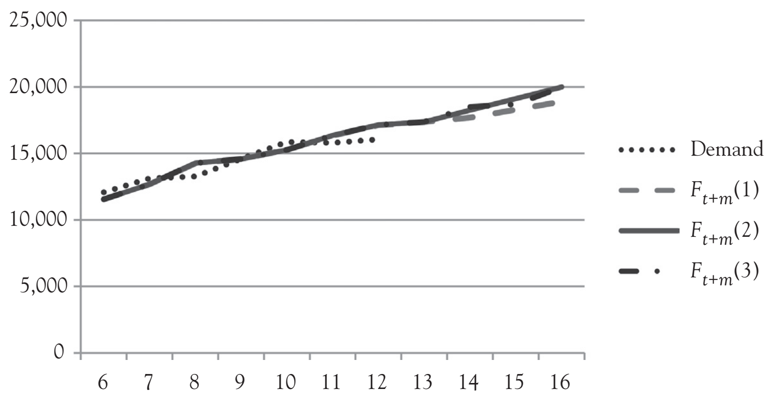


Figure 4.3 DM: plots of forecasts against actual demand

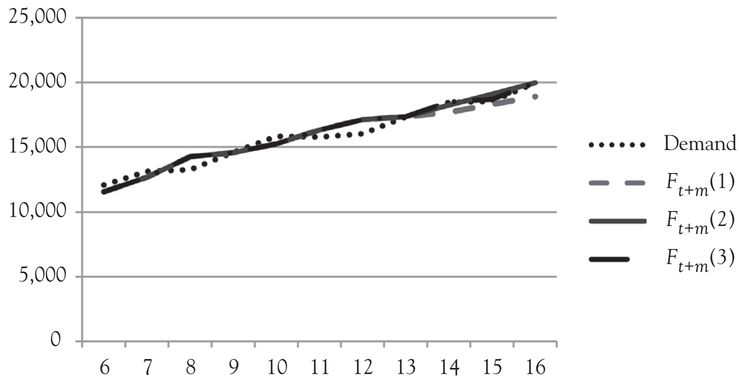


Figure 4.4 Plots of forecasts against actual demand using holdout sample

demand is large enough to split the historical data." With that comment, we proceed to experiment with the holdout sample technique.

Dr. App says, "Suppose that actual data from September 2014 through June 2015 are available, we can then use data from 2011 through June 2014 for forecasting and hold data from September 2014 through June 2015 for evaluations and adjustments. This hypothetical case is illustrated in Figure 4.4, where we see that the third technique of alternating the time trend mimic the actual data most closely."

Dr. App then encourages us to experiment with longer-term alternatives when we get home, for example, by taking the average of a three-period trend and alternating it with another three-period trend.

Double Exponential Smoothing

We learn that there are two approaches to DE: (1) Brown's double exponential smoothing or Brown's one-parameter double exponential smoothing, hereby called Brown's DE, and (2) Holt's two-parameter double exponential smoothing, hereby called Holt's DE.

Brown's DE

The Brown's DE approach uses a single smoothing parameter, a, for both the smoothing and the trend equations (Brown 1963). To obtain Brown's

DE forecasts, we need to calculate the simple exponential smoothing (ES) (E'), the second ES (E''), and then add a trend equation. For all these calculations, the condition that $0 < a < 1$ still holds.

First, we need to review the ES equation from Chapter 2. Cita volunteers to write it on the board. Arti says that she has read the section and volunteers to extend Cita's equation to the following smoothing equations:

$$\begin{aligned} E'_t &= aA_t + (1-a)F_t = aA_t + (1-a)E'_{t-1} \\ E''_t &= aE'_t + (1-a)E''_{t-1} \end{aligned} \tag{4.10}$$

Fin says he has also read the section carefully and can provide the forecast equation as:

$$F_{t+m} = g_t + T_t\, m$$

where

$$g_t = E'_t + (E'_t - E''_t)$$

T is the trend, which is written as

$$T_t = [a/(1-a)] * (E'_t - E''_t). \tag{4.11}$$

Dr. App is very pleased with the class' initiative for active learning. She then asks if anyone can provide a suggestion for the initial value E' in Equation 4.10.

Rea says he believes the initial E' value can be calculated in multiple ways as discussed in the ES technique section in Chapter 2: (1) to take the average of a subset of the actual values, (2) to take the first value of the actual data, or (3) to follow Montgomery et al. (2008) and take the average of all values in the available data.

Dr. App praises him on his correct comment and says, "For simplicity we choose Rea's option (2), $E'_1 = A_1$, and hence

$$E''_1 = E'_1 = A_1 \text{ and } g_1 = A_1$$

Table 4.2 Brown's DE: calculating the forecast values using a = 0.5

(1)	(2)	(3)	(4)	(5)	(6)
T	A_t	E'_t	E''_t	T_t	F_{t+1}
1	20	$= A_1 = 20$	$= A_1 = 20$	0.5/0.5 * (20 − 20) = 0	
2	30	0.5 * 30 + 0.5 * 20 = 25	0.5 * 25 + 0.5 * 20 = 22.5	1 * (25 − 22.5) = 2.5	20 + 0 = 20
3	40	0.5 * 40 + 0.5 * 25 = 32.5	0.5 * 32.5 + 0.5 * 22.5 = 27.5	1 * (32.5 − 27.5) = 5	25 + 25 − 22.5 + 2.5 = 30
4	35	0.5 * 35 + 0.5 * 32.5 = 33.8	0.5 * 33.8 + 0.5 * 27.5 = 30.7	1 * (33.8 − 30.7) = 3.1	32.5 + 32.5 − 27.5 + 5 = 42.5
5	40				33.8 + 33.8 − 39.7 + 3.1 = 40

Also, we can choose $T_1 = \frac{a}{1-a}(E_1' - E_1'') = 0$

Alternatively, $T_1 = \frac{a}{1-a}(A_2 - A_1)$ (4.12)

We will now work on an example." We see that Table 4.2 displays calculations and forecasts for the first five periods of the data in Table 4.1 using the smoothing factor $a = 0.5$, $E'_1 = A_1$, and $T_1 = 0$. This demonstration can be followed using a handheld calculator. For simplicity, the calculations in this figure are rounded off to one decimal place.

Sol points out that the value in column (3) is only E', and the forecast value is in column (6), which is a good point as it reminds us that we are using DE instead of the ES technique.

Dr. App also mentions that it takes longer for the series to settle using DE but it will predict the market movements better than the ES thanks to the combination of the two ES calculations and the added trend.

We then move to Holt's DE technique.

Holt's DE

We learn that in Holt's DE, two different smoothing parameters are used (Holt 1957). The first parameter, *a*, is used for the smoothing equation and can still be experimented with various factors between zero and one:

$$E_t = aA_t + (1 - a)\,(E_{t-1} + T_{t-1}), \text{ where } 0 < a < 1 \tag{4.13}$$

The second parameter, *b*, is used to smooth the trend T_t, which is calculated as:

$$T_t = b(E_t - E_{t-1}) + (1 - b)\,T_{t-1}, \text{ where } 0 < b < 1 \tag{4.14}$$

Dr. App says that we can also try to use various factors for parameter *b* and that the equation for the multiperiod forecasts is quite simple:

$$F_{t+m} = E_t + T_t\,m \tag{4.15}$$

The initial value for T can be either of the following:

$$T_1 = A_2 - A_1, \text{ or } T_1 = \frac{(A_n - A_1)}{n-1},$$

where n is the number of the actual observations, or

$$T_1 = \frac{(A_2 - A_1) + (A_4 - A_3)}{2}, \text{ or}$$

$$T_1 = \frac{(A_2 - A_1) + (A_3 - A_2) + (A_4 - A_3)}{3}. \tag{4.16}$$

Dr. App then asks the class to make suggestions for the initial value of E. We all guess correctly that this initial value for E can be calculated following Rea's suggestions in the section "Brown's DE."

To help us understand the concept, we study Table 4.3, which displays the same dataset from Table 4.2 and shows the steps to calculate Holt's DE. Similar to the reports in Table 4.2, the calculations in this figure are also rounded off to one decimal place.

We work on the problem using a handheld calculator and applying Equation 4.15 with the two parameters specified as $a = 0.5$, $b = 0.4$, and $E_1 = A_1$. We notice again that the value in column (3) is only E_t, and the forecast value is in column (5).

Dr. App says that Holt's DE process takes even longer to settle due to its use of two different parameters.

Table 4.3 Holt's DE: calculating the forecast values using a = 0.5 and b = 0.4

(1)	(2)	(3)	(4)	(5)
T	**A**	E_t (a = 0.5)	T_t (b = 0.4)	F_{t+1}
1	20	= A_1 = 20	[(30 – 20) + (35 – 40)]/2 = 2.5	
2	30	0.5 * 30 + 0.5 * (20 + 2.5) = 26.3	0.4(26.3 – 20) + 0.6 * 2.5 = 4	20 + 2.5 = 22.5
3	40	0.540 + 0.5 * (26.3 + 4) = 35.2	0.4(35.2 – 26.3) + 0.6 * 4 = 6	26.3 + 4 = 30.3
4	35	0.5 * 35 + 0.5 * (35.2 + 6) = 38.1	0.4(38.1 – 35.2) + 0.6 * 6 = 4.8	35.2 + 6 = 41.2
5	40			38.1 + 4.8 = 42.9

Cita raises her hand and asks, "Does that mean that Holt's DE technique produces better results than Brown's DE technique?" Dr. App responds, "No, it does not necessarily work that way. Evaluations and monitoring are crucial to find the best model with the smallest error measurements, as we learned in Chapter 3."

She then asks the class to open the Excel file and go to the applied section.

Excel Applications

We find that Figure 4.5 displays the data and calculations from the file Ch04.xls, Fig.4.8, to obtain Brown's DE forecasts using the smoothing factor $a = 0.3$ and $E_1 = A_1$. The following steps are needed for this exercise:

In cell E3, type = 0.3 * D3 + (1 – 0.3) * E2 and press Enter

Copy and paste the formula in cell E3 into cells E4 through E13 and F3 through F13

In cell G2, type = (0.3/0.7) * (E2 – F2) and press Enter

Copy and paste the formula in cell G2 into cells G3 through G13

In cell H3, type = E2 + (E2 – F2) + G2 and press Enter

Copy and paste the formula in cell H3 into cells H4 through H14 to obtain F_{t+1}

	A	B	C	D	E	F	G	H
1		Quarter	t	Demand A_t	E'	E"	T_t	F_{t+m}
2		9/30/2011	1	7812	7812	7812	0	
3		12/30/2011	2	8654	8064	7888	76	7812
4		3/30/2012	3	9407	8467	8061	174	8317
5		6/30/2012	4	9625	8814	8287	226	9047
6		9/30/2012	5	11114	9504	8652	365	9567
7		12/30/2012	6	12080	10277	9140	487	10721
8		3/30/2013	7	13131	11133	9738	598	11901
9		6/30/2013	8	13275	11776	10349	611	13126
10		9/30/2013	9	14588	12619	11030	681	13814
11		12/30/2013	10	15852	13589	11798	768	14890
12		3/30/2014	11	15782	14247	12533	735	16148
13		6/30/2014	12	16044	14786	13209	676	16696
14		9/30/2014	13					17040
15		12/30/2014	14					17716
16		3/30/2015	15					18392
17		6/30/2015	16					19068

Sheet1 Sheet2 Sheet3

Figure 4.5 Brown's DE with a = 0.3

To obtain F_{t+m} using the last period trend:

In cell H15, type =$E$13 + ($E$13 − $F$13) + C3 * $G$13 and press Enter

Copy and paste the formula in cell H15 into cells H16 and H17

For the average trend and the alternative trends, refer to the section on "Excel Applications" under "Double Moving Averages"

Suddenly, Arti exclaims, "Oh, how come the value in cell E3 of Figure 4.5 is different from that in cell E3 of Figure 2.6?"

We all look at the cells and see that she is correct.

Dr. App assures us, "Yes, the one in Figure 4.5 represents the term E'_2, which could be the forecast for period three (F_3) if you use the ES technique whereas the one in Figure 2.6 represents the forecast for period two (F_2). If you wish to scrutinize this point, you can compare the two figures and find that in Figure 4.5, the value in cell E3 is $E'_2 = 8064$, which equals the forecast for period three $F_3 = 8064$ in Figure 2.6, using the same smoothing factor $a = 0.3$. The forecasts of Brown's DE are in column H." Now it is clear to us.

We then experiment with Holt's DE using the smoothing factor $a = 0.7$, $b = 0.4$, and $E_1 = A_1$. Figure 4.6 displays the data from the file

	A	B	C	D	E	F	G	H
1		Quarter	t	Demand A_t	E_t (a = 0.7)	T_t (b = 0.4)	F_{t+m}	
2		9/30/2011	1	7812	7812	530		
3		12/30/2011	2	8654	8401	554	8342	
4		3/30/2012	3	9407	9105	614	8955	
5		6/30/2012	4	9625	9469	514	9719	
6		9/30/2012	5	11114	10620	769	9982	
7		12/30/2012	6	12080	11642	870	11389	
8		3/30/2013	7	13131	12684	939	12512	
9		6/30/2013	8	13275	13098	729	13623	
10		9/30/2013	9	14588	14141	855	13827	
11		12/30/2013	10	15852	15339	992	14996	
12		3/30/2014	11	15782	15649	719	16331	
13		6/30/2014	12	16044	15926	542	1636 Row: 1	
14		9/30/2014	13				16468	
15		12/30/2014	14				17010	
16		3/30/2015	15				17552	
17		6/30/2015	16				18094	

Fig.4.2 Fig.4.3 Fig.4.4 Fig.4.5 Fig.4.8 **Fig.4.9**

Figure 4.6 Holt's DE with a = 0.7 and b = 0.4

Ch04.xls, Fig.4.9, for this exercise. We learn that the following steps should be carried out:

In cell E3, type = 0.7 * D3 + (1 – 0.7) * E2 and press Enter
Copy and paste the formula in cell E3 into cells E4 through E13
In cell F2, type = ((D3 – D2) + (D5 – D4))/2 and press Enter
In cell F3, type = 0.4 * (E3 – E2) + (1 – 0.4) * F2 and press Enter
Copy and paste the formula in cell F3 into cells F4 through F13
In cell G3, type = E2 + F2 and press Enter
Copy and paste the formula in cell G3 into cells G4 through G14
The forecast value for period 13 is in cell G14

To obtain F_{t+m} using the last period trend:

In cell G15, type= $E$13 + C3 * $F$13 and press Enter
Copy and paste the formula in cell G15 into cells G16 and G17

Dr. App then says, "You can experiment with the average trend, the alternative trends, and various smoothing factors for Brown's DE and Holt's DE when you get home."

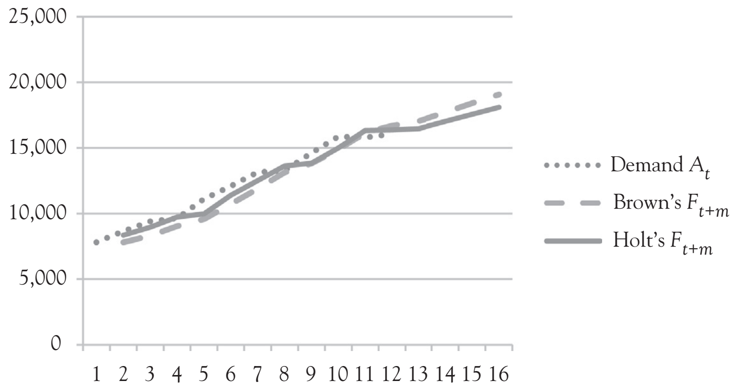


Figure 4.7 DE: plots of the forecasts against actual demand

We also chart the actual demand against Brown's DE and Holt's DE forecasts using the data from the file Ch04.xls, Fig.4.10, and display them in Figure 4.7.

Dr. App concludes this section by reminding us that the holdout sample is again an appropriate strategy for evaluations in these multiperiod forecasts. She also says that Dr. Theo has recovered from his flu and is very happy to return to the class to lead us into the following section.

Interval Forecasts

Dr. Theo says that we have only learned to make point forecasts, that is, one value for each period. We learn that a point forecast does not account for the inherent fluctuation in the market. Obtaining an interval forecast will allow us to state with a certain level of confidence that the actual value will likely fall between the upper and lower bounds of this range.

Concept

Alte gives us an example from her Alcorner store. She used to buy zippers from an online source on a weekly basis. Since the shipping takes two weeks to arrive, she had to place the order on a rolling scheme: the package needed for the third week has to be ordered by the first week of a month, the package needed for the fourth week has to be ordered by the second week, and so on. Last month, she read the section on multiperiod forecasts and, based on her calculations, decided to place one large order of zippers for four weeks in place of her usual rolling scheme orders. However, by week four, she had run out of zippers and had to rush to a local store to buy them at twice the online cost. Hence, she guessed that the uncertainty is what she did not account for in her four-week forecast.

Dr. Theo thanks her for sharing her experience with us and says that she should have used interval forecasts instead of point forecasts for placing her orders. In Chapter 3, we learned that the mean absolute error (MAE) can be used to evaluate a forecast model. This same MAE can be used to construct an interval forecast. First, Equation 3.4 in Chapter 3 is for one-period MAE. For multiperiod forecasts, the uncertainty increases over time, so the simplest approximation is to adjust the MAE to MAE_{t+m}:

$$MAE_{t+m} \approx \frac{\sum_{t=1}^{T} |A_t - F_t|}{T - m} \tag{4.17}$$

where m is the number of periods ahead to be forecasted.

The standard error of the forecast, $se(f)_{t+m}$, then can be written as:

$$se(f)_{t+m} \approx 1.25 * MAE \approx 1.25 * \frac{\sum_{t=1}^{T} |A_t - F_t|}{T - m} \tag{4.18}$$

To calculate interval forecasts, a Z-critical value of a normal distribution, Z_c, which is similar to the one we learned in Chapter 3, is given so that:

$$P\left[F - Z_c * se(f)_{t+m} \leq A \leq F + Z_c * se(f)_{t+m}\right] = 1 - \alpha \tag{4.19}$$

The interval is called a $100(1 - \alpha)\%$ confidence interval. For example, if we choose $\alpha = 0.05$, then the confidence interval is 95 percent, that is, we are 95 percent confident that the actual value will fall between the upper and lower bounds of the interval.

We then break into groups to work on the problem for Alte's store. Alte tells us the four-week forecast for Alcorner is $F = 50$. We follow the instruction in Chapter 3 and find that the sum of $|A_t - F_t|$ is 96. Alte's data contain 16 periods, but her forecast is four periods ahead, that is, $m = 4$. Thus, the standard error of the forecast for the fourth period is:

$$se(f)_{t+4} = 1.25 * \frac{96}{16-4} = 10$$

We decide to choose $\alpha = 0.05$. Similar to the procedure in Chapter 3, $\alpha/2 = 0.025$, so a normal distribution table will give $Z_c = Z_{(0.975)} = 1.96$, and the 95 percent confidence interval forecast for Alte's store is:

$$50 \pm 1.96 * 10 = (30.4; 69.6) \approx (30; 70)$$

Thus, we are 95 percent confident that the number of zippers needed for her store will be between 30 and 70. Alte is very happy. From now on, she will make sure to order 70 zippers instead of 50 for her four-week supply.

Dr. Theo is quite pleased and says that he will continue to guide us through the next section because the Excel application is simple.

Excel Application

The interval forecasts have a lower bound and an upper bound, so we need the regions for the two tails of a normal distribution. For example, if $\alpha = 0.05$, then:

$$\alpha/2 = 0.025, \text{ so } 1 - \alpha/2 = 0.975.$$

After that, all we have to do is to type in any cell

= NORMSINV(1 – α/2) and press Enter.

For example, type

= NORMSINV(0.975) and press Enter. This yields 95 percent Z-critical value = 1.9599 ≈ 1.96.

Once the Z-critical is obtained, we can calculate the interval forecasts using the Excel mathematical operations introduced in Chapter 1.

Dr. Theo concludes the section by saying that this technique is the simplest one. In Chapters 5 and 6, we will learn more sophisticated techniques of calculating interval forecasts.

Exercises

1. The file Maui.xls contains data on visitor arrivals at Maui in Hawaii. Perform the DM(4) procedure on an Excel spreadsheet with the last trend value as the long-term trend for F_{t+m} up to $m = 4$. Construct columns in Excel similar to the ones in Figure 4.2.
2. Use the dataset Maui.xls to perform Brown's DE procedure on an Excel spreadsheet with $a = 0.7$ and the long-term trend as an average of all previous trends for F_{t+m} up to $m = 4$. Construct columns in Excel similar to the ones in Figure 4.5.
3. Use a handheld calculator to perform the DM(3) on the first six observations in the dataset Sales.xls. Organize the results into a table similar to the one in Table 4.1.
4. Use a handheld calculator to perform Brown's DE with $a = 0.2$ and then Holt's DE with $a = 0.2$ and $b = 0.7$ on the first four observations in the dataset Sales.xls. Organize the results into tables similar to the ones in Tables 4.2 and 4.3.
5. The file Electricity.xls contains data on demand for electricity in Hawaii. Perform Holt's DE on an Excel spreadsheet with $a = 0.3$ and $b = 0.6$ and with the last period trend as the long-term trend for F_{t+m} up to $m = 4$.
   a. Construct columns in Excel similar to the ones in Figure 4.6.
   b. Construct a 95 percent confidence interval for the three-period forecast ($m = 3$).

CHAPTER 5

Simple Linear Regressions

Rea tells the class a story about his Realmart Company. A customer asked him, "You provided me with numerous sale prices of the past and nothing else. Do you really believe that these past prices are the only factors that affect future prices in the real estate market?" Rea admitted that many other factors might affect consumer spending and the subsequent home prices, but he did not know how to incorporate these determinants into his analysis. He promised the customer that in three weeks he would find out how to account for these factors.

Ex says that he too heard that China's yuan had appreciated 10 percent against the U.S. dollars from 2010 to 2013 before it tumbled 2 percent in 2014. He wants to how this exchange rate fluctuation will affect China's spending on the U.S. exports so that he can advice his boss.

Luckily, this week and next week we are going to learn one of the techniques used in associative relations—the linear regression analysis. Dr. Theo tells us that once we finish with this chapter, we will be able to:

1. Explain the concept of an econometric model used in associative analyses.
2. Develop models for simple linear regressions and discuss conditions for using them.
3. Discuss the econometric forecasting approach using simple linear regressions.
4. Analyze numerous methods of evaluations and adjustments in simple linear regressions.
5. Perform regressions and obtain forecasts using Excel.

Basic Concept

Chapters 5 and 6 depart from time series analyses to present one of the associative analyses, the linear regressions, in which the performance of a variable depends on other variables in addition to its past performance. A linear regression is performed on an econometric equation, which contains a dependent variable on the left-hand side and one or more explanatory variables on the right-hand side. Other types of associative analyses, the nonregression models, will be discussed in Chapters 8, 9, and 10.

Econometric Models

Dr. Theo says that an econometric model answers the question "by how much" whereas an economic model reflects general theory. For example, economic theory posits that personal consumption (CONS) depends on personal income:

$$\text{CONS} = a_1 + a_2\,\text{INCOME} \tag{5.1}$$

where a_1 is a constant representing the average consumption by a person with no income, INCOME = 0, and a_2 is the marginal propensity to consume, which is the change in consumption due to a unit change in personal income.

An econometric model is developed so that you can collect data and estimate the value of *a*'s:

$$\text{CONS} = a_1 + a_2\,\text{INCOME} + e \tag{5.2}$$

where a_1 and a_2 are the intercept and the slope of the regression line, respectively. The error term e captures the random component of CONS. We can generalize the model in Equation 5.2 to any variables:

$$y = a_1 + a_2 x + e \tag{5.3}$$

For cross-sectional data, the six classic assumptions for a simple linear regression are:

i. The model is $y_i = a_1 + a_2 x_i + e_i$
ii. $E(e_i) = E(y_i) = 0$
iii. $\text{Var}(e_i) = \text{Var}(y_i) = \sigma^2$
iv. $\text{Cov}(e_i, e_j) = \text{Cov}(y_i, y_j) = 0$ for $i \neq j$
v. x_i is not random and must take at least two different values
vi. $e_i \sim N(0, \sigma^2)$; $y_i \sim ([a_1 + a_2 x_i], \sigma^2)$

If assumptions (i) through (v) hold, then the Gauss–Markov theorem states that the simple linear regression will produce the best linear unbiased estimators (BLUE) using the least squares technique, which is also called the ordinary least squares (OLS) estimation. The OLS technique minimizes the sum of the squared errors to the regression line, hence the name *least squares*. If assumption (vi) holds, then test results are valid. In practice, assumption (vi) only needs to hold in approximation.

Under the central limit theorem, when the sample size is sufficiently large, the error terms and the OLS estimators have a distribution that approximates a normal distribution.

Cita then asks, “So how large is large enough?” Dr. Theo commends her on the question and says that the justification of *sufficiently large* is a matter of interpretation, but a cross sectional dataset of 30 observations or a time series dataset of 20 observations is usually considered sufficient.

For time-series data, assumptions (i), (ii), (iii), (iv), and (vi) are still applied except that the subscript i is changed to t, and the subscript j is changed to z. Assumption (v) changes to:

v. y_t and x_t are stationary random variables, and e_t is independent of current, past, and future values of x_t.

We learn that for Chapters 5 and 6, a stationary series is one that is neither explosive nor wandering aimlessly. More discussions on this concept will come in Chapter 7.

Estimations

The simple linear regression is applied when y is a linear combination of the parameters in Equation 5.3. For example, the model $y = a_1 + a_2 x^2 + e$ is considered linear because the squared term is not for parameters a_1 and a_2. Thus, we can let $w = x^2$ so that the model is $y = a_1 + a_2 w + e$. Similarly, the model $y = a_1 + a_2 \ln(x) + e$ is linear in logarithm if we set $w = \ln(x)$.

The estimated version of Equation 5.3 for a particular sample is:

$$\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_i, \quad y_i = \hat{y}_i + \hat{e}_i = \hat{a}_1 + \hat{a}_2 x_i + \hat{e}_i \tag{5.4}$$

Specific values for the parameters $\hat{a}_1$ and $\hat{a}_2$ are called point estimates of the OLS regressions, and the general estimators can be calculated as:

$$\hat{a}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

and

$$\hat{a}_1 = \bar{y} - \hat{a}_2 \bar{x} \tag{5.5}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of x and y, respectively.

For example, Table 5.1 displays a small dataset of three observations and the steps needed to calculate the numerator and denominator in the formula for $\hat{a}_2$.

Hence, $\hat{a}_2 = 3/6 = 1/2 = 0.5$, and $\hat{a}_1 = 2 - (1/2) * 1 = 2 - (1/2) = 3/2 = 1.5$, so the equation for the regression line becomes

$$\hat{y}_i = 1.5 + 0.5\, x_i$$

Table 5.1 Calculating numerator and denominator in formula for $\hat{a}_2$

Variable	x	$\bar{x}$	y	$\bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})\ (y - \bar{y})$
	2	1	2	2	$(2 - 1)^2 = 1$	$(1)(2 - 2) = 0$
	−1	1	1	2	$(-1 - 1)^2 = 4$	$(-2)\ (1 - 2) = 2$
	2	1	3	2	$(2 - 1)^2 = 1$	$(1)(3 - 2) = 1$
where		$\bar{x} = (2 - 1 + 2)/3$;		$\bar{y} = (2 + 1 + 3)/3$	$\sum (x - \bar{x})^2 = 6$	$\sum (x - \bar{x})(y - \bar{y}) = 3$

where 1.5 is the intercept and 0.5 is the slope of the line. Suppose y is weekly consumption and x is weekly income per person, and both are in hundreds of dollars. Then the results imply that:

i. The weekly consumption of a person with no income is \$150 (= 1.5 * \$100).
ii. A \$100 increase in weekly income raises consumption by \$50 (= 0.5 * \$100).

To calculate interval estimates, a t-distribution for a sample of N observations is given as:

$$t = \frac{\hat{a}_k - a_k}{se(\hat{a}_k)} \sim t_{(N-2)} \text{ for } k = 1 \text{ or } 2 \text{ in simple regression}$$

where

$N - 2$ = the degrees of freedom (df) for the simple linear regressions

$\hat{a}_k$ = the coefficient (parameters) to be estimated

$se(\hat{a}_k)$ = the standard error of the coefficient estimate.

Using the same procedure learned in Chapter 4 for interval forecasts, we have

$$P[\hat{a}_k - t_c * se(\hat{a}_k) \le a_k \le \hat{a}_k + t_c * se(\hat{a}_k)] = 1 - \alpha \quad (5.6)$$

Equation 5.6 provides a formula for calculating an interval estimator of a_k.

For example, suppose the sample size is $N = 32$ (df = 30), $\hat{a}_2 = 0.5$, and se $(\hat{a}_2)$= 0.2.

For a 95 percent confidence interval, $\alpha = 0.05$, so $\alpha/2 = 0.025$, and typing = TINV(0.05, 30) into any Excel yields $t_c = t_{(0.975,\ 30)} = 2.042$. Hence, the interval estimate for a_2 is:

$$0.5 \pm 2.042 * 0.2 = (0.0916; 0.9084).$$

Predictions and Forecasts

Predictions

Predictions are used only when cross-sectional data are available and so we can only make predictions for the near future. For example, if the data in Table 5.1 are cross sectional, the prediction for y when $x = 4$ (in hundreds of dollars) can be calculated using Equation 5.4:

$$\hat{y} = 1.5 + 0.5 * 4 = 1.5 + 2 = 3.5 \text{ (\$ hundreds)} = \$350.$$

Thus, a person with a weekly income of $400 will spend $350 on average.

Interval prediction accounts for any uncertainty in the future. Let $\hat{y}_1 = \hat{a}_1 + \hat{a}_2 x_1$, then the standard error of the forecast $se(f)$ can be approximated as in Kmenta (2000):

$$se(f) = s\sqrt{1 + \frac{1}{N} + \frac{(x_1 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}; \quad s = \sqrt{\frac{\sum \hat{e}_i^{\,2}}{N-K}} = \sqrt{\frac{\text{SSE}}{N-K}} \qquad (5.7)$$

where

s = the standard error of the regression.

SSE = the sum of the squared errors (called the residuals in Excel).

The interval prediction is calculated in similar manner as in Equation 5.6:

$$P[\hat{y}_1 - t_c se(f) \leq y_1 \leq \hat{y}_1 + t_c se(f)] = 1 - a \qquad (5.8)$$

For example, if $se(f) = 0.1$, $N = 32$, and $\alpha = 0.05$, then $t_c = 2.042$, so the prediction for weekly consumption is:

$$3.5 \pm 2.042 * 0.1 = (3.2958; 3.7042) = (\$329.58; \$370.42)$$

Therefore, we predict with 95 percent confidence that a person with a weekly income of $400 will spend between $329.58 and $370.42 weekly.

Forecasts

For time series data, we can write the model with two different periods:

$$y_t = a_1 + a_2 x_{t-1} + e_t \qquad (5.9)$$

Multiperiod forecasts can be performed if x's are known for more than one period ahead. For example, if the data in Table 5.1 are time series, and x's are known with $x_t = 2$ while $x_{t+1} = -1$, then the forecasts for y are:

$$y_{t+1} = 1.5 + 0.5 * 2 = 2.5$$

$$y_{t+2} = 1.5 + 0.5 * (-1) = 1$$

Dr. Theo tells us that if the future values of x's are unknown, one of the techniques from the previous chapters can be used. This is a *forecast of forecast* technique, where the independent variable has been forecasted before the dependent variable is forecasted. For example, if x's in Table 5.1 are $x_{t-2} = 2$, $x_{t-1} = -1$, and $x_t = 2$, then MA(3) can be calculated to obtain $x_{t+1} = (2 - 1 + 2)/3 = 1$, and the forecasts for y are:

$$y_{t+1} = 1.5 + 0.5 * 2 = 2.5$$

$$y_{t+2} = 1.5 + 0.5 * 1 = 2$$

We learn that one-period interval forecasts can be calculated using Equation 5.8, adapted for time series data with the estimated equation $\hat{y}_{t+1} = \hat{a}_1 + \hat{a}_2 x_t$ and $N = T$ (Pindyck and Rubinfeld 1998). For multiperiod forecast, he says that the simplest way to obtain an approximation of Equation 5.7 is:

$$se(f)_{t+m} \approx s\sqrt{1 + \frac{1}{T} + \frac{(x_{t+m} - \bar{x})^2}{\sum (x_t - \bar{x})^2}}; \quad s \approx \sqrt{\frac{\sum \hat{e}_i^{\,2}}{T - 2 - m}} = \sqrt{\frac{\text{SSE}}{T - 2 - m}} \tag{5.10}$$

Dr. Theo also tells us that Dr. App is away at a conference, so he will work with us in the following section.

Excel Applications

Cross-Sectional Data

Rea has collected data on personal income (INCOME) and residential-property investment (INV) for 50 states in the United States and

Washington, DC, in 2012. He tells us that the dataset is too large to display but is available in the file Ch05.xls, Fig.5.2 and Fig.5.3, and that the units are in thousands of dollars. We open the data and follow these steps to perform a regression of INV on INCOME (i.e., INV and INCOME are the dependent and independent variables, respectively):

Click on Data and then Data Analysis on the Ribbon
Select Regression in the list instead of Descriptive Statistics and click OK
A dialog box will appear as shown in Figure 5.1
In the Input Y Range box, enter B1:B52
In the Input X Range box, enter C1:C52
Choose Labels and Residuals
Check the Output Range button and enter F1
Click OK. Another dialogue box will appear
Click OK to overwrite the data and obtain the regression results

The Excel Summary Output (henceforth called the results) is displayed in Figure 5.2.

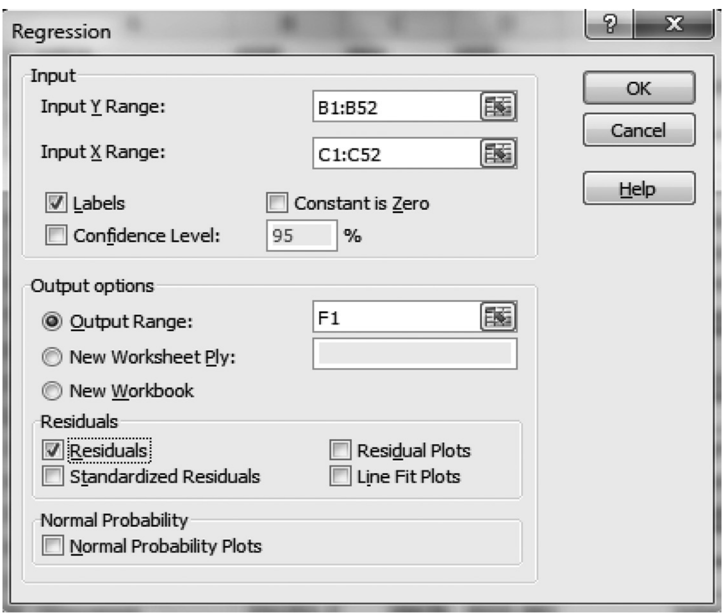


Figure 5.1 Performing regression: commands in dialog box

	A	B	C	D	E	F	G	H	I
1		SUMMARY OUTPUT							
2		*Regression Statistics*							
3		Multiple R	0.970314951						
4		R Square	0.941511104						
5		Adjusted R Squ	0.940317454						
6		Standard Error	11150.20507						
7		Observations	51						
8		ANOVA							
9			*df*	*SS*	*MS*	*F*	*Significance F*		
10		Regression	1	98064951117	98064951117	788.7658626	7.25245E-32		
11		Residual	49	6092026586	124327073.2				
12		Total	50	1.04157E+11					
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
14		Intercept	-3613.949784	2031.289072	-1.779141055	0.08141981	-7695.978003	468.0784355	
15		INCOME	0.135985478	0.00484193	28.08497574	7.25245E-32	0.126255256	0.1457157	

Sheet4 / Sheet1 / **Sheet2** / Sheet3

Figure 5.2 Cross-sectional data: simple linear regression results

Data Source: Bureau of Economic Analysis.com (2014).

From these results, the estimated equation can be written as:

$$\text{INV}_i = -3613 + 0.136 \text{ INCOME}_i$$

To calculate the prediction value for INV in 2013, we need to substitute a particular value of INCOME into this equation. It turns out that Excel automatically calculates predicted values and reports them next to the residuals. For example, you can find the predicted INV for Connecticut in cell G25 of the data file, which is 26419.19 (in thousands of dollars) = $26,419,190. Note that Excel also reports the upper and lower 95 percent bounds for the coefficient estimates in cells G14 through H15. Commands for interval forecasts will be provided in the following section.

Time-Series Data

Ex has collected data on China–United States real exchange rate (EXCHA) and exports from the United States to China (EXPS) in millions of dollars for the period 1981–2012 from the International Monetary Fund (IMF) website. The hypothesis is that the exchange rate affects exports, so we regress EXPS_t on EXCHA_{t-1} using the data in the file Ch05.xls, Fig.5.4:

Click on Data and then Data Analysis on the Ribbon
Select Regression in the list and click OK
A dialog box will appear

	A	B	C	D	E	F	G	H
1		SUMMARY OUTPUT						
2		*Regression Statistics*						
3		Multiple R	0.374690088					
4		R Square	0.140392662					
5		Adjusted R Squ	0.111739084					
6		Standard Error	22651.0473					
7		Observations	32					
8		ANOVA						
9			*df*	*SS*	*MS*	*F*	*Significance F*	
10		Regression	1	2513866001	2513866001	4.899655558	0.034609442	
11		Residual	30	15392098311	513069943.7			
12		Total	31	17905964312				
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
14		Intercept	-9929.614766	15531.36747	-0.639326498	0.527462585	-41648.89875	21789.66922
15		$EXCHA_{t-1}$	4807.421043	2171.847788	2.213516559	0.034609442	371.9161262	9242.925959

Sheet1 Sheet2 Sheet3

Figure 5.3 Time series data: simple linear regression results

Data Source: IMF.com: IMF Data and Statistics (2014).

In the Input Y Range box, enter B1:B33
In the Input X Range box, enter C1:C33
Choose Labels and Residuals
Check the Output Range button and enter P1
Click OK, and a dialogue box will appear. Click OK to overwrite the data

The results of the estimated coefficients are displayed in Figure 5.3.

In the data file, the forecast value of $EXPS_{2013}$ is in cell Q56. The value can be verified by the following equation:

$$EXPS_{2013} = -9929.6148 + 4807.4210\ EXCHA_{2012}$$
$$= -9929.6148 + 4807.4210 * 6.98$$
$$\approx 23{,}626 \text{ (in millions of dollars).}$$

Obtaining Interval Forecasts

To obtain the interval forecasts for this series:

In cell B36, type = TINV(0.05,30) and press Enter
(this is the 95 percent *t*-critical value, which is 2.042)
Copy the Predicted EXPS and the Residuals in cells Q25 through R56
Paste these values into cells D3 through E34

In cell F3, type = E3^2 and press Enter (this is the squared error)

Copy and paste this formula into cells F4 through F34

In cell F35, type = SUM(F3:F34), then press Enter (this is the sum of the squared errors)

In cell F36 type = F35/(32 – 2) and press Enter

(This is s^2, we will take the square-root of it in cell L3 later)

Copy and paste-special the value in cell F36 into cells G3 through G34

In C34, type = AVERAGE(C2:C33), then press Enter (this is the average of x, called xbar)

Copy and paste-special the value in cell C34 into cells H3 through H34

In cell I3, type = (C2 – H3)^2 and press Enter (this is $[x_t - \text{xbar}]^2$)

Copy and paste the formula in cell I3 into cells I4 through I34

In cell I35, type =SUM(I3:I34), then press Enter (this is the sum of $[x_t - \text{xbar}]^2$)

Copy and paste-special the value in cell I35 into cells J3 through J34

(This is the sum of $[x_t - \text{xbar}]^2$ for all cells in question)

In cell K3, type = G3 * (1+(1/32) + (I4/J3)), then press Enter (this is var[f])

In cell L3, type =K3^0.5 and press Enter (this is *se*[f])

Copy and paste the formulas in cell K3 and L3 into cells K4 through L34

In cell M3, type =D3 – 2.042 * L3, then press Enter (this is the lower 95 percent bound)

In cell N3, type =D3 + 2.042 * L3, then press Enter (this is the upper 95 percent bound)

Copy and paste the formulas in cell M3 and N3 into cells M4 through N34

These are the lower 95 percent and upper 95 percent bounds for all forecast values

The interval forecasts for 2013 are displayed in cell M34 and N34, which are:

(Lower 95 percent bound: −42,295.11; and upper 95 percent bound: 89,547.48)

Since export values are nonnegative, the interval should be rewritten as:

(Lower 95 percent bound: 0; and upper 95 percent bound: 89,547.48),

or

Interval forecast_{2013} = (0; 89,547.48)

Alternatively, you can take the standard error of the regression (*s*) from cell C6 in Figure 5.3 or from cell Q7 in the file Ch05.xls, Fig.5.4.

Evaluations and Adjustments

We learn that the evaluation methods in linear regression are quite different from the time series analyses. First, we have to perform tests to see whether or not the estimated coefficients are statistically significant or have the expected values. We then have to perform tests on several common problems encountered in econometric forecasting concerning the errors. Dr. Theo reminds us that under the central limit theorem, when the sample size is sufficiently large, the errors and the OLS estimators have a distribution that approximates a normal distribution; so test results are valid.

Testing Estimated Coefficients

To verify the statistical significance or the expected values of estimated coefficients, *t*-tests are performed.

Procedure

i. State the hypotheses

H_0: $a_k = c$ where c is a constant, which is our conjecture

H_a: $a_k > c$, or H_a: $a_k < c$, or H_a: $a_k \neq c$.

ii. The test statistic

$$t_{\text{STAT}} = t = \frac{\hat{a}_k - c}{se(\hat{a}_k)} \sim t_{(N-2)} \text{ for } k = 1 \text{ or } 2 \text{ in simple regression.} \quad (5.11)$$

iii. The rejection region: the critical *t*-value shows the boundary of this region. The significance levels of the test are often at 1, 5, or 10 percent.

iv. Decision: If the *t*-statistic value falls into one of the two rejection regions, then we reject the null hypothesis. Otherwise, we do not reject the null.

A test with the null hypothesis (H_0) stated as $a_2 = 0$ is called a test of significance because if a_2 is zero, there is no significant relationship between the dependent and independent variables, so the latter does not help in predicting the former.

Examples of *t*-Tests

Arti offers an example: Yesterday she performed a regression using data for a class of 40 students at her school. The dependent variable is their spending on music lessons (SPEND), and the independent variable is their income (INCOME). She found the following relationship between the two: SPEND = 0.10 * INCOME, and se ($\hat{a}_2$) = 0.02. Dr. Theo wants us to test the significance of the slope, that is, whether or not the slope is zero against the alternative hypothesis that the slope is positive. We find that the degrees of freedom is $N - 2 = 38$, and we need to perform the tests in the four standard steps as follows.

Right-Tail Test (>): we want to test the alternative hypothesis $a_2 > 0$.

i. H_0: $a_2 = 0$; H_a: $a_2 > 0$
ii. $t_{\text{STAT}} = t_{(N\text{-}2)} = (0.10 - 0)/0.02 = 5$
iii. We decide to choose $\alpha = 0.05$, so $t_c = t_{(0.95,\ 38)} = 1.686$ from a *t*-table.
   Dr. Theo says that Excel always reports a two-tail critical value, so to find *t*-critical for one tail test, type into any cell = TINV(2α, df), then press Enter.
   For example, type = TINV(0.10, 38) and press Enter, this yields 1.68595 ≈ 1.69.
iv. Decision: Since $t_{(N\text{-}2)} > t_{(0.95,\ 38)}$, we reject the null, meaning $a_2 > 0$, and implying that the students' income has a positive effect on their music-lesson expenditures.

Left-Tail Test (<): Dr. Theo wants us to test the alternative hypothesis $a_2 < 0.14$

i. H_0: $a_2 = 0.14$; H_a: $a_2 < 0.14$
ii. $t_{\text{STAT}} = t_{(N\text{-}2)} = (0.10 - 0.14)/0.02 = -2.0$
iii. We continue to choose $\alpha = 0.05$, so $t_c = t_{(0.95,\ 38)} = 1.686$, and $-t_c = -1.686$.

iv. Decision: $t_{STAT} < -t_c$, so we reject the null, meaning $a_2 < 0.14$ and implying that the students tend to spend less than 14 percent of their rising income on music lessons.

Two-Tail Test: This time, Sol wants to test the alternative $a_2 \neq 0$, and Dr. Theo agrees.

i. H_0: $a_2 = 0$; H_1: $a_2 \neq 0$

ii. $t_{STAT} = t_{(N\text{-}2)} = (0.10 - 0)/0.02 = 5$

iii. Since this is a two-tail test, Dr. Theo reminds us to use $\alpha/2 = 0.025$, and so $t_c = t_{(0.975,\ 38)} = 2.024$

For a two-tail test, we have to type into any cell = TINV(α, df). Hence, we type = TINV(0.05, 38), then press Enter. Excel gives us 2.02439.

iv. Decision: Since $t_{STAT} > t_c$, we reject once more the null, meaning $a_2 \neq 0$ and implying that the students' income helps in predicting their music-lesson expenditures.

Error Diagnostics

Dr. Theo then discusses two common problems with the errors. The first is heteroskedasticity, which can occur with either cross-sectional or time-series data. The second is autocorrelation, which is a specific problem in time-series regressions. We find that the two problems have the same consequences:

i. The OLS estimators are no longer the BLUE.

ii. The standard errors are incorrect, so statistical inferences are not reliable.

Heteroskedasticity

This problem refers to the violation of the classic assumption (iii), which states that the variance of the errors is a constant. Errors exhibiting changing variance are said to be heteroskedastic. To see the problem, we all look at the original equation:

$$y_i = a_1 + a_2 x_i + e_i \tag{5.12}$$

Dr. Theo reminds us that we need to obtain $\text{var}(e_i) = \sigma^2$. If $\text{var}(e_i) = \sigma_i^2$ (note the subscript *i*), the dispersion changes when the identity changes, and we have a heteroskedasticity problem.

For example, if $\sigma_i^2 = \sigma^2 \sqrt{x_i}$, where $\sigma^2 = 150$, and the variable x_i changes from \$100 to \$144, then $\sigma_1^2 = \sigma^2 \sqrt{x_1} = 150 * 10 = 1500$, but $\sigma_2^2 = \sigma^2 \sqrt{x_2} = 150 * 12 = 1800$. Thus, the variance is no longer a constant. The same problem could occur with time-series data.

To this point, Fin says, "Oh yes, suppose we want to perform a regression of food expenditures on food prices, then the variance of the errors might change if it depends on the prices."

Dr. Theo praises him and says that to detect heteroskedasticity, a Lagrange Multiplier (LM) test is usually performed (testing a variance function). Theoretically, given the model Equation 5.12, we hypothesize that the variance is a function of a variable *w*:

$$\text{var}(e_i) = \sigma_i^2 = E(e_i^2) = f(c_1 + c_2 w)$$

If only c_1 is significant, then c_2 is zero, and var(*e*) is a constant, so the hypotheses are:

$$H_0\text{: } c_2 = 0;\ H_a\text{: } c_2 \neq 0$$

There are several LM tests. We are only required to learn the White version, which let $w = x$ and uses the chi-squared distribution, $\chi^2_{(K-1)}$, where K is the number of estimated coefficients (parameters) and $(K - 1)$ is the degree of freedom (df). To perform this LM test, we need to:

Estimate the original equation: $y_i = a_1 + a_2 x_i + e_i$

Obtain $\hat{e}_i$ and generate $\hat{e}_i^2$ then estimate the variance function:

$$\hat{e}_i^2 = c_1 + c_2 x_i + v_i \tag{5.13}$$

Using the same four-step procedure for any test:

i. H_0: $c_2 = 0$; H_a: $c_2 \neq 0$
ii. Calculate $\text{LM}_{\text{STAT}} = N * R^2$

iii. Find $\chi_c^2 = \chi_{(K-1)}^2$ using either a chi-square distribution table or Excel
iv. If $LM_{STAT} > \chi_c^2$, we reject the null hypothesis, meaning c_2 is different from zero and implying that the heteroskedasticity exists.

For example, estimating $\hat{e}_i^2 = c_1 + c_2 x_i + v_i$ yields $R^2 = 0.32$, $N = 40$, so $LM_{STAT} = 40 * 0.32 = 12.8$. The 5 percent critical value of $\chi_{(2-1)}^2$ is 3.84. Hence, we reject the null, meaning that c_2 is different from zero and implying that the data has a heteroskedasticity problem.

We learn that to find critical values of chi-squared distribution in Excel, we need to type in any Excel cell = CHIINV(α, df). For example, type = CHIINV(0.05,1) and press Enter. This yields 3.8415 ≈ 3.84.

Autocorrelation

This problem is also called serial correlation and occurs when the classic assumption (iv) for time series, $\text{Cov}(e_t, e_z) = 0$ for $t \neq z$, is violated. The common form of autocorrelation is:

$$y_t = a_1 + a_2 x_{t-1} + e_t \text{ and } e_t = r e_{t-1} + v_t \text{, so:} \tag{5.14}$$

$$y_t = a_1 + a_2 x_{t-1} + r e_{t-1} + v_t \tag{5.15}$$

For the LM test on autocorrelation, we use $\hat{e}_{t-1}$ and $\hat{v}_t$ from Equation (5.14):

$$y_t = a_1 + a_2 x_{t-1} + r\hat{e}_{t-1} + \hat{v}_t \text{, but}$$

$$y_t = \hat{a}_1 + \hat{a}_2 x_{t-1} + \hat{e}_t \text{ so}$$

$$\hat{a}_1 + \hat{a}_2 x_{t-1} + \hat{e}_t = a_1 + a_2 x_{t-1} + r\hat{e}_{t-1} + \hat{v}_t$$

$$\hat{e}_t = (a_1 - \hat{a}_1) + (a_2 - \hat{a}_2) x_{t-1} + r\hat{e}_{t-1} + \hat{v}_t$$

$$\hat{e}_t = c_1 + c_2 x_{t-1} + r\hat{e}_{t-1} + \hat{v}_t, \tag{5.16}$$

where

$$c_1 = a_1 - \hat{a}_1 \text{, and } c_2 = a_2 - \hat{a}_2$$

Hence, to perform the LM test, we need to:

Estimate the original equation: $y_t = a_1 + a_2 x_{t-1} + e_t$

Obtain $\hat{e}_t$ and generate $\hat{e}_{t-1}$

Estimate Equation 5.16 and obtain R^2 for the four-step LM test:

i. H_0: $r = 0$; H_a: $r \neq 0$
ii. Calculate $LM_{STAT} = T * R^2$
iii. Find $\chi_c^2 = \chi_{(J)}^2$ using either a chi-square distribution table or Excel. J is the degree of freedom and equals the number of restrictions, in this case $J = 1$ because we only have to test e_{t-1}.
iv. If $LM_{STAT} > \chi_c^2$, we reject the null hypothesis, meaning r is different from zero and implying that autocorrelation exists.

Next, we work on an example: Suppose estimating model Equation 5.16 yields $R^2 = 0.26$, and the sample size $T = 35$, then $LM_{STAT} = 35 * 0.26 = 9.1$. In this case, $J = 1$, and $\chi_{(1)}^2 = 3.84$. Since $LM_{STAT} > \chi_{(1)}^2$, we reject the null, meaning $r \neq 0$ and implying that the autocorrelation exists.

Adjustments

An Insignificant Coefficient

If a t-test reveals that a_2 is not significantly different from zero, then x does not help predict y, so the original model needs adjustments. First, a new independent variable similar to the original might be chosen. For example, if you regress labor productivity on high-school enrollments and find that a_2 is statistically insignificant, you might want to replace the dataset on enrollments with a dataset on labor force with high-school education.

Second, you can transform the original model to a new model by changing x to x', for example, $x' = x^2$, or $x' = \ln(x)$, or also change y to $y' = \ln(y)$. When both y and x are changed to logarithmic form, the model is called a log–log model. The advantage of using the log–log model is threefold. First, it allows the use of a linear regression technique on a nonlinear model, which might help in predicting the nonlinear trend of the series. Second, it reduces the volatility of the data, especially in a time series.

Finally, a_2 becomes the elasticity of y with respect to x. Elasticity is a useful forecast approach in business as discussed in Froeb and McCann (2010).

To this point, Alte offers an example of using elasticity in forecasting: Her Alcorner store is located next to a hair-style service on a university campus. She heard that there will be a 1 percent increase in faculty salary this year. She has read this chapter and decided to estimate a model relating logarithm of spending on hair-style service (HAIR) to logarithm of income (INCOME). Her regression results are:

$$\ln(\text{HAIR})_t = 0.2 + 0.2 \ln(\text{INCOME})_{t-1}$$

Thus, she was able to predict that this 1 percent rise in income will increase spending on hair styling by 0.2 percent. She advised the hair-styling owner to stock up his inventory, and he was very grateful.

Forecasting with Heteroskedasticity

We learn that the generalized least squares (GLS) instead of OLS estimators are often used to correct for heteroskedasticity. There are several approaches to GLS estimations. The simplest one is to divide both sides of Equation 5.12 by $\sqrt{x_i}$:

$$y_i' = \frac{y_i}{\sqrt{x_i}}; \; x_{i1}' = \frac{1}{\sqrt{x_i}}; \; x_{i2}' = \frac{x_i}{\sqrt{x_i}}; \; e_i' = \frac{e_i}{\sqrt{x_i}}. \qquad (5.17)$$

Regress the transformed equation:

$$y_i' = a_1 x_{i1}' + a_2 x_{i2}' + e_i' \qquad (5.18)$$

Note that the model no longer has a constant because x_{i1}' changes with each observation. The problem is solved because $\text{var}(e_i') = \text{var}\left(\frac{e_i}{\sqrt{x_i}}\right) = \frac{1}{x_i}\text{var}(e_i) = \frac{1}{x_i}\sigma^2 x_i = \sigma^2$.

The predicted values are for y', so we need to multiply $\hat{y}'$ by $\sqrt{x_i}$ to obtain $\hat{y}$ for forecasts. The interval forecasts then can be calculated as usual. Dr. Theo advises us that sometimes transforming the model by $\sqrt{x_i}$

alone is not enough. In that case, we can transform the model once more by dividing both sides of the model by $\hat{\sigma}$ so that

$$\text{var}(e_i') = \text{var}\left(\frac{e_i}{\sigma}\right) = \frac{1}{\sigma^2}\text{var}(e_i) = \frac{1}{\sigma^2}\sigma^2 = 1 \text{ (a constant).}$$

Forecasting with Autocorrelation

We also learn that we should use GLS estimators to obtain corrected coefficient estimates when an autocorrelation problem exists. Given:

$$y_t = a_1 + a_2 x_{t-1} + e_t \quad \text{and } e_t = r e_{t-1} + v_t \text{, so:}$$

$$y_t = a_1 + a_2 x_t + r e_{t-1} + v_t$$

Retrogress one period: $y_{t-1} = a_1 + a_2 x_{t-1} + e_{t-1}$, so: $e_{t-1} = y_{t-1} - a_1 - a_2 x_{t-1}$

$$y_t = a_1 + a_2 x_t + r (y_{t-1} - a_1 - a_2 x_{t-1}) + v_t$$

$$= a_1 + a_2 x_t + r y_{t-1} - ra_1 - ra_2 x_{t-1} + v_t$$

$$y_t = a_1 (1 - r) + a_2 x_t + r y_{t-1} - a_2 r x_{t-1} + v_t$$

$$y_t - r y_{t-1} = a_1 (1 - r) + a_2 (x_t - r x_{t-1}) + v_t.$$

Hence,

$$y_t' = a_1 x_{t1}' + a_2 x_{t2}' + v_t ,$$

where

$$y_t' = y_t - r y_{t-1},\ x_{t1}' = 1 - r,\ x_{t2}' = x_t - rx_{t-1} \tag{5.19}$$

Note again that the model no longer has a constant because x_{t1}' changes over time. Estimating model Equation 5.19 will yield a BLUE estimator because the troublesome error, *e*, is substituted. Empirically, *r* is nonexistent, and we need the estimated autocorrelation coefficient of the errors $\hat{r}_k$:

$$\hat{r}_k = (\sum_{t=k+1}^{T} \hat{e}_t \hat{e}_{t-k}) / \sum_{t=1}^{T} \hat{e}_t^2; \text{ for } k = 1,\ \hat{r}_1 = (\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}) / \sum_{t=1}^{T} \hat{e}_t^2 \tag{5.20}$$

The predicted values are for $y'_t = y_t - r\, y_{t-1}$, so we need to calculate $\hat{y}_t = \hat{y}_t + \hat{r} y_{t-1}$, where y_{t-1} is the actual value of y at period $(t - 1)$. The forecast values are then calculated and the interval forecasts can be obtained as usual.

Other Measures

p-Value

A probability value is abbreviated as a *p*-value and indicates the probability that a random variable falls into the rejection region. For example, a *p*-value = 0.02 implies that we reject the null at a 2 percent significance level (between 1 percent and 5 percent significance levels).

We find that most econometric software, including Excel, report *p*-values. This is great news because we can look at *p*-values of the estimated coefficients and avoid looking at the *t*-table or calculating *t*-critical values. We learn that we can reject the null if the *p*-value $\leq \alpha$. For example, if we choose $\alpha = 0.05$, then we reject the null if the *p*-value ≤ 0.05. These are commonly used values for a test of significance:

If the *p*-value ≤ 0.01:	the estimated coefficient is highly significant
If $0.01 <$ the *p*-value ≤ 0.05:	the estimated coefficient is significant
If $0.05 <$ the *p*-value ≤ 0.10:	the estimated coefficient is weakly significant
If the *p*-value > 0.10:	the estimated coefficient is insignificant

Dr. Theo tells us to look at Figure 5.2 so that we can see how easy it is to interpret *p*-values: The estimated coefficient of the intercept is weakly significant (with *p*-value = 0.08) whereas that of INCOME is highly significant (with *p*-value = $7.25 * 10^{-32}$). In Figure 5.3, the estimated coefficient of the intercept is insignificant (with *p*-value = 0.53) whereas that of EXCHA_{t-1} is significant (with *p*-value = 0.035).

R-squared (R^2)

An R^2 value measures the model goodness-of-fit. Given an estimation equation, $y = b_1 + b_2 x + e$, we want to know how much the variation in *y* can be explained by the variation in x. In the section "Basic Concept" of this chapter, we have:

$$y_i = \hat{y}_i + \hat{e}_i \rightarrow y_i - \overline{y} = \hat{y}_i - \overline{y} + \hat{e}_i$$

Hence, we can write $\sum (y_i - \overline{y})^2 == \sum (\hat{y}_i - \overline{y})^2 + \sum \hat{e}_i^2$, (5.21)

where

$\sum (y_i - \overline{y})^2$ = the total sum of squares (SST)

$\sum (\hat{y}_i - \overline{y})^2$ = the sum of squares of the regression (SSR)

$\sum \hat{e}_i^2$ = the sum of squared errors (SSE)

R^2 is the coefficient of determination and is defined as:

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{5.22}$$

A model is a perfect fit if $R^2 = 1$. In reality, $0 < R^2 < 1$ and is reported by all econometric packages. For example, in Figure 5.2, $R^2 = 0.94$, which is a very good fit and implies that 94 percent of the variation in residential-property investment can be explained by personal income. In Figure 5.3, $R^2 = 0.14$, which is not a very good fit and implies that only 14 percent of the variation in exports from the United States to China is explained by the exchange rate between the two countries.

Dr. Theo points out that Excel reports all three measures, SSR, SSE, and SST, in its summary output, with the error called as *Residual* in Excel. For example, in Figure 5.3, they are in cells D10, D11, and D12, respectively. We are happy to hear that R-squared is always reported by all econometric packages, so we will not have to calculate it.

Standard Errors

Dr. Theo reminds us that two standard errors are introduced in the section "Predictions" of this chapter. The first is the standard error of the forecast, $se(f)$, and the second is the standard error of the regression, (s). Holding other factors constant, the smaller these values become, the better it is for the fitness of the regression and the forecasts. The standard error of the regression is conveniently reported in most econometric software. The $se(f)$ can be calculated as discussed in the section "Excel Applications" under "Predictions and Forecasts."

To conclude the section, Dr. Theo tells us that several measures in Chapter 3, such as MAE and RMSE, can be used to evaluate models in Chapters 5 and 6 as well.

Excel Applications

Dr. App has just returned from her conference and will work with us in this section.

Testing Heteroskedasticity

The yearly data on per capita income and consumption are from the U.S. Bureau of Economic Analysis and are available in the file Ch05.xls, Fig.5.5. First, we regress CONS on INCOME:

Click on Data and then Data Analysis on the Ribbon
Select Regression and click OK
In the Input Y Range box, enter B1:B34
In the Input X Range box, enter A1:A34
Choose Labels and Residuals
Check the Output Range button and enter F1
Click OK and then OK again to override the data range
Copy and paste the residuals (e) from cells H24 through H57 into cells C1 through C34
Generate e-squared (e^2) by typing = C2(^2) into cell D2, then press Enter
Copy and paste this formula into cells D3 through D34

Next, we need to regress e^2 on INCOME:

Click on Data and then Data Analysis on the Ribbon
Select Regression and click OK
In the Input Y Range box, enter D1:D34
In the Input X Range box, enter A1:A34

	A	B	C	D	E	F	G	H
1		*Regression Statistics*						
2		R Square	0.438897401					
3		Observations	33					
4		ANOVA						
5			*df*	*SS*	*MS*	*F*	*ignificance F*	
6		Regression	1	1.92913E+13	1.93E+13	24.24836	2.67E-05	

Sheet5 / Sheet6 / Sheet7 / Sheet8 / Sheet9

Figure 5.4 Sections of regression results for heteroskedasticity test

Choose Labels
Check the Output Range button and enter P1
Click OK and then OK again to override the data range

Figure 5.4 shows a section of the second regression with the number of observations and R^2.

From this figure, $N = 33$ and $R^2 = 0.4389$, so $LM_{STAT} = 33 * 0.4389 = 14.48$.

Typing =CHIINV(0.05,1) into any cell gives you $\chi_c^2 = \chi_{(2-1)}^2 = 3.84$.

Since $LM_{STAT} > \chi_c^2$, we reject the null hypothesis.

Testing Autocorrelation

The same yearly data on per capita income and consumption are used for this test and are available in the file Ch5.xls, Fig.5.6.

First, we regress CONS on INCOME:

Click on Data and then Data Analysis on the Ribbon
Select Regression and click OK
In the Input Y Range box, enter B1:B34
In the Input X Range box, enter A1:A34
Choose Labels and Residuals
Check the Output Range button and enter G1
Click OK and then OK again to override the data range
Copy and paste the residuals (*e*) from cells I24 through I57 into cells C1 through C34
Generate $e(t-1)$ by copy and paste cells C2 through C34 into cells D3 through D35
Copy and paste INCOME in column A into column E

	A	B	C	D	E	F	G	H
1		*Regression Statistics*						
2		R Square	0.717453501					
3		Observations	32					
4			*df*	*SS*	*MS*	*F*	*ignificance F*	
5		Regression	2	20180025.12	10090013	36.81899	1.1E-08	

Sheet5 / Sheet6 / Sheet7 / Sheet8 / Sheet9

Figure 5.5 Sections of regression results for autocorrelation test

Next, we regress the Residuals (*e*) on *e*(*t*–1) and INCOME:

Click on Data and then Data Analysis on the Ribbon
Select Regression and click OK
For Input Y Range, enter B3:B34
For Input X Range, enter D3:E34
Uncheck the box Labels, that is, do not use Labels
Check the Output Range button and enter Q1
Click OK and then OK again to override the data range

Figure 5.5 shows a section of this second regression with the number of observations and R^2.

From this figure, $T = 32$ and $R^2 = 0.7175$, so $LM_{STAT} = 32 * 0.7175 = 22.96$.

Typing =CHIINV(0.05,1) into any cell gives you $\chi_c^2 = \chi_{(1)}^2 = 3.84$.

Since $LM_{STAT} > \chi_c^2$, we reject the null hypothesis.

Forecasting with Heteroskedasticity

The yearly data on per capita income and consumption are used again in this demonstration and are available in the file Ch05.xls, Fig.5.7. We learn to perform the following steps:

In cell C2, type = A2^(1/2), then press Enter (this is $INCOME^{1/2}$)
Copy and paste the formula into cells C3 through C34
In cell D2, type = B2/C2 (this is CONS′)
Copy and paste the formula into cells D3 through D34
In cell E2, type = 1/C2 (this is X1′)
Copy and paste the formula into cells E3 through E34
In cell F2, type = A2/C2 (this is INCOME′)
Copy and paste the formula into cells F3 through F34

Next, we need to regress CONS′ on X1′ and INCOME′:

Go to Data Analysis and choose Regression
In the Input Y Range box, enter D1:D34
In the Input X Range box, enter E1:F34
Check the box Labels, Constant is Zero, and Residuals
(Note: Make sure that you check the box Constant is Zero, because the model no longer has a constant.)
Check the Output Range button and enter J1
Click OK and then OK again to obtain the regression results

The regression results are displayed in Figure 5.6. From this figure, the estimated equation is:

$$\text{CONS}'_t = 1381\ \text{X1}'_t + 0.044\ \text{INCOME}'_t \qquad (5.23)$$

Because CONS′ = CONS/(INCOME)$^{1/2}$, to obtain predicted values of CONS:

Copy and paste the values in cells K25 through K58 into cells G1 through G34
In cell H2, type = G2 * C2, then press Enter (this is CONS = [CONS′] * [INCOME]$^{1/2}$)
Copy and paste the formula into cells H3 through H34

	A	B	C	D	E	F	G	H
1		SUMMARY OUTPUT						
2		*Regression Statistics*						
3		Multiple R	0.940892994					
4		R Square	0.885279625					
5		Adjusted R Square	0.849320904					
6		Standard Error	5.956818656					
7		Observations	33					
8		ANOVA						
9			*df*	*SS*	*MS*	*F*	*Significance F*	
10		Regression	2	8488.488502	4244.244	119.6111	5.0717E-15	
11		Residual	31	1099.994343	35.48369			
12		Total	33	9588.482845				
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
14		Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
15		X1*	1381.06374	571.3444667	2.417217	0.021714	215.799018	2546.328463
16		INCOME*	0.043951488	0.013539593	3.246145	0.002806	0.01633731	0.071565671

Sheet8 Sheet9 Sheet10 Sheet11 Sheet1 She

Figure 5.6 Forecasting with heteroskedasticity: regression results

We learn that once the predicted values are obtained, we can calculate interval forecasts as usual.

Forecasting with Autocorrelation

The same yearly data for the section "Testing Autocorrelation" are used here and are available in the file Ch05.xls, Fig.5.8. The residuals *e* in column C and their lagged values $e(t-1)$ in cells D3 through D35 are from Figure 5.5. The commands for this section will continue from that step.

In cell E3, type = C3^2 and press Enter (this is e^2)
Copy and paste the formula into cells E4 through E34
In cell F3, type = C3 * D3 and press Enter (this is $e_t * e_{t-1}$)
Copy and paste the formula into cells F4 through F34
In cell E35, type = SUM(E3:E34) and press Enter (this is the sum of e^2)
Copy and paste the formula into cells F35 (this is the sum of $e_t * e_{t-1}$)
In cell G3, type = F35/E35, then press Enter (this is $\hat{r}_1$, called *r*-hat in the Excel file)
Copy and paste-special the value in cell G3 into cells G4 through G34
Copy and paste the values in cells B2 through B34 into cells H3 through H35
(this is the lagged values of CONS)
Copy and paste the values in cells A2 through A34 into cells I3 through I35
(this is the lagged values of INCOME)
In cell J3, type = B3 – (G3 * H3), then press Enter (this is CONS′)
Copy and paste the formula into cells J4 through J34
In cell K3, type = 1 – G3 and press Enter (this is X1′)
Copy and paste the formula into cells K4 through K34
In cell L3, type = A3 – (G3 * I3), then press Enter (this is INCOME′)
Copy and paste the formula into cells L4 through L34

Next, we need to regress CONS′ on X1′ and INCOME′:

Go to Data Analysis and choose Regression and click OK
In the Input Y Range box, enter J2:J34
In the Input X Range box, enter K2:L34

	A	B	C	D	E	F	G	H
1		SUMMARY OUTPUT						
2		*Regression Statistics*						
3		Multiple R	0.965485704					
4		R Square	0.932162644					
5		Adjusted R Square	0.896568066					
6		Standard Error	290.265203					
7		Observations	32					
8		ANOVA						
9			*df*	*SS*	*MS*	*F*	*Significance F*	
10		Regression	2	34732335.75	17366168	206.1171	7.1956E-18	
11		Residual	30	2527616.642	84253.89			
12		Total	32	37259952.39				
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
14		Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
15		X1*	3736.199152	234.672736	15.92089	3.56E-16	3256.93349	4215.46482
16		INCOME*	0.001238651	0.003000395	0.412829	0.682667	-0.00488897	0.00736627

Sheet10 / Sheet1 / Sheet2 / Sheet3

Figure 5.7 Forecasting with autocorrelation: regression results

Check the boxes Labels, Constant is Zero, and Residuals

Check the Output Range button and enter P1

Click OK and then OK again to obtain the regression results displayed in Figure 5.7.

From this figure, the estimated equation is:

$CONS'_t = 3736\ X1'_{t-1} + 0.0012\ INCOME'_{t-1}$

Because $CONS'_t = CONS_t - r\ CON_{t-1}$, to obtain predicted values of $CONS_t$:

Copy and paste the values in cells Q25 through Q57 into cells M2 through M34

In cell N3, type = M3 + G3 * H3, then press Enter (this is $CONS_t = CONS'_t + r\ CON_{t-1}$)

Copy and paste the formula into cells N4 through N34.

We learn that once the predicted values are obtained, we can calculate point and interval forecasts as usual.

Exercises

1. Data on employment (EMP) and residential investment (INV) for 50 states and Washington, DC, in 2012 are in the file Emp.xls.
   a. Regress INV on EMP (i.e., y = INV, x = EMP) and provide comments on the results, including the significances of a_1 and a_2, R^2, and the standard error of regression.

b. Write the estimated equation and find the point prediction for investment.

2. Data on the CPI growth ($CPIG_t$) and unemployment rates ($UNEMP_{t-1}$) from the end of January 2012 through the end of February 2014 are in the file Unemp.xls.

a. Regress $CPIG_t$ on $UNEMP_{t-1}$ (y = CPIG, x = UNEMP) and provide comments on the results, including the significances of $\hat{a}_1$ and $\hat{a}_2$, R^2, and the standard error of regression.

b. Write the estimated equation and calculate the one-period point forecast for CPIG (at the end of March 2014) using either a handheld calculator or Excel, showing every step of your calculations.

3. Use the results in Exercise 2 to calculate the one-period interval forecast for CPIG (at the end of March 2014) at a 95 percent confidence interval using either a handheld calculator or Excel, showing every step of your calculations.

4. Use the results in Exercise 2 to test the following hypotheses at a 1 percent significance level:

a. The slope is −3 against the alternative hypothesis that the slope is smaller than −3.

b. The slope is zero against the alternative hypothesis that the slope is different from zero.

Write the testing procedure in four standard steps similar to those in the section "Evaluations and Adjustments" of this chapter. The calculations of the *t*-statistics might be performed using a handheld calculator or Excel.

PART III

Advanced Forecast Techniques

This part contains two chapters:

- Chapter 6 Multiple Linear Regressions
- Chapter 7 Advanced Time Series Techniques

CHAPTER 6

Multiple Linear Regressions

Having learned the concept of econometric forecasting, Rea is looking forward to the multiple linear regression technique so that he can show the determinants of home prices to his customers at Realmart. Fin also tells us that the U.S. stock market has performed very well since 2010, and he wonders how this factor affects consumer spending. Ex then says that incomes are rising worldwide thanks to the recoveries from the global recession and adds that he wishes to see how they will affect the U.S. exports. Dr. Theo tells us that all these issues will be discussed this week and that once we finish with the chapter, we will be able to:

1. Develop models for multiple linear regressions and discuss conditions for using them.
2. Discuss the econometric forecasting approach using multiple linear regressions.
3. Analyze numerous methods of evaluations and adjustments.
4. Describe and address common problems in panel-data forecasting.
5. Perform regressions and obtain forecasts using Excel.

We learn that this chapter will involve two or more explanatory variables.

Basic Concept

In business and economics, we often see more than one factor affecting the movement of a market. Hence, a new model needs to be introduced.

Econometric Model

An economic model with more than one determinant of consumption (CONS) will look like this:

$$\text{CONS} = a_1 + a_2\,\text{INCOME} + a_3\,\text{STOCKP} \tag{6.1}$$

where CONS and INCOME are the same as in Chapter 5, and STOCKP is the average stock price, which might affect consumer wealth and subsequent spending. The interpretation of a_1 and a_2 is the same as in Chapter 5, and a_3 represents the change in consumption due to one unit change in the average stock price.

Converting the economic model in Equation 6.1 into an econometric model yields:

$$\text{CONS}_i = a_1 + a_2\,\text{INCOME}_i + a_3\,\text{STOCKP}_i + e_i \tag{6.2}$$

The generalized version of this econometric model for cross-sectional regressions is:

$$y_i = a_1 + a_2\,x_{i2} + \ldots + a_k x_{ik} + e_i \tag{6.3}$$

where y is the dependent variable, and xs are usually called explanatory variables or the regressors instead of independent variables, because the presence of more than one x implies that the xs might not be completely independent of each other.

The six classic assumptions in multiple linear regressions for cross-sectional data are as follows:

i. The model is $y_i = a_1 + a_2\,x_{i2} + \ldots + a_k x_{ik} + e_i$
ii. $E(e_i) = E(y_i) = 0$
iii. $\text{Var}(e_i) = \text{Var}(y_i) = \sigma^2$
iv. $\text{Cov}(e_i, e_j) = \text{Cov}(y_i, y_j) = 0$ for $i \neq j$
v. Each x_{ik} is not random, must take at least two different values, and xs are not perfectly correlated to each other (called the multicollinearity problem)
vi. $e_i \sim N(0, \sigma^2)$; $y_i \sim ([a_1 + a_2\,x_{i2} + \ldots + a_k x_{ik}], \sigma^2)$

Dr. Theo reminds us, "Assumption (v) only requires that xs are not 100 percent correlated to each other. In practice, any correlation of less than 90 percent can be acceptable depending on the impact of the correlation on a specific problem."

Regarding time-series data, assumption (v) changes to:

a. y and xs are stationary random variables, and e_t is independent of current, past, and future values of xs.
b. when some of the xs are lagged values of y, e_t is uncorrelated to all xs and their past values.

If assumptions (i) through (v) hold, then the multiple linear regressions will produce the best linear unbiased estimators (BLUE) using the ordinary least squares (OLS) technique. If assumption (vi) holds, then test results are valid. The central limit theorem concerning the distribution of the errors and the OLS estimators continues to apply.

Estimations

The estimated version of Equation 6.3 on a particular sample is:

$$\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_{i2} + \ldots + \hat{a}_k x_{ik}, \quad y_i = \hat{y}_i + \hat{e}_i = \hat{a}_1 + \hat{a}_2 x_{i2} + \ldots + \hat{a}_k x_{ik} + \hat{e}_i \tag{6.4}$$

Point Estimates:

Suppose that estimating Equation 6.2 yields the following results:

$$\text{CONS} = 1.5 + 0.5 * \text{INCOME} + 0.02 \text{ STOCKP} \tag{6.5}$$

where the units continue to be in hundreds of dollars, then the results imply the following:

i. Weekly consumption of a person with no income is \$150.
ii. Holding the stock price constant, \$100 increase in weekly income raises weekly consumption by \$50.
iii. Holding the personal income constant, \$100 increase in the average stock price increases weekly consumption by \$2 (= 0.02 * 100).

Interval Estimates:

We learn that the equation for interval estimate is the same as Equation 5.8, which is rewritten here:

$$P[\hat{a}_k - t_c * se(\hat{a}_k) \leq a_k \leq \hat{a}_k + t_c * se(\hat{a}_k)] = 1 - \alpha \qquad (6.6)$$

The only change is that t-critical bears $(N - K)$ degrees of freedom instead of $(N - 2)$ because we have K parameters (coefficients) to be estimated in multiple regressions.

Predictions and Forecasts

Predictions

We then work on an example by substituting values of personal income and price level into Equation 6.5 and find that a person with a weekly income of $400 when the average stock price is $1,000 can expect a weekly consumption of:

$$\text{CONS}_i = 1.5 + 0.5 * 4 + 0.02 * 10 = 3.7 \text{ (\$ hundreds)} = \$370 \qquad (6.7)$$

Dr. Theo says that interval predictions can also be made for multiple regressions with similar formulas as those in Equation 5.9, except for the estimated variance of the error terms:

$$se(f) = s\sqrt{1 + \frac{1}{N} + \frac{(x_1 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}; \quad s = \sqrt{\frac{\sum \hat{e}_i^2}{N - K}} = \sqrt{\frac{\text{SSE}}{N - K}} \qquad (6.8)$$

The interval prediction is also calculated in a similar manner as in Equation 5.10:

$$P[\hat{y}_1 - t_c se(f) \leq y_1 \leq \hat{y}_1 + t_c se(f)] = 1 - \alpha \qquad (6.9)$$

In the consumption example in Equation 6.7, if $N = 33$ so that the df = 33 − 3 = 30, $se(f)$ = 0.1, then the 95 percent confidence interval prediction is

$$3.7 \pm 2.042 * 0.1 = (3.4958; 3.9042) = (\$349.58; \$390.42)$$

Hence, we predict with 95 percent confidence that a person with a weekly income of $400 will spend between $349.58 and $390.42 weekly when the stock price is added to the econometric model.

Forecasts

For the multiple linear regressions using time series data, the econometric model with one-period lag is written as:

$$y_t = a_1 + a_2 x_{t-1,2} + \ldots + a_k x_{t-1,k} + e_t \tag{6.10}$$

For example, the forecast for period $(t + 1)$ is:

$$\text{CONS}_{t+1} = 1.5 + 0.5 * \text{INCOME}_t + 0.02 * \text{STOCKP}_t$$

Multiple linear regressions also have great advantage over simple regressions when more than one lag is involved because multiperiod forecasts can be obtained without the need of using MA or exponential smoothing (ES) techniques. For example, if we have:

$$\text{CONS}_{t+1} = 5 + 0.4 * \text{INCOME}_t + 0.02 * \text{STOCKP}_t + 0.1 * \text{INCOME}_{t-1} + 0.01\ \text{STOCKP}_{t-1}$$

then two-period forecasts can be obtained, and this model is called a distributed lagged (DL) model. Another model that includes lagged-dependent variables in addition to other explanatory variables is called autoregressive distributed lagged (ARDL) model. For example:

$$\text{CONS}_{t+1} = 100 + 0.4 * \text{INCOME}_t + 0.2 * \text{CONS}_t$$

The ARDL model allows long-term forecasts thanks to the lagged-dependent variable (hence the name *autoregressive*). Assuming that income remains constant over time, once CONS_{t+1} is obtained, substitute CONS_{t+1} into the model and continue the next substitutions to find:

$$\text{CONS}_{t+2} = 100 + 0.4 * \text{INCOME}_t + 0.2 * \text{CONS}_{t+1}$$

$$\text{CONS}_{t+3} = 100 + 0.4 * \text{INCOME}_t + 0.2 * \text{CONS}_{t+2}$$

Suppose $INCOME_t$ = \$4000 monthly and $CONS_t$ = \$2000 for the first month, then:

$$CONS_{t+1} = 100 + 0.4 * 4000 + 0.2 * 2000 = 100 + 1600 + 400 = 2100$$

$$CONS_{t+2} = 100 + 0.4 * 4000 + 0.2 * 2100 = 100 + 1600 + 420 = 2120$$

$$CONS_{t+3} = 100 + 0.4 * 4000 + 0.2 * 2120 = 100 + 1600 + 424 = 2124$$

This process can be extended far into the future. It is called *recursion by the law of iterated projections* (henceforth called the recursive principle) and is proved formally in Hamilton (1994). Since income is not changing every month, the model is quite realistic and convenient for long-term forecasts.

Dr. Theo then reminds us that one-period interval forecasts can be calculated using Equation 6.8, which is adapted for time series data with $N = T$. He also says that the simplest approximation to Equation 6.8 for multiple periods ahead is:

$$se(f)_{t+m} \approx s\sqrt{1 + \frac{1}{T} + \frac{(x_{t+m} - \bar{x})^2}{\sum (x_t - \bar{x})^2}}; \quad s \approx \sqrt{\frac{\sum \hat{e}_i^2}{T - K - m}} = \sqrt{\frac{\text{SSE}}{T - K - m}} \tag{6.11}$$

We learn that this is only one of the several ways to approximate the multiperiod interval forecasts. More discussions of this topic will be presented in Chapter 7.

Excel Applications

Dr. App just returned from the conference and is very happy to see us again. She says that the procedures for performing multiple linear regressions are very similar to those for simple linear regressions except for the correlation analysis.

Cross-Sectional Data

The dataset is again offered by Rea and is available in the file Ch06.xls, Fig.6.1 and Fig.6.2. Data on financial aid by local governments (LOCAID) are added to the original data on personal income (INCOME) and residential-property investment (INV) in Chapter 5. The units are in thousands of dollars. The dependent variable is INV, and the two explanatory variables are LOCAID and INCOME. First, we carry out a correlation analysis:

Go to Data then Data Analysis on the Ribbon
Select Correlation instead of Regression and click OK
A dialog box will appear as shown in Figure 6.1

In the Input Range, enter C1:D52
Check the box Labels in the First Row
Check the button Output Range and enter O1 then click OK
A dialogue box will appear, click OK to overwrite the data

The result reveals that the correlation coefficient between INCOME and LOCAID is 0.8981, which is quite high but acceptable to perform a regression (in the data file, you can find this correlation coefficient in cell P3. We copy and paste it into cell G3 in Figure 6.2).

Next, we perform a regression of INV on LOCAID and INCOME:

Go to Data then Data Analysis, select Regression then click OK
The input Y range is B1:B52, the input X range is C1:D52
Check the boxes Labels and Residuals
Check the button Output Range and enter F1 and click OK
A dialogue box will appear, click OK to overwrite the data

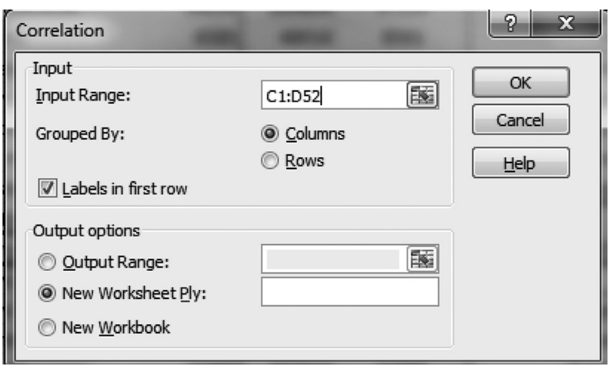


Figure 6.1 Dialog box for correlation analysis

	A	B	C	D	E	F	G	H
1		SUMMARY OUTPUT					*INCOME*	*LOCAID*
2		*Regression Statistics*				INCOME	1	
3		Multiple R	0.981390095			LOCAID	0.898097701	1
4		R Square	0.963126519					(Ctrl)
5		Adjusted R Squ	0.961590124					
6		Standard Error	8945.006001					
7		Observations	51					
8		ANOVA						
9			*df*	*SS*	*MS*	*F*	*Significance F*	
10		Regression	2	1.00316E+11	50158173675	626.8742667	3.99153E-35	
11		Residual	48	3840630353	80013132.36			
12		Total	50	1.04157E+11				
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
14		Intercept	-5141.883518	1654.818714	-3.107218619	0.00316933	-8469.119542	-1814.647495
15		INCOME	0.093909531	0.008832121	10.63272745	3.27114E-14	0.076151362	0.111667699
16		LOCAID	0.384619525	0.072508012	5.304510718	2.83677E-06	0.238832396	0.530406653

Sheet9 / Sheet10 / Sheet1 / **Sheet2** / Sheet3

Figure 6.2 Cross sectional data: multiple regression results

The regression results are given in Figure 6.2.

From these results, the estimated equation can be written as:

$$\text{INV}_i = -5141 + 0.0939\ \text{INCOME}_i + 0.3846\ \text{LOCAID}_i$$

The predicted values are next to the residuals in the data file, and interval prediction can be calculated using Equations 6.8 and 6.9.

Time Series Data

Ex again shares with us the dataset, which is available in the file Ch06.xls, Fig.6.3. Data on Real GDP (RGDP) for China are added to data on China–United States real exchange rate (EXCHA) and exports from the United States to China (EXPS) for the period 1981–2012. The new data on RGDP are from World Bank's World Development Indicators (WDI) website and are in billions of dollars, which are converted to millions of dollars before adding to the original data. The dependent variable is EXPS_t, and the two explanatory variables are RGDP_{t-1} and EXCHA_{t-1}. We first perform a correlation analysis:

Go to Data then Data Analysis, select Correlation then click OK
In the Input Range, enter C1:D33
Check the box Labels in the First Row
Check the button Output Range and enter O1 then click OK
A dialogue box will appear, click OK to overwrite the data

	A	B	C	D	E	F	G	H	I
1		SUMMARY OUTPUT					*EXCHAt-1*	*RGDPt*	
2		*Regression Statistics*				EXCHAt-1	1		
3		Multiple R	0.98612422			RGDPt	0.422599245	1	
4		R Square	0.972440977						
5		Adjusted R Squ	0.970540355						
6		Standard Error	4125.074206						
7		Observations	32						
8		ANOVA							
9			*df*	*SS*	*MS*	*F*	*Significance F*		
10		Regression	2	17412493434	8706246717	511.6434739	2.42006E-23		
11		Residual	29	493470879	17016237.21				
12		Total	31	17905964312					
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
14		Intercept	-970.0660929	2844.64137	-0.341015252	0.735550491	-6788.010944	4847.878759	
15		EXCHAt-1	-649.6924223	436.4082004	-1.48872643	0.147354591	-1542.24741	242.8625652	
16		RGDPt	0.019323617	0.000653051	29.58975467	3.23657E-23	0.017987978	0.020659256	

Sheet4 Sheet1 Sheet2 **Sheet3**

Figure 6.3 Time series data: multiple linear regression results

Data Source: IMF.com: IMF Data and Statistics (2014); World Bank.com (2014).

The results of the correlation and the estimated coefficients are displayed in Figure 6.3. The correlation between INCOME and LOCAID is 0.4226, which is quite low, so the regression results are reliable. Next, we perform a regression of $EXPS_t$ on $RGDP_{t-1}$ and $EXCHA_{t-1}$:

Go to Data then Data Analysis, select Regression and click OK
The input Y range is B1:B33, the input X range is C1:D33
Check the boxes Labels and Residuals
Check the button Output Range and enter F1 then click OK
A dialogue box will appear, click OK to overwrite the data

From the results, the estimated equation is:

$$EXPS_{t+1} = -970.0661 - 649.6924\ EXCHA_t + 0.0193\ RGDP_t$$

You can find the forecast value of $EXPS_{2013}$ in cell G57.

Again, the predicted values for EXPS are next to the residuals in the data file, and interval forecasts can be calculated using Equation 6.11.

Evaluations and Adjustments

We learn that a *t*-test can still be used to evaluate each of the estimated coefficients. Additionally, testing for the joint significance of two or more coefficients is performed using an *F*-test. For example, in Figure 6.3, the coefficient of EXCHA is no longer significantly different from zero, but EXCHA might be jointly significant with RGDP. In that case, EXCHA should not be eliminated from the model even though its coefficient is statistically insignificant.

F-*Tests*

Dr. Theo says that *F*-tests are used to evaluate the joint significance of two or more coefficients or the significance of a model.

Testing the Joint Significance

Consider an equation for multiple regressions with three explanatory variables:

$$\text{INV} = a_1 + a_2\ \text{INCOME} + a_3\ \text{LOCAID} + a_4\ \text{FED} + e \qquad (6.12)$$

where INCOME and LOCAID are the same as in the section "Predictions and Forecasts" of this chapter, and FED is federal tax credits to residential-property investment. The model in Equation 6.12 is called an unrestricted model. Suppose the regression results for Equation 6.12, with the standard errors of the corresponding coefficients in parentheses, are as follows:

$$\text{INV} = -23 + 0.086\ \text{INCOME} + 0.32\ \text{LOCAID} + 0.12\ \text{FED}$$

(*se*) (−11) (0.01) (0.04) (0.09)

Number of Observations = 51

$\text{SSE}_U = 150$

where *U* stands for unrestricted.

From these results, the coefficient of FED is not statistically significant. However, if FED and LOCAID are jointly significant, or FED and INCOME are jointly significant, then FED should not be eliminated from the model. To test the joint significance of FED and LOCAID, perform a regression on the restricted model:

$$\text{INV} = a_1 + a_2\ \text{INCOME} + e$$

Suppose the results for this restricted model are as follows:

$$\text{INV} = -12 + 0.986\ \text{INCOME} \qquad (se)\ (-5)\ (0.02)$$

$$\text{SSE}_R = 180$$

where R stands for restricted.

The test is performed in four steps as follows:

i. The hypotheses
H_0: $a_3 = a_4 = 0$; H_a: a_3 and a_4 are jointly significant.

ii. The F-statistic

$$F_{\text{STAT}} = F = \frac{(\text{SSE}_R - \text{SSE}_U)/J}{\text{SSE}_U/(N-K)} \sim F_{(J,\ N-K)} \qquad (6.13)$$

where J is the number of restrictions, in this case $J = 2$ (a_3 and a_4), and $(N - K)$ is the degrees of freedom (df), in this case the df = 47 (= 51 – 4). Hence,

$$F_{\text{STAT}} = \frac{(180-150)/2}{150/47} = \frac{15}{3.19} = 4.7$$

iii. The F-critical value, F_c, can be found from any F-distribution table. For example, if $\alpha = 0.05$ is $F_c = F_{(0.95,\ 2,\ 47)} \approx 3.195$
In Excel, to obtain F_c type = FINV(α, J, $N - K$) and then press Enter.
For example, type = FINV(0.05, 2, 47) then press Enter, this will give 3.195.

iv. Decision: Since $F_{\text{STAT}} > F_c$, we reject the null, meaning the two coefficients are jointly significant and implying that we should include FED in the regression.

Testing the Model Significance

To this point, Arti raises her hand and asks, “What if all coefficients of the explanatory variables in a model are zero? Cita also asks, “Does that mean that the model does not predict anything?” Dr. Theo praises them for raising the issue and says that in this case the model should be revised. Thus, the four-step procedure for the test is as follows:

i. H_0: All a_k are zeros for $k = 2, 3, \ldots, K$; H_a: at least one $a_k \neq 0$.
For example, in Equation 6.12, we can state that

$$H_0: a_2 = a_3 = a_4 = 0; \; \mathrm{H_a}: \text{at least one } a_k \neq 0.$$

ii. The F-statistic:

$$F_{\mathrm{STAT}} = F = \frac{(\mathrm{SST} - \mathrm{SSE})/J}{\mathrm{SSE}/(N-K)} \sim F_{(J, N-K)} \tag{6.14}$$

J is the number of restrictions. In the test for model significance, only a_1 is not included in the null hypothesis, so $J = K - 1$, and in the preceding example, $J = 4 - 1 = 3$.

iii. F-critical: This step is similar to the test for the joint significance.

iv. Decision: If $F_{\mathrm{STAT}} > F_c$, we reject the null hypothesis, meaning at least one $a_k \neq 0$ and implying that the model is significant.

Fligh then asks, “Can anyone analyze the relationship between t-tests and F-tests?” Mo volunteers to address the issue. Here is his discussion.

F-tests and t-tests are both testing for the significance or the expected values of the estimated coefficients. The comparison and contrast of F-test to t-test are as follows:

1. In F-test we have joint hypotheses.
2. For the test with $a_k \neq 0$, the t-test is a two-tail test whereas the F-test is a one-tail test.
3. F distribution has J numerator df and $(N - K)$ denominator df.
4. When $J = 1$, the t- and F-tests are equivalent.

Dr. Theo is very pleased saying that all four points are correct and that we can now move to the next section.

Error Diagnostics

Heteroskedasticity and Autocorrelation

The testing procedures on these two problems in multiple linear regressions are similar to those in simple linear regressions except that the hypotheses are stated for multiple coefficients.

Testing Heteroskedasticity

Estimate the original equation $y_i = a_1 + a_2 x_{i2} + \ldots + a_K x_{iK} + e_i$
Obtain $\hat{e}_i$ and generate $\hat{e}_i^2$ then estimate the variance function:

$$\hat{e}_i^2 = \mathrm{c}_1 + c_2 w_{i2} + \ldots + c_s w_{ik} + \mathrm{v}_i \qquad (6.15)$$

Obtain R^2 for the $LM_{\mathrm{STAT}} = N * R^2$
Using the same four-step procedure for any test, state the hypotheses as:
H_0: $c_2 = c_3 = \ldots = c_k = 0$; H_a: at least one c is not zero
The next three steps are similar to the one for simple regressions with df = $K - 1$

Testing Autocorrelation

Estimate equation $y_t = a_1 + a_2 x_{t2} + \ldots + a_k x_{tk} + e_t$
Obtain $\hat{e}_i$ and generate $\hat{e}_{t-1}, \hat{e}_{t-2}, \ldots, \hat{e}_{t-k}$
Estimate the equation

$$\hat{e}_t = c_1 + c_2 x_t + r_1 \hat{e}_{t-1} + r_2 \hat{e}_{t-2} + \ldots + r_k \hat{e}_{t-k} + \hat{v}_t \qquad (6.16)$$

Obtain R^2 for the $LM_{\mathrm{STAT}} = T * R^2$
The hypotheses are stated as

H_0: $r_1 = r_2 = \ldots = r_k = 0$; H_a: at least one $r \neq 0$

The next three steps are similar to the one for simple regressions, where df = the number of restrictions.

Testing Endogeneity

Dr. Theo says that endogeneity is also called the problem of endogenous regressors and occurs when assumption (v), which states that *x* is not random, is violated. Given the equation

$$y_i = a_1 + a_2 x_{i2} + \ldots + a_k x_{ik} + e_i \tag{6.17}$$

Suppose that x_{i2} is random then x_{i2} might be correlated with the error term. In this case the model has an endogeneity problem. The consequences of the endogeneity are serious:

i. OLS estimators are biased in small samples and do not converge to the true values even in a very large sample (the estimators are inconsistent).
ii. The standard errors are inflated, so *t*-tests and *F*-tests are invalid.

To detect endogeneity, a Hausman test must be performed. Dr. Theo says that the original Hausman test requires knowledge of matrix algebra, so he teaches us the modified Hausman test, which is discussed in Kennedy (2008).

The theoretical justification of the modified Hausman test is simple. In Equation 6.17, the exogenous variables are x_js, where $x_j = x_{i1}, x_{i3}, .., x_{ik}$. When the endogenous variable x_{i2} is regressed on these x_js, the part of x_{i2}, which is explained by x_js, will be factored out. The rest of x_{i2} is explained by the residual, v_i, from the estimation:

$$x_{i2} = b_1 + b_2 x_{i1} + b_3 x_{i3} + .. + b_k x_{ik} + v_i \tag{6.18}$$

The v_i from Equation 6.18 can be added to Equation 6.17 in the subsequent regression:

$$y_i = a_1 + a_2 x_{i2} + \ldots + a_k x_{ik} + c v_i + e_i \tag{6.19}$$

A *t*-test on the estimated coefficient of v_i is then performed. If this coefficient is not statistically different from zero, then x_i is not correlated with the error term, and the model does not have an endogeneity problem. Hence, the steps to perform the test are as follows:

i. Regress x_{i2} on all x_js and obtain the residuals $\hat{v}$.
ii. Include the residuals in the subsequent regression of Equation 6.17.
iii. Perform the *t*-test on the coefficient of $\hat{v}$ with the hypotheses: H_0: $c = 0$; H_a: $c \neq 0$.
iv. If $c \neq 0$, then the null hypothesis is rejected, meaning x_i is correlated with the error and implying that the model has an endogeneity problem.

Dr. Theo then asks the class, "Suppose adding $\hat{v}$ into Equation 6.19 for the second regression yields $\hat{c} = 2.5$ with its standard error $se(\hat{c}) = 2.5$, and $N = 32$. Can you find the *t*-statistic and make your decision?" We are able to calculate $t_{STAT} = 2.5/2.5 = 1$. Thus, c is not statistically different from zero, and we do not reject the null, meaning that x_2 is not correlated with the error and implying that the model does not have an endogeneity problem.

Adjustments

Heteroskedasticity and Autocorrelation

We learn that adjustments for these two problems are similar to the cases of simple linear regressions except that the procedures are performed on multiple explanatory variables.

Forecasting with Endogeneity

Dr. Theo then says that endogeneity problems are corrected by using the method of moments (MM). Theoretically, the purpose of the MM estimation is to find a variable w to use as a substitute for x. Theoretically, MM estimators will satisfy the condition that $\text{cov}(w, e) = 0$. Empirically, w almost satisfies this condition by minimizing $\text{cov}(w, e)$, and w is called the instrument variable (IV). The MM estimators are the IV estimators. In the following section, we continue to assume that x_{i2} is the endogenous variable.

The instrumental variable w has to satisfy two conditions:

i. w is not correlated with e, so that cov(w, e) = 0.
ii. w is strongly (or at least not weakly) correlated with x_{i2}.

In the first stage, we perform a regression of the endogenous variable x_{i2} on all exogenous variables, including the IV, which is w_i:

$$x_{i2} = b_1 + b_3 x_{i3} + b_4 w_i + \ldots + v_i \tag{6.20}$$

We then estimate the original equation with the predicted value of x_{i2}, $\hat{x}_{i2}$, in place of x_{i2}. Because of this second stage, the IV estimators are also called two stages least squares (2SLS) estimators:

$$y_i = a_1 + a_2 \hat{x}_{i2} + a_3 x_{i3} + \ldots + a_k x_{ik} + e_i \tag{6.21}$$

The problem is solved because $\hat{x}_{i2}$ does not contain v_i, $\hat{x}_{i2} = b_1 + b_3 x_{i3} + \ldots + b_4 x_{ik}$, hence $\hat{y}_i$ does not contain v either, $\hat{y}_i = a_1 + a_2 \hat{x}_{i2} + a_3 x_{i3} + \ldots + a_k x_{ik}$.

Other Measures

Goodness of Fit

Although an R^2 value is still reported by Excel, using more than one variable decreases df. Hence, an adjusted R^2 value is a better measure to account for the decreasing df:

$$\text{Adjusted } R^2 = \bar{R}^2 = 1 - \frac{\text{SSE}/(N-K)}{\text{SST}/(N-1)} \tag{6.22}$$

Model Specification

Dr. Theo reminds us that multiple regression models have more than one explanatory variable, so the issue of how many variables should be included in the model becomes important.

An omitted variable causes significant bias of the estimated coefficients. For example, if we remove FED from Equation 6.12 reasoning that FED is not statistically significant, then the estimated coefficients will be biased because FED and LOCAID are jointly significant. Hence eliminating FED will cause an omitted variable.

Including irrelevant variables will not significantly bias the estimated coefficients, but may inflate the variances of your coefficient estimates, so the tests are less reliable. For example, adding social assistance (SOCIAL) to Equation 6.12 will yield these results:

$$\text{INV} = -22 + 0.085\ \text{INCOME} + 0.33\ \text{LOCAID} + 0.12\ \text{FED} + 0.08\ \text{SOCIAL}$$

$$(se) \quad (-11)\ (0.055) \quad (0.08)\ (0.09)\ (0.07)$$

where the standard error of the coefficient estimate for INCOME is greatly inflated. In this case, the *t*-test result implies that the coefficient of INCOME is not statistically significant whereas it actually would be significant had we removed SOCIAL from Equation 6.12.

We now see that choosing a correct model is crucial. Dr. Theo says that we might want to use a piecewise-downward approach starting from all theoretically possible variables with all available data. *F*- and *t*-tests then are used to eliminate the highly insignificant variables. He says that we can also use a piecewise-upward approach, which starts from a single explanatory variable. However, the downward approach is preferable because this approach avoids the omitted variable problem that might arise if you use the piecewise-upward approach.

Excel Applications

Dr. App reminds us that the *F*-tests can be performed using a handheld calculator except for the critical value, the command of which has already been given in the section "*F*-Tests." Also, Excel applications for the heteroskedasticity and autocorrelation tests have been given in Chapter 5. Hence, she only discusses Excel applications for the endogeneity test and correction.

Data on sale values ($SALE_t$), advertisement expenditures (ADS_t), income (INC_t), and values of sale coupons ($COUP_{t-1}$) are from the file Ch06.xls, Fig.6.4 and Fig.6.5. The model is as follows:

$$SALE_t = a_1 + a_2 INC_t + a_3 ADS_t + e_t \tag{6.23}$$

However, we suspect that ADS_t is endogenous and so we prefer using $COUP_{t-1}$ as an IV for ADS_t. Since $COUP_{t-1}$ is in period $(t-1)$, it is not correlated to e_t. Since most companies include sale coupons with their advertisements and calculate the coupon values before sending out advertisements, $COUP_{t-1}$ is correlated with ADS_t. This makes $COUP_{t-1}$ a good IV for ADS_t. To perform the modified Hausman test for endogeneity, we first regress ADS on COUP and INC:

Go to Data then Data Analysis, select Regression and click OK
The input Y range is E1:E35, the input X range is C1:D35
Check the Labels box
Check the button Output Range and enter J1 then click OK
A dialogue box will appear, click OK to overwrite the data

The results are displayed in Figure 6.4.

The results support our argument that $COUP_{t-1}$ is correlated with ADS_t because the estimated coefficient of $COUP_{t-1}$ has the p-value = 0.00124 < 0.05. Additionally, by the classic assumptions, $COUP_{t-1}$ is not correlated to e_t. Therefore, $COUP_{t-1}$ satisfies both conditions to be a good IV.

Next, copy the Residuals from cells L25 through L59 and paste into cells F1 through F35

	A	B	C	D	E	F	G	H
1			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
2		Interce	2508.214957	664.9607615	3.771974382	0.000686	1152.0185	3864.411372
3		COUPt	0.380526249	0.107075729	3.553804873	0.00124	0.1621439	0.598908639
4		INCt	-0.051123988	0.015223712	-3.358181508	0.00209	-0.082173	-0.02007502

Sheet1 | Sheet2 | Sheet3 | Sheet4 | Book1

Figure 6.4 Qualification of $COUP_{t-1}$ as an instrument variable

	A	B	C	D	E	F	G	H
1			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
2		Intercept	-7454.031656	891.2925815	-8.36317031	2.46703E-09	-9274.293945	-5633.769366
3		INCt	0.223655875	0.018993221	11.77556317	8.93071E-13	0.184866542	0.262445207
4		ADSt	0.105698126	0.188271964	0.561411927	0.578686865	-0.278804521	0.490200774
5		Residuals	0.538281667	0.223354687	2.409985997	0.022292484	0.082130543	0.994432792

Sheet1 Sheet2 Sheet3 Sheet4 Book1

Figure 6.5 Modified Hausman test results

Finally, perform the regression of the original equation with the Residuals ($\hat{v}_t$) added:

$$\text{SALE}_t = a_1 + a_2\text{INC}_t + a_3\text{ADS}_t + c\hat{v}_t + e_t$$

Go to Data then Data Analysis, select Regression and click OK
The input Y range is B1:B35, the input X range is D1:F35
Check the Labels box
Check the button Output Range and enter N22
Click OK and then OK again to overwrite the data

The main results displayed in Figure 6.5, reveal that the estimated coefficient of $\hat{v}_t$, called *Residuals*, is significant with a *p*-value of 0.02 (you can find this value in cell R41 in the Excel file), so the endogeneity problem exists.

To correct this endogeneity problem:

Copy and paste Predicted ADS from cells K25 through K59 into cells G1 through G35
Second, copy INC in column D and paste INC into column H next to Predicted ADS
Finally regress the original equation with Predicted ADS in place of ADS:

$$\text{SALE}_t = a_1 + a_2\text{INC}_t + a_3\text{Predicted ADS} + e_t$$

Go to Data then Data Analysis, select Regression then click OK
The input Y range is B1:B35, the input X range is G1:H35
Check the box Labels
Check the button Output Range and enter A37
Click OK and then OK again to overwrite the data

This will correct the endogeneity problem. F-tests should be performed on the joint significance and the model significance before point and interval forecasts can be calculated.

Forecasts with Panel Data

Concept

Dr. Theo reminds us that panel data comprise cross-sectional identities over time and advises us to carry out forecasts using these data whenever they are available. Arti asks, "What do we gain from using panel data?" Alte volunteers to explain. Here is her discussion, "First, the sample size is enlarged. For example, if we have a time series dataset on TV sales in Vietnam for 16 months and another dataset on TV sales in Cambodia for the same 16 months, then combining the two datasets gives us 32 observations. Second, being able to observe more than one identity over time provides us with additional information on the characteristics of the market. For example, we can understand demand for TV in Indochina by studying their sales in Cambodia and Vietnam. Finally, we are able to carry out comparative study over time. For example, we can compare demand in Vietnam with demand in Cambodia and develop different strategies to increase sales in each country."

Dr. Theo commends her on a good analysis and says that performing forecasts on a panel dataset is simple if the two identities share the same behavior, for example, if sales in Vietnam and Cambodia share the same market characteristics, then the coefficient estimates will be the same for the two countries. In that case, we only have to stack one dataset above the other as shown in Figure 1.1, columns G through I, of Chapter 1, and perform an OLS estimation called *pooled OLS* to enjoy a dataset of 32 observations.

Most of the time, the two markets do not share the same behavior. In this case, the classic OLS estimators using all 32 observations are biased, and the diagnostic tests are invalid. Hence, panel-data techniques are needed. One way to write the forecast equation is:

$$y_{it} = a_{1i} + a_{2i}x_{2it} + a_{3i}x_{3it} + e_{it} \qquad (6.24)$$

Dr. Theo reminds us to note the subscript "i" in a_{1_i} indicates different intercepts across the identities. For example, Vietnam is different from Cambodia. Another way to write the relations is to also allow differences in the slopes.

Forecasting with Fixed Effect Estimators

Dr. Theo then introduces us to a way to solve the problem by using the "fixed effect" estimators. For introductory purposes, he discusses the difference in intercepts first. Theoretically, the adjustment is made by taking the deviation from the mean, so we first take an average of Equation 6.24:

$$\bar{y}_{it} = a_{1i} + a_2\bar{x}_{2it} + a_3\bar{x}_{3it} + \bar{e}_{it} \tag{6.25}$$

Next, we subtract Equation 6.25 from Equation 6.24:

$$y_{it} - \bar{y}_{it} = a_2(x_{2it} - \bar{x}_{2it}) + a_3(x_{3it} - \bar{x}_{3it}) + (e_{it} - \bar{e}_{it})$$

Then we can perform forecasts on the transformed model using OLS:

$$\tilde{y}_{it} = a_2\tilde{x}_{2it} + a_3\tilde{x}_{3it} + \tilde{e}_{it} \tag{6.26}$$

where

$$\tilde{y}_{it} = y_{it} - \bar{y}_{it};\ \tilde{x}_{2it} = (x_{2it} - \bar{x}_{2it});\ \tilde{x}_{3it} = (x_{3it} - \bar{x}_{3it});\ \tilde{e}_{it} = (e_{it} - \bar{e}_{it})$$

The fixed-effect problem is solved because the intercept is removed from Equation 6.26.

Empirically, the most convenient and flexible way to perform forecasts on a fixed-effect model is to use the least squares dummy variable (LSDV) approach whenever the number of identities is not too large (less than 100 identities are manageable). To obtain the LSDV estimators, we generate a dummy variable for each of the identities. For example, if we have eight different identities, then

$$D_{1i} = \begin{cases} 1 & i = 1 \\ 0 & \text{otherwise} \end{cases} \quad D_{2i} = \begin{cases} 1 & i = 2 \\ 0 & \text{otherwise} \end{cases} \dots D_{8i} = \begin{cases} 1 & i = 8 \\ 0 & \text{otherwise} \end{cases}$$

Then Equation 6.24 can be adjusted as:

$$y_{it} = a_{11} D_{1i} + a_{12} D_{2i} + \ldots + a_{18} D_{8i} + a_2 x_{2it} + a_3 x_{3it} + e_{it} \qquad (6.27)$$

Dr. Theo reminds us that the constant is removed, so the regression does not have a constant.

We then go back to the example of the TV sales in Vietnam and Cambodia, we need to add two dummies: D_V = Vietnam and D_C = Cambodia. Thus, Equation 6.23 can be written as:

$$\text{SALE}_{it} = a_{11}D_V + a_{12}D_C + a_2\text{INC}_{it} + a_3\text{ADS}_{it} + e_{it}$$

Ex asks, "What if the two markets might also differ over time?" Dr. Theo says, "Then adding time dummies to the equation will solve the problem:"

$$\text{SALE}_{it} = a_{11}D_V + a_{12}D_C + a_2\text{INC}_{it} + a_3\text{ADS}_{it} + b_{11}t_1 + b_{12}t_2 + \ldots + b_{1T}t_T + e_{it}$$

Fin asks, "How about the differences in the slopes of the two markets." Alte suggests, "Then we can add the slope dummies to the equation:"

$$\begin{aligned}\text{SALE}_{it} = {} & a_{11}D_V + a_{12}D_C + a_2\text{INC}_{it} + a_3\text{ADS}_{it} + b_{11}t_1 + b_{12}t_2 + \\ & + \ldots + b_{1T}t_T + c_1(D_V * \text{INC}_{it}) + c_2(D_C * \text{INC}_{it}) + \\ & + d_1(D_V * \text{ADS}_{it}) + d_2(D_C * \text{ADS}_{it}) + e_{it}\end{aligned}$$

Dr. Theo praises her for the correct answer. He then says that point and interval forecasts can be obtained in the same manner as those in the section on "Forecasts" of this chapter.

Dr. Theo also tells us that three other approaches to correct for the problem of different characteristics in panel data are first differencing, random effect, and seemingly unrelated estimations. The first differencing approach is similar to the one used in time series analysis for AR(1) models, which will be discussed in Chapter 7. The other two methods are beyond the scope of this book. He encourages us to read an econometric book if we are interested in learning more techniques on panel data.

Detecting Different Characteristics

We want to test if the equations for Vietnam and Cambodia have identical coefficients so that OLS can be performed. This is a modified F-test for which the restricted model is:

$$\text{SALE}_{it} = a_{11} + a_2\text{INC}_{it} + a_3\text{ADS}_{it} + e_{it}$$

The unrestricted model needs only one dummy because a_{11} already catches one country's effect:

$$\text{SALE}_{it} = a_{11} + a_{12}D_C + a_2\text{INC}_{it} + a_3\text{ADS}_{it} + e_{it}$$

If the two countries share the same trait, then $a_{12} = 0$, so the hypotheses are stated as:

$$H_0\text{: } a_{12} = 0;\ H_a\text{: } a_{12} \neq 0.$$

Under assumptions of equal error variances and no error correlation, this F-test is similar to the F-tests discussed in the section on "Evaluations and Adjustments" of this chapter:

$$F_{\text{STAT}} = \frac{(\text{SSE}_R - \text{SSE}_U)/J}{\text{SSE}_U/(NT-K)} \text{ and } F_c = F_{(J,NT-K)} \qquad (6.28)$$

where NT = the number of observations.

Dr. Theo reminds us that if we reject H_0, then the two equations do not have identical coefficients, and a panel-data technique is needed for estimations and forecasts.

Excel Applications

Ex shares with us a yearly dataset on per capita income and imports for Australia, China, and South Korea from the years 2004 to 2012 from the World Bank website. Since one lagged variable is generated, we have data for the years 2005–2012 to perform the regressions and tests. We find that the data are from the file Ch06.xls, Fig.6.6.

Testing Different Characteristics

The restricted model is:

$$\text{IMPS}_{it} = a_{11} + a_2\text{PERCA}_{i,t-1} + e_{it}$$

where IMPS is imports, and PERCA is per capita income.

The unrestricted model is:

$$\text{IMPS}_{it} = a_{11} + a_{1A}D_A + a_{1C}D_C + \text{PERCA}_{i,t-1} + e_{it} \qquad (6.29)$$

D_A and D_C are the dummies for Australia and China, respectively.

To perform the cross-equation test on the three countries, we first estimate the restricted model by regressing IMPS on PERCA:

Go to Data then Data Analysis, select Regression and click OK
The input Y range is E1:E25, the input X range is G1:G25
Check the box Labels
Check the button Output Range and enter L1
Click OK and then OK again to overwrite the data

Second, we regress IMPS on PERCA, D_A and D_C:

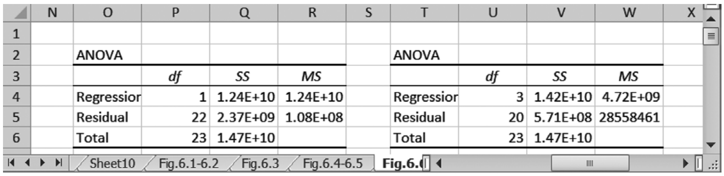
Go to Data then Data Analysis, select Regression and click OK
The input Y range is E1:E25, the input X range is G1:I25
Check the box Labels
Check the button Output Range and enter L20
Click OK then OK again to overwrite the data

The ANOVA sections with the sums of the squared errors for the two models are displayed in Figure 6.6: SEE_R in cell Q5 and SSE_U in cell V5 (in the Excel file, they are in cells N13 and N32, respectively).

From the results in Figure 6.6, the four steps for the test are as follows:

i. H_0: $a_{1A} = a_{1C} = 0$; H_a: $a_{1A} \neq 0$, or $a_{1C} \neq 0$, or both $\neq 0$.

ii. $F_{\text{STAT}} = \dfrac{(2.37 * 10^9 - 5.71 * 10^8)/2}{5.71 * 10^8 / (24 - 4)} = 31.51.$



	N	O	P	Q	R	S	T	U	V	W	X
1											
2		ANOVA					ANOVA				
3			df	SS	MS			df	SS	MS	
4		Regressior	1	1.24E+10	1.24E+10		Regressior	3	1.42E+10	4.72E+09	
5		Residual	22	2.37E+09	1.08E+08		Residual	20	5.71E+08	28558461	
6		Total	23	1.47E+10			Total	23	1.47E+10		

Sheet10 / Fig.6.1-6.2 / Fig.6.3 / Fig.6.4-6.5 / Fig.6.(

***Figure 6.6** ANOVA sections of the results for the restricted and unrestricted models*

iii. We decide to use $\alpha = 0.05$. Typing = FINV (0.05, 2, 20) into any cell in Excel gives us $F_c = 3.49$.

iv. F-statistic is greater than F-critical, so we reject the null, meaning at least one pair of coefficients is different and implying that a panel-data technique is needed.

Forecasting with Panel Data

We find that the data are from the file Ch06.xls, Fig.6.7. We are going to perform an LSDV estimation, so the regression equation is:

$$\text{IMPS}_{it} = a_{1A}D_A + a_{1C}D_C + a_{1K}D_K + \text{PERCA}_{i,t-1} + e_{it}$$

We learn to perform the following steps:

Go to Data then Data Analysis, select Regression and click OK
The input Y range is E1:E25, the input X range is G1:J25
Check the boxes Labels and Constant is Zero, and check the Residuals button
Check the button Output Range and enter L1
Click OK and then OK again to overwrite the data

The main results are reported in Figure 6.7.

Note that the three dummies are employed to control for the fixed effects. To recover the intercept for predictions and forecasts, recall the theoretical equation:

$$\bar{y}_{it} = a_{11} + a_2\bar{x}_{2it}, \text{ so:}$$

$$a_{11} = \bar{y}_{it} - a_2\bar{x}_{2it} \tag{6.30}$$

	df	SS	MS	F	gnificance F	
Regression	4	60965840119	1.52E+10	533.6933	3.17E-19	
Residual	20	571169211.3	28558461			
Total	24	61537009331				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
PERCAt-1	0.167632075	0.228798968	0.732661	0.472265	-0.30963	0.644898
DA	8977.572976	9054.416782	0.991513	0.33328	-9909.61	27864.76
DC	74346.83895	2005.852971	37.06495	6.56E-20	70162.7	78530.97
DK	38957.43691	4778.343714	8.152916	8.69E-08	28989.99	48924.89

Figure 6.7 Main regression results for Equation 6.28

Once the intercept is recovered, substitute it into the estimated equation to calculate the point and interval forecasts as usual. We also learn that more issues and solutions using panel data are discussed in Baltagi (2006).

Exercises

1. The file RGDP.xls contains data on RGDP, CONS, INV, and EXPS. The data are for the United States from the first quarter in 2006 to the first quarter in 2014. Given RGDP as the dependent variable:
   a. Perform a correlation analysis for the three explanatory variables.
   b. Perform a multiple regression of RGDP on the other three variables. Provide comments on the results, including the significances of a_1, a_2, and a_3, R^2, adjusted R^2, and the standard error of regression.
2. Write the estimated equation for the regression results for Exercise 1, enter the standard errors below the estimated coefficients, adding the adjusted R^2 next to the equation. Obtain the point forecast for the second quarter of the year 2014 based on this equation using a handheld calculator.
3. Use the results in Exercise 1 and carry out an additional regression on a restricted model as needed to test the joint significance of INV and EXPS at a 5 percent significance level. Write the procedure in four standard steps similar to those in the section "Evaluations and Adjustments" of this chapter. The calculations of the F-statistics might be performed using a handheld calculator or Excel.
4. The file Exps.xls contains data on EXPS, the EXCHA, and RGDP. The data are for the United States from the first quarter in 2006 to the first quarter in 2014. Given EXPS as the dependent variable:

a. Perform a multiple regression of EXPS on the other two variables.
b. Construct a 99 percent confidence interval for the one-period forecast (for Quarter 2, 2014).
c. Obtain point forecasts for two periods by applying the MA(3) for explanatory variables using either a handheld calculator or Excel.

5. The file Growth.xls contains data on GDP growth (GROW), INV, and money growth (MONEY) for four regions A, B, C, and D, in 10 years. Assuming that the four regions differ in intercepts only:
   a. Perform a regression of GROW on INV and MONEY using the LSDV technique.
   b. Provide point and interval forecasts for two periods ahead ($m = 2$).

CHAPTER 7

Advanced Time Series Techniques

Fligh raises a question in today's class, "Dr. Theo, so far my boss has been satisfied with my demand forecasts for Flightime Airlines. However, he says that the holiday season is coming and that he expects high volumes of visitor arrivals by air. So he has asked me to adjust for the seasonal and cyclical effects but I don't know how to estimate them." Dr. Theo assures him that this week we will learn how to adjust for these two components. Some of the estimations require regression techniques learned in Chapters 5 and 6. He says that once we finish with this chapter, we will be able to:

1. Discuss the concept of decomposing a time series.
2. Analyze triple exponential technique.
3. Explain the AR(p) and ARMA/ARIMA (p, d, q) models.
4. Apply Excel while obtaining forecasts using the models learned in (1), (2), and (3).

The seasonal and cyclical components of a time series will be discussed in several sections of the chapter.

Decomposition

Dr. Theo reminds us that we were introduced to four components of a time series in Chapter 2. Each of these components now can be isolated and calculated before they are combined together to form the final forecasts.

Concept

In Chapter 6, we assume an additive model for multiple linear regressions. In this chapter, we assume a multiplicative model for decomposing a time series:

$$A_t = T_t * R_t * S_t * C_t \tag{7.1}$$

where
A_t = the actual data at time t
T_t = the trend component at time t
R_t = the random component at time t
S_t = the seasonal component at time t
C_t = the cyclical component at time t

Three-Component Decomposition

For introductory purposes, Dr. Theo first teaches us a decomposition technique that can be performed without the cyclical component similar to the one in Lawrence, Klimberg, and Lawrence (2009). The three-component model can be written as:

$$A_t = T_t * R_t * S_t \tag{7.2}$$

The decomposition process is performed in four steps:

1. Construct seasonal-random indices (SR_t) for individual periods.
2. Construct a composite seasonal index (S_t).
3. Calculate the trend-random value (TR_t) for each period.
4. Calculate the forecast values by replacing A_t with F_t in Equation 7.2: $F_t = TR_t * S_t$.

Dr. Theo then discusses each of these four steps in detail.

Constructing Seasonal-Random Indices for Individual Periods The seasonal-random component is written as:

$$SR = S_t * R_t = A_t / T_t \tag{7.3}$$

Equation 7.3 implies that our main task in step (1) is to obtain the de-trended data. The following technique consists of calculating a center moving average (CMA). Since data come in either quarterly or monthly, the exact center is either between the second quarter and the third quarter, or between June and July, respectively.

Suppose we have five years of quarterly data, making a total of 20 quarters. Let $Q1$ denote quarter 1, $Q2$ denote quarter 2, and so on. Then the following steps should be performed for quarterly data:

First, calculate the early moving average: EMA = $(Q1 + Q2 + Q3 + Q4)/4$.

Second, calculate the late moving average: LMA = $(Q2 + Q3 + Q4 + Q5)/4$.

Third, calculate the trend T_t = central moving average = CMA = (EMA + LMA)/2.

Finally, calculate $SR_t = A_t / T_t$ following Equation 7.3.

For example, data for A_3, T_3 for $Q3$, are provided, and SR_3 are calculated as follows:

Time	A_3	$CMA_3 = T_3$	SR_3
$Q3_{year1}$	64.5	65.8625	64.5/65.8625 = 0.9793
$Q3_{year2}$	59.4	60.3875	59.4/60.3875 = 0.9836
$Q3_{year3}$	70.1	68.85	70.1/68.85 = 1.0182
$Q3_{year4}$	70.8	70.75	70.8/70.75 = 1.0007
$Q3_{year5}$	72.4	73.0375	72.4/73.0375 = 0.9913

Ex then asks, "How about monthly data?" Dr. Theo answers, "You can analyze monthly data in the same manner, except that 12-month moving averages will be calculated."

Constructing a Composite Seasonal Index Dr. Theo continues, "Since the seasonal-random indices are for individual quarters, and since there is some randomness in the series, an average seasonal index (AS) must be calculated."

Dr. Theo then asks us to calculate the average seasonal index for the third quarter (AS_3), and we are able to find the answer as follows:

$$AS_3 = (0.9793 + 0.9836 + 1.0182 + 1.0007 + 0.9913)/5 \approx 0.9946$$

Dr. Theo continues that 1.00 is the average index, so the sum of the four quarters must be 4.00:

$$AS_1 + AS_2 + AS_3 + AS_4 = 4.00$$

However, this is not the case most of the time due to the randomness of a time series. Hence, an adjusted average has to be calculated by scaling down all values to obtain a composite index (S_Q) for each quarter, with S_1 for $Q1$, S_2 for $Q2$, and so on. For example, if the sum of four quarters is 4.1347, then the equation for scaling down the preceding third quarter is derived as follows:

$$\frac{0.9946}{4.1347} = \frac{S_3}{4} \rightarrow S_3 = \frac{0.9946 * 4}{4.1347} \approx 0.9622$$

We learn that an index lesser than 1.00 indicates a low season, and an index greater than 1.00 implies a high season. Also, some researchers multiply the index by 100 to obtain it in hundreds.

Calculating the Trend-Random Values (TR_t) In the section "Constructing Seasonal-Random Indices for Individual Periods," we estimate each period trend so that we can divide the data by the trend to obtain the seasonal-random index. In this step, we need to obtain trend-random values by deseasonalizing the data:

$$TR_t = T_t * R_t = A_t/S_t \tag{7.4}$$

Dr. Theo says that the regression technique is used for this purpose, and the econometric equation for the trend values is written as:

$$TR_t = b_1 + b_2\, t + e_t \tag{7.5}$$

where TR_t = the predicted value of the TR_t line at time t

b_1 = the intercept of the TR_t line

b_2 = the slope of the TR_t line

For example, suppose the regression results on the quarterly data yield this equation:

$$TR_Q = 62.9 + 0.49 * t$$

Then the forecasted TR_t for period 3 is:

$$TR_3 = 62.9 + 0.49 * 3 = 64.37$$

Obtaining Forecast Values We learn that the forecast values can be obtained by replacing A_t with F_t in Equation 7.2:

$$F_t = TR_t * S_t \tag{7.6}$$

For example, the composite seasonal index for *Q3* is $S_3 = 0.9622$ and the trend $TR_3 = 64.37$, therefore:

$$F_3 = 64.37 * 0.9622 = 61.94$$

Dr. Theo reminds us to note the subscript t in Equation 7.6 for F_t instead of $t + 1$ as in Chapters 2 through 4. In the decomposition technique, we assume that the cycle will repeat itself in the years ahead so that the calculated indices can be used to forecast for a whole new year or years ahead instead of one-period forecasts. In the preceding example, data are available up to the fourth quarter of the fifth year, so we can use the calculated indexes to obtain forecasts for the four quarters in the sixth year, the seventh year, and so on. This will become clear when we get to the section on "Excel Applications."

Adding the Cyclical Component

Dr. Theo now adds the cyclical component, so the four-component model is:

$$A_t = T_t * R_t * S_t * C_t \tag{7.7}$$

where C_t is the cyclical component at time t.

He reminds us that calculations of the composite seasonal index S_Q and time-random index TR_Q are the same as earlier except that a cyclical component is assumed. Hence, we can use the same S_Q and TR_Q. The next step is to calculate the cyclical-random index, which can be obtained using the following equation:

$$CR_t = C_t * R_t = A_t/(S_t * TR_t) \tag{7.8}$$

Once CR_t is calculated, the cyclical component C is isolated from CR (i.e., we are derandomizing the cyclical component) by calculating moving averages of the CR data to smooth out the randomness. The order of the moving average depends on the length of the business cycle. For example, if MA(5) is selected, then the results are put in the center of each five-period horizon. The values of MA(3) can be adopted for shorter cycles.

Once the cyclical component C is obtained, the random component R can be isolated:

$$R_t = CR_t/C_t \tag{7.9}$$

A random index of 1.00 implies no randomness in the series. Dr. Theo reminds us that this random component is not needed to calculate the forecast values because it has been incorporated into TR_t. This random component is just to show how random the time series is. Finally, forecast values are calculated using Equation 7.7 with F_t in place of A_t:

$$F_t = TR_t * S_t * C_t \tag{7.10}$$

To conclude this theoretical section, Dr. Theo reminds us that an additive model for the decomposition technique is possible and is discussed in Gaynor and Kirkpatrick (1994).

Excel Applications

Fligh is collecting data on tourism and shares with us one of his datasets on hotel occupancy in Maui, Hawaii. The data are from the first quarter of 2008 through the second quarter of 2013 and are available in the file Ch07.xls, Fig.7.1.

Dr. App tells us that it is more convenient to obtain the regression results for the TR_t line first so that the steps of decomposition calculations can be grouped together in one figure.

Estimating the TR_t Line

We learn that we need to perform the following steps:

Go to Data then Data Analysis, select Regression then click OK
Enter the input *Y* range C1:C23 and the input *X* range D1:D23
Check the Labels box and the button at the Output Range box
Enter F1 into the box then click OK and then OK again to overwrite the data.

Figure 7.1 displays the results of the regression with the hotel occupancy as a dependent variable and the time period as an independent variable. From this figure, you can see that the coefficient estimate of the TR_t line is positive and statistically significant, implying an upward trend.

The data in this figure also reveal that the high season is in the first quarter and the low season is in the second quarter. Additionally, the hotel occupancy reached a high value of 80.9 percent in the first quarter

	A	B	C	D	E	F	G	H	I
1		SUMMARY OUTPUT							
2		*Regression Statistics*							
3		Multiple R	0.449853431						
4		R Square	0.202368109						
5		Adjusted R Square	0.162486515						
6		Standard Error	6.529018057						
7		Observations	22						
8		ANOVA							
9			*df*	*SS*	*MS*	*F*	*Significance F*		
10		Regression	1	216.3043732	216.3043732	5.07422313	0.035672751		
11		Residual	20	852.5615359	42.62807679				
12		Total	21	1068.865909					
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
14		Intercept	62.9025974	2.881695821	21.82832655	2.02032E-15	56.89148525	68.91371	
15		Time	0.494240542	0.21940857	2.252603634	0.035672751	0.036562285	0.951919	
16									

Sheet16 / Sheet17 / Sheet18 / Sheet1 / **Sheet2** / Sheet3

Figure 7.1 Regression results for the TR_t line

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

of 2008, dropped to a trough of 56.3 percent in the second quarter of 2009, and it again rose to 80.2 percent in the first quarter of 2013. Since the seasonal pattern is clear, decomposition is the appropriate technique.

Three-Component Decomposition

Constructing Seasonal-Random Indexes for Individual Quarters (SR_t)

The file Ch07.xls, Fig.7.2, shows the same quarterly data on the Maui hotel occupancy as those in the file Ch07.xls, Fig.7.1. We must proceed as follows:

In cell E4, type = (D2 + D3 + D4 + D5)/4 and press Enter
Copy and paste the formula into cells E5 through E23
In cell F4, type = (D3 + D4 + D5 + D6)/4 and press Enter
Copy and paste the formula into cells F5 through F23
In cell G4, type = (E4 + F4)/2 and press Enter
Copy and paste the formula into cells G5 through G23
In cell H4, type = D4/G4 and press Enter
Copy and paste the formula into cells H5 through H23

The seasonal-random indexes for individual quarters is displayed in column H of Figure 7.2.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Time	Quarter	Hotel (A)	EMA	LMA	CMA	SR	AS	S	S	TR	Forecast
2		1	Q1-2008	80.9									
3		2	Q2-2008	66.5									
4		3	Q3-2008	64.5	67.975	63.75	65.8625	0.979313	0.994619	0.962225	0.962225	64.3853	61.95316
5		4	Q4-2008	60	63.75	61.2	62.475	0.960384	0.947296	0.916443	0.916443	64.8796	59.45843
6		5	Q1-2009	64	61.2	59.925	60.5625	1.05676	1.126465	1.089777	1.089777	65.3738	71.24288
7		6	Q2-2009	56.3	59.925	59.375	59.65	0.943839	1.066283	1.031555	1.031555	65.868	67.94649
8		7	Q3-2009	59.4	59.375	61.4	60.3875	0.983647	4.134663	4	0.962225	66.3623	63.85544
9		8	Q4-2009	57.8	61.4	63.4	62.4	0.926282			0.916443	66.8565	61.2702
10		9	Q1-2010	72.1	63.4	66.075	64.7375	1.113729			1.089777	67.3508	73.39732
11		10	Q2-2010	64.3	66.075	68.1	67.0875	0.95845			1.031555	67.845	69.98584
12		11	Q3-2010	70.1	68.1	69.6	68.85	1.018155			0.962225	68.3392	65.75772
13		12	Q4-2010	65.9	69.6	69.975	69.7875	0.944295			0.916443	68.8335	63.08197
14		13	Q1-2011	78.1	69.975	70.15	70.0625	1.114719			1.089777	69.3277	75.55177
15		14	Q2-2011	65.8	70.15	70.5	70.325	0.935656			1.031555	69.822	72.02518
16		15	Q3-2011	70.8	70.5	71	70.75	1.000707			0.962225	70.3162	67.66
17		16	Q4-2011	67.3	71	71.75	71.375	0.942907			0.916443	70.8104	64.89374
18		17	Q1-2012	80.1	71.75	72.15	71.95	1.113273			1.089777	71.3047	77.70621
19		18	Q2-2012	68.8	72.15	73.025	72.5875	0.947822			1.031555	71.7989	74.06452
20		19	Q3-2012	72.4	73.025	73.05	73.0375	0.991272			0.962225	72.2932	69.56228
21		20	Q4-2012	70.8	73.05	74.05	73.55	0.96261			0.916443	72.7874	66.70551
22		21	Q1-2013	80.2	74.05	55.95	65	1.233846			1.089777	73.2816	79.86066
23		22	Q2-2013	72.8	55.95	38.25	47.1	1.545648			1.031555	73.7759	76.10386
24		23									0.962225	74.2701	71.46456
25		24									0.916443	74.7644	68.51729
26		25									1.089777	75.2586	82.01511
27		26									1.031555	75.7528	78.14321

Sheet12 Sheet13 Sheet14 Sheet1 Sheet2 Sheet3

Figure 7.2 Decomposition without the cyclical componet: obtaining forecast values

Obtaining a Composite Seasonal Index (S) Cells I4 through I7 display the calculations of the average indexes AS_Q, and cell I8 displays their sum. We need to perform the following steps:

In cell I4, type = (H4 + H8 + H12 + H16 + H20)/5 and press Enter
Copy and paste the formula into cells I5 through I7
In cell I8 type = SUM (I4:I7) and press Enter
In cell J4 type = I4 * 4/$I$8 and press Enter
Copy and paste the formula into cells J5 through J7
Copy and paste the formula in cell I8 into cell J8 to see that the sum is 4.00.
Copy and paste-special the values in cells J4 through J7 into cells K4 through K27

(We assume that the cycle will repeat itself in the next four quarters.)

Dr. App reminds us that the indexes in cells J4 through J7 are the composite seasonal indexes S_Q. Thus, you can paste repeatedly and extend the periods to $t = 26$ for forecast values.

Calculating the Trend-Random Values (TR_t) The TR_t values are calculated by applying the regression results for the TR_t line in Figure 7.1 into the 26 periods from 1 to 26. Hence,
In cell L4, type = 62.9026 + 0.49424 * B4 and press Enter
Copy and paste the formula into cells L5 through L27.

Obtaining Forecast Values ($S_t * TR_t$) We learn that the following steps must be performed:

In cell M4, type = K4 * L4 and press Enter
Copy and paste the formula into cells M5 through M27.
The forecast values for the next four periods are displayed in cells M24 through M27.

Dr. App reminds us that we can extend the forecasts into the long-term future by simply extending data in columns K and L as far as you need.

Adding Cyclical Component

We find that the file Ch07.xls, Fig.7.3, retains most of the data from the file Ch07.xls, Fig.7.2, except for the calculations of the seasonal indexes. We proceed with this exercise as follows:

In cell G4, type = D4/(E4 * F4) and press Enter
Paste the formula into cells G5 through G23
Copy and paste-special the values in cells G20 through G23 into cells G24 through G27
In cell H6, type = (G4 + G5 + G6 + G7 + G8)/5 and press Enter
Copy and paste the formula into cells H7 through H25
In cell I6, type = G6/H6 and press Enter
Copy and paste the formula into cells I7 through I25
(The results in column I are only used to show the randomness of the data. They are not used for forecasting)
In cell J6, type = E6 * F6 * H6 and press Enter
Copy and paste the formula into cells J7 through J25

The forecast values for the next two periods are in cells J24 and J25.

	A	B	C	D	E	F	G	H	I	J
1		Time	Quarter	Hotel (A)	S	TR	CR	C	R	Forecast
2		1	Q1-2008	80.9						
3		2	Q2-2008	66.5						
4		3	Q3-2008	64.5	0.96222	64.3853	1.04111			
5		4	Q4-2008	60	0.91644	64.8796	1.00911			
6		5	Q1-2009	64	1.08978	65.3738	0.89834	0.94147	0.95418	67.0734
7		6	Q2-2009	56.3	1.03155	65.8680	0.82859	0.92193	0.89876	62.6416
8		7	Q3-2009	59.4	0.96222	66.3623	0.93023	0.91657	1.01490	58.5279
9		8	Q4-2009	57.8	0.91644	66.8565	0.94336	0.92065	1.02467	56.4086
10		9	Q1-2010	72.1	1.08978	67.3508	0.98232	0.96814	1.01465	71.0590
11		10	Q2-2010	64.3	1.03155	67.8450	0.91876	0.99103	0.92707	69.3581
12		11	Q3-2010	70.1	0.96222	68.3392	1.06603	1.00910	1.05642	66.3563
13		12	Q4-2010	65.9	0.91644	68.8335	1.04467	0.99535	1.04955	62.7888
14		13	Q1-2011	78.1	1.08978	69.3277	1.03373	1.02088	1.01258	77.1295
15		14	Q2-2011	65.8	1.03155	69.8220	0.91357	1.01509	0.89999	73.1122
16		15	Q3-2011	70.8	0.96222	70.3162	1.04641	1.01232	1.03368	68.4935
17		16	Q4-2011	67.3	0.91644	70.8104	1.03708	0.99136	1.04612	64.3328
18		17	Q1-2012	80.1	1.08978	71.3047	1.03081	1.01680	1.01377	79.0118
19		18	Q2-2012	68.8	1.03155	71.7989	0.92892	1.01980	0.91089	75.5307
20		19	Q3-2012	72.4	0.96222	72.2932	1.04079	1.01323	1.02720	70.4826
21		20	Q4-2012	70.8	0.91644	72.7874	1.06138	0.99839	1.06310	66.5979
22		21	Q1-2013	80.2	1.08978	73.2816	1.00425	1.02076	0.98382	81.5187
23		22	Q2-2013	72.8	1.03155	73.7759	0.95659	1.02488	0.93337	77.9972
24		23			0.96222	74.2701	1.04079	1.01345	1.02698	72.4259
25		24			0.91644	74.7644	1.06138	1.00392	1.05724	68.7859
26		25			1.08978	75.2586	1.00425			
27		26			1.03155	75.7528	0.95659			

Sheet12 Sheet13 Sheet14 Sheet1 Sheet2 Sheet3

Figure 7.3 Decomposition with cyclical component: obtaing forecast values

Triple Exponential Smoothing

The first approach to the triple exponential smoothing is called Holt–Winters exponential smoothing (HWE) thanks to Winters (1960) who modified the original model by Holt (1957). The HWE uses a seasonal index similar to the decomposition technique (DT). Different from the DT, the HWE adds a third equation to the existing two equations of the double exponential smoothing (DE) instead of decomposing other components of the time series.

The second approach to the triple exponential smoothing is called the higher-order exponential smoothing (HOE) because it adds a nonlinear term to the trend equation. The nonlinear term could be in quadratic, cubic, logarithmic, or any other form depending on the curvature of the series.

Concept

Holt–Winters Exponential Smoothing

This model comprises three equations with three parameters. The first parameter, *a*, is used to smooth out the original series:

$$E_t = a(A_t/S_{t-L}) + (1 - a) * (E_{t-1} + T_{t-1}),\ 0 < a < 1, \qquad (7.11)$$

where
L = the length of the cyclical component
S = the seasonal index
$L = 4$ if quarterly data are used
$L = 12$ if monthly data are used (Lapin 1994). The subscript $(t - L)$ indicates that the seasonal factor is considered from L periods before period t

At this point, Arti asks, "Does that mean that the HWE uses the individual seasonal indexes instead of the composite ones?" Dr. Theo commends her for the correct observation and provides us with an example from Figure 7.3, where the actual data on hotel occupancy for the first quarter of 2012 should be divided by the seasonal index for the first quarter of 2011:

$$E_{Q1,\,2012} = a(A_{Q1,\,2012}/S_{Q1,\,2011}) + (1 - a) * (E_{Q4,\,2011} + T_{Q4,\,2011})$$

The second parameter, b, is used to smooth out the trend T_t, which is the same as the one in Chapter 4:

$$T_t = b(E_t - E_{t-1}) + (1 - b)\ T_{t-1},\ 0 < b < 1 \qquad (7.12)$$

The third parameter, c, is used to smooth out the seasonal changes:

$$S_t = c(A_t/E_t) + (1 - c)\ S_{t-L},\ 0 < c < 1 \qquad (7.13)$$

The equation for the multiple period forecasts is:

$$F_{t+m} = (E_t + T_t\ m) * S_{t-L+m} \qquad (7.14)$$

Because the three distinct parameters are used in the three equations, this approach is also called three-parameter exponential smoothing.

Higher-Order Exponential Smoothing

This approach is appropriate when the curvature of the trend is observed. Adding a quadratic term is suitable if the curve is convex, that is, the trend rises at increasing rates over time. If the curve is concave, adding a logarithmic term is more suitable than a quadratic one.

Alte asks, "How can we know the curvature of the time series?" Dr. Theo replies, "Constructing a time series plot will help you recognize the shape of the curve and select the model." He then says that this curve fitting processes already smooth out for the seasonal and cyclical components, so there is no need to go through the de-seasoning and de-randomizing processes as in decomposition technique. Thus, the forecast model with a quadratic term added is:

$$F_{t+1} = b_1 + b_2\ t + b_3\ t^2 + e_{t+1} \qquad (7.15)$$

where F_{t+1} = the forecasted value of the series at time $(t + 1)$
b_1 = the intercept of the trend curve = the initial value of the actual data
b_2 = the slope of the linear trend
b_3 = the slope of the nonlinear trend

If the trend equation is extended to allow a logarithmic term, then Equation 7.15 becomes:

$$F_{t+1} = b_1 + b_2\, t + b_3 \ln(t) + e_{t+1} \tag{7.16}$$

In Equations 7.15 and 7.16, there are a total of three parameters to be estimated, so the technique is also called the triple exponential or three-parameter exponential smoothing technique. We learn that estimating Equation 7.15 or Equation 7.16 to obtain the three parameters and the predicted values in one regression is the best approach of forecasting.

Dr. Theo tells us that researchers also try higher-order polynomial models (Brown 1963; Montgomery, Jennings, and Kulahci 2008) and hence we have the name *higher-order exponential smoothing*.

Excel Applications

Holt–Winters Exponential Smoothing

The file Ch07.xls, Fig.7.4, displays the same data from Fligh as those in the file Ch07.xls, Fig.7.2, and the calculations using the smoothing factors a = 0.4, b = 0.6, and c = 0.7. The data are on hotel occupancy in Maui, Hawaii from the first quarter of 2008 through the second quarter of 2013. The initial exponential smoothing value is chosen as the first actual value, and the first trend value is calculated by averaging the first two trend values. We learn to proceed as follows:

	A	B	C	D	E	F	G	H	I	J
1		Time	Quarter	Hotel (A)	S	E (a = 0.4)	T (b = 0.6)	S(c = 0.7)	Forecast	
2		1	Q1-2008	80.9						
3		2	Q2-2008	66.5						
4		3	Q3-2008	64.5	0.979313					
5		4	Q4-2008	60	0.960384					
6		5	Q1-2009	64	1.05676					
7		6	Q2-2009	56.3	0.943839	56.3	-6.1			
8		7	Q3-2009	59.4	0.983647	54.38191	-3.59086	1.058386		
9		8	Q4-2009	57.8	0.922586	54.54833	-1.33649	1.029843		
10		9	Q1-2010	72.1	1.107314	59.21808	2.267256	1.169301		
11		10	Q2-2010	66.3	0.97914	64.98921	4.369581	0.99727		
12		11	Q3-2010	69.1	0.989263	69.71478	4.583172	0.988921	73.40841	
13		12	Q4-2010	68.9	0.96991	74.45133	4.6752	0.924581	76.51521	
14		13	Q1-2011	78.1	1.083781	75.68833	2.612276	1.054499	92.52276	
15		14	Q2-2011	69.8	0.95682	75.49519	0.929028	0.940935	78.08686	
16									75.57753	

Sheet12 / Sheet13 / Sheet14 / Sheet1 / Shee

Figure 7.4 Holt-Winters exponential smoothing: obtaining forecast values

In cell G7, type = ((D7 – D6) + (D5 – D4))/2, then press Enter
In cell F8, type = 0.4 * (D8/E4) + 0.6 * (F7 + G7), then press Enter
Copy and paste the formula into cells F9 through F15
In cell G8, type = 0.6 * (F8 – F7) + 0.4 * G7, then press Enter
Copy and paste the formula into cells G9 through G15
In cell H8, type = 0.7 * (D8/F8) + 0.3 * E4, then press Enter
Copy and paste the formula into cells H9 through H15
In cell I12, type = (F11 + G11) * H8, then press Enter
Copy and paste the formula into cells I13 through I16
The one period forecast is in cell J16

For multiperiod forecasts, follow the same alternative procedures as those in Chapter 4. The advantage of the HWE is that it can be used for data that exhibit seasonal patterns in addition to the two deterministic and random components as in moving average and DE techniques.

Higher-Order Exponential Smoothing

Cita is doing research on the behavior of the firms and shares with us data on labor force from the Department of Business, Economic Development, and Tourism in Hawaii. Data are for the first quarter of 2006 through the second quarter of 2013 and are available in the file Ch07.xls, Fig.7.5. To select an appropriate model for regression, we need to construct a time series plot. Figure 7.5 displays this plot, which shows a concave curve and

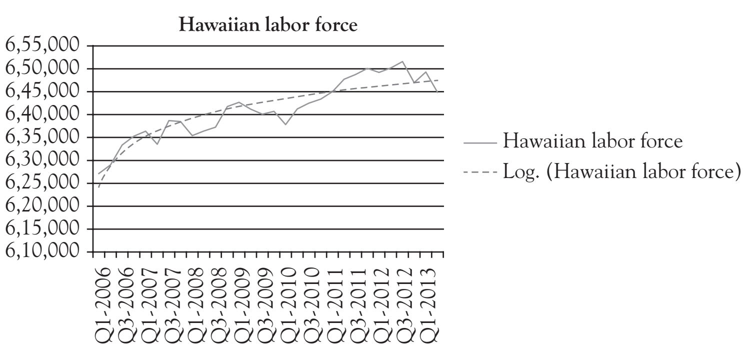


Figure 7.5 Selecting regression model: time series plot

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

implies that adding a logarithmic trend to the forecast model appears to be the best strategy.

Hence, the econometric model is written as:

$$\text{LABOR}_{t+1} = b_1 + b_2\,\text{TIME}_t + b_3\,\ln(\text{TIME}_t) + e_{t+1}$$

We learn to perform the following steps with the data in the file Ch07.xls, Fig.7.6:

Go to Data then Data Analysis, select Regression and click OK
The input *Y* range is D1:D31, the input *X* range is B1:C31
Check the boxes Labels and Residuals
Check the button Output Range and enter G21 then click OK
(We place the results into cell G21 to make the next steps of calculations convenient)
A dialogue box will appear, click OK to overwrite the data

Dr. App reminds us to click on Residuals so that predicted values are reported in the Excel results. Sections of the regression results and data (from time period t = 26 through t = 30) are displayed in Figure 7.6. We also notice that the actual and predicted values from Excel results end at period t = 30, where a dark line marks the end in cell E31. The advantage of this approach is that multiperiod forecasts can be easily obtained.

	A	B	C	D	E	F	G	H	I	J	K
27	Q2-2012	26	3.2580965	650,200	647781.3		Standard E	2403.998			
28	Q3-2012	27	3.2958369	651,600	648305.4		Observati	30			
29	Q4-2012	28	3.3322045	647,000	648825.6						
30	Q1-2013	29	3.3672958	649,300	649342		ANOVA				
31	Q2-2013	30	3.4011974	644,900	649855.1			*df*	*SS*	*MS*	*F*
32	Q3-2013	31	3.4339872		650364.9		Regressio	2	1.04E+09	5.2E+08	90.05249
33	Q4-2013	32	3.4657359		650871.8		Residual	27	1.56E+08	5779206	
34	Q1-2014	33	3.4965076		651375.8		Total	29	1.2E+09		
35	Q2-2014	34	3.5263605		651877.1						
36	Q3-2014	35	3.5553481		652376			*Coefficient*	*andard Err*	*t Stat*	*P-value*
37	Q4-2014	36	3.5835189		652872.5		Intercept	627587.3	1743.053	360.0506	3.04E-51
38	Q1-2015	37	3.6109179		653366.8		TIME	415.2572	129.8899	3.196993	0.003526
39	Q2-2015	38	3.6375862		653859		Ln(TIME)	2884.312	1344.574	2.145149	0.041093

Sheet1 Sheet2 Sheet3

Figure 7.6 Higher-order exponential smoothing: obtaining forecast values

The values in cells E32 through E39 are multiperiod forecasts. To proceed with this exercise, we must follow these steps:

Copy and paste the formula in cell C31 into cells C32 through C39
In cell E32, type = $H$37 + $H$38 * B32 + $H$39 * C32, then press Enter
Copy and paste this formula into cells E33 through E39

Dr. App reminds us that we can extend the forecasts as far as we wish into the future. The results from Figure 7.6 also confirm our intuition that a combination of a logarithmic trend and linear trend is suitable for this time series: coefficient estimates of both variables are statistically significant with *p*-values of TIME and ln(TIME) equal 0.0035 and 0.041, respectively.

Dr. Theo says that sometimes our selection of a model might not be suitable. If the model is not statistically significant, then we will need adjustments as discussed in Chapters 5.

A Brief Introduction to AR and ARIMA Models

The name ARIMA sounds so pretty that we are curious to learn about the model. The autoregressive (AR) and autoregressive integrated moving average (ARIMA) models belong to time series analyses instead of associative analyses because they involve only the dependent variable and its own lags instead of an outside explanatory variable.

Concept

AR and ARIMA models explore the characteristics of a time series dataset that often has the current value correlated with its lagged values and can be used for multiple period forecasts.

AR models

A dependent variable in an AR model can contain numerous lags. The model is denoted as AR(p), where p is the number of lags. When a dependent variable is correlated to its first lag only, the model is called an autoregressive model of order one and is denoted as AR(1):

$$y_t = ay_{t-1} + e_t \tag{7.17}$$

If $|a| < 1$, then the series is stationary because the series gradually approaches zero when t approaches infinity. When $|a| \geq 1$, the series is nonstationary: If $a > 1$, then the series explodes when t approaches infinity. If $|a| = 1$, the series is said to follow a random walk because it is wandering aimlessly upward and downward with no real pattern:

$$y_t = y_{t-1} + e_t \tag{7.18}$$

At this point, Ex asks, "Why is the process called a random walk?" Fin raises his hand and offers a story to explain the concept, "A drunken man behind the wheel was stopped by a police officer, who then asked him to walk a straight line. Of course the man could not do it. He just wandered aimlessly one step to the right, one step to the left, going forward one step, and then going backward one step. Hence, the best you can guess of his next step is to look at his previous step and add some random error to it. This is exactly what we see in Equation 7.18."

Dr. Theo thanks Fin for the fun story and says that e_t is assumed to be independent with a mean of zero and a constant variance. He then continues, "If all lagged variables are stationary as in Equation 7.17, regressions using OLS can be performed, and multiple period forecasts can be obtained by the same recursive principle discussed in Chapter 6. For example, a model with a constant and two lagged values can be estimated as:

$$\hat{y}_t = \hat{a}_1 + \hat{a}_2 y_{t-1} + \hat{a}_3 y_{t-2}$$

$$\hat{y}_{t+1} = \hat{a}_1 + \hat{a}_2 \hat{y}_t + \hat{a}_3 y_{t-1}$$

$$\hat{y}_{t+2} = \hat{a}_1 + \hat{a}_2 \hat{y}_{t+1} + \hat{a}_3 \hat{y}_t$$

Hence, long-term forecasts can be obtained in a very convenient manner."

Point Forecasts

We then break into groups to work on the following example. Estimating a stationary AR(2) model of Galaxy phone sales from a telephone shop in our city gives the following results:

$$SALE_t = 0.54\ SALE_{t-1} + 0.48\ SALE_{t-2}$$

Sale values are known as $SALE_t$ = \$4800 and $SALE_{t-1}$ = \$4700. Thus:

$$SALE_{t+1} = 0.54\ SALE_t + 0.48\ SALE_{t-1}$$

$$= 0.54 * 4800 + 0.48 * 4700 = 4848$$

$$SALE_{t+2} = 0.54\ SALE_{t+1} + 0.48\ SALE_t$$

$$= 0.54 * 4848 + 0.48 * 4800 \approx 4922$$

$$SALE_{t+3} = 0.54\ SALE_{t+2} + 0.48\ SALE_{t+1}$$

$$= 0.54 * 4922 + 0.48 * 4848 \approx 4985$$

Dr. Theo tells us that when a series follows a random walk, we might obtain significant regression results from completely unrelated data, which is called a spurious regression. He assures us that if one of the lagged variables has unit coefficient ($a_k = 1$), taking the first difference of the equation can turn it into a stationary series. For example, taking the first difference of Equation 7.18 yields a first differencing model:

$$\Delta y_t = y_t - y_{t-1} = e_t \qquad (7.19)$$

Δy_t is a stationary series because e_t is an independent random variable with a mean of zero and a constant variance. Any series that can be made stationary by taking the first difference is said to be integrated of order one and is denoted as an I(1). Any series similar to Δy_t is said to be integrated of order zero and is denoted as an I(0). This characteristic can be applied in forecasting an AR(p) model, in which (p − 1) series are stationary.

For example, an AR(3) model that has the first lagged series nonstationary is as follows:

$$y_t = y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + e_t, \text{ where } |a_2| < 1 \text{ and } |a_3| < 1$$
$$\Delta y_t = y_t - y_{t-1} = a_2 y_{t-2} + a_3 y_{t-3} + e_t \tag{7.20}$$

Estimation can be performed on Equation 7.20 using the OLS technique. Once coefficient estimates are obtained, the predicted value of Δy_t can be calculated. Since y_{t-1} is known, y_t can be calculated, which allows for multiperiod forecasts.

We then work on the next example. A restaurant estimates this equation for its sale values:

$$\Delta SALE_t = SALE_t - SALE_{t-1} = 0.12\ SALE_{t-2} + 0.08\ SALE_{t-3}$$

Sale values are known as $SALE_t$ = \$4200, $SALE_{t-1}$ = \$4100, and $SALE_{t-2}$ = \$4000. Hence:

$$\Delta SALE_{t+1} = SALE_{t+1} - SALE_t = 0.12\ SALE_{t-1} + 0.08\ SALE_{t-2}$$
$$= 0.12 * 4100 + 0.08 * 4000 = 812$$
$$SALE_{t+1} = 812 + SALE_t = 4200 + 812 = 5012$$
$$\Delta SALE_{t+2} = SALE_{t+2} - SALE_{t+1} = 0.12\ SALE_t + 0.08\ SALE_{t-1}$$
$$= 0.12 * 4200 + 0.08 * 4100 = 832$$
$$SALE_{t+2} = 832 + SALE_{t+1} = 832 + 5012 = 5844$$

Dr. Theo then reminds us that the first differenced model can be extended to panel data. The suitable case for using first differencing in panel data is that the error term follows a random walk (Wooldridge 2013). In this case, taking the first difference serves two purposes: (i) to eliminate the intercept a_{1i} presented in Equation 6.24, and (ii) to make Δe become an I(0):

$$e_{it} = e_{i,t-1} + v_{it}$$
$$y_{it} = a_{1i} + a_{2i}x_{2it} + a_{3i}x_{3it} + e_{it} \tag{7.21}$$
$$y_{it} - y_{i,t-1} = a_{2i}(x_{2it} - x_{2i,t-1}) + a_{3i}(x_{3it} - x_{3i,t-1}) + (e_{it} - e_{i,t-1})$$

Interval Forecasts

The formula for the standard error of the regression, s, in AR(p) models is still the same:

$$s = \sqrt{\frac{\text{SSE}}{N - K}}$$

However, the formulas for the standard error of the forecast, $se(f)$, can be calculated in a more precise manner than the ones in the previous chapters and are shown in Hill, Griffiths, and Lim (2011):

$$se(f)_{t+1} = s = \sqrt{\frac{\text{SSE}}{N - K}}$$

$$se(f)_{t+2} = \sqrt{s^2(\hat{a}_2^2 + 1)} \tag{7.22}$$

$$se(f)_{t+3} = \sqrt{s^2\left[\ (\hat{a}_2^2 + \hat{a}_3)^2 + \hat{a}_2^2 + 1\ )\ \right]}$$

For AR(1) models, the first two $se(f)$ values are the same as in Equation 7.22, but the $se(f)$ for the interval forecasts can be extended into m periods ahead as shown in Griffiths, Hill, and Judge (1993):

$$se(f)_{t+m} = \sqrt{s^2[\hat{a}_2^2 + \hat{a}_2^4 + \ldots + \hat{a}_2^{2(m-1)} + 1]} \tag{7.23}$$

As an example, suppose that SSE = 1260 and T = 40. The point forecasts from the previous example are:

$$\text{SALE}_{t+1} = 0.54\ \text{SALE}_t + 0.48\ \text{SALE}_{t-1} = 0.54 * 4800 + 0.48 * 4700 = 4848$$

$$\text{SALE}_{t+2} = 0.54\ \text{SALE}_{t+1} + 0.48\ \text{SALE}_t = 0.54 * 4848 + 0.48 * 4800 \approx 4922$$

$$\text{SALE}_{t+3} = 0.54\ \text{SALE}_{t+2} + 0.48\ \text{SALE}_{t+1} = 0.54 * 4922 + 0.48 * 4848 \approx 4985$$

The standard errors of the forecast errors are:

$$s^2 = \frac{1260}{40-2} = 33.1579; \quad se(f)_{t+1} = s = \sqrt{33.1579} = 5.7583$$

$$se(f)_{t+2} = \sqrt{33.1579 * (0.54^2 + 1)} = \sqrt{33.1579 * 1.2916} = 6.5442$$

$$se(f)_{t+3} = \sqrt{33.1579 * \left[(0.54^2 + 0.48)^2 + 1.2916\right]}$$

$$= \sqrt{33.1579 * 1.8870} = 7.91$$

Thus, the interval forecasts for a 95 percent confidence interval are:

$$\text{SALE}_{t+1} = 4848 \pm 5.7583 * 2.024 = (4836; 4860)$$

$$\text{SALE}_{t+2} = 4922 \pm 6.5442 * 2.024 = (4909; 4935)$$

$$\text{SALE}_{t+3} = 4985 \pm 7.91 * 2.024 = (4969; 5001)$$

ARMA versus ARIMA

When a dependent variable is correlated to a lagged value of its errors, the model is called a moving average of order one and is denoted as MA(1):

$$y_t = a\, e_{t-1} \tag{7.24}$$

Combining an AR(1) and an MA(1) gives an ARMA(1, 1) model:

$$y_t = a_1 y_{t-1} + e_t + a_2\, e_{t-1} \tag{7.25}$$

If the series can be made stationary by taking the first difference, the ARMA(1, 1) becomes an ARIMA(1, 1, 1), where the letter I in the middle of ARIMA stands for *integrated of order one*.

The model can have any order. Hence, a general ARIMA model can be written as an ARIMA (p, d, q), where p is the order of autoregressive, d is the order of integration (differencing), and q is the order of moving average. Once an ARIMA model is specified, the value for period ($t + 1$)

can be forecasted by substituting the value for period t into Equation 7.17 or Equation 7.18 and continue for the subsequent periods using the recursive principle.

For choosing p and q in the ARIMA (p, d, q) model, a model search process called Box–Jenkins procedure is often followed. The Box–Jenkins procedure consists of three steps. The first step is the *model identification*, in which a model is derived based on economic theories. The second step is the *estimation*, in which regressions are performed. The third step is the *diagnostic checking*, in which tests are carried out and residual plots are produced to evaluate the robustness of the model. The procedure is then repeated until the best model is obtained.

To conclude the section, Dr. Theo says that the topic of ARIMA models requires an in-depth knowledge of time series analysis and is beyond the scope of this textbook. He encourages us to read a book written specifically for time series modeling if we are interested in ARIMA models and the Box–Jenkins procedure.

Testing for Stationarity

We learn that the test is called the Dickey–Fuller test or unit-root test because a series is stationary if $a < 1$. The null hypothesis is that y is nonstationary. The calculations of the statistic values are the same as those of the t-statistics. However, when y is nonstationary, the variance of y is inflated and so the distribution is no longer a t-distribution but follows a τ (tau) distribution. The statistic is therefore called τ-statistic (tau-statistic). The original equation to derive the test is:

$$y_t = ay_{t-1} + e_t$$

Subtracting y_{t-1} from both sides yields:

$$y_t - y_{t-1} = (a - 1)\, y_{t-1} + e_t$$

$$\Delta y_t = c\, y_{t-1} + e_t$$

where

$$c = a - 1 \tag{7.26}$$

If $a = 1$ then $c = 0$, so Equation 7.26 is very convenient to test because t-statistics in all quantitative packages are for a test of significance with the null hypothesis for $c = 0$. The four steps of the Dickey–Fuller test are similar to those in Chapter 5 with the hypotheses:

$$H_0: c = 0; H_a: c < 0$$

To perform the Dickey–Fuller test, the model in Equation 7.26 is often extended to allow for a constant term and a trend. The model with the constant term is:

$$\Delta y_t = a_1 + c\, y_{t-1} + e_t \tag{7.27}$$

The model that adds the trend is:

$$\Delta y_t = a_1 + c\, y_{t-1} + b\, t + e_t \tag{7.28}$$

Dr. Theo says that these three models are usually estimated concurrently so that the most appropriate model is selected based on whether or not the constant term or the trend is significant. The τ-critical values for Dickey–Fuller tests are organized into tables that are three pages long in Fuller (1976, 371–3). For instructional purposes, Table 7.1 displays the most important ones based on OLS estimates for large samples and are reformatted to fit the preceding models.

Since τ-statistics have the same values as t-statistics, you can look for the standard errors from Excel Summary Outputs and calculate the t-statistics using the same formula for those in Chapter 5. Hence, the four standard steps of the Dickey–Fuller tests are as follows.

Table 7.1 Critical values for Dickey–Fuller τ distribution

Significance Level	0.01	0.025	0.05	0.10
For model (7.26)	−2.58	−2.23	−1.95	−1.62
For model (7.27)	−3.43	−3.12	−2.86	−2.57
For model (7.28)	–3.96	−3.66	−3.41	−3.12

Source: Adapted from Fuller (1976, 373).

i. H_0: $c = 0$; H_a: $c < 0$.
ii. Calculate τ-statistics = *t*-statistics.
iii. Find τ-critical.
iv. Decision: If |τ-statistics| > |τ-critical| (or τ-statistics < τ-critical), we reject H_0, meaning $c < 0$, implying the model is stationary.

Dr. Theo reminds us that we can extend the preceding models to allow more lags and still use the critical values listed in Table 7.1.

Excel Applications

Dr. App starts this section by informing us that we will need special statistical packages to perform estimations and obtain forecasts with ARIMA (*p*, *d*, *q*) models. Hence, she only provides demonstrations for the AR(*p*) model in the following section.

Sol is doing research on the effect of the solar energy on electricity companies. She shares with us a monthly dataset on the sales of electricity for commercial facilities in Kauai from the Department of Business, Economic Development and Tourism in Hawaii. The data are available in the file Ch07.xls, Fig.7.8.

Testing Stationarity

We first estimate Equation 7.26 by regressing $\Delta ELECT_t$ on $ELECT_{t-1}$ without a constant term:

Go to Data then Data Analysis, select Regression and click OK
The input *Y* range is E1:E99, the input *X* range is D1:D99
Check the boxes Labels and Constant is Zero
Check the button Output Range and enter I1 and click OK
Click OK again to overwrite the data

We learn that Figure 7.7 displays coefficient estimates and their *t*-statistics for the three regressions of Equations 7.26, 7.27, and 7.28. Panel 7.8a is for Equation 7.26. Panels 7.8b and 7.8c are for Equations 7.27

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Panel 7.8a					Panel 7.8b					Panel 7.8c			
2			*Coefficients*	*andard Err*	*t Stat*			*Coefficients*	*andard Err*	*t Stat*			*Coefficients*	*andard Err*	*t Stat*
3		Intercept	0	#N/A	#N/A		Intercept	7442740	1774675	4.193861		Intercept	9533512	2013667	4.734404
4		ELECTt-1	-0.00309	0.005924	-0.52177		ELECTt-1	-0.32099	0.075998	-4.22364		TIME	-10082.4	4848.309	-2.07957
5												ELECTt-1	-0.38889	0.081539	-4.76941

Sheet4 / Sheet1 / Sheet5 / Sheet2 / Sheet3

Figure 7.7 Regression results for the Dickey–Fuller tests

Data Source: Department of Business, Economic Development, and Tourism: State of Hawaii (2014).

and 7.28, respectively. We then continue with the second regression of ΔELECT_t on ELECT_{t-1} for Equation 7.27, which has the constant term added to the model:

Go to Data then Data Analysis, select Regression and click OK
The input *Y* range is E1:E99, the input *X* range is D1:D99
Check the box Labels (this time do *not* check on the box Constant is Zero)
Check the button Output Range and enter N1 and click OK
Click OK again to overwrite the data and obtain the results in Panel 7.8b

Finally, we perform the third regression of ΔELECT_t on ELECT_{t-1} and TIME for Equation 7.28:

Go to Data then Data Analysis, select Regression and click OK
The input *Y* range is E1:E99, the input *X* range is C1:D99
Check the box Labels
Check the button Output Range and enter S1 and click OK
Click OK to overwrite the data and obtain the results in Panel 7.8c

From the results in Panel 7.8a, the null hypothesis is not rejected, implying that the model is not stationary. From the results in Panel 7.8b and 7.8c, the null hypotheses are rejected for the second and the third models, implying the stationarity of each model. Panel 7.8c also shows that the trend is significant. Hence, forecasting should be performed using the third model in Panel 7.8c.

Forecasting with AR Models

Since we choose Equation(7.28), of which the results are displayed in Panel 7.8c, using the recursive principle to project one period forward yields:

$$\Delta ELECT_{t+1} = 9{,}533{,}512 - 10082\ TIME - 0.3889\ ELECT_t$$

The data shows that $ELECT_{Feb\text{-}06} = 21{,}761{,}678$, so

$$\begin{aligned}\Delta ELECT_{Mar\text{-}06} &= 9{,}533{,}512 - 10082 * 1 - 0.3889 * 21{,}761{,}678 \\ &= 1{,}060{,}313\end{aligned}$$

$$\begin{aligned}ELECT_{Mar\text{-}06} &= ELECT_{Feb\text{-}06} + \Delta ELECT_{Mar\text{-}06} = 21{,}761{,}678 + 1{,}060{,}314 \\ &= 22{,}821{,}991\end{aligned}$$

To obtain the multiperiod forecasts in Excel, we must perform the following steps:

In cell F3, type = 9533512 – 10082 * C2 – 0.3889 * B2 and press Enter
(Alternatively, you can enter the cell numbers in cells T17, T18, and T19, respectively)
In cell G3, type = B2 + F3 and press Enter
In cell F4, type = 9533512 – 10082 * C3 – 0.3889 * G3
Copy and paste the formula into cells F5 through F103
(Ignore the values at this moment, as the formula in cell G4 needs to be adjusted)
In cell G4 type = G3 + F4
Copy and paste the formula into cells G4 through G103 for multiperiod forecast

Dr. App concludes, "You can extend the forecasts into long-term future by extending columns F and G. You can also obtain the interval forecasts by typing the formulas in the section on "Concept" under "A Brief Introduction to AR and ARIMA Models" in any Excel cells using the mathematical operations learned throughout this book. However, a handheld calculator is doing just as well and so no Excel application is introduced for interval forecasts."

Exercises

1. The file Hawaii.xls contains data on the number of visitor arrivals to the Big Island of Hawaii from Quarter 1, 2008, to Quarter 2, 2013. Obtain point forecasts for the next four quarters (Quarter 3, 2013, through Quarter 2, 2014) using the decomposition technique without the cyclical component.
2. Use the dataset in (1) to obtain point forecasts for the next four quarters (Quarter 3, 2013, through Quarter 2, 2014) using the decomposition technique with the cyclical component and MA(5) to de-randomize the cyclical component.
3. The file Molokai.xls contains data on oil consumption in Molokai, Hawaii. Perform the Dickey–Fuller tests on Equations 7.26, 7.27, and 7.28.
4. Use the dataset in (3) to obtain point forecasts for the next four periods by performing a regression on model (7.28) and by using the recursive principle to project the model forward.
5. The file Kauai.xls contains data on energy consumption (ENER) in Kauai for the period from February 2006 through March 2014.
   a. Estimate the AR(1) model by regress $ENER_t$ on $ENER_{t-1}$
   b. Obtain the first-period interval forecast ($m = 1$) following the formula in Equation 7.22 and the subsequent example.

PART IV

Business and Economic Applications of Forecasting

This part contains three chapters:

- Chapter 8 Business Models
- Chapter 9 Economic Models
- Chapter 10 Business Cycles and Rates of Changes

CHAPTER 8

Business Models

Sol comes to the class today with exciting news. A new customer at her Solarists store wants to sign a contract with her company in a joint production process. The customer will start a new company called Photoics that installs the photovoltaic systems for the city residents while her company will provide solar panels. For inventory planning, her boss wanted her to forecast the demand of the photovoltaic systems. Unfortunately, data on the photovoltaic demand for the city are very limited because this is a new product. How can she solve the problem?

Dr. Theo is very glad that she raises the question. He says that one of the sections in this chapter will address her problem. Several business models of the associative analyses using either nonregression or regression techniques will be introduced. Upon completion of the chapter, we will be able to:

1. Analyze the concept of operational forecasting.
2. Describe each of the financial forecast techniques.
3. Explain the two diffusion models to forecast sales and demand for a new product.
4. Apply Excel commands while forecasting the models in (1), (2), and (3).

We are looking forward to learn these practical models.

Operational Forecasting

Operational forecasting involves various departments in a company such as manufacturing, input purchases, marketing, sales, and so on. Krueger (2008) introduced a running forecast technique, of which a simplified version is summarized in this section.

Running Forecasts

Dr. Theo explains, "In a running process, forecasts are calculated for the coming multiple periods but are revised each period throughout the horizon. For example, a company can start from January and calculate forecasts for the next 12 months. The full cycle repeats by the following January. However, the forecasts are revised each month throughout the next 12 months. This is another *forecast of forecast* technique, where demand for a company's product has been forecasted using any of the techniques introduced in the previous chapters. Based on the demand forecasts, the company develops operational plans to order inputs, deliver its products, or post advertisements for its coming products."

Dr. Theo then asks, "Is there any of you who is familiar with the technique?" Arti raises her hand to share her experience. Her Artistown School usually orders books from nationwide suppliers to sell to its students. In the past, they ordered exactly the same number of books every month. Since enrolling in this class, she has been able to forecast the student demand more accurately. She now looks forward to learning this *forecast of forecast* technique so that she can efficiently place book orders. For the instructional purposes, she shows us a full dataset in the file Ch08.xls, Fig.8.3.

Dr. Theo picks the first four months of the dataset and makes a simplification by rounding the number of books to the nearest tens. This subset is displayed in Table 8.1 and can be calculated using a handheld calculator.

Table 8.1 Artistown School: balance sheet before book orders are placed

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Month	**Forecast**	**Inventory**	**Reserve**	**Balance**	**Order**	**Arrive**
December		100				
January	30	100 – 30 = 70	30 * 0.5 = 15	70 – 15 = 55	0	0
February	40	70 – 40 = 30	40 * 0.5 = 20	30 – 20 = 10	0	0
March	50	30 – 50 = –20	50 * 0.5 = 25	–20 –25 = –45	0	0
April	70	–20 – 70 = –90	70 * 0.5 = 35	–90 – 35 = –125	0	0

Arti points out that in this figure, column (2) reports her forecasts of book demands from January through April. She also says that she calculated these forecasts in the December of the previous year. In column (3) she begins with 100 books, which were the remaining inventory from December. Each of the other values in column (3) is the projected remaining inventory at the end of each month.

Column (4) represents the reserve of books for emergency when demand from the students suddenly increases. At Artistown School, the reserve is 50 percent of the monthly demand. Column (5) is the balance if no book is ordered and so no book will arrive at the studio. For this reason, the values in columns (6) and (7) are all zeros. Arti wants us to help her fill out these two columns.

We look at columns (3) and (4) of this figure and focus on the values in March to discover that there is a shortage of 20 books to satisfy the forecast demand of 50 and another shortage of 25 books for the reserve in March. Hence, the total shortage is 45 books in column (5) if no shipment arrives in March. This total shortage will rise to 125 books if no shipment arrives in April. Dr. Theo points out that the variables in columns (6) and (7) depend on the values in columns (2) through (5). Arti says her school often plans two months ahead for the number of books to be ordered because it takes six to eight weeks for the books to arrive.

We break into groups to discuss the solutions to the problem. Table 8.2 reports these forecasts of book orders to be placed in January and February. The forecasts can be made for a whole year and can be adjusted each month. Because there will be a shortage of 45 books in March, this amount needs to be ordered in January so that they can arrive by March. These 45 books will be added to the remaining inventory of 30 books from February to make an inventory of 75 books. This new book total satisfies both the demand of 50 books in March and the reserve of 25 books. Hence, the balance in March becomes zero.

In February, if the school follows the value on the last row of column (5) in Table 8.1, which is −125, and orders 125 books, then there will a surplus of 45 books, which was already ordered in January. Hence, the school needs to adjust the forecast and reduces its order to 80 books (= 125 − 45), which adds to the 25 books remaining from March to make

Table 8.2 Artistown School: balance sheet after book orders are placed

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Month	**Forecast**	**Inventory**	**Reserve**	**Balance**	**Order**	**Arrive**
December		100				
January	30	100 – 30 = 70	30 * 0.5 = 15	70 – 15 = 55	45	0
February	40	70 – 40 = 30	40 * 0.5 = 20	30 – 20 = 10	80	0
March	50	75 – 50 = 25	50 * 0.5 = 25	25 – 25 = 0		45
April	70	105 – 70 = 35	70 * 0.5 = 35	35 – 35 = 0		80

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Month	Forecast	Inventory	Reserve	Balance		Inventory	Reserve	Balance	Order	Arrive	Total Order
2		December		103				103					
3		January	31	72	16	57		72	16	57	45		
4		February	42	30	21	9		30	21	9	80		
5		March	50	-20	25	-45		25	25	0	34	45	45
6		April	70	-90	35	-125		35	35	0	42	80	125
7		May	46	-136	23	-159		23	23	0	27	34	159
8		June	43	-179	22	-201		21.5	22	0	49	42	201
9		July	32	-211	16	-227		16	16	0	55	27	227
10		August	43	-254	22	-276		21.5	22	0	60	49	276
11		September	51	-305	26	-331		25.5	26	0	39	55	331
12		October	57	-362	29	-391		28.5	29	0	29	60	391
13		November	45	-407	23	-430		22.5	23	0	30	39	430
14		December	34	-441	17	-458		17	17	0	.	29	458
15		January	31	-472	16	-488		15.5	16	0	.	30	488

Sheet1 Sheet2 Sheet3

Figure 8.1 Artistown School: running forecasts of book orders

105 books by April, just enough to satisfy both the demand of 70 books in April and the reserve of 35 books.

We now see that we can continue to obtain our forecasts in a rolling process and so it is called the *running forecast* technique.

Excel Application

Arti reminds us that Figure 8.1 displays the full dataset of her Artistown School for 13 months and that the data and commands are available in the file Ch08.xls, Fig.8.3.

Columns B through F are equivalent to columns (1) through (5) in Table 8.1, of which columns (6) and (7) are not displayed in Figure 8.1 because they all contain zero values. Columns H through L are equivalent to columns (3) through (7) of Table 8.2, of which columns (1) and (2) are already in columns B and C in Figure 8.1. Column M is added to calculate cumulative purchases over time. Initially, only the demand forecasts and the remaining inventory in December are known. We need to perform the following steps:

In cell D3, type = D2 – C3 and press Enter
In cell E3, type = 0.5 * C3 and press Enter
In cell F3, type = D3 – E3 and press Enter
Copy and paste the formula in cells D3, E3, and F3 into cells D4 through F15
Copy and paste-special the values in cells D2 through D4 into cells H2 through H4
In cell H5, type = L5 + H4 – C5 and press Enter
Copy and paste the formula in cell H5 into cells H6 through H15
(Ignore the temporary values, which will be adjusted gradually in a running process)
Copy and paste-special the values in cell E2 through E15 into cells I2 through I15
In cell J3, type = H3 – I3 and press Enter
Copy and paste the formula in cell J3 into cells J4 through J15
(Ignore again the temporary values, which will be adjusted gradually)
In cells K3 and M5, type the number 45
In cell L5, type = K3 then press Enter
Copy and paste the formula in cell L5 into cells L6 through L15
In cell M6, type = M5 + L6
Copy and paste the formula in cell M6 into cells M7 through M15
In cell K4 type = ABS(F6) – M5 then press Enter
Copy and paste the formula in cell K4 into cells K5 through K13
(The last order will be in November for the following January)
The final results in cells J5 through J15 of Figure 8.1 should be all zeros

Financial Forecast

Financial forecasting covers any subject related to financial markets such as expected interest rates, yields to maturity in a bond market, rates of returns by holding a stock, assets and liabilities, prices and exchange rates, and so on. Since the topics are numerous, Dr. Theo says we only discuss the most frequently used models in businesses.

Bond Markets

In this section, we discuss two groups of models: The first is based on expectations theory, which assumes that an investor cares only about the expected yield to maturity (YTM) of a bond; the second group adds uncertainty to the interest rate.

Calculating Expected YTM

When a company invests in a bond market, the future payment is:

$$\text{FV} = \text{PV} * (1 + i)^T \tag{8.1}$$

where FV is the future value, PV is the present value, which is also the price of the bond, i is the interest rate or the expected YTM if the bond is held until it matures, T is the number of years. We can solve for the other variables from Equation 8.1:

$$\text{PV} = \text{FV}/(1 + i)^T$$

$$i = \left(\frac{\text{FV}}{\text{PV}}\right)^{1/T} - 1 \tag{8.2}$$

The term $(1 + i)$ is called the discount factor, and i is the interest rate, which is the expected YTM in bond markets and which is also called the rate of discount.

At this point, Fin raises his hand to offer an example. A customer came to him to invest \$200,000 in a bond market. He offered her a bond that will pay \$225,000 in three years. The customer then asks him, "So what is my expected yield per year?" He was able to calculate the YTM for the customer as follows:

$$i = \left(\frac{225{,}000}{200{,}000}\right)^{1/3} - 1 = \sqrt[3]{1.125} - 1 \approx 1.04 - 1 \approx .04 \approx 4\%$$

The customer was very happy and decided to buy the bond. Fin then says that there are two complicated cases that we cannot calculate the yield so easily. Dr. Theo asks him if he can share his experience with the class. Fin is very happy to oblige, and here is his analysis.

The first case is a fixed payment security, in which a payment on the security is the same every year so that the principal is amortized. The equation for this security is:

$$P = \frac{F}{1+i} + \frac{F}{(1+i)^2} + \ldots + \frac{F}{(1+i)^T} = F\left[\frac{1-[1/(1+i)]^T}{i}\right] \tag{8.3}$$

where P is the principal of the bond, and F is the future payment.

At this point, Dr. Theo interrupts Fin to remind us that the derivation of Equation 8.3 is in Appendix 8.A. Fin then continues with his discussion.

In this case, you cannot easily solve for i, so the best strategy is to solve for P/F:

$$\frac{P}{F} = \left[\frac{1-[1/(1+i)]^T}{i}\right] \tag{8.4}$$

Once this ratio is obtained, educational guesses and adjustments have to be made to come up with an expected YTM.

The second case is the coupon bond, which pays a regular interest payment until the maturity date, when the face value (V) is repaid. The equation for this security is:

$$\begin{aligned} P &= \frac{F}{1+i} + \frac{F}{(1+i)^2} + \ldots + \frac{F}{(1+i)^T} + \frac{V}{(1+i)^T} \\ &= F\left[\frac{1-[1/(1+i)]^T}{i}\right] + \frac{V}{(1+i)^T} \end{aligned} \tag{8.5}$$

where P is the price of the bond, F is the future payment, and V is the face value.

In this case, you cannot even solve for P/F and will have to start guessing with the original equation. For Equation 8.4 or Equation 8.5, the guess-and-adjustment process is best worked out on an Excel spreadsheet.

Dr. Theo says that it is very true and that Dr. App will show us the Excel applications later.

Interest Rate Forecasts

Dr. Theo says that in the previous section, we assume that investors only care about their returns. In reality, investors also worry about the risk incurred by a rise in interest rates, which cause the prices of bonds to fall. The uncertainty increases over time, so a term premium is added to the interest rates on long-term bonds.

To account for the risk, we have to forecast the interest rate instead of using a fixed rate. Currently, there are three models for interest rate forecasting (IRF). The first is based on growth theory, which links the interest rate to real GDP growth. The second is based on monetary theory, which links the interest rate to inflation. And the last one is based on financial theory, which links the interest rate to the volatility of the financial market.

Dr. Theo says that we will combine all three into an econometric model using multiple linear regressions. Hence, the equation is written as:

$$\mathrm{INT}_{t+1}^{i} = a_1 + a_2\mathrm{GDPG}_t + a_3\mathrm{INF}_t + a_4(\mathrm{INT}_t - \mathrm{int}_t) + e_{t+1} \quad (8.6)$$

where

INT_{t+1}^{i} = the long-term interest rate on bond *i* at time *t*

GDPG_t = the growth rate of real GDP at time *t*

INF_t = the inflation, measured by the growth rate of CPI at time *t*

INT_t = the average long-term interest rate (5–30-year U.S. Treasuries)

int_t = the short-term interest rate (usually on three-month T-bills)

$(\mathrm{INT}_t - \mathrm{int}_t)$ = the term premium, which measures the volatility of the market

We learn that we can estimate Equation 8.6 as discussed in Chapter 6 to obtain point and interval forecasts and that alternative measures of the interest rate determinants are in Orphanides and Williams (2011).

Stock Market

We are very happy to get to this section. We all think that stock markets are fascinating because of their high returns, random behavior, and competitiveness. It turns out that stock markets are not completely random. There are some patterns that help us predict the market values. Dr. Theo says that this section will introduce the capital asset pricing model (CAPM), the arbitrage pricing theory (APT) model, and the dividend discount model (DDM).

Capital Asset Pricing Model

The CAPM formulates the expected return on a particular stock as a dependent variable of the market risk premium, which is the difference between the average return to the stock market as a whole and the interest rate on a risk-free bond. First introduced by Treynor (1962), a simple version of the model can be constructed as an econometric model for a linear regression:

$$R_t^i = r_t + \beta^i (R_t - r_t) + e_t^i \tag{8.7}$$

where

R_t^i = the expected return on stock i at time t

r_t = the yield (rate of return) on a risk-free bond, often a treasury bond (T-bond)

R_t = average return to the stock market as a whole

$(R_t - r_t)$ = the market risk premium, which measures the volatility of the market

β^i = the estimated coefficient of the market risk premium

All assumptions on simple linear regressions hold. There are also five assumptions on the market:

i. The market is competitive.
ii. There are no transaction costs or taxes.
iii. The investors have all information regarding investment choices.
iv. All investors can borrow and lend without changing the interest rate.

Bollerslev, Engle, and Wooldridge (1988) turn this model into a forecast model, of which a simple version is introduced here:

$$R_{t+1}^{i} = r_t + \beta^i (R_t - r_t) + e_{t+1}^{i} \tag{8.8}$$

where R_{t+1}^{i} is the forecast value for the following period. Using this model, an expected return of a stock can be forecasted. For example, if your regression result yields $\beta^i = 0.3$, and at a particular time t you find $R_t = 11\%$, $r_t = 1\%$, then you can calculate the forecast value:

$$R_{t+1}^{i} = 1\% + 0.3 * (11\% - 1\%) = 4\%.$$

Thus, the expected return of the stock in the next period is 4 percent. The estimated coefficient of β^i reveals the direction and volatility of a stock. If $\beta^i > 0$, the stock moves in the same direction with the market, and if $\beta^i < 0$, the stock moves in the opposite direction with the market. Additionally, if $\beta^i < |1|$, the stock is less volatile than the market, and if $\beta^i > |1|$, the stock is more volatile than the market as a whole.

APT Model

The APT model is an extension of the CAPM and was developed by Ross (1976). This model allows for more than one explanatory variable. For example, the stock price of an agricultural sector might depend on climate changes and the quality of land. This can be written as an econometric model for multiple regressions:

$$R_t^{i} = r_t + \beta_1^i (R_t - r_t) + \beta_2^i X_{2t} + \ldots + e_t^{i} \tag{8.9}$$

where X is any factor that affects the expected return of the stock in question in addition to the market risk premium. Based on the forecast version for the CAPM model, the APT model can also be written as:

$$R_{t+1}^{i} = r_t + \beta_1^i (R_t - r_t) + \beta_2^i X_{2t} + \ldots + e_{t+1}^{i} \tag{8.10}$$

where R_{t+1}^{i} is again the forecast value for the following period. The regression results then can be used to calculate the expected return of the stock in the following period.

Ex raises his hand and offers an example: His company's stock depends positively on the income (INC) of its trading partners and negatively on the profits (PRO) of its trading competitors. Performing a regression, he finds that $\beta^i_{\text{INC}} = 0.4$ for the INC coefficient and $\beta^i_{\text{PRO}} = 0.2$ for the PRO coefficient. This year, the growth rate of INC is 3 percent and the growth rate of PRO is 2 percent. Using $R_t = 11\%$, $r_t = 1\%$, and $\beta^i = 0.3$, he forecasts the expected return as:

$$R^i_{t+1} = 1 + 0.3 * (11\% - 1\%) + 0.4 * 3\% - 0.2 * 2\% = 4.8\%$$

We are very impressed with Ex's example, which gives us a feel of a real-life situation. Dr. Theo then moves to the next model.

Dividend Discount Model

The DDM utilizes the present value formulas in the section on "Bond Markets" to calculate a company's expected price based on the investment value theory by Williams (1938). First introduced by Gordon and Shapiro (1956), the model was modified by Gordon (1959) and so was often called the Gordon growth model. The main idea is that a company's stock price is worth the sum of all of its future dividend payments. The equation for calculating the expected price of a stock is:

$$P = \frac{D_1}{1+i} + \frac{D_2}{(1+i)^2} + \ldots + \frac{D_T}{(1+i)^T} \tag{8.11}$$

where

P = the expected price of a company's stock

D_T = the expected dividend paid by the company at the end of time T

T = the number of time periods

i = the investor's discount rate

The investor's discount rate is subjective. For example, if you want to receive at least 5 percent return from investing in any security, then your discount rate is 5 percent.

If the actual stock price exceeds the expected one, the stock is overvalued. In this case, you might want to move some of your shares in this stock to a different stock. If the actual stock price is below this expected price, the stock is undervalued, and you can predict that the price of this stock will rise. Therefore, it might be profitable to buy some shares of the stock. For this reason, the expected stock price is also called the intrinsic value or fundamental value of a stock.

Assuming that the company's profits and subsequent dividends are growing at a constant rate over time:

$$P = \left(\frac{1+G}{i-G}\right) * D_0 \tag{8.12}$$

where
D_0 = the initial dividend
G = growth rate of the company's profits (Π) and dividends (D)

D_0 is known at the beginning, and G can be forecasted using any technique introduced in the previous chapters.

At this point, Cita asks, "How can you obtain Equation 8.12 from Equation 8.11?" Dr. Theo says that the derivation is in Appendix 8.B.

Sol then gives an example: As a part of her retirement plan, her company offers her either investing in the company stock or investing in a savings account that pays 0.3 percent per quarter. The initial quarterly dividend paid by her company four quarters ago is $2 per share, her discount rate is 1 percent per quarter, and the average growth rate of her company's profits over the past four quarters has been 0.5 percent per quarter. We are able to calculate the expected price of her company's stock as:

$$P = \left(\frac{1+0.5}{1-0.5}\right) * \$2 = \$6.00$$

Dr. Theo then tells us to log on to *The Wall Street Journal* website and look at her company stock price. We find that it is listed on the stock exchange for $2.50 today. Thus, we know that the stock is undervalued, and there is a high probability that its price will rise in the future. Sol is very excited and decides to choose the stock option without delay.

To conclude the section, Dr. Theo reminds us that the assumption of a constant growth rate for both profits and the dividends might not hold in the long run. Thus, the errors might be large, and it is safer to use this model only for short-run forecasts.

Excel Applications

Dr. App reminds us that Figure 8.2 provides a demonstration for bond markets. Columns A through D display calculations for Equation 8.4, and columns F and G display calculations for Equation 8.5.

Expected YTM

Alte offers an example from her Alcorner for the case of fixed payments: A business loan of $25,000 is provided by a local bank to her business. She will repay the bank a fixed amount of $5,000 each year for six years. Dr. App tells us to find out the bank's YTM at the end of the sixth year. We find the data in the file Ch08.xls, Fig.8.4, and proceed as follow:

In cell C2, type = A2/B2 and press Enter (hence $P/F = 5$ as shown in cell C2)
(Try the first calculation with any interest rate, e.g., $i = 10\% = 0.1$)
In cell D2, type = (1 − (1/(1 + 0.1))^6)/0.1 and press Enter
Copy and paste the formula in cell D2 into cells D3 through D9 so that you can try various rates
(The answer, $P/F = 4.35526$, is too low, so try to reduce i to 0.09)
Double click on cell D3 and change 0.1 to 0.09 and press Enter
Continue to reduce i by 0.01 gradually from cells D4 through D6 where you will see the value 4.91732 which is close to 5
In cell D7, change i from 0.06 to 0.055 and press Enter
In cell D8, change i further to 0.054 and press Enter
In cell D9, change i further to 0.0545 and press Enter

G9 fx =(1200*(1-(1/(1+0.0721))^5)/0.0721)+(20000/(1+0.0721)^5)

	A	B	C	D	E	F	G
2	25,000	5,000	5	4.355260699		19,000	16967.37058
3				4.48591859			17666.20924
4				4.622879664			18402.91599
5				4.76653966			19179.96051
6				4.917324326			18786.23453
7				4.995530309			18942.45737
8				5.01140422			19021.19987
9				5.003457456			19013.30657

Sheet1 Sheet2 Sheet3

Figure 8.2 YTM of the fixed payment security and the coupon bond

Now you obtain roughly 5.0035, which is at $i = 0.0545$
Hence, the bank's YTM is 5.45%

Cita provides an example for the case of a coupon bond: She considers buying a bond that pays $1,200 per year for five years and then repays the face value of $20,000 at the end of the fifth year. The price of the bond is $19,000. She asks us to find the YTM for her.

We try the first calculation again with an arbitrary interest rate $i = 10\% = 0.1$. We find the data in columns F and G of the file Ch08.xls, Fig.8.4, and proceed as follows:

In cell G2, type = (1200 * (1 − (1/(1 + 0.1))^5)/0.1) + (20000/((1/(1 + 0.1))^5) and press Enter
Copy and paste cell G2 into cells G3 through G9
Continue to reduce i by 0.01 gradually from cells G4 through G9
Cell G9 shows roughly 19,013, and the calculation bar shows $i = 0.0721 = 7.21\%$

Interest Rate Forecasts

Dr. App instructs us to perform a regression of model (8.6) using the data in the file Ch08.xls, Fig.8.5, and following the commands in the previous chapters. Figure 8.3 displays a section of the regression results and the forecasts for the interest rate on a 10-year U.S. Treasury Bond. We choose the output range of G21 to place the regression results close to cells B36 because the command for the forecasting starts in this cell. From this figure, the estimated equation is:

$$\begin{aligned} \text{INT}^i_{t+1} = 0.2307 + 0.4726 * \text{GDPG}_t + 1.0971 * \text{INF}_t \\ + 0.8282 * (\text{INT}_t - \text{int}_t) \end{aligned}$$

We find that the forecasts of the explanatory variables are already calculated by Dr. App for our convenience. We only have to perform the following steps to obtain the forecasts:

In cell B36, type = \$H\$37 + \$H\$38 * C36 + \$H\$39 * D36 + \$H\$40 * E36 then press Enter

Copy and paste this formula into cells B37 through B40

	A	B	C	D	E	F	G	H	I	J	K
34	2012	1.8	2.14	0.87	1.71		Total	33	328.8944		
35	2013	2.35	2.45	0.09	2.29						
36	2014	3.927868	2.35	0.09	3.003888			*Coefficients*	*andard Err*	*t Stat*	*P-value*
37	2015	5.140786	2.48	1.10	3.056328		Intercept	0.230655	1.18478	0.194682	0.846955
38	2016	5.544367	2.87	1.12	3.294577		$GDPG_t$	0.472619	0.169194	2.793361	0.008998
39	2017	6.380664	2.96	1.13	4.239747		INF_t	1.097075	0.14597	7.515745	2.23E-08
40	2018	6.725954	3.12	1.15	4.538865		$(INT_t\text{-}int_t)$	0.828201	0.341804	2.423026	0.021637

Fig.8.3 Fig.8.4 **Fig.8.5** Fig.8.6 APT Fig.8.7 Fig.8.8

Figure 8.3 Forecasting interest rates: regression results and forecasts

Data Source: Federal Reserve.com (2014).

	F	G	H	I	J	K	L	M	N
5		Adjusted R Square	0.33983753						
6		Standard Error	0.68825243						
7		Observations	21						
8		ANOVA							
9			*df*	*SS*	*MS*	*F*	*gnificance F*		
10		Regression	1	5.350615567	5.350616	11.29557	0.003282		
11		Residual	19	9.000136647	0.473691				
12		Total	20	14.35075221					
13			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
14		Intercept	-0.14144304	0.182518213	-0.77495	0.447905	-0.52346	0.240572	
15		(Rt - rt)	0.61506599	0.183006952	3.360889	0.003282	0.232028	0.998104	

Fig.8.3 Fig.8.4 Fig.8.5 **Fig.8.6** APT Fig.8.7 Fid

Figure 8.4 Forecasting expected return using CAPM model: regression results

Forecasting Stock Markets

Mo offers us daily data on the prices of his company Motorland stock (MOT). He also shares with us daily data on the market risk premium. The data are for the period from May 1 to May 30, 2014, and are available in the file Ch08.xls, Fig.8.6. We perform a regression of model (8.8) for the CAPM model and display a section of the regression results in Figure 8.4.

The estimated coefficient of β^i is 0.6151, so the estimated equation is:

$$R^i_{t+1} = -0.1414 + 0.6151 * (R_t - r_t)$$

We then use this equation to calculate the expected returns for May 29 through June 1, 2014: $R_{5/29}$(MOT) = 0.5607, $R_{5/30}$(MOT) = 0.5939, and $R_{6/1}$(MOT) = 0.5380. The values $R_{5/29}$(MOT) and $R_{5/30}$(MOT) are for evaluation by comparing them to the actual values.

We see that the results reveal a mean absolute percentage error (MAPE) of roughly 30 percent, which is too large. Dr. App says, "This is the very reason that the model needs to be extended to allow more variables than just the market risk premium." The regression results also show that the MOT moves in the same direction with the general market ($\beta^i > 0$), and the former is less volatile than the latter ($\beta^i < 1$).

We then forecast this stock market using the APT model as shown in (8.10). Since the motorcycle production also depends on aluminum and petroleum prices, these two variables are added to the econometric model with MET as the average daily price of pressed metal companies and PETRO as the average daily price of petroleum companies on the stock exchange. The regression results, which are shown in the file Ch08.xls, on the sheet APT, yield the estimated equation as $R^i_{t+1} = -0.0376 - 0.1724\ \text{MET}_t + 0.6519 * (R_t - r_t) - 0.4847\ \text{PETRO}_t$.

Substituting the coefficient estimates into this equation, we obtain the forecasts for May 29 through June 1, 2014: $R_{5/29}$(MOT) = 0.7350, $R_{5/30}$(MOT) = 0.6915, and $R_{6/1}$(MOT) = 0.8558. The results reveal a MAPE of roughly 20 percent, which is an improvement compared to the results using the CAPM model. Since stock markets are more volatile than bond markets, this forecast result can be deemed as acceptable.

The estimated coefficients for the two regressions of models (8.8) and (8.10) are quite similar, except that the coefficient for the market risk premium in (8.10) is no longer statistically significant. This implies that the market risk premium might not be the most important determinant of the expected return on the MOT. However, an *F*-test reveals that the estimated coefficients are jointly significant, so we should not exclude the variable market risk premium from model (8.10) in the regressions and calculations of the forecasts.

Finally, we come to the section that discusses the production adoption process needed for Sol's forecasts of the photovoltaic demand in the city.

Diffusion Models on Sales and Demand

We learn that the first diffusion model was introduced by Rogers (1962), refined by Bass (1969), and extended by Lawrence and Lawton (1981)

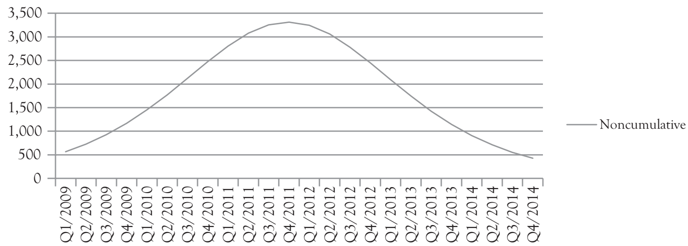


Figure 8.5 Noncumulative distribution of the product adoption process

to forecast the process of how a new product will be adopted in a population. The models allow a forecaster to generate an entire product life cycle from a few data observations such as the number of the previous buyers and the total market size. Key parameters and their changes can be calculated to obtain the potential demand of a new product prior to the existence of historical sale data.

Sol is very happy that she will not need to collect a long series of the historical data to analyze photovoltaic demand. She shares with us one of the datasets from various cities in the nation to examine the pattern of the product adoption process. We chart the data from the file Ch08.xls, Fig.8.7, and display the plot in Figure 8.5.

From this figure, we see that the adoption process has a noncumulative distribution that follows a nearly normal distribution. Dr. Theo tells us that this is the model's assumption, which might not hold, so the diffusion models usually produce large errors. From this assumption, the cumulative distribution of the adoption is in the form of an S curve. We also chart the cumulative distribution using the data from the file Ch08.xls, Fig.8.8, and display the plot in Figure 8.6.

From this figure, we see that the cumulative distribution of the adoption process is indeed in the form of an S-curve for this case.

Bass Model

The Bass model is based on the assumption that the timing of a consumer's initial purchase is related to the number of previous buyers.

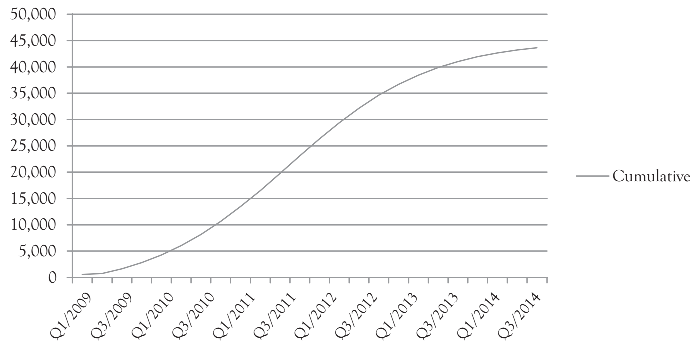


Figure 8.6 Cumulative distribution of the product adoption process

Adopters are classified into various categories depending on the timing of their adoptions. The adopters who make the decision to adopt a product early and independently of other adopters are called the innovators. Bass (1969) lists five classes of adopters from the earliest to the latest: (1) innovators, (2) early adopters, (3) early majority, (4) late majority, and (5) laggards. The innovators and the early adopters are the ones who create a diffusion process that results in purchasing by the later adopters. All classes of adopters after innovators are considered imitators.

$P(T)$ is the probability that an initial purchase will be made at time T given that no purchase has yet been made. Then a function with $P(T)$ as the dependent variable can be written as:

$$P(T) = p + (q/m) * N(T) \tag{8.13}$$

where

P = the coefficient of innovation, which depends on factors affecting innovators

q = the coefficient of imitation, which depends on factors affecting imitators

m = the total market potential during the considered period

$N(T)$ = is the number of previous buyers up to time T

Factors affecting innovators can be advertisements or personal preferences. Factors affecting imitators can be interaction with the innovators or peer pressure. Extended surveys in the past have shown that the average value of p often falls between 0.01 and 0.03, and that the average value of q often falls between 0.3 and 0.5.

Let $f(T)$ be the likelihood of a purchase at time T, then sales at time T is:

$$S(T) = mf(T) = P(T) * [m - N(T)] \tag{8.14}$$

Combining Equations 8.13 and 8.14 yields:

$$S(T) = [p + (q/m) * N(T)] * [m - N(T)], \text{ so:}$$

$$S(T) = p\, m + (q - p) * N(T) - (q/m) * [N(T)]^2 \tag{8.15}$$

For example, suppose the total market potential for the iOS8 phones in San Francisco is 120,000 people, the number of previous buyers in the second quarter of 2014 is 10,000 people, quarterly $p = 0.02$, and quarterly $q = 0.4$. Then the potential number of sales in the third quarter is:

$$\begin{aligned} S(Q3) &= (0.02) * 120 + (0.4 - 0.02) * (10) - (0.4/120) * 10^2 \\ &= 2.4 + 3.8 - 0.3333 = 5.8667 \text{ (in thousands)} \\ &= 5{,}8667 \text{ (iOS8 phones).} \end{aligned}$$

Lawrence–Lawton Model

Dr. Theo reminds Sol to take notes carefully on this model because it is more refined than the original Rogers and Bass models, so it will help her obtain forecasts of a whole production cycle with only a few observations. Lawrence, Klimberg, and Lawrence (2009) analyze in details five steps of the diffusion process, which is summarized here:

i. Awareness: The potential buyers become aware of the innovation.
ii. Interest: The potential buyers seek additional information.
iii. Evaluation: Enough information is gathered for judgments.
iv. Trial: Samples of the innovation are provided for trying out.
v. Adoption: The adoption process takes place.

The model introduced by Lawrence and Lawton (1981) is for a cumulative unit of sales, $S(T)$, to the end of period T:

$$S(T) = \frac{N + N_0}{[1 + (N / N_0)e^{-p_d T}]} - N_0 \quad (8.16)$$

where N_0 = cumulative number of adopters at time T_0
N = total market potential buyers
p_d = a diffusion-rate parameter, which is the speed that the new idea spreads from one consumer to the next

We break into groups to work on this example: If $p_d = 0.4$, there is a 40 percent possibility that one consumer will tell another consumer in the market about the new product. Suppose the total market for a new brand of camcorders in Korea is 10,000 people, the cumulative number of adopters at the beginning is 1,000 people, the time horizon is two years, and the yearly diffusion rate $p_d = 0.5$, then:

$$S(T) = \frac{10+1}{[1+(10/1)e^{-0.5*2}]} - 1 = \frac{11}{1+10*e^{-1}} - 1 = \frac{11}{1+10/2.7183} - 1$$
$$= \frac{11}{4.6788} - 1 = 2.351 - 1 = 1.351$$

Hence, the cumulative unit of sales for the two-year period is 1,351 camcorders.

In reality, we have to estimate N_0 based on the trial sale of the first period $S_1 = S(1)$, that is, the number of sales is the same as the cumulative sales in the first period:

$$N_0 = \frac{\mathrm{NS}_1 e^{-p_d}}{[N(1 - e^{-p_d}) - S_1]} \quad (8.17)$$

The number of sales in the following periods is:

$$S_2 = S(2) - S(1),$$
$$S_3 = S(3) - S(2), \ldots,$$
$$S_T = S(T) - S(T\text{–}1) \quad (8.18)$$

For example, suppose yearly $p_d = 0.7$, $N = 400{,}000$ people; $S_1 = S(1) = 35{,}000$ for a new brand of computers initially sold in Shanghai, then you can calculate N_0 and $S(2)$:

$$N_0 = \frac{400{,}000 * 35{,}000 * e^{-0.7}}{[400{,}000 * (1 - e^{-0.7}) - 35{,}000]} = 41{,}789$$

$$S(2) = \frac{400{,}000 + 41{,}789}{[1 + (400{,}000 / 41{,}789)e^{-0.7\times 2}]} - 41{,}789 = 89{,}680$$

That is, the cumulative number of sales for the two-year period is 89,680 and hence the sale forecast for the second year alone is:

$S_2 = S(2) - S_1 = 89{,}680 - 35{,}000 = 54.680$ new computers

Excel Application

Since the Bass model can be easily calculated using the mathematical operations for Excel introduced in Chapter 1, Dr. App only provides us with one Excel application for the Lawrence–Lawton model here. Photovoltaic electricity came into existence recently and has developed rapidly in the city. Sol shares with us her data on the photovoltaic permits issued by the city government. Figure 8.7 reveals that the total market potential is $N = 55{,}000$ in cell C2.

	A	B	C	D	E	F
1	Quarter	T	N	p(d)	S_T (Cumulative)	PV Permit
2	Q1/2014	1	55,000	0.150000	1,766	1,766
3	Q2/2014	2	14,181		3,680	1,914
4	Q3/2014	3			5,739	2,058
5	Q4/2014	4			7,932	2,193
6	Q1/2015	5			10,247	2,315
7	Q2/2015	6			12,666	2,419
8	Q3/2015	7			15,168	2,502
9	Q4/2015	8			17,727	2,559
10	Q1/2016	9			20,316	2,589
11	Q2/2016	10			22,906	2,590
12	Q3/2016	11			25,468	2,562
13	Q4/2016	12			27,975	2,507

Sheet1 Sheet2 Sheet3

Figure 8.7 Forecasting photovoltaic sales

Since a photovoltaic company does not apply for a permit from the county until a contract with a home owner is signed, this number of permits is a good proxy for the market sales and demand and can be used to forecast future sales and demand.

The sale number in the first quarter of 2014 is the number of permits issued from January 2 to March 30, $S_1 = S(1) = 1{,}766$, and is displayed in cell E2. Since this is the first period used in the forecast, it is also the first quarterly sale value, which is displayed in cell F2. A survey of the early adopters in the city reveals the quarterly diffusion rate parameter $p_d = 0.15$, which is displayed in cell D2.

We find the data in the file Ch08.xls, Fig.8.9, and proceed as follows:

In cell C3, type = (C2 * E2 * (EXP(–D2)))/(C2 * (1 – (EXP(–D2))) – E2) and press Enter
(this is the formula to calculate N_0)
In cell E3, type = (($C$2 + $C$3)/(1 + ($C$2/$C$3) * EXP(–$D$2 * B3))) – $C$3 and press Enter
(this is the formula to calculate $S(2)$ the cumulative units of sales)
In cell F3, type = E3 – E2 and press Enter
(this is the formula to calculate S_2, the forecasted sales for quarter 2)
Copy and paste the formulas in cells E3 and F3 into cells E4 through F13
The forecasts up to the fourth quarter of 2016 are in cells F3 through F13

Dr. App then tells us that monitoring and adjustments are crucial because the actual data provided by Sol are limited to a quarter, so evaluations for long-term forecasts cannot be performed. She confirms Dr. Theo's remark that the model might produce very large errors if the assumption of nearly normal distribution is violated.

Exercises

1. A CD store in Hong Kong has the forecasts on its CD demand displayed in Table 8.3.

   Use an Excel spreadsheet or a handheld calculator to fill in the blank spaces and construct another table similar to Figure 8.1 with the reserve values equal 50 percent of the forecast values.

Table 8.3 Forecasts for CD demand

Week	Forecast	Inventory	Reserve	Balance	Order	Arrive
Week 1		90				
Week 2	20					
Week 3	30					
Week 4	45					
Week 5	55					

2. A company considers buying a coupon bond that pays $600 per year for six years and then repays the face value of $10,000 at the end of the sixth year. The price of the bond is $9,200. Use an Excel spreadsheet to find out the YTM of this bond.
3. A regression on the return of an automobile company (AUTO) stock on the market risk premium and the changes in oil prices of a petroleum company (PETRO) yields $\beta_1^i = 0.75$ for the coefficient of the market risk premium and $\beta_2^i = -0.2$ for the coefficient of PETRO. Additionally, the average return on the stock market as a whole is 9 percent yearly, the average yield on a three-year T-bond is 1 percent yearly, and the growth rate of PETRO price is 4 percent yearly. Use a handheld calculator to forecast the expected return on AUTO stock in the next period.
4. Suppose the total market potential for a new Galaxy tablet in Dubai is 9 million people, the number of previous buyers in the third quarter of 2014 is 1 million people, quarterly $p = 0.02$, and quarterly $q = 0.4$ for the Bass model. Use a handheld calculator to forecast the potential number of sales in the fourth quarter of 2014.
5. The Lawrence–Lawton model:
   a. A town has the cumulative number of adopters of the iCloud at time $T_0 = 1300$. The market size = 12,000 and the diffusion rate parameter = 0.6. Forecast the yearly cumulative unit of sales up to the end of year 2.
   b. Given $N = 400{,}000$ people in a city, $S_1 = 30{,}000$ iClouds, $p_d = 0.7$, use a handheld calculator to calculate N_0, forecast cumulative sales $S(2)$, and sale forecast in year 2, S_2.
   c. Use an Excel spreadsheet to calculate $S(1) = S_t$ and annual sale forecasts S_T for $T = 3$ through $T = 15$.

Appendixes

The following appendixes provide the derivations of Equations 8.3 and 8.12 in the section on "Financial Forecast."

Appendix 8.A Deriving Equation 8.3

$$P = \frac{F}{1+i} + \frac{F}{(1+i)^2} + \ldots + \frac{F}{(1+i)^T} = F\left[\frac{1}{1+i} + \frac{1}{(1+i)^2} + \ldots + \frac{1}{(1+i)^T}\right]$$

$$= \text{FS} \qquad \text{(A.1)}$$

where

$$S = \left[\frac{1}{1+i} + \frac{1}{(1+i)^2} + \ldots + \frac{1}{(1+i)^T}\right]$$

Let $x = \dfrac{1}{1+i}$ then $S = x + x^2 + \ldots + x^T$ (A.2)

Multiplying both sides of Equation A.2 by (1–x) yields:

$$S(1-x) = (1-x)(x + x^2 + \ldots + x^T) = 1 - x^2 + x^2 - \ldots - x^{T+1} = 1 - x^{T+1}$$

$$S = \frac{x}{1-x} - \frac{x^{T+1}}{1-x} = \frac{1/1+i}{1-(1/1+i)} - \frac{(1/1+i)^{T+1}}{1-(1/1+i)} = \frac{1}{1+i-1} - \frac{(1/1+i)^T}{1+i-1}$$

(A.3)

$$S = \frac{1-(1/1+i)^T}{i} \qquad \text{(A.4)}$$

Combining Equations A.4 and A.1 yields:

$P = \text{FS} = F\left[\dfrac{1-[1/(1+i)]^T}{i}\right]$, which is the expression in Equation 8.3.

Appendix 8.B Deriving Equation 8.12

$$P = \frac{D_1}{1+i} + \frac{D_2}{(1+i)^2} + \ldots + \frac{D_T}{(1+i)^T} \tag{B.1}$$

Because the growth rate is G,

$$D_1 = D_0(1+G); \quad D_2 = D_1(1+G) = D_0(1+G)^2 \tag{B.2}$$

Combining Equations B.1 and B.2 yields:

$$P = \frac{D_1}{1+i} + \frac{D_2}{(1+i)^2} + \ldots + \frac{D_T}{(1+i)^T}$$

$$= D_0\left[\frac{1+G}{1+i} + \left(\frac{1+G}{1+i}\right)^2 + \ldots + \left(\frac{1+G}{1+i}\right)^T\right] = D_0 S \tag{B.3}$$

Let $x = \frac{1+G}{1+i}$ then $S = x + x^2 + \ldots + x^T$ (B.4)

From Equation A.3 $S = \frac{x}{1-x} - \frac{x^{T+1}}{1-x} = \frac{x}{1-x}$ when T approaches infinitive, so Equation B.4 becomes:

$$S = \frac{x}{1-x} = \frac{(1+G)/(1+i)}{1-(1+G)/(1+i)} = \frac{1+G}{1+i-1-G} = \frac{1+G}{i-G} \tag{B.5}$$

Combining Equations B.3 and B.5 yields:

$P = D_0 S = \left(\frac{1+G}{i-G}\right) * D_0$, which is the expression in Equation 8.12.

CHAPTER 9

Economic Models

Cita has just returned from an important meeting with the city council. The city recently established an Economic Competitiveness Office with Cita as the chair. The office is in charge of helping the private firms in three regards: (1) determining their production allocations based on the profit maximization or cost minimization principle subject to uncertainty in resources and demands, (2) calculating the changes in their input and final-good demands due to changes in technologies, and (3) making decisions on their rental prices of land to guarantee a competitive market for city land use. Cita is worried because her knowledge of economic theories is only at undergraduate level. Dr. Theo assures her that several models based on economic theories will be discussed this week. Once we finish with the chapter, we will be able to:

1. Explain the product allocation model based on economic principles.
2. Analyze changes in input and final-consumption demands of a product.
3. Develop a model for a private firm on land-use forecasts.
4. Calculate trip-distribution forecasts for the city.
5. Apply Excel while analyzing the topics in (1), (2), (3), and (4).

Most of us do not know much about economic theories, so we all look forward to gaining new knowledge.

Production Forecasts

We learn that the two production models are production allocations and input–output models. The production–allocation model helps firms forecast how much of each good to produce in the near future subject to uncertainty in the resources. The input–output model has many applications. In this chapter, it is used to forecast what will happen to the total input demand for a good if one of the technical coefficients changes and what will be the changes in the demand for final consumption of a good if one of the inputs changes.

Production Allocations

Given its historical data and interval forecasts on labor, capital, and market demand, a firm needs to forecast how much of each good it should produce the next period so that it can order inputs for its production and make a delivery plan of its production to markets. The firm is operating based on the profit maximization or cost minimization principle.

We learn that the model was first introduced by Leonid Kantorovich (1940), a Soviet mathematician and economist, during World War II to plan military expenditures based on the cost minimization principle.

The classical model for production allocations comprises three parts:

1. A linear function to be maximized (such as profits) or minimized (such as costs)
2. A set of constraints such as labor, capital, land, natural resources, and so on
3. A set of production limits that depend on market demand or quotas

The only difference between the preceding model and a forecast model for production allocations is that the forecast model faces uncertainty in resources or demand or both. Hence, this allocation forecast problem can be solved using linear programing, also called linear optimization. At this point, Dr. Theo asks our class to share our experiences if we have any. Mo says that his cousins run a factory in Europe. He puts forth a problem regarding the factory and shows us how it can be solved using a handheld calculator before using Excel.

The Ammonia Division of his cousins' fertilizer company in Europe produces ammonium sulfate (*SU*) and ammonium salts (*SA*). Their historical data reveal that the profit rates of the two products are different:

$$\text{Profit per unit (euros per ton)} \quad SU \quad €40$$
$$SA \quad €25 \qquad (9.1)$$

Their production time is also different:

Time (minutes per ton): *SU* Two minutes
SA One minute (9.2)

Additionally, the maximum limits of weekly production quantities depend on the quotas set by the provincial authority, which faces uncertainty on the quota for SA:

Maximum limits (tons): *SU* 5,500
SA 6,600–7,200 (9.3)

Another uncertainty is that there might be 120 to 150 worker hours available for the Ammonia Division four weeks from this week. Given this information, their task is to forecast how many tons of each product they should produce four weeks from now to maximize the division's total profit.

Dr. Theo is quite pleased and tells us to work on the problem. He also decides to call the quantities of ammonium sulfate X tons and that of ammonium salts Z tons for easy applications in Excel later. Based on Equations 9.1 through 9.3, he instructs us to perform the following steps:

1. The linear function: Since the profits per unit are €40 for ammonium sulfate (X) and €25 for ammonium salts (Z), we are able to write the linear function for the profit maximization as:

$$\text{Maximize } 40X + 25Z$$

2. The constraints: Because 120 to 150 worker hours are available, we need to write an equation for the total worker hours, which equal the hours to make a unit of each product times quantities. We then change minutes to hours and write the constraint as follows:

   Time (hours per ton): For *SU* (X) Two minutes = 1/30 hours per ton
   For *SA* (Z) One minute = 1/60 hours per ton

   Constraints (hours): $(1/30)X + (1/60)Z \leq 150$; or
   $(1/30)X + (1/60)Z \leq 120$

Dr. Theo tells us that the constraints can be written as $120 \le [(1/30)X + (1/60)Z] \le 150$. Here they are written separately for Excel Solver in the next section on “Excel Application”.

3. Production limits: We combine the aforementioned maximum limits set by the provincial authority and the nonnegative conditions for the quantities:

$$0 \le X \le 5{,}500$$
$$0 \le Z \le 7{,}200; \quad \text{or } 0 \le Z \le 6{,}600$$

In sum, the problem is:

Maximize $40X + 25Z$

$$\text{Subject to } (1/30)X + (1/60)Z \le 150; \ \text{ or}$$
$$(1/30)X + (1/60)Z \le 120 \qquad (9.4)$$
$$0 \le X \le 5{,}500; 0 \le Z \le 7{,}200; \ \text{ or } 0 \le Z \le 6{,}600$$

Using a handheld calculator, we first calculate the profit per hour for *SU* ($\prod_{SU}$) and *SA* ($\prod_{SA}$). From the constraint, one unit of *SU* (*X*) is produced in 1/30 hours whereas one unit of *SA* (*Z*) is produced in 1/60 hours, so profit per hour for each product is:

$$\Pi_{SU} = \frac{1}{1/30}(\text{ton / hr}) * 40 \text{ (euro/ ton)} = 30 \text{ (ton / hr)} * 40 \text{(euro/ ton)}$$
$$= 1200 \text{ (euro/ hr)}$$
$$\Pi_{SA} = \frac{1}{1/60}(\text{ton / hr}) * 25 \text{ (euro/ ton)} = 60 \text{ (ton / hr)} * 25 \text{(euro/ ton)}$$
$$= 1500 \text{ (euro/ hr)}$$

Hence, *SA* is more profitable to produce than *SU*, and Mo’s cousins want to produce up to its maximum limit of 7,200 units, which require 120 worker hours (= 7,200/60), or a maximum of 6,600 units, which requires 110 worker hours (= 6,600/60).

If they have 150 worker hours, then the remaining 30 to 40 worker hours is for *SU*, which will come off at 900 to 1,200 units (= 30 tons per hour * 30 hours, or = 30 tons per hour * 40 hours). If they have 120 worker hours, then the remaining 0 to10 hours is for *SU*, which will come off at 0 to 300 units. In brief, their forecasts for four weeks from now are:

Ammonium salts: between 6,600 and 7,200 units

Ammonium sulfate:

With 150 worker hours: between 900 and 1,200 units

With 120 worker hours: between 0 and 300 units

"Wow, that is a long problem," we exclaim. Dr. Theo says that if we have many constraints, it will be complicated to solve by hand, so Excel is very convenient.

Input–Output Demands

Dr. Theo says there is another production forecast model based on input–output demand. Although the idea of linking various input sectors to final output sectors goes back to the 19th century, Leontief (1986) introduced the modern version that has been used widely at the present time. The model is realistic because each sector in the economy produces goods that are used as both inputs and final demands. For example, the agriculture sector provides apples and oranges as final goods for the consumers, but also as inputs for the manufacturing sector to produce apple and orange juices.

Given an economy with n sectors, each sector produces a good i that can be used as inputs for several sectors in addition to being used as a final good. Let x_{ij} be the quantity of output of good i used by sector j and x_j the product of sector j; then the technical coefficient for good j is defined as:

$$a_{ij} = x_{ij} / x_j$$

Let y be the input demands for $j = 1, 2, \dots\ n$ sectors. If the economy can produce m inputs, then:

$$\begin{aligned} y_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{m1}X_{m1} \\ &\dots\dots\dots\dots\dots\dots\dots\dots \qquad \text{, where } X_i \text{ is the total output of good } i. \\ y_n &= a_{1n}X_1 + a_{2n}X_n + \dots + a_{mn}X_{mn} \end{aligned} \qquad (9.5)$$

If A_i is the collection of all a_{ij} used as inputs for the n sectors and F_i the final consumption of good i, then the total output of good i is:

$$X_i = A_i X_i + F_i \tag{9.6}$$

The input–output model has numerous applications. Dr. Theo tells us that we only address a simple problem of two goods that do not require the knowledge of matrix operations. He encourages the students who are interested in a full input–output model to take a course on the general equilibrium computational model. For the problem of two goods, the equations are

$$y_i = a_{i1}X_1 + a_{i2}X_2$$

$$F_i = X_i - y_i \text{ for } i = 1 \text{ and } 2$$

Cita has carried out research on the subject and shares an example on two sectors in the American economy with the class. The dataset for the output of the two sectors, agriculture (A) and manufacturing (M), is from the Bureau of Economic Analysis website:

Total output (billions of dollars)

Agriculture: $X_A = 420$

Manufacturing: $X_M = 5{,}419$

Their estimated technical coefficients are reported as follows:

	Output 1: Agriculture (A)	**Output 2: Manufacturing (M)**
Input 1: Agriculture (A)	0.11	0.05
Input 2: Manufacturing (M)	0.22	0.27

We first calculate the total input demand for each sector:

$$y_A = A_A X_A = 0.11 * 420 + 0.05 * 5{,}419$$

$$= 46.2 + 270.95 = 317.15 \text{ (\$ billions) = input demand for } A.$$

$$y_M = A_M X_M = 0.22 * 420 + 0.27 * 5{,}419$$

$$= 92.4 + 1463.13 = 1555.53 \text{ (\$ billions) = input demand for } M.$$

We also calculate the original demand for final consumption of the agricultural product:

$$F_A = X_A - y_A = 420 - 317.15 = 102.85 \text{ (\$ billions)}$$

Dr. Theo then asks us to recalculate the total input demand and the demand for final consumption of the agricultural product if the technical coefficient for the agricultural product used in the manufacturing sector falls 10 percent, that is, from 100 percent to 90 percent.

We recalculate the new total input demand for the agricultural product:

$$y_{A'} = 0.11 * 420 + 0.05 * (0.90) * 5419$$
$$= 46.2 + 243.86 = 290.06 \text{ (\$ billions)}$$

We are also able to recalculate the new demand for the final consumption of the agricultural product:

$$F_{A'} = X_A - y_{A'} = 420 - 290.06 = 129.94 \text{ (\$ billions)}$$

Thus, the change in final consumption of the agricultural product is 27.09 billion dollars (= 129.94 – 102.85).

Excel Application

Dr. App tells us that the exercise in the section on "Input–Output Demands" can be repeated using Excel mathematical operations. However, they are simple enough to solve using a handheld calculator. Regarding large matrices of input–output, we will need special software and so they are beyond the scope of this book. Hence, this section only uses Excel to solve the same problem in the section on "Production Allocations." We can add as many constraints as we wish to solve without too much trouble compared to solving by hand. First, we need to install another "Add-Ins" tool.

For Microsoft Office (MO) 1997–2003:

Go to Data Tools, click on Add-Ins from the drop down menu
Click on Solve Add-in option from the new drop down menu and click OK
Whenever you need this tool, click on Data Tools and then click on Solver

For MO 2007:

Click on the Office logo at the top left that you have to click on to open any file
Click on Excel Options at the bottom center
Click on Add-Ins from the menu at the bottom of the left column in the Excel Options window
The View and Manage Microsoft Office Add-Ins window will appear
In this window, click on Go at the bottom center
A new dialog box will appear, check the Solve Ad-in box and then click OK
Whenever you need this tool, click on Data and then Solver on the Ribbon

For MO 2010:

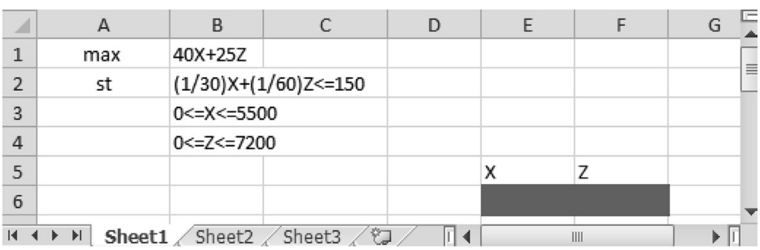
Click on File and then Options at the bottom left column
Click on Add-Ins from the menu at the bottom of the left column in the Excel Options window
The View and Manage Microsoft Office Add-Ins window will appear
In this window, click on Go at the bottom center
A new dialog box will appear, check the Solve Ad-in box and then click OK
Whenever you need this tool, click on Data and then Solver on the Ribbon

Figure 9.1 displays the Excel spreadsheet, which is available in the file Ch09.xls, Fig.9.1. We learn that we must proceed as follows:

Go through cells A1 through B4 to make sure that all mathematical expressions from System 9.4 are there.
(Note that *max* stands for *maximize*, and *st* stands for *subject to*)
Highlight the solution space in cells E6 and F6 (where Solver will report the solutions).

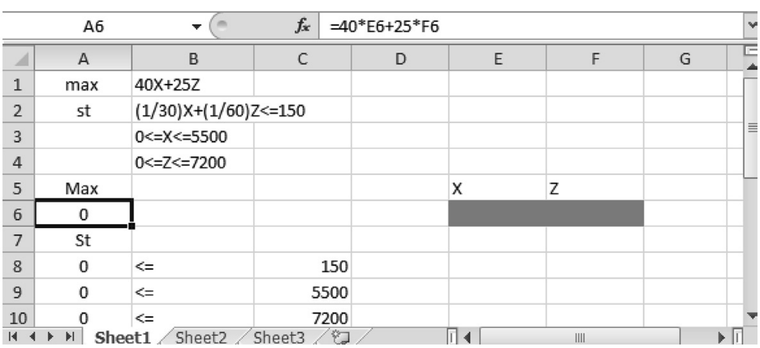
We then open the file Ch09.xls, Fig.9.2, because we need to enter the linear function into Excel. We find that Figure 9.2 shows what the Excel page looks like and what we should do:

In cell A5 type Max (Maximize)
In cell A6 type = 40 * E6 + 25 * F6 and press Enter



	A	B	C	D	E	F	G
1	max	40X+25Z					
2	st	(1/30)X+(1/60)Z<=150					
3		0<=X<=5500					
4		0<=Z<=7200					
5					X	Z	
6							

Figure 9.1 The problem and solution space in Excel



A6 =40*E6+25*F6

	A	B	C	D	E	F	G
1	max	40X+25Z					
2	st	(1/30)X+(1/60)Z<=150					
3		0<=X<=5500					
4		0<=Z<=7200					
5	Max				X	Z	
6	0						
7	St						
8	0	<=	150				
9	0	<=	5500				
10	0	<=	7200				

Figure 9.2 Entering the linear function and the constraints

Dr. App tells us to ignore what we see on the Excel spreadsheet temporarily because we have not opened the Solver tool). Here are the next steps:

In cell A7 type st (subject to)
In Cell A8, type = (1/30) * E6 + (1/60) * F6 and press Enter
In Cell B8, type <= and press Enter; in Cell C8, type 150 and press Enter
In Cell A9, type = E6 and press Enter
In Cell B9, type <= and press Enter; in Cell C9 type 5500 then press Enter
In Cell A10, type = F6 and press Enter
In Cell B10, type <= and press Enter; in Cell C10, type 7200 and press Enter

Do not worry about the nonnegative condition and the zero values on the page; Excel will automatically adjust the condition and the values later.

Next, we are going to open the Solver
Click on Data on the Ribbon and then click on Solver

Figure 9.3 displays the Solver Parameters dialog box and the selected parameters.

In the box Set Objective enter A6
Check the button Max if it has not been checked
In the box By Changing Variable Cells enter E6:F6
Click on the box Subject to the Constraints and then click on Add
The Add Constraint dialog box will appear

Figure 9.4 displays this dialog box and the selected constraints.

In the box Cell Reference enter A8
Click on the arrow of the next box to choose <=; in the box Constraint enter C8
Once you finish entering this constraint, click Add to add the next two constraints

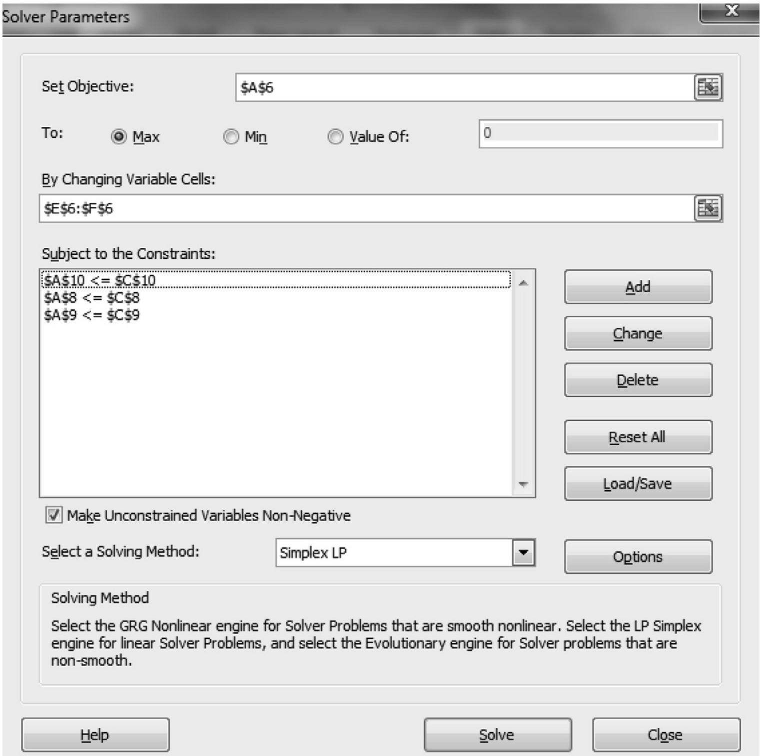


Figure 9.3 The solver parameters in the solver dialog box

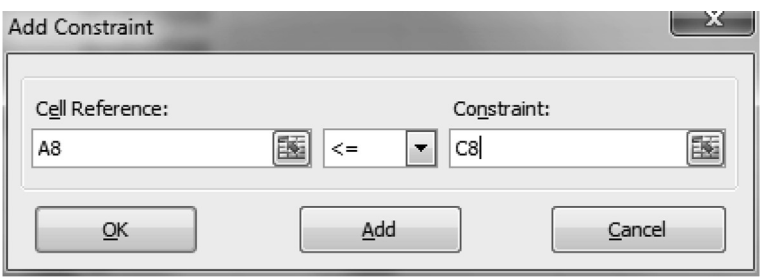


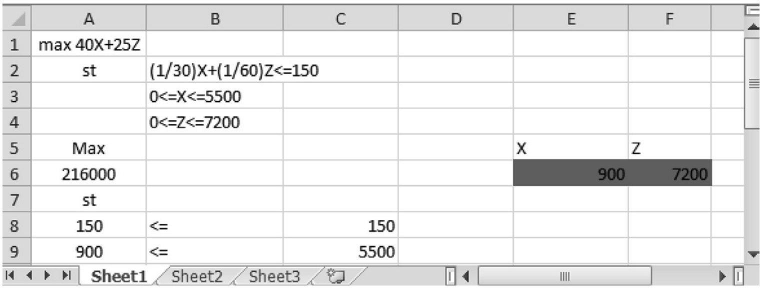
Figure 9.4 The add constraint dialog box with the first constraint entered

In the box Cell Reference enter A9

Click on the arrow of the next box to choose <=

In the box Constraint enter C9 and click Add

In the box Cell Reference enter A10



	A	B	C	D	E	F
1	max 40X+25Z					
2	st	(1/30)X+(1/60)Z<=150				
3		0<=X<=5500				
4		0<=Z<=7200				
5	Max				X	Z
6	216000				900	7200
7	st					
8	150	<=	150			
9	900	<=	5500			

Sheet1 Sheet2 Sheet3

Figure 9.5 The first production allocation for ammonium division

Click on the arrow of the next box to choose <=; in the box Constraint enter C10

After the last constraint is entered, click OK to return to the Solver shown in Figure 9.3

Use the arrow in the Select a Solving Method box to choose Simplex LP (linear program)

Click on Solve at the bottom right next to Close; a new dialog box will appear

Select Keep Solver Solution and click OK

We now see the quantity 900 and 7200 appearing in cells E6 and F6 in Figure 9.5 or in the file Ch09.xls, Fig.9.5.

We learn that we should repeat the same procedures for the other constraints of the 120 worker hours and production limit of 6,600 units.

Gravity Models

We are delighted to learn that we are going to modify a model that is based on Newton's law of gravitational attraction:

$$f_{1,2} = k * m_1 * m_2 / d_{1,2}^2 \qquad (9.7)$$

where $f_{1,2}$ = the gravitational attraction force between any two objects

m_1 = the mass of the first object

m_2 = the mass of the second object

$d_{1,2}$ = is the distance between the two

The parameter k is a measure of gravitational attraction and has to be decided case by case based on the characteristics of the masses. For example, if the gravitational attraction between the two is one half of their masses, then the parameter $k = 0.5$.

Gravity models have many applications. In econometrics, it is used to predict bilateral trade between two countries, which depends positively on their income and negatively on their distance in addition to several other determinants. Dr. Theo says that we only discuss two cases in forecasting: land use and trip distribution. He also reminds Cita that the land-use model is closely related to her job at the Economic Competitiveness Office.

Land-Use Forecasts

We learn that the early studies on land distribution focused on the population density in the cities and found that urban population is crowded around the downtown area and gradually spread out in the more remote areas. However, in this class, we focus on a model introduced by Dunn (1954) for land-use forecasting in both urban and rural areas. This model is simple but has useful applications and is related to the gravity model in that one of the factors that affect the land use (called determinants of the land use) is the distance to the market:

$$R_{t+1} = Q_t(P_t - C_t) - Q_t T_t d \tag{9.8}$$

where

R_{t+1} = the rental price per unit of land at time $(t + 1)$

Q_t = the quantity of production at time t

P_t = the market price per unit of production at time t

C_t = the production cost per unit of production at time t

T_t = the unit transportation cost at time t

d = the distance from the firm's headquarter to the market

From this equation, the first group on the right-hand side is the profit from the production process, that is, QP is total revenue and QC the total cost. The second group on the right-hand side is the cost of transportation to the market for any firm. Hence, the right-hand side is the net profit of the firm, and the breakeven point is where the rental price per unit of land equals the net profit of the firm. Data on *Q*, *P*, *C*, *T*, and *d* are collected for the current period, and the forecast of the rental price for the next period is calculated.

At this point, Fligh says that he can offer an example from his friend, who is the manager of a firm that produces tractors in the city. His firm produces 100 tractors per month that can be sold for $100,000 each. The cost to produce a tractor is $60,000. The transportation cost is $100 per mile per tractor and the distance to the tractor dealer is 100 miles.

Dr. Theo is very pleased and asks us to substitute these data into Equation 9.8 to forecast his rental price next month.

$$\begin{aligned} R &= 100 * (100{,}000 - 60{,}000) - 100 * 100 * 100 \\ &= 4{,}000{,}000 - 1{,}000{,}000 = \$\ 3{,}000{,}000 \end{aligned}$$

Hence, his rental price for the next month should be roughly $3,000,000 in order to break even and survive the competition from other companies.

Trip-Distribution Forecasts

We learn that the trip distribution model is popular in transportation forecasts. The model was initially used to understand how people accept a job offer or remain in a job, and lately it has been used to predict the pattern of transportation between two regions, often called the origin and destination. We start with a simple version of the Voorhees (1956) model for trip-distribution forecasting with one origin where workers (W) live and one destination where the jobs (J) are offered. In this case, the forecast model is:

$$T_{OD} = \frac{aW_O bJ_D}{C_{OD}^c} \tag{9.9}$$

where T_{OD} = the number of round trips between the origin and the destination

W_O = the number of workers in the origin
J_D = the number of jobs at the destination
C_{OD} = the travel costs between the two regions, such as money (m) or distance (d)

The parameters a, b, c are friction factors that have to be calibrated based on actual surveys, for example, a is the friction factor for the workers and shows the reluctance of the people who make the trips to a job based on their family conditions at home, b is the friction factor for the jobs and reveals how attractive a job is, and c is the friction factor for the travel costs and indicates how much people are willing to travel to various distances. The calibration of the trip-distribution model involves adjustments of these friction factors.

Given this model, we want to forecast how many round trips the workers will make from their hometown to their workplaces. Levinson and Kumar (1994, 1995) suggest several models for friction factors; the most commonly used model is:

$$a = \frac{1}{\sum J_D d_{OD}}; \quad b = \frac{1}{\sum W_O d_{OD}} \tag{9.10}$$

If a survey found that the most significant cost between two regions is the distance, and the relation is

$$C_{OD}^c = d_{OD}^2 \tag{9.11}$$

then substituting Equations 9.10 and 911 into Equation 9.9 will allow us to calculate T_{OD}:

$$T_{OD} = \frac{aW_O bJ_D}{d_{OD}^2} \tag{9.12}$$

Dr. Theo then extends the original model for two regions to a model for four regions. The friction factors can now be written as:

$$a_1 = \frac{1}{J_1 d_{11} + J_2 d_{21}}; \quad a_2 = \frac{1}{J_1 d_{12} + J_2 d_{22}}$$

$$b_1 = \frac{1}{W_1 d_{11} + W_2 d_{21}}; \quad b_2 = \frac{1}{W_1 d_{12} + W_2 d_{22}} \tag{9.13}$$

The number of round trips between two regions is then calculated using these equations:

$$\begin{aligned} T_{11} &= a_1 W_1 b_1 J_1 / d_{11}^2; \quad T_{12} = a_1 W_1 b_2 J_2 / d_{12}^2; \\ T_{21} &= a_2 W_2 b_1 J_1 / d_{21}^2; \quad T_{22} = a_2 W_2 b_2 J_2 / d_{22}^2 \end{aligned} \tag{9.14}$$

Arti raises her hand and says her friend works for the county office and has a problem that needs to be solved. The four regions in her province are: Ami (A), Bero (B), Cira (C), and Dore (D). Data for the number of workers (W), the number of jobs (J), and the distances (d) between them are provided in Table 9.1.

A survey of the friction factors between any two regions in her provinces yields the following equations:

$$a_i = \frac{1}{\sum J / d_{OD}}; \quad b_j = \frac{1}{\sum W / d_{OD}} \tag{9.15}$$

$$T_{OD} = \frac{aW_O bJ_D}{C_{OD}^c}; \quad C_{OD}^c = d_{OD} \tag{9.16}$$

We are more than glad to help Arti's friend find the solutions. The calculations can be done using a handheld calculator and then Excel. First, we use this information to calculate the friction factors:

$a_1 = 1/(40/24 + 32/28) = 1/(1.67 + 1.14) = 1/2.81 = 0.3559$

$a_2 = 1/(40/18 + 32/25) = 1/(2.22 + 1.28)= 1/3.5 = 0.2857$

Table 9.1 Workers/jobs and distance among four regions

Workers/jobs (in thousands)	W (in miles)	J	Distance d_{ij}	1 (A)	2 (B)
1 (C)	30	40 (A)	1 (C)	24	18
2 (D)	35	32 (B)	2 (D)	28	25

$$b_1 = 1/(30/24 + 35/28) = 1/(1.25 + 1.25) = 1/2.5 = 0.4$$

$$b_2 = 1/(30/18 + 35/25) = 1/(1.67 + 1.4) = 1/3.07 = 0.3257$$

We then forecast the number of round trips between two regions by substituting these data into Equation 10.16:

$$T_{11} = (0.3559 * 0.4 * 30 * 40)/24 \approx 7.12 \text{ (thousand)}$$
$$\approx 7{,}120 \text{ trips between A and C}$$

$$T_{12} = (0.3559 * 0.3257 * 30 * 32)/18 \approx 6.19 \text{ (thousand)}$$
$$\approx 6{,}190 \text{ trips between B and C}$$

$$T_{21} = (0.2857 * 0.4 * 35 * 40)/28 \approx 5.71 \text{ (thousand)}$$
$$\approx 5{,}710 \text{ trips between A and D}$$

$$T_{22} = (0.2857 * 0.3257 * 35 * 32)/25 \approx 4.17 \text{ (thousand)}$$
$$\approx 4{,}170 \text{ trips between B and D}$$

Arti is very happy. She is going to show the solutions to her friend.

Dr. Theo reminds us that trip-distribution forecasts for more than four regions requires knowledge of matrix algebra and that a comprehensive analysis of N regions is presented in Tsekeris and Stathopoulos (2006) for our reference.

Excel Application

Dr. App tells us that the exercise in the section on "Land-Use Forecasts" is simple enough to solve using a handheld calculator. Thus, we only use an Excel spreadsheet to solve the problem in the section on "Trip-Distribution Forecasts". Figure 9.6 displays data on the number of workers, number of jobs, and the distances.

Dr. App tells us to open the file Ch09.xls, Fig.9.7, and enter the following commands:

In cell H1, type = 1/((C2/E2) + (C3/E3)) and press Enter
In cell H2, type = 1/((C2/F2) + (C3/F3)) and press Enter
In cell H3, type = 1/((B2/E2) + (B3/E3)) and press Enter
In cell H4, type = 1/((B2/F2) +(B3/F3)) and press Enter

	A	B	C	D	E	F	G	H
1	W/J	W	J	Distance	1	2	a_1	0.355932
2	1	30	40	1	24	18	a_2	0.285533
3	2	35	32	2	28	25	b_1	0.4
4							b_2	0.326087

Figure 9.6 Data and calculations of friction factors

	A	B	C	D	E	F	G	H	I	J
1	W/J	W	J	Distance	1	2	a_1	0.355932	T_{11}	7.118644
2	1	30	40	1	24	18	a_2	0.285533	T_{12}	6.190125
3	2	35	32	2	28	25	b_1	0.4	T_{21}	5.71066
4							b_2	0.326087	T_{22}	4.171265

Figure 9.7 Data and calculations of friction factors

Figure 9.7 displays the same data and calculations of round trips between pairs of cities. We open the file Ch09.xls, Fig.9.8, and proceed as follows:

In cell J1, type = (H1 * H3 * B2 * C2)/E2 and press Enter
In cell J2, type = (H1 * H4 * B2 * C3)/F2 and press Enter
In cell J3, type = (H2 * H3 * B3 * C2)/E3 and press Enter
In cell J4, type = (H2 * H4 * B3 * C3)/F3 and press Enter

Dr. App asks us to compare the results using the Excel spreadsheet with those using a handheld calculator. We find that they are the same once the values on the Excel spreadsheet are rounded off to two decimal places.

Exercises

1. The Pan Division of a kitchen-appliance manufacturer produces two kinds of pans: large (*L*) and small (*S*). Their historical data reveal that the profit rates of the two products are:

Profit per unit (dollars per pan)	*L*	$4.0
	S	$2.2

Their production time is:

Time (minutes per unit): L Two minutes

S One minute

From the firm's plan, between 180 to 200 worker hours will be available for this Pan Division next week. Additionally, the maximum limits of weekly production quantities depend on the quotas set by the plant leaders:

Maximum limits (units): L 6,000

S 8,400

Use a handheld calculator and then an Excel spreadsheet to forecast how many units of each product this Pan Division should produce next week to maximize the division's total profit.

2. Data on two manufacturing sectors in an economy are given in the following text:

Total output (millions of dollars)

Meat processing: $X_M = 250$

Vegetable processing: $X_V = 180$

Their estimated technical coefficients are reported as follows:

	Output 1: Meat (M)	**Output 2: Vegetables (V)**
Input 1: Meat (M)	0.20	0.10
Input 2: Vegetables (V)	0.10	0.05

a. Calculate the total input demand for each sector.

b. Calculate the original demand for final consumption in the vegetables sector.

c. Recalculate the total input demand and the demand for final consumption if the technical coefficient for the vegetable products used in the meat processing sector falls 50 percent (from 100 to 50 percent).

3. Given the trip-distribution model involving four regions—A, B, C, D—with workers and jobs in thousands and distances in miles, the equations of the friction factors are as follows:

$$a_i = \frac{1}{\sum J / d_{AB}}; \quad b_j = \frac{1}{\sum W / d_{AB}}$$

Table 9.2 Workers, jobs, and distances

Workers/jobs (thousands)	W	J	Distance d_i (miles)	1 (A)	2 (B)
1 (C)	24	30 (A)	1 (C)	22	16
2 (D)	26	34 (B)	2 (D)	28	18

The equations for the round trips between two regions are:

$$T_{11} = a_1 W_1 b_1 J_1 / d_{11}; \quad T_{12} = a_1 W_1 b_2 J_2 / d_{12}; \quad T_{21} = a_2 W_2 b_1 J_1 / d_{21};$$

$$T_{22} = a_2 W_2 b_2 J_2 / d_{22}$$

The data for the regions are provided in Table 9.2.

Use a handheld calculator and then an Excel spreadsheet to calculate the friction factors and forecast the number of round trips between any two regions.

CHAPTER 10

Business Cycles and Rates of Change

Fligh has a conversation with Dr. Theo before class. His company, Flightime Airlines, is engaged in a fierce competition with a new arrival, Skylight, which started very small but has attracted customers away from his Flightime by undercutting airfares. Currently, Flightime is still ahead of Skylight in their respective sales. However, he heard that sales at Skylight have been growing at a rate of 7.5 percent monthly whereas sales at his Flightime have only grown at a rate of 2 percent. He is worried that Skylight will soon catch up with his company's sales and wants to find out when that will happen if their respective rates remain at 7.5 and 2 percent.

Dr. Theo tells him this issue of *catching up* will be one of the topics for this week. Upon completing the chapter, we will be able to:

1. Describe the concept of business cycles and economic leading indicators.
2. Construct a diffusion index for these indicators to predict the economic turning point.
3. Explain the catching-up models using the rates of change concept.
4. Develop various investment strategies.
5. Obtain forecasts for (2), (3), and (4) using Excel.

Forecasts on the turning points of the economy use business cycle measurements and rates of change. Forecasts on catching up and investment choices are based on rates of change only.

Turning Points

We have learned that a time series often has a cyclical component. The difference between any cyclical component of any time series and the business cycles is that the latter fluctuate around a long-run trend in real gross domestic product (RGDP). When the RGDP is approaching a peak, the economy is in an expansionary period. When the RGDP is approaching a trough, the economy is in a recessionary period. The technique we are going to learn helps us forecast the turning point, when the economy changes its direction. The technique can also be applied to the business forecast of a company's turning point based on the company's leading factors.

Theoretically, an economy is in recession (having a turning point) if its RGDP falls continuously for six consecutive months. Also, an economy is in a period of expansion if its RGDP rises continuously for six consecutive months. In reality, RGDP might go down for four months, inch up slightly in the fifth month then decline again in the next two months, making it hard to conclude whether the economy is having a turning point.

To help forecast a turning point in the economy, various measures called economic indicators are developed. The Conference Board, a global nonprofit organization for businesses, publishes the Global Business Cycle Indicators for each month on its website. The list always comprises indices of the leading, the coincident, and the lagging indicators, as well as their rates of change.

Concept

The leading indicators occur before a turning point and so provide a warning of a possible change in the direction of the economy. The coincident indicators happen concurrently with a change in the direction of the economy, and the lagging indicators measure factors that change after the economy has already followed a new pattern. Only leading indicators help with forecasting and will be discussed in this chapter. The most recent leading indicators published by the Conference Board are as follows:

Average weekly hours in the manufacturing sector
Average weekly initial claims for unemployment insurance

Manufacturers' new orders of consumer goods and materials
Institute for Supply Management (ISM) index of new orders
Manufacturers' new orders of nondefense capital goods
Building permits for new private housing units
Stock prices for 500 common stocks
Leading credit index
Interest rate spread for the 10-year Treasury bonds minus the federal funds target
Average consumer expectations for business conditions

Fin then tells us that some of the leading indicators affect the economy very strongly. For that reason, the business reporters, who are allowed to enter the locked rooms in Washington, DC, where the statistics are released, often write frantically in their 30-minute allowance before breaking their stories to the world. In the meantime, the tension level is also high worldwide, where corporate leaders, business managers, financial advisors, and individual investors are staring at their computer screens waiting eagerly for the news release. This is what Fin encounters on most of his weekdays.

Dr. Theo thanks him for sharing his experience and then summarizes the general opinion of several experts on the leading indicators to help us understand the concept. He says that the detailed analyses can be found in Rogers (2009), Baumohl (2008), or Ellis (2005).

Average Weekly Hours in the Manufacturing Sector

This series is a leading indicator because any adjustment to the working hours of existing employees is usually made in advance of a new hire or layoff. The indicator measures new jobs created, the unemployment rate, average hourly earnings, and the length of the average workweek. Since consumer spending rises with employment, this is a much anticipated indicator for each month and can be considered a highly sensitive series for forecasting.

Since the indicator contains information about both job and wage growth, several economists choose to focus more on the wage data than on the employment data, arguing that it is the income that drives spending instead of employment. This argument has gained attention recently due to the slow recovery period from 2009 to 2013 when employment was growing but wages remained flat.

The data are reported by the U.S. Department of Labor on the first Friday of each month at 8:30 a.m. Eastern time for the previous month.

Initial Jobless Claims

This series is reported weekly and measures the number of people filing first-time claims for state unemployment insurance. It is more sensitive than other measures of unemployment because laid-off workers usually file claims immediately either to receive unemployment benefits or to look for a new job if they are not qualified for the benefits. The series can be considered highly sensitive. However, the series is also very volatile and requires revisions by taking four-week moving averages to smooth out the volatility in claims.

The data are reported by the U.S. Department of Labor every Thursday at 8:30 a.m. Eastern time for the previous week.

Manufacturers' New Orders of Consumer Goods and Materials

This information is considered a leading indicator because it reflects the changes in consumer demand and hence leads to changes in actual production. Its official name is the Preliminary Report on Manufacturers' Shipments, Inventories, and Orders and is a measure of shipments (sales), inventories, and orders at the manufacturing level.

The data are reported by the U.S. Bureau of the Census during the first week of the month at 8:30 a.m. Eastern time for the previous two months. Since the data are two months old, the index is not considered a good prediction of the economy and can be considered a low sensitive series.

ISM Index of New Orders

This index is compiled by the ISM and is one of the first released each month that is believed to have a high impact on the markets. It is a leading indicator because orders of inputs have to be placed in advance of the production. The indicator is based on a survey of purchasing at roughly 300 industrial companies and is considered the best indicator for the manufacturing sector.

Surprisingly, the series does not help as much in forecasting as expected. Perhaps the index is calculated from nine subindexes: new orders, production, employment, supplier deliveries, inventories, prices, new export orders, imports, and backlog of orders. Some of these subindexes are determined only once the economy has already settled in a clear pattern instead of leading the economy. Hence, the series can be considered moderately sensitive.

The data are released by the ISM on the first business day of the month at 10 a.m. Eastern time for the previous month.

Manufacturers' New Orders for Nondefense Capital Goods

This information is considered a leading indicator because investing in capital goods requires long-term plans, so firms do not place new orders unless they realize positive changes in actual production and rising demand.

The data are reported by the U.S. Bureau of the Census during the first week of the month at 8:30 a.m. Eastern time for the previous two months. Since the data are also two months old, the index is not considered a good prediction of the economy and can be considered a low sensitive series.

Building Permits for New Private Housing Units

This information leads the economy because a construction firm does not apply for a building permit until a contract for a new building in the near future is signed with a customer. Since construction is a large investment, it is supposed to have a large impact on the economy. Officially named Housing Starts and Building Permits, the indicator also measures the number of buildings already under construction in addition to the number of permits.

The series is a leading indicator of home sales and spending in general. However, consumers and firms usually do not invest until the economy has already shown signs of improvement from a recession. Also, in an economic expansion, building permits and new houses continue to be built until a recession looms large. Hence, the series can only be considered moderately sensitive.

The data are released by the U.S. Bureau of the Census around the 18th day of the month at 8:30 a.m. Eastern time for the previous month.

Stock Prices of 500 Common Stocks

This dataset is from the Standard & Poor's 500 (S&P 500) and is considered a leading indicator because changes in stock prices reflect the investor's expectations for the future of the economy. The S&P 500 incorporates the 500 largest companies in the United States and so its price change is a good measure of the movements in the stock prices. The S&P 500 stock prices are posted on *The Wall Street Journal* website and various other websites every weekday. So it is very timely and highly sensitive.

The index is reported in the third or fourth week of each month by the Conference Board based on the previous month's data.

Leading Credit Index

This index is a new indicator at the Conference Board, introduced to replace the previous indicator of Money Supply, which no longer predicts the turning point of the RGDP very well but has trailed behind the economy since 2008. The credit index is constructed based on the several subindicators that predict the movements in the financial markets and can lead the economy.

The indicator was initially released on January 26, 2012, by the Conference Board for the previous month and has been reported in the third or fourth week of each month by the board since then. It has been welcomed by the professional world as a highly sensitive series.

Interest Rate Spread of the 10-Year Treasury Less Federal Funds Target

This measure is usually referred to as the term spread, which is highly sensitive because a long-term bond has to yield a higher rate than a short-term one in order to attract any investor at all due to the risks involved in holding a long-term bond. It is a leading indicator because right before a recession, the spread between the short-term and long-term bonds becomes closer. Then the yield of long-term bonds declines sharply during a recession, which causes a negative spread.

The difference between these two series can also be studied in a yield curve, which is upward-sloping over time in the normal situation with the yield of the short-term bond much smaller than that of the long-term. A change in the direction of this curve, called an inverted yield curve, can predict a recession in the economy. Recently this indicator has lost favor among the professionals because it failed to predict the 1990–91 and the 2007–08 recessions. Hence, the series can only be considered moderately sensitive.

Interest rates are posted on *The Wall Street Journal* website and various other websites every weekday.

Average Consumer Expectations for Business Conditions

This information is the only leading indicator that is based solely on expectations. The indicator measures consumer confidence and leads the economy because it can indicate an increase or decrease in consumer spending that affects the demand side of the economy. Unfortunately, consumer confidence changes quite often depending on daily news regarding stock markets and the world economy, so the series can be considered moderately sensitive.

The indicator is released by the University of Michigan in two rounds: The first round is on the 15th of the month (a preliminary reading), and the second round is on the last business day of the month (a final reading) at 10 a.m. Eastern time for the current month.

Constructing Diffusion Indexes

Dr. Theo reminds us that the Conference Board also calculates a composite index that averages the 10 leading indicators and adjusts for seasonal fluctuation. A simple idea is to look at this composite index. If it goes down for three consecutive months, one might predict a coming recession and vice versa.

For example, the Conference Board reported on December 22, 2014, that the Leading Economic Index for the United States increased 0.8 percent in September, followed by a 0.6 percent increase in October and a 0.6 percent increase in November. This is a consistent three-month rise and might predict an economic expansion in the near future.

At this point, Fin raises his hand and asks, "Is predicting a turning point so easy?" Dr. Theo answers, "No, actually this rule is too simple to give a reliable prediction in reality, especially in predicting a recession. Additionally, if you come from a developing country, the 10 preceding indicators might not fit your country's conditions, and you might want to track a different set of indexes. Hence, a diffusion index for each month will take into account the rates of change in all relevant indicators and help you construct your own indexes instead of looking at the single composite index provided by the Conference Board."

There are several approaches to calculating the diffusion indices. Dr. Theo says he has compiled the steps presented in the following text based on the suggestions of the Conference Board:

i. Calculate the rate of change for each indicator:

$$\%\Delta = \frac{I_t - I_{t-1}}{I_{t-1}} * 100\% = \left(\frac{I_t}{I_{t-1}} - 1 \right) * 100\% \qquad (10.1)$$

where I is an individual index for a leading indicator

For example, if an index changes from 20 to 19 going from January to February, then the rate of change between the two months is:

$$\%\Delta = \frac{19 - 20}{20} * 100\% = \left(\frac{19}{20} - 1 \right) * 100\% = -5(\%)$$

ii. Ascribe an assigned value (AV) to each change:

1 to a positive change

0.5 to the unchanged one

0 to a negative change

This raises a question of how much change should be considered a true change.

Table 10.1 Calculating a diffusion index

Indicators	1	2	3	4	5	6	7	8	Diffusion index
January	20	30	15	40	28	14	30	40	
February	19	30.1	16	39	30	15	31	40	
(i) %ΔΔ	–5	0.3	6.7	–2.5	7.1	7.1	3.3	0	
(ii) AV	0	0.5	1	0	1	1	1	0.5	
(iii) Index:	[(0 + 0.5 + 1 + 0 + 1 + 1 + 1 + 0.5)/8] * 100								= 62.5

The Conference Board suggests that a positive movement is any rise that is equal to or greater than 0.5 percent, and a negative movement is any fall that is equal to or greater than 0.5 percent. This implies that any change less than 0.5 percent constitutes an unchanged variable.

iii. Sum the AV calculated in step (ii), divide the sum by the number of indexes, and multiply the ratio by 100.

We then break into groups to work on an example: Table 10.1 shows the following monthly indexes of eight leading indicators in a hypothetical economy between January and February 2013 and the steps to calculate a diffusion index for February. The results show an index of 62.5, which is above the average index of 50 and implies a stable economic condition.

Dr. Theo says that in his opinion, you have to continue to track the diffusion indexes for at least five months. If the diffusion indexes are below 50 for 80 percent of the time, the economy has a high probability of landing in a recession and vice versa. Additionally, if more than seven out of 10 individual indexes are falling, the economy looks gloomy. The severity of the gaps in the diffusion indices also have to be taken into account, for example, a negative gap of 20 points from the average of 50 points implies a severe condition compared with a gap of five points.

Arti then asks, “We are talking about economic recessions and expansions. But then why is the cycle called a business instead of an economic cycle?” Alte offers to answer. Here is her explanation: “The up and down cycle of the economy depends on the business activities. When businesses are growing, they start to invest in new capital, which causes the economy to expand for several years until it reaches a peak. Then businesses are no longer growing, and there is no additional capital. Gradually, the plants and capital are worn out, and the economy is approaching a trough. This cycle is renewed when depleted capital is replaced by a new one."

"Excellent," Dr. Theo commends her. He then continues with the discussion of the diffusion index by saying, "The Conference Board also suggests that you construct a six-month diffusion index. In this case, you list all six months first and see if an individual index is going up or down at the end of the six-month period."

Dr. Theo believes that a six-month diffusion index is not much of a difference from the Conference Board composite index because it provides only one index for the whole six-month period. Moreover, he does not believe this diffusion index is of much help in forecasting because the economy is already in a recession after six months of continuous declining RGDP, so it is too late to predict the situation. Hence, he suggests, "If you wish to have longer-term indices, you can construct two-month or three-month indexes, which give more room to think ahead of a possible turning point."

One remaining issue with diffusion indexes is that the Conference Board and current researchers treat them equally. Dr. Theo also believes some weighted approaches should be used to account for the difference in the level of sensitivity in the released data. He says that scrutinizing Excel applications will clarify this point.

Excel Applications

Dr. App shows us Figure 10.1, which displays the data from the file Ch10.xls, Fig.10.2, on 10 indexes of the leading indicators for a hypothetical economy from January through July 2014.

We learn that the following steps must be performed:

In cell C4, type = ((C3 – C2)/C2) * 100 and press Enter
Copy and paste the formula into cells D4 through L4
In cell C7, type = ((C6 – C3)/C3) * 100 and press Enter

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Indicators	1	2	3	4	5	6	7	8	9	10	Diffusion Index
2		**January**	31	41	27	25	33	40	38	36	27	43	
3		**February**	30	45	28	25	34	38	37	34	28	46	
4		%Δ	-3.226	9.7561	3.7037	0	3.0303	-5	-2.632	-5.556	3.7037	6.9767	
5		Value	0	1	1	0.5	1	0	0	0	1	1	55
6		**March**	29	42	25	25	35	33	37	32	24	45	
7		%Δ	-3.333	-6.667	-10.71	0	2.9412	-13.16	0	-5.882	-14.29	-2.174	
8		Value	0	0	0	0.5	1	0	0.5	0	0	0	20
9		**April**	28	42	26	25	36	33	37	30	25	45	
10		%Δ	-3.448	0	4	0	2.8571	0	0	-6.25	4.1667	0	
11		Value	0	0.5	1	0.5	1	0.5	0.5	0	1	0.5	55
12		**May**	27	40	21	26	35	34	32	28	23	38	
13		%Δ	-3.571	-4.762	-19.23	4	-2.778	3.0303	-13.51	-6.667	-8	-15.56	
14		Value	0	0	0	1	0	1	0	0	0	0	20
15		**June**	25	38	24	27	34	29	30	27	22	40	
16		%Δ	-7.407	-5	14.286	3.8462	-2.857	-14.71	-6.25	-3.571	-4.348	5.2632	
17		Value	0	0	1	1	0	0	0	0	0	1	30
18		**July**	24	35	25	25	35	25	27	26	22	40	
19		%Δ	-4	-7.895	4.1667	-7.407	2.9412	-13.79	-10	-3.704	0	0	
20		Value	0	0	1	0	1	0	0	0	0.5	0.5	30

Sheet1 Sheet2 Sheet3

Figure 10.1 Obtaining diffusion indexes for February through July 2014

Copy and paste the formula into cells D7 through L7 and into the following cells:
Cells C10 through L10
Cells C13 through L13
Cells C16 through L16
Cells C19 through L19
Enter values 0, 0.5, or 1 by looking at the percentage changes

For example, the value in cell C4 is −3.226. This implies a fall of 3.226 percent, which is more than a 0.5 percent decrease, so enter 0. Once you finish entering all values of 1, 0.5, or 0:

In cell M5, type = ((C5 + D5 + E5 + F5 + G5 + H5 + I5 + J5 + K5 + L5)/10) * 100 and press Enter
Copy and paste the formula from cell M5 into cells M8, M11, M14, M17, and M20

The diffusion indexes are in column M.

The results reveal that only four out of six indexes are below 50, which is only 67 percent below 50 instead of the 80 percent benchmark mentioned by Dr. Theo in his theoretical section. Alte asks, "Should we still send a warning that the economy is heading into a recession?"

Table 10.2 Assigning different weights across leading indicators

Indicator	1	2	3	4	5	6	7	8	9	10
Sensitivity	H	H	L	M	L	M	H	H	M	M
Weight	2	2	1	1.5	1	1.5	2	2	1.5	1.5

Dr. App says, "Here is my interpretation of the indexes: The two indexes above 50 are just 5 points from the 50 level, whereas the ones below 50 show gaps of 20 to 30 points. Based on this observation alone, I would rather send a warning on a high probability of a recession."

Dr. App now comes to Dr. Theo's last point in the section on "Constructing Diffusion Indexes": Should we use a weighted approach to constructing diffusion indexes? Dr. App says that she also believes in a weighted-index approach. However, she warns us, "This is the technique suggested by this class' instructors, so take it at your own risk." Here is the idea: We can assign different weights across the leading indicators based on their own level of sensitivity, either high (H), low (L), or moderate (M). For example, Table 10.2 lists the leading indicators reported by the Conference Board, their sensitivity, and possible weights.

Hence, the total of the weighted indexes are:
Total = 2 + 2 + 1 + 1.5 + 1 + 1.5 + 2 + 2 + 1.5 + 1.5 = 16

The steps of calculations are as follows:

i. Calculate the rate of changes for each indicator using Equation 10.1 as usual.
ii. Ascribe a value of 1 to a positive change, 0.5 to the unchanged one, and 0 to a negative change. After that multiply the values by 2 and 1.5 for the highly sensitive and moderately sensitive indicators, respectively.
iii. Sum the value calculated in step (ii), divide the sum by 16, and multiply the ratio by 100.

Figure 10.2 displays the data from the file Ch10.xls, Fig.10.4, as well as the new calculations. From the results, we are now more confident to

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Indicators	1	2	3	4	5	6	7	8	9	10	Diffusion Index
2		Weight	2	2	1	1.5	1	1.5	2	2	1.5	1.5	
3		**January**	31	41	27	25	33	40	38	36	27	43	
4		**February**	30	45	28	25	34	38	37	34	28	46	
5		%Δ	-3.2258	9.7561	3.7037	0	3.0303	-5	-2.6316	-5.5556	3.7037	6.97674	
6		Value	0	2	1	0.75	1	0	0	0	1.5	1.5	48.4375
7		**March**	29	42	25	25	35	33	37	32	24	45	
8		%Δ	-3.3333	-6.6667	-10.714	0	2.94118	-13.158	0	-5.8824	-14.286	-2.1739	
9		Value	0	0	0	0.75	1	0	1	0	0	0	17.1875
10		**April**	28	42	26	25	36	33	37	30	25	45	
11		%Δ	-3.4483	0	4	0	2.85714	0	0	-6.25	4.16667	0	
12		Value	0	1	1	0.75	1	0.75	1	0	1.5	0.75	48.4375
13		**May**	27	40	21	26	35	34	32	28	23	38	
14		%Δ	-3.5714	-4.7619	-19.231	4	-2.7778	3.0303	-13.514	-6.6667	-8	-15.556	
15		Value	0	0	0	0.75	0	1.5	0	0	0	0	14.0625
16		**June**	25	38	24	27	34	29	30	27	22	40	
17		%Δ	-7.4074	-5	14.2857	3.84615	-2.8571	-14.706	-6.25	-3.5714	-4.3478	5.26316	
18		Value	0	0	1	1.5	0	0	0	0	0	1.5	25
19		**July**	24	35	25	25	35	25	27	26	22	40	
20		%Δ	-4	-7.8947	4.16667	-7.4074	2.94118	-13.793	-10	-3.7037	0	0	
21		Value	0	0	1	0	1	0	0	0	0.75	0.75	21.875

Figure 10.2. Obtaining diffusion indexes using weighted indicators

conclude that the economy is heading into a recession because all six diffusion indices are below 50 with many substantially so.

We then work with Dr. Theo on several models based on rates of change.

Models Based on Rates of Change

We learn that we can apply the rates of change concept to models in this section to forecast catching-up games and investment choices. The original ideas are from Thirlwall (2003) and are modified to fit the business problems in this section.

Catching Up

Rates of change can be used to study convergence (or catching up) among countries or companies. For example, the sales of Fligh's company have grown at a monthly rate of 2 percent whereas those of Skylight have grown at a monthly rate of 7.5 percent. How long would it take Skylight to catch up with Flightime at their respective growth rates?

In economic theory, there is an unconditional convergence where developing countries will catch up with the developed countries even if the former do not grow faster than the latter. There are three arguments for this unconditional convergence.

First, the neoclassical growth theory assumes diminishing returns to capital. A developed country, which has accumulated a large stock of capital per worker over time, will eventually reach a long-run equilibrium called the steady state, where the country's growth rate of capital per worker will be zero. This will allow a developing country to catch up and the two will converge to the same level of RGDP.

Second, since knowledge and technology are considered public goods, there are positive externalities flowing from developed to developing countries. These knowledge and technology transfers and spillovers allow the developing countries to harvest benefits from the developed countries without costs because the former do not have to fund the research and development (R&D) that results in new knowledge and technology.

Third, the progress from an underdeveloped economy to a developed one always goes through an industrialization process. During this process, the factors of production and resources are automatically moving from the agricultural sector with low productivity to the manufacturing and service sectors with high productivity. Since resources cannot be shifted forever, developing countries will have greater shifts of resources and be able to catch up.

Empirically, there is no clear evidence of unconditional convergence. There are certain conditions for the convergence to occur. The conditions for a country could be high investment in education, a higher savings rate, or good governance.

The following section modifies the original ideas to forecast catching points between two companies. The conditions for a company to catch up with another company could be corporate restructuring, high investment in capital, rigorous R&D funding, or simply launching an aggressive campaign.

Catching Up to Current Level

Given an average growth rate in the sales of a small company, we can find how long it would take a small company to catch up with a large company, whose sales do not grow. The relationship between the former and the latter can be expressed as:

$$\text{SALE}_{Lt} = \text{SALE}_{St}\,(1 + \text{GROW}_S)^T \qquad (10.2)$$

where

$SALE_{Lt}$ = the current sales of the large company at time t

$SALE_{St}$ = the current sales of the small company at time t

$GROW_S$ = the sale growth rate of the small company

T = the time it takes for the small company to catch up with the large one

Taking the natural logarithm of Equation 10.2 yields:

$$\ln SALE_{Lt} = \ln SALE_{St} + T\ln(1+GROW_S)$$
$$T\ln(1+GROW_S) = \ln SALE_{Lt} - \ln SALE_{St}$$

Solving for T to obtain:

$$T = \frac{\ln SALE_{Lt} - \ln SALE_{St}}{\ln(1+GROW_S)} = \frac{\ln \dfrac{SALE_{Lt}}{SALE_{St}}}{\ln(1+GROW_S)} \quad (10.3)$$

To this point, Fligh raises his hand and asks, "Can we apply this problem to calculate how long it would take for Skylight to catch up with Flightime if my company does not grow?" Dr. Theo replies, "Sure, let's work on the problem."

Fligh provides the class with the information: His company's sales this month are \$134,840, and Skylight's sales are \$45,040. We already know that monthly sale growth at Skylight is 7.5 percent, so

$$T = \frac{\ln \dfrac{134,840}{45,040}}{\ln(1+0.075)} = 15.27 \approx 15$$

Hence, it would take Skylight roughly 15 months at a growth rate of 7.5 percent to catch up with Flightime if the latter does not grow.

Catching Up at Future Level

Dr. Theo then says, "But Flightime is also growing, so we need to revise Equation 10.2 to find how long it would take for the gap between the two companies to be eliminated in the future." The relationship between the former and the latter can be expressed as:

$$\text{SALE}_{Lt}\,(1 + \text{GROW}_L)^T = \text{SALE}_{St}\,(1 + \text{GROW}_S)^T \qquad (10.4)$$

where GROW_L = the sale growth rate of the large company

Taking the natural logarithm of Equation 10.4 yields:

$$\ln \text{SALE}_{Lt} + T\ln(1+\text{GROW}_L) = \ln \text{SALE}_{St} + T\ln(1+\text{GROW}_S)$$
$$T[\ln(1+\text{GROW}_S) - \ln(1+\text{GROW}_L)] = \ln \text{SALE}_{Lt} - \ln \text{SALE}_{St}$$

Solving for T to obtain:

$$T = \frac{\ln \dfrac{\text{SALE}_{Lt}}{\text{SALE}_{St}}}{\ln(1+\text{GROW}_S) - \ln(1+\text{GROW}_L)} \qquad (10.5)$$

Dr. Theo then says, "Let's work on the problem between Flightime and Skylight again." We already know that the monthly sale growth at Fligh's company is 2 percent, and the monthly sale growth at Skylight is 7.5 percent; hence:

$$T = \frac{\ln \dfrac{134{,}840}{45{,}040}}{\ln(1+0.075) - \ln(1+0.02)} = 21.02 \approx 21$$

Thus, it takes Skylight roughly 21 months to catch up with Flightime's sales. Fligh exclaims, "That is fast! I will report this result to my boss so that he can pursue an aggressive campaign to increase sales. Otherwise, we might soon go bankrupt."

Required Growth Rate to Reach a Target

Dr. Theo continues, "In solving the problem in Equation 10.5, we assume that Skylight is growing at a constant rate of 7.5 percent. Suppose that

Flightime continues to grow at a 2 percent rate, but Skylight sets a target to catch up with the Flightime in 25 months instead of 21 months. It then needs to calculate the required growth rate to reach that target." We can start from Equation 10.4:

$$SALE_{Lt}\,(1 + GROW_L)^T = SALE_{St}\,(1 + GROW_S)^T$$

$$(1+GROW_S)^T = \frac{SALE_{Lt}(1+GROW_L)^T}{SALE_{St}}$$

$$1+GROW_S = \left[\frac{SALE_{Lt}(1+GROW_L)^T}{SALE_{St}}\right]^{1/T}$$

$$GROW_S = \left[\frac{SALE_{Lt}(1+GROW_L)^T}{SALE_{St}}\right]^{1/T} - 1 \qquad (10.6)$$

We work on the problem again by substituting 25 months, sales for the two airlines, and the sale growth of 2 percent for Flightime into Equation 10.6:

$$GROW_S = \left[\frac{134{,}840(1+0.02)^{25}}{45{,}040}\right]^{1/25} - 1 = 0.0657 \approx 6.6\%$$

Thus, Skylight has to grow at a monthly rate of 6.6 percent to eliminate the gap in 25 months.

Required Growth Rate to Keep a Constant Gap

Dr. Theo discusses the last case of the catching-up games: Suppose Skylight is running out of resources to compete with Flightime at this moment and wants to know what growth rate will keep the gap between the two companies constant until a certain date in the future, say 25 months from now. In this case, the equation is:

$$SALE_{Lt} - SALE_{St} = SALE_{Lt}\,(1 + GROW_L)^T - SALE_{St}\,(1 + GROW_S)^T \qquad (10.7)$$

$$\text{SALE}_{St}(1+\text{GROW}_S)^T = \text{SALE}_{Lt}(1+\text{GROW}_L)^T - \text{SALE}_{Lt} + \text{SALE}_{St}$$

$$(1+\text{GROW}_S)^T = \frac{\text{SALE}_{Lt}(1+\text{GROW}_L)^T - \text{SALE}_{Lt} + \text{SALE}_{St}}{\text{SALE}_{St}}$$

$$\text{GROW}_S = \left[\frac{\text{SALE}_{Lt}(1+\text{GROW}_L)^T - \text{SALE}_{Lt} + \text{SALE}_{St}}{\text{SALE}_{St}}\right]^{1/T} - 1 \tag{10.8}$$

We work on the airline problem a last time:

$$\text{GROW}_S = \left[\frac{134{,}840(1+0.02)^{25} - 134{,}840 + 45{,}040}{45{,}040}\right]^{1/25} - 1$$
$$= 0.0438 \approx 4.4\%$$

Therefore, Skylight can keep the gap at a constant value for 25 months if its sales grow at a monthly rate of 4.4 percent.

Investment Choices

Cita shares with us a problem from the city council. The mayor wants to make decisions on various investment possibilities. He wants Cita to help with finding the right mix of strategies and time horizons to maximize the city's welfare in the future. Cita provides us the following information:

The city economy has a capital output ratio of 1.5, that is, $K/Y = 1.5$, and so $Y = K/1.5$.

The initial capital stock is $390 million.

The depreciation rate is 10 percent ($d = 0.1$), so the capital for the next period = $K(1 - \text{d})$ unless a new investment is made.

Consumption is chosen as the measure of the city welfare.

The social discount rate is 1 percent ($i = 0.01$).

For starting, we assume no government spending; hence:

$$C = \frac{Y - I}{(1+i)^t}\,;\ t = 0, 1, 2, 3, 4.$$

> The mayor is newly elected in November and has four years ahead as his time horizon. He has to make a decision between two investment scenarios: the first is to invest 8 percent of the output and have more output value for current consumption. The second is to invest 9 percent of the output and have less output value for current consumption but more for future consumption.

The mayor wants to make a decision on the plan that will maximize the city's welfare in four years. We work on the calculations and display them in Table 10.3.

The results show that the second plan yields more capital than the first one: by the end of the first year, the first plan has \$410.8 million (= 390 + 20.8) whereas the second plan has \$413.4 million (= 390 + 23.4). Hence, by the end of the second year, the second plan already yields a better consumption value of \$233.2 million (= $[261.4 - 23.5]/[1.01]^2$) compared with \$232.9 million (= $[258.3 - 20.7]/[1.01]^2$) for the first plan. Additionally, output and capital stock grow at faster rates with the second plan thanks to the high rate of investment, opening a venue for future growth. However, if the mayor has only one year to go, he should choose the first plan with the 8 percent investment rate because by the end of the first year, the first plan yields a better consumption value of \$224.5 million compared to \$223.5 million for the second one.

Dr. Theo tells us that Excel applications conducted by Dr. App will add government spending and use more time horizons.

Excel Applications

Dr. App says that the problems in the section on "Catching Up" can be solved faster with a handheld calculator, so she only shows us Excel demonstrations for the investment problems in the section on "Investment Choices." The problem is to choose between two investment plans: the first one invests 8 percent of the output, and the second invests 16 percent of the output. The depreciation rate is 10 percent, and the discount rate is 3 percent. The capital output ratio equals 3, and government spending is fixed at \$10 million. Figure 10.3 shows data from the

Table 10.3 Making investment decision

Year	Plan 1 (I/Y = 8%; d = 0.1; i = 1%)					Plan 2 (I/Y = 9%; d = 0.1; i = 1%)				
	K	***K*(1–*d*)**	**Y**	***I***	**C**	***K***	***K*(1–*d*)**	**Y**	***I***	**C**
0	390	390	260	20.8	239.2	390	390	260	23.4	236.6
1	410.8	369.72	246.5	19.7	224.5	413.4	372.1	248	22.3	223.5
2	430.5	387.5	258.3	20.7	233.0	435.7	392.2	261.4	23.5	233.2
3	451.2	406.1	270.7	21.7	241.7	459.3	413.3	275.6	24.8	243.4
4	472.8	425.6	283.7	22.7	250.8	484.1	435.6	290.4	26.1	254.0

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Plan A		(Y = K/3)	(I/Y = 8%)		i=3%	Plan B		(Y = K/3)	I/Y = 16%)		i=3%
2	Time	K	K*(1-d)	Y	I	Gov	C	K	K*(1-d)	Y	I	Gov	C
3	0	450.00	450.00	150.00	12.00	10.00	128.00	450.00	450.00	150.00	24.00	10.00	116.00
4	1	462.00	415.80	138.60	11.09	10.00	114.09	474.00	426.60	142.20	22.75	10.00	106.26
5	2	473.09	425.78	141.93	11.35	10.00	113.65	496.75	447.08	149.03	23.84	10.00	108.57
6	3	484.44	436.00	145.33	11.63	10.00	113.21	520.60	468.54	156.18	24.99	10.00	110.91
7	4	496.07	446.46	148.82	11.91	10.00	112.76	545.58	491.03	163.68	26.19	10.00	113.27
8	5	507.97	457.18	152.39	12.19	10.00	112.31	571.77	514.60	171.53	27.45	10.00	115.66
9	6	520.17	468.15	156.05	12.48	10.00	111.86	599.22	539.30	179.77	28.76	10.00	118.09
10	7	532.65	479.38	159.79	12.78	10.00	111.40	627.98	565.18	188.39	30.14	10.00	120.54
11	8	545.43	490.89	163.63	13.09	10.00	110.94	658.12	592.31	197.44	31.59	10.00	123.03
12	9	558.52	502.67	167.56	13.40	10.00	110.48	689.71	620.74	206.91	33.11	10.00	125.55
13	10	571.93	514.74	171.58	13.73	10.00	110.02	722.82	650.54	216.85	34.70	10.00	128.10

Sheet1 Sheet2 Sheet3

Figure 10.3 Obtaining forecasts on future consumption from two investment plans

file Ch10.xls, Fig.10.6. Dr. App tells us to proceed with the following steps, ignoring the results in each column for a while because they will be gradually adjusted:

In cell D3, type = C3/3 and press Enter

(Note that the values in cells B3, C3, H3, and I3 are initial capital with $d = 0$ for $t = 0$)

Copy and paste the formula from cell D3 into cells D4 through D13 and J3 through J13

In cell E3, type = D3 * 0.08 and press Enter

Copy and paste the formula from cell E3 into cells E4 through E13

In cell K3, type J3*0.16 and press Enter

Copy and paste the formula into cells K4 through K13

In cell G3, type = (D3 – E3 – F3)/1.03^(A3) and press Enter

Copy and paste the formula from cell G3 into cells G4 through G13 and M3 through M13

In cell B4, type =B3 + E3 and press Enter

Copy and paste the formula from cell B4 into cells B5 through B13 and H4 through H13

In cell C4, type =B4 * (1 – 0.1) and press Enter

Copy and paste the formula from cell C4 into cells C5 through C13 and I4 through I13

The forecast values for future consumption are in columns F and M.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Plan A		(Y = K/3)	(I/Y = 8%)		i=4%	Plan B		(Y = K/3)	I/Y = 16%)		i=4%
2	Time	K	K*(1-d)	Y	I	Gov	C	K	K*(1-d)	Y	I	Gov	C
3	0	450.00	450.00	150.00	12.00	10.00	128.00	450.00	450.00	150.00	24.00	10.00	116.00
4	1	462.00	415.80	138.60	11.09	10.00	112.99	474.00	426.60	142.20	22.75	10.00	109.83
5	2	473.09	425.78	141.93	11.35	10.00	111.48	496.75	447.08	149.03	23.84	10.00	115.94
6	3	484.44	436.00	145.33	11.63	10.00	109.97	520.60	468.54	156.18	24.99	10.00	122.30
7	4	496.07	446.46	148.82	11.91	10.00	108.49	545.58	491.03	163.68	26.19	10.00	128.94
8	5	507.97	457.18	152.39	12.19	10.00	107.02	571.77	514.60	171.53	27.45	10.00	135.87
9	6	520.17	468.15	156.05	12.48	10.00	105.56	599.22	539.30	179.77	28.76	10.00	143.10

Sheet1 Sheet2 Sheet3

Figure 10.4 Future consumption from two investment plans with discount rate = 4 percent

From the results, Plan A will not achieve a higher consumption value than plan B until the end of period 4. Thus, if the time horizon is three years, Plan A is a better option than Plan B.

At this point, Mo raises his hand and asks, "Comparing Figure 10.3 with Table 10.3, I guess that the higher the discount rate, the longer it takes to maximize the consumption when we choose a higher investment rate. Is that true?"

Dr. App says, "This guessing turns out to be true only for low discount rates, which are roughly between 1 and 3 percent. When the discount rate is more than 3 percent, different results can occur. Therefore, you will have to experiment with more than two plans before drawing your own decisions."

She then gives us another example: Figure 10.4 displays the data from the file Ch10.xls, Fig.10.7, and the results for a similar problem with the discount rate of 4 percent. The steps of entering formulas are the same as those in Figure 10.3 except for the ones in columns G and M, where the discount factor has to be changed to 1.04. In this figure, Plan B surpasses Plan A at the end of period 2, sooner than the results in Figure 10.3, which has a discount rate of 3 percent.

From the results, Plan A is only superior to Plan B when the one-year horizon is taken using the discount rate of 4 percent.

To conclude the chapter, Dr. App says that while an individual discount rate is subjective, we can use the interest rate in the market as a proxy for a social discount rate. She also says that the decision of using short-term or long-term interest rates depends on the targeted time horizons of a project and the subjective preferences on the future consumption of each community.

Exercises

1. The file Diffusion.xls provides data on the individual indexes of the leading indicators for January through July. Construct the first diffusion index for February using a handheld calculator then continue with the indexes for the other months using Excel commands.
2. Artistown's profit is $108,100 and grows at an annual rate of 2 percent while another school's profit is $269,500 and grows at an annual rate of 1 percent. How long would it take for the gap between the two schools to be eliminated in the future?
3. Use the same information in Exercise (2) except the growth rate of Artistown. Find out how fast Artistown has to grow annually to eliminate the gap in two years.
4. Use the same information in Exercise (2) except the growth rate of Artistown. Find out the growth rate of Artistown that will keep the gap between the two schools the same in the next two years.
5. Consider three different investment ratios of 0, 8, and 40 percent. Assume that the capital output ratio is 2, and that there is no depreciation rate, government spending, or discount rate. The initial capital stock equals 300.
   a. Construct a table similar to Figure 10.3 up to eight years using a handheld calculator or an Excel spreadsheet.
   b. When does the second policy become superior to the first in welfare enhancement? When does the third policy become superior to the second?
   c. What could have made the results more realistic?

CHAPTER 11
Conclusion

Our class is approaching the end. Our professors are very anxious to summarize the important points so that the knowledge gained during this class will benefit the students for a lifetime.

Mo asks to tell us a story he saw on the Internet: "Aristotle (384–322 BC) raised the question of whether the chicken comes first or the egg comes first and concluded that both must have always existed. Thurman and Fisher (1988) tried to answer the question by performing a test using data on chickens that lay eggs except for commercial broilers and on eggs produced in the United States from 1930 to 1983. They found that the egg comes first." He then asks, "Does that mean that the egg causes the chicken to come into existence?"

Dr. Theo thanks Mo for raising a good question. He says that this will be one of the topics discussed today but he first wants to summarize the techniques learned throughout this course.

Theoretical Summary

Dr. Theo's summarization comprises two broad categories.

Pros and Cons of the Techniques

Moving Averages and Double Moving Averages—Chapters 2 to 4

Moving averages (MA) and double moving averages (DM) are the simplest techniques in forecasting. They are easy to implement and are the most cost-effective method, so they are good starting steps to forecasting. However, the simplicity of the models, including the weighted moving averages (WM) model with integer weights, might render inaccurate forecasts. The techniques usually produce a mean absolute percentage error (MAPE) of roughly 12 to 15 percent for one-period forecasts and 20 to 25 percent for four-period forecasts.

Exponential Smoothing and Double Exponential Smoothing— Δ Chapters 2 to 4

These are more sophisticated techniques than the MAs because the weight (or weights in the double exponential smoothing [DE]) can be adjusted continuously from zero to one. Hence, the exponential smoothing (ES) technique often provides more accurate forecasts than the MA technique. However, the ES and DE models can only cover a limited range of data because they do not account for the seasonal and cyclical components of a time series. The techniques usually produce an MAPE of roughly 10 to 12 percent for one-period forecasts and 18 to 22 percent for four-period forecasts.

Advanced Time Series Analysis—Chapter 7

The decomposition and triple ES both incorporate the seasonal and cyclical components of a time series into their models and so can cover a wider range of data and longer-term forecasts than the preceding two techniques. The triple ES can even handle nonlinear trends, and the AR and ARMA/ARIMA models can handle a time series that follows a random walk. They are also able to reduce errors, with an MAPE of roughly 8 to 10 percent for one-period forecasts and 15 to 17 percent for four-period forecasts. Their weakness is that they still make forecasts based only on past performances of a time series.

Linear Regression Technique—Chapters 5 and 6

A widely applied technique in associative analyses, a linear regression, irrespective of whether it is simple or multiple linear regressions, often produces a smaller MAPE than those obtained from MA and ES techniques because the technique aims to minimize the errors (the least squared approach). Additionally, it takes into account more determinants of a market than the time series itself. A good linear regression model can produce an MAPE of roughly 8 to 10 percent for one-period forecasts and 15 to 17 percent for four-period forecasts. However, when the relationship between the dependent variable and the explanatory ones are not linear in parameters, advanced knowledge of nonlinear regressions is needed to adjust the model accordingly.

Business Models—Chapter 8

These are very helpful models because they are developed specifically for businesses and have enjoyed wide applications in the business world. The running forecast model helps with inventory and ordering. The financial models help with investing, and the diffusion model is good for product adoption forecasts. The weakness of the running forecast model is that its applications are limited to inventory and ordering plans. The weakness of the other two groups is that the errors are large with an MAPE of roughly 14 to 18 percent for one-period forecasts and 20 to 25 percent for four-period forecasts.

Economic Models—Chapter 9

Between the two groups of the economic models, namely, production models and gravity models, the production models have a stronger base in economic theory. Hence, the production models produce smaller errors than the business models, with an MAPE of roughly 10 to 12 percent for one-period forecasts and 18 to 22 percent for four-period forecasts. The gravity models usually result in an MAPE's range similar to that for business models. Their weakness lies in their applications, which are much more limited than those of the business models.

Business Cycles and Rates of Change—Chapter 10

The turning-point technique is popular in all disciplines whereas the models based on rates of change are currently used in the realm of public policy. However, they can be applied to a business environment, for example, forecasting a turning point of a business or catching-up games between two companies. Both groups are for long-term forecasts, but the models based on rates of change are for longer terms than the turning point models. The weakness of these models is that the errors are large, with an MAPE of roughly 15 to 20 percent for one-period forecasts, 20 to 25 percent for four-period forecasts, and 25 to 45 percent for 12-period forecasts.

Cautions on Forecast Results

Dr. Theo then emphasizes that there are several pitfalls to avoid.

Mistaking Correlation for Causality

This is the very issue raised by Mo about the chickens and the eggs. Dr. Theo explains that although the correlation between two variables makes them move in the same direction, stating that a variable *X* causes a variable *Y* to change is a common mistake in forecasting. In fact, variable *X* does not need to cause variable *Y* in order to help predict *Y*. Hence, producing a model that has predictive power is the most important goal in forecasting. Using a holdout sample is one way to test the predictive power of a model. The second way is to carry out an *F*-test on the regression model. The third way is to perform a Granger Causality test, which is the test on the chickens and the eggs mentioned by Mo.

Dr. Theo then says, "However, the fact that the egg comes first might not imply that the egg causes the chicken to come into existence. That is one of the reasons I did not teach the Granger Causality test in this class. Another reason is that I cannot teach a whole econometric course in a forecasting course. You can find details of the test in Ramanathan (1998) or Stock and Watson (2007), which are introductory textbooks in econometrics."

We now understand Dr. Theo's point: It is a good habit to make evaluations on the *predictive power* of a model, which is crucial, instead of focusing on the *causal relation*, which is not important, in making forecasts and discussing the forecast results.

Spurious Precision

Forecasters and decision makers are often confused between precision and accuracy. Due to the uncertainty of the future, all point forecasts carry large errors, especially when the forecasts are several periods away from the current one. Hence, interval forecasts should be more important than point forecasts, and a good forecaster should report a confidence interval instead of a single point. Use past experiences and current market conditions to make a decision on whether we should follow an upper or lower bound of an interval forecast for ordering inputs and delivering our products.

Theory-less Forecasting

Naïve forecasters may extrapolate past data without understanding any plausible theory behind them. For example using an AR(p) model without understanding the model structure and the conditions to use it can lead to a spurious regression due to nonstationarity. When this occurs, the results show strongly significant coefficients but the model still fails to predict the future. Hence, most sound forecasts are grounded not only in empiricism but also in theory. If the results seem to contradict a theory, a respecification process should be carried out to adjust the model.

Overreliance on Forecasts

Forecasters are sometimes subjective and can overestimate or underestimate the future. For example, a production manager, who holds an optimistic view of a company, can overestimate the forecast values. In contrast, a sales person who receives bonuses for his sales might underestimate the expected values so that his actual sales will exceed the forecasts in the future. For this reason, decision makers who take forecast values literally can be in for a great surprise. Thus, good decision makers should be cautious and communicate with related parties to adjust accordingly.

Ignoring Global Market Movements

The Financial Crisis of 2008–2009 revealed that the era of independent economies no longer exists. No matter how closed or strong an economy is and how sophisticated our forecast model, excluding international market movements might lead to huge errors in our results. We learn that we should utilize the Global Indicator Program provided by the Conference Board to assist us in understanding global financial conditions and adjusting our forecasts as analyzed in Manini and Ozyildirim (2013).

To conclude his remarks, Dr. Theo emphasizes that regardless of what we are trying to forecast, a combination of several quantitative techniques and qualitative judgments usually provides the best results.

Empirical Summary

Dr. App starts her section by sharing her experience in applying the forecast concepts into data analyses.

Data Issues

Collecting Data

If someone offers us a historical dataset, we are very lucky. In many instances, we have to search for one. To locate a dataset it is best to inquire within our company's other departments, partner companies, or firms that we have to order inputs from, or firms to which we deliver outputs. We can also contact federal and local government offices. For example, Cita's office has data on private firms in the city. We learn that we can also ask the Department of Energy, Department of Urban and Regional Planning, or the Department of Business and Economic Development in our city.

Dr. App then says, "In the worst situation, create your own dataset."

At this point Arti offers her own experience, "Two years ago, before opening my art school, I conducted a quantitative survey by sending out survey forms asking people for specific values that they can fill in my questionnaires on their income, the number of children in their households, their preferences in visual and performing arts, and so on. Based on the information, I compiled a dataset for myself on the demand for instructions in the arts. I did not know how to forecast at that time, but I was able to perform some statistical analysis and came up with an estimate of the potential revenue for my school."

Alte also raises her hand and says, "Last year, when I was considering to buy the alteration business at Alcorner, I went to the site and sat on a bench outside the store with a handful of my beads. Whenever a customer was walking out of the store with the store's shopping bag, I moved a bead from my right pocket to my left one. At the end of the day, I multiplied the number of beads with a spending average of $10 for each paying customer to make one data point. I did this every day for two weeks as an approximation of the demand for alterations and came up with a dataset. I eventually decided to purchase the business."

Dr. App thanks them for sharing their experiences with the class and moves on to the next subject. In the following section is what she says.

Cleaning the Data

Once you have successfully downloaded a dataset, do the following steps:

1. Eliminate all text except the labels on the top row.
2. For cross-sectional data or time series data, refer to Chapter 1 commands to transpose the data from horizontal to vertical arrangements.
3. For panel data: You can only transpose the time series section of each identity. You then copy and paste the time series for each identity gradually into Excel.
4. However, you can still refer to Chapter 1 to see how panel data should be organized.
5. Time series periods should be arranged from the lowest to the highest values. Refer to Chapter 1 for the sort commands.
6. A missing observation is often marked with a dot (.) or the letters N/A (nonapplicable). Delete the dot or the letters from the cell.

Missing Observations

Excel cannot handle missing observations. In cross-sectional data analyses, the best solution is to eliminate the observation completely. For example, suppose you want to regress productivity on education and investment. You have data on productivity and investment for 50 states and Washing, DC, but data on education for Washington, DC, is missing. The best course of action is to eliminate Washington, DC, completely from the data analyses.

In time series analysis, eliminating one period creates a gap in the series. For example, Mo wants to regress motorcycle sales on income and has data on motorcycle sales from January 2012 to December 2012, but data on the income of the city residents for June 2012 are missing. In this case, instead of eliminating data for June, the best strategy is to calculate an approximated value of the city resident income in June by averaging the income values in May and July. He can use this average to fill in the missing value for June.

Changing Units

We learn that if we only need to interpret the effect of an explanatory variable on the dependent variable while holding other variables constant, then the units of the explanatory variables should not matter. Since we need to calculate the predicted values in forecasting, we will have great difficulty if too many units are used. For example, if consumption is in dollars, income is in hundreds of dollars, and the stock prices are in thousands of dollars, then before calculating we will have to change the values of the coefficient estimates. Hence, changing all units into a single unit before performing regressions will make the calculations much easier.

Logarithmic Functions

Dr. App also reminds us that the logarithm of zero is undefined as shown in any college algebra textbook. Hence, if a cross-sectional dataset has a zero number, we can eliminate the data point before performing the regressions. If a time series dataset has a zero number, we can replace the zero with a small number. In order for the number to be small enough so that it does not bias the results, we need to scale up the dataset.

For example, if the units are in thousands of dollars and the other values are in the range of 5 through 10, we can change them to dollars so that the values are in the range of 5,000 through 10,000 and then add 1.0 to the whole series so that 5,000 becomes 5,001 and the zero becomes 1.0.

Preliminary Analysis

Preliminary steps of data analysis such as sketching a time series plot and performing descriptive statistics before applying any forecast technique are very important. The fact that we obtain a reliable dataset does not imply that we can use all observations for estimations. For example, the maximum value in descriptive statistics might reveal an extremely high sale volume due to a recent promotion. This value will cause an over-estimation of future sales. Any outlier should be eliminated before data analyses are performed. After a model is developed and forecast values are obtained, add these outliers back to the periods to which they belong. Other values reported in the descriptive statistics are also important as discussed in Chapter 1.

Technical Summary

Finally, Dr. App provides us with a summary table of all techniques introduced in this course and goes over each technique with us. Her table is displayed in Table 11.1.

Dr. Apps concludes her section by telling us that this textbook is an introduction to forecasting. Although it might be enough for simple problems, she hopes that we continue our learning in the future by reading more advanced books whenever we have the time. She also hopes that we combine the quantitative knowledge we have learned in this class with our common sense and qualitative judgment to obtain the best forecasts for our businesses.

We thank Dr. Theo and Dr. App for a course that has provided valuable knowledge that we can apply to our daily lives. Many of our classmates came to this country as political refugees with empty hands but have been doing quite well thanks to their entrepreneurial spirits, the determination to learn, and their eagerness to apply what they have learned into managing their businesses. We know that we can do the same.

It is time to say goodbye. We all wish the professors and each other good luck. We also wish you all the best for your future and hope that you have enjoyed this course as much as we have.

Table 11.1 Summarization of methods learned in this course

Technique	Using conditions	Remarks	Book sections	Illustrations Excel application
Part I: Basics				
Simple MA	Data exhibit neither seasonal nor cyclical components	Treating all periods with equal weight One-period forecasts	Chapter 2, section on "Moving Averages"	Illustration: Figure 2.4 Data and commands: Ch02.xls, Fig.2.4
WM	Data exhibit neither seasonal nor cyclical components	Different weights are assigned using integers One-period forecasts	Chapter 2, section on "Moving Averages"	Illustration: Figure 2.5 Data and commands: Ch02.xls, Fig.2.5
ES	Data exhibit neither seasonal nor cyclical components	One flexible weight $0 < a < 1$ One-period forecasts	Chapter 2, section on "Exponential Smoothing"	Illustration: Figure 2.6 Data and commands: Ch02.xls, Fig.2.6
Part II: Intermediate Forecast Techniques				
DM	Data exhibit neither seasonal nor cyclical components	Combining two MAs and a trend equation Multiperiod forecasts	Chapter 4, section on "Double Moving Average"	Illustrations: Table 4.1, and Figures 4.1, 4.2, and 4.3 Data and commands: Ch04.xls, Fig.4.3
DE	Data exhibit neither seasonal nor cyclical components	Two parameters E: $0 < a < 1$ T: $0 < b < 1$ Multiperiod forecasts	Chapter 4, sections "Brown's DE" and "Holt's DE" under "Double Exponential Smoothing"	Illustrations: Tables 4.2 and 4.3; Figures 4.5, 4.6, and 4.7 Data and commands: Ch04.xls, Fig.4.8 and Fig.4.9
Simple linear regressions	Data are stationary Six classic assumptions hold	Number of explanatory variable: 1 Multiperiod forecasts	Chapter 5, sections "Basic Concept" and "Predictions and Forecasts"	Illustrations: Table 5.1, Figures 5.1, 5.2, and 5.3 Data and commands: Ch05.xls, Fig.5.2, Fig.5.3, and Fig.5.4

(*Continued*)

Table 11.1 Summarization of methods learned in this course (Continued)

Technique	Using conditions	Remarks	Book sections	Illustrations Excel application
Part III: Advanced Forecast Techniques				
Multiple linear regressions	Data are stationary Six classic assumptions hold	Number of explanatory variables: two or more Multiperiod forecasts	Chapter 6, sections "Basic Concept," "Predictions and Forecasts," and "Forecasts with Panel Data"	Illustrations: Figures 6.1, 6.2, 6.3, 6.6, and 6.7 Data and commands: Ch06.xls, Fig.6.2, Fig.6.3, Fig.6.6, and Fig.6.7
Decomposition	Data exhibit seasonal and cyclical components	Multiplicative model Multiperiod forecasts	Chapter 7, section on "Decomposition"	Illustrations: Figures 7.1, 7.2, and 7.3 Data and commands: Ch07.xls, Fig.7.1, Fig.7.2, and Fig.7.3
Holt–Winters exponential smoothing (HWE)	Data exhibit seasonal and cyclical components	Three parameters: E: $0 < a < 1$ T: $0 < b < 1$ S: $0 < c < 1$	Chapter 7, section on "Triple Exponential Smoothing"	Illustrations: Figure 7.4 Data and commands: Ch07.xls, Fig.7.4
Higher-order exponential smoothing (HOE)	Data are stationary	Three parameters One equation in polynomial or log form Linear in parameters	Chapter 7, section on "Triple Exponential Smoothing"	Illustrations: Figures 7.5 and 7.6 Data and commands: Ch07.xls, Fig.7.5 and Fig.7.6
AR (p), ARIMA (p, d, q)	Stationary data: use original AR Random walk: use differenced models	Univariate p = number of lagged dependent variables	Chapter 7, section on "Brief Introduction to AR and ARIMA Models"	Illustrations: Table 7.7 and Figure 7.7 Data and commands: Ch07.xls, Fig.7.8

Part IV: Business and Economic Applications of Forecasting				
Operational forecasting	Demand forecasts are available as prerequisites	Forecasts of forecasts in a running process	Chapter 8, section on "Operational Forecasting"	Illustrations: Tables 8.1 and 8.2; Figure 8.1 Data and commands: Ch08.xls, Fig.8.3
Bond: yield to maturity (YTM) and interest rate forecasts (IRFs)	YTM: assumes constant i IRF: similar to regressions	A guess and adjustment process on an Excel spreadsheet	Chapter 8, sections "Bond Markets" and "Excel Applications" under "Financial Forecast"	Illustrations: none Data and commands: Ch08.xls, Fig.8.4 and Fig.8.5
Stock: expected returns and prices	CAPM and ATP models: similar to regressions DDM: constant growth rates of ΔI and D	CAPM and ATP model: similar to regressions DDM: for short-run forecast only	Chapter 8, sections "Stock Market" and "Excel Applications" under "Financial Forecast"	Illustrations: none Data and commands: Ch08.xls, Fig.8.6
Diffusion models: Bass and Lawrence–Lawton	Assumptions on shapes of *pdf* and cumulative distributions hold	Generate an entire product life cycle from several initial data points; for long-term forecasts	Chapter 8, section on "Diffusion Models on Sales and Demand"	Illustrations: Table 8.3 and Figures 8.5 and 8.6 Data and commands: Ch08.xls, Fig.8.9
Production forecasts	Profit maximization or cost minimization principles	Short-term forecasts on production allocations subject to resource constraints	Chapter 9, section on "Production Forecasts"	Illustrations: none Data and commands: Ch09.xls, Fig.9.1, Fig.9.2, Fig.9.3, Fig.9.4, and Fig.9.5
Gravity models	Physic principles of gravity between two objects	Short-term forecasts of efficient land-use and trip distributions	Chapter 9, section on "Gravity Models"	Illustrations: Table 9.1 Data and commands: Ch09.xls, Fig.9.7

(*Continued*)

Table 11.1 Summarization of methods learned in this course (Continued)

Technique	Using conditions	Remarks	Book sections	Illustrations Excel application
Turning points	Data on economic leading indicators are available	Long-term forecasts: current technique on diffusion indices gives equal weights to all indicators	Chapter 10, section on "Turning Points"	Illustrations: Tables 10.1 and 10.2 Data and commands: Ch10.xls, Fig.10.2 and Fig.10.4
Rates of change: catching up or investment choices	Assumptions: all decisions depend on growth rates of profit, revenue, or investment	Long-term forecasts; four scenarios for catching-up games; more realistic with discount rates for investment choices	Chapter 10, section on "Models Based on Rates of Change"	Illustrations: Table 10.3 Data and commands: Ch10.xls, Fig.10.6 and Fig.10.7

References

Baltagi, B.H. 2006. "Forecasting with Panel Data." Discussion Paper Series 1: Economic Studies, Deutsche Bundesbank no. 25/2006, pp. 1–38.

Bass, F. 1969. "A New Product Growth Model for Consumer Durables." *Management Science* 15, no. 5, pp. 215–27.

Baumohl, B. 2008. *The Secrets of Economic Indicators*. NJ: Prentice Hall.

Bollerslev, T., R.F. Engle, and J. Wooldridge. 1988. "A Capital Asset Pricing Model with Time Varying Co-variances." *Journal of Political Economy* 96, pp. 116–31.

Bradley, J.V. 1968. *Distribution Free Statistical Tests.* NJ: Herman H.J. Lynge & Søn A/S.

Brown, R.G. 1963. *Smoothing Forecasting and Prediction of Discrete Time Series*. NJ: Prentice Hall.

Brown, S. 2014. Measures of Shape: Skewness and Kurtosis. TC3: Statistics. http://www.tc3.edu/instruct/sbrown/stat/shape.htm

Bureau of Economic Analysis.com. 2014. Regional Economic Accounts. http://bea.gov/regional/index.htm

Davidson, R., and J. MacKinnon. 2004. *Econometric Theory and Method.* NY: Oxford University Press.

Department of Business, Economic Development, and Tourism: State of Hawaii. 2014. http://dbedt.hawaii.gov/

Dunn, S., Jr. 1954. *The Location of Agricultural Production*. Gainesville, Florida: University of Florida Press.

Ellis, J.H. 2005. *Ahead of the Curve*. MA: Harvard Business School Press.

Federal Reserve.com. 2014. http://www.federalreserve.gov/releases/h15/data.htm

Froeb L.M., and T.B. McCann. 2010. *Managerial Economics: A Problem Solving Approach.* OH: South-Western.

Fuller, W.A. 1976. *Introduction to Statistical Time Series*. NY: John Wiley and Sons.

Gaynor, P.E., and R.C. Kirkpatrick. 1994. *Introduction to Time-Series Modeling and Forecasting in Business and Economics.* NY: McGraw-Hill.

Gordon, M.J. May 1959. "Dividends, Earnings and Stock Prices." *Review of Economics and Statistics* 41, no. 2, pp. 99–105

Gordon, M.J., and E. Shapiro. 1956. "Capital Equipment Analysis: The Required Rate of Profit." *Management Science* 3, no. 1, pp. 102–10.

Greene, W.H. 2012. *Econometric Analysis.* NJ: Prentice Hall.

Griffiths, W.E, R.C. Hill, and G.G. Judge. 1993. *Learning and Practicing Econometrics.* MA: John Wiley & Sons.

Hamilton, J.D. 1994. *Time Series Analysis*. NJ: Princeton University Press.

Heizer J., and Render B. 2011. *Operations Management.* NJ: Prentice Hall.

Hill, R.C., W.E. Griffiths, and G.C. Lim. 2011. *Principle of Econometrics*. NJ: Wiley and Son.

Holt, C.C. 1957. "Forecasting Seasonals and Trends by Exponential Weighted Moving Averages." Office of Naval Research Memorandum, No. 52. PA: Carnegie Institute of Technology.

Hyndman, R., and G. Athanasopoulos. 2014. *Forecasting: Principle and Practice*. KY: Open-Access Textbooks.

IMF.com: IMF Data and Statistics. 2014. http://www.imf.org/external/data.htm

IMF.com: Direction of Trade Statistics. 2014. http://elibrary-data.imf.org/FindDataReports.aspx?d=33061&e=170921

Kantorovich, L.V. 1940. "A New Method of Solving Some Classes of Extreme Problems." *Doklady Akad Science USSR* 28, pp. 211–14.

Kennedy, P. 2008. *A Guide to Econometrics*. MA: Wiley-Blackwell.

Kmenta, J. 2000. *Elements of Econometrics.* MI: The University of Michigan Press.

Krueger, R. 2008. *Business Forecasting: A Practical, Comprehensive Resources for Managers and Practitioners*. KY: Krueger

Lapin, L.L. 1994. *Quantitative Methods for Business Decisions*. FL: Dryden Press.

Lawrence, K., and. W.H. Lawton. 1981. "Application of Diffusion Models: Some Empirical Results." In *New Product Forecasting*, eds. Y. Win, V. Mahajan, and R. Cardozo, 529–41. KY: Lexington Books.

Lawrence, K.D., R.K. Klimberg, and S.M. Lawrence. 2009. *Fundamentals of Forecasting Using Excel.* NY: Industrial Press, Inc.

Leontief, W.W. 1986. *Input-Output Economics.* 2nd ed. NY: Oxford University Press.

Levinson, D., and A. Kumar. 1994. "The Rational Locator: Why Travel Times Have Remained Stable." *Journal of the American Planning Association* 60, no. 3, pp. 319–32

Levinson, D., and A. Kumar. 1995. "A Multi-modal Trip Distribution Model." *Transportation Research Record* 1466, pp. 124–31

Manini, J.-C., and A. Ozyildirim. 2013. "Using the Indicators Approach to Understand Financial Conditions and Financial Instability. In *Understanding Business Cycles*, ed. A. Ozyildirim. NY: The Conference Board.

Mason, S.J. 2004. "Cross Validation and Other 'Out-of-Sample' Testing Strategies." Paper presented at the AMS Short Course on Significance Testing Model Evaluation and Alternatives, International Research Institute for Climate Prediction, Earth Institute of Columbia University, NY.

Michaelsen, J. 1987. "Cross-Validation in Statistical Climate Forecast Models." *Journal of Applied Meteorology and Climatology* 26, pp. 1589–600.

Montgomery D.C., C.L. Jennings, and M. Kulahci. 2008. *Introduction to Time Series and Forecasting.* NJ: Wiley Inter-Sciences.

National Center for Education Statistics.com. 2014. https://nces.ed.gov/

Orphanides, A., and J.C. Williams. May 2011. "Monetary Policy Mistakes and the Evolution of Inflation Expectations." NBER Working Paper No. 17080, National Bureau of Economic Research, pp. 1–44.

Pindyck, R., and D.L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts*. MA: Irwin McGraw-Hill.

Poirier, D.L. 1996. *Intermediate Statistics and Econometrics.* MA: The MIT Press.

Ramanathan, R. 1998. *Introductory Econometrics with Applications.* FL: The Dryden Press.

Rogers, E.M. 1962. *Diffusion of Innovations*. NY: Free Press.

Rogers, R.M. 2009. *Economic Indicators.* NY: Alpha Books.

Ross, S. 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13, no. 3, pp. 341–60.

Stock, J.H., and M.W. Watson. 2007. *Introduction to Econometrics*. MA: Pearson Addition Wesley.

Thirlwall, A.P. 2003. *Growth and Development*. NY: Palgrave Macmillan.

Thurman, W.N., and M.E. Fisher. May 1988. "Chickens, Eggs, and Causality, or Which Came First?" *American Journal of Agricultural Economics*, pp. 237–8.

Treynor, J.L. 1962. *Toward a Theory of Market Value of Risky Assets*. A final version was published in 1999, in *Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics*, 15–22, ed. Robert A. Korajczyk. London: Risk Books.

Tsekeris, T., and A. Stathopoulos. 2006. "Gravity Models for Dynamic Transport Planning: Development and Implementation in Urban Networks." *Journal of Transport Geography* 14, no. 2, pp. 152–60.

U.S. Department of Agriculture.com. 2014. http://www.ers.usda.gov/Data/Macroeconomics/

Voorhees, A.M. 1956. *A General Theory of Traffic Movement.* New Haven, CT: Institute of Traffic Engineers.

Williams, J.B. 1938. *The Theory of Investment Value*. MA: Harvard University Press.

Winters, P.R. 1960. "Forecasting Sales by Exponentially Weighted Moving Averages." *Management Science* 6, no. 3, pp. 324–42.

Wooldridge, J.M. 2013. *Introductory Econometrics: A Modern Approach.* OH: South-Western.

World Bank.com. 2014. World Development Indicators. http://data.worldbank.org/data-catalog/world-development-indicators

Index

OTHER TITLES FROM THE ECONOMICS COLLECTION

Philip Romero, The University of Oregon and
Jeffrey Edwards, North Carolina A&T State University, Editors

- *Fiscal Policy within the IS-LM Framework* by Shahdad Naghshpour
- *Monetary Policy within the IS-LM Framework* by Shahdad Naghshpour
- *Building Better Econometric Models Using Cross Section and Panel Data* by Jeffrey A. Edwards
- *Basel III Liquidity Regulation and Its Implications* by Mark Petersen and Janine Mukuddem-Petersen
- *Saving American Manufacturing: The Fight for Jobs, Opportunity, and National Security* by William R. Killingsworth
- *What Hedge Funds Really Do: An Introduction to Portfolio Management* by Philip J. Romero and Tucker Balch
- *Advanced Economies and Emerging Markets: Prospects for Globalization* by Marcus Goncalves and Jose Alves
- *Comparing Emerging and Advanced Markets: Current Trends and Challenges* by Marcus Goncalves and Harry Xia
- *Learning Basic Macroeconomics: A Policy Perspective from Different Schools of Thought* by Hal W. Snarr
- *The Basics of Foreign Exchange Markets: A Monetary Systems Approach* by William D. Gerdes
- *Learning Macroeconomic Principles Using MAPLE* by Hal W. Snarr
- *Macroeconomics: Integrating Theory, Policy and Practice for a New Era* by David G. Tuerck
- *Emerging and Frontier Markets: The New Frontline for Global Trade* by Marcus Goncalves and Jose Alves
- *Doing Business in Emerging Markets: Roadmap for Success* by Marcus Goncalves, Jose Alves, and Rajabahadur V. Arcot

THE BUSINESS EXPERT PRESS DIGITAL LIBRARIES

EBOOKS FOR BUSINESS STUDENTS

Curriculum-oriented, born-digital books for advanced business students, written by academic thought leaders who translate real-world business experience into course readings and reference materials for students expecting to tackle management and leadership challenges during their professional careers.

POLICIES BUILT BY LIBRARIANS

- *Unlimited simultaneous usage*
- *Unrestricted downloading and printing*
- *Perpetual access for a one-time fee*
- *No platform or maintenance fees*
- *Free MARC records*
- *No license to execute*

The Digital Libraries are a comprehensive, cost-effective way to deliver practical treatments of important business issues to every student and faculty member.

Seeing the Future

How to Build Basic Forecasting Models

Tam Bang Vu

This book guides you through an enjoyable journey, step by step, into the future. A team of fictional characters is introduced to share their learning and working experiences with the readers. In the beginning of the book, you will take the first step by learning the most basic models for one-period forecasts based on past performance of a market. You will also learn how to evaluate your newly built models. Next, you will progress further into intermediate-level models, including multi-period forecasts based on past performance of a market or based on an external factor. It also introduces interval forecasting, which allows you to obtain a range of forecast values instead of a single value in the future.

In the second half, you will familiarize yourself with advanced models that provide multi-period forecasts based on multiple internal or external factors. Toward the end, you will learn several applied models in business and economics that will facilitate you with practical applications related to real life situations. The last chapter summarizes all models introduced in this book and provides a table of references for finding the most important concepts, tables, and figures in the book so that you can recall every step of your adventure.

Tam Bang Vu holds an MM from Manhattan School of Music and a PhD in economics from the University of Hawaii-Manoa. She is an associate professor, the economics chair at the College of Business and Economics, and the interim director of the MA program in China-US Relations at the University of Hawaii-Hilo. She is also a cooperating graduate faculty member at the University of Hawaii-Manoa. She had been a sole proprietor of various businesses for more than 20 years before her academic career.

ECONOMICS COLLECTION

Philip J. Romero and Jeffrey A. Edwards, *Editors*

ISBN: 978-1-60649-790-6
90000
9 781606 497906

BUSINESS EXPERT PRESS