

VUZF UNIVERSITY OF FINANCE, BUSINESS AND
ENTREPRENEURSHIP

Olha Mezentseva
Oksana Ilienکو
Oleksii Kolesnikov
Olena Savielieva
Dmytro Lukianov

**FORECASTING MODELING AND ANALYTICS
OF ECONOMIC PROCESSES**

Sofia, 2020

VUZF University of Finance, Business and Entrepreneurship
Publishing House “St. Grigorii Bogoslov”

Chief Editor: Assoc. Prof. Dr. Grigorii Vazov

Deputy Chief Editor: Assoc. Prof. Dr. Radostin Vazov

Members:

Prof. Doctor of Economic Sciences Metodi Hristov, Prof. Dr. Radoslaw Grabowski, Prof. Dr. Daniela Bobeva, Prof. Dr. Emilia Milanova, Prof. Dr. Virginia Zhelyazkova, Prof. Dr. Grigor Dimitrov, Assoc. Prof. Dr. Yakim Kitanov, Assoc. Prof. Dr. Desislava Yosifova, Assoc. Prof. Dr. Stanislav Dimitrov, Assoc. Prof. Dr. Krassimir Todorov, Assoc. Prof. Dr Daniela Ilieva

Recommended for publication by the Editorial Board of the VUZF University of Finance, Business and Entrepreneurship, 12 August 2020

Reviewers:

Igor Britchenko – Doctor of Economics, Professor, Department of Finance, VUZF University of Finance, Business and Entrepreneurship, Sofia, Bulgaria

Maksym Bezpartochnyi – Doctor of Economics, Professor at the Department of Economics and Marketing, National Aerospace University named after N. Zhukovsky “Kharkiv Aviation Institute”, Kharkiv, Ukraine; Professor at the University Economics and Social, Przeworsk, Poland

Jelena Caiko – Doctor of Technical Sciences, Professor, ISMA University, Riga, Latvia

D 94 Forecasting modeling and analytics of economic processes / O. Mezentseva, O. Iliencko, O. Kolesnikov, O. Savielieva, D. Lukianov. Sofia: VUZF Publishing House “St. Grigorii Bogoslov”, 2020. – 252 p.

The book will be useful for economists, finance and valuation professionals, market researchers, public policy analysts, data analysts, teachers or students in graduate-level classes. The book is aimed at students and beginners who are interested in forecasting modeling and analytics of economic processes and want to get an idea of its implementation.

© O.Mezentseva, O.Iliencko, O.Kolesnikov, O.Savielieva, D.Lukianov, 2020

© VUZF Publishing House “St. Grigorii Bogoslov”, 2020

ISBN 978-954-8590-88-4

TABLE OF CONTENTS

INTRODUCTION	5
CHAPTER 1. MATHEMATICAL AND SIMULATION MODELS IN BUSINESS ECONOMICS	9
CHAPTER 2. LIBRARY FUNCTIONS FOR BUSINESS ECONOMICS	15
2.1 Model Demand Equation	22
CHAPTER 3. ECONOMIC IMPACT MODELS	29
CHAPTER 4. FISCAL IMPACT MODELS	51
CHAPTER 5. APPLICATIONS FOR BUSINESS	61
5.1 Business Valuation and Damages Estimation	64
5.2 The Role of Economist and Accountant in Business Valuation	68
CHAPTER 6. APPLICATIONS FOR FINANCE	102
6.1 Auto Depreciation	108
CHAPTER 7. MODELING LOCATION AND RETAIL SALES	114
7.1 Applying Economic Theory to Sales Location	116
7.2 Drive-Time Analysis	119
7.3 Applying a Gravity Model to the Distance-Sales Analysis	125
7.4 Monte Carlo Data on Retail Sales	129
7.5 Nonlinear Optimization and Gravity Model Estimation	133
7.6 Gravity and Polynomial Models	135

CHAPTER 8. DETERMINING SALES AND MARKET AREAS	140
8.1 Defining Market Areas	142
CHAPTER 9. APPLICATIONS FOR MANUFACTURING	148
CHAPTER 10. FUZZY LOGIC BUSINESS APPLICATIONS...	155
CHAPTER 11. THE NATURE OF REGRESSION ANALYSIS	171
11.1 Historical origin of the term regression	171
11.2 Statistical versus deterministic relationships	175
11.3 Types of Data	178
11.4 The Accuracy of Data	183
CHAPTER 12. TWO-VARIABLE REGRESSION ANALYSIS.	188
12.1 The sample regression function (SRF)	197
CHAPTER 13. APPLICATIONS OF ANALYTICS IN THE BANKING SECTOR.....	203
13.1 Logistic Regression Equation	209
13.2 Regression Model	214
13.2 Logistic Regression Model Using R	220
13.3 Model Building and Interpretation of Training and Testing Data	226
CONCLUSIONS	231
REFERENCES	234

INTRODUCTION

The primary objective of this book is to help economists, statisticians, developers, engineers, and data analysts who are well versed in writing codes; have a basic understanding of data and statistics; and are planning to transition to a data scientist profile. The most challenging part is practical and hands-on knowledge of building predictive models and machine learning algorithms and deploying them in industries to address industrial business problems. This book will benefit the reader in solving the business problems in various industrial domains by sharpening their analytical skills in getting practical exposure to various predictive model and machine learning algorithms

This book focuses on industrial business problems and practical analytical approaches to solve those problems by implementing predictive models and machine learning techniques using MATLAB, and R analytical languages.

Data analytics has become part and parcel of any business in today's world. In fact, it has evolved into an industry in itself. Vast numbers of software platforms are available for data extraction, scrubbing, analysis, and visualization. Some of these platforms are specialized for carrying out one of the above-listed aspects of data analytics, while others offer a generalist tool to carry out almost all tasks ranging from data scrubbing to visualization.

Who may read this book:

1. Economists, finance and valuation professionals, market researchers, public policy analysts, and other practitioners whose occupation demands the ability to model market behavior under a variety of real-world conditions
2. Teachers or students in graduate-level classes in applied economics, particularly those who cover topics such as fiscal and economic impact analysis, fuzzy logic applications, retail sales analysis, and the integration of geographic information systems, as this is one of the very few books to rigorously cover these subjects.

Prerequisite Knowledge

We presume the reader has the following knowledge:

1. A good grounding in the laws of microeconomics and a

familiarity with their application. Most practitioners and graduate students in economics, finance, public policy, and related disciplines should either be able to acquire this or should have already done so. Most undergraduates will need additional training.

2. A working understanding of mathematics, including (for some chapters) an understanding of the calculus used in comparative statistics. The completion of a graduate course in math for economists would be helpful, but is not mandatory.

3. For many of the chapters, a familiarity with (but not necessarily expertise in) the MATLAB software environment. Most of the applications presented here could be accomplished in other software environments, but we provide examples of applications in MATLAB.

4. For some chapters, an awareness (but not necessarily a working knowledge) of geographical information systems.

What You Will Learn

- Introduction to analytics and data understanding.
- How to approach industrial business problems with an analytical approach.
- Practical and hands-on knowledge in building predictive model and machine learning techniques.
- Building the analytical strategies.

Economic and fiscal impact models as well as tax policy and forecasting tax revenue are covered in Chapter 3, Chapter 4, and Chapter 5. These chapters provide practitioners in these fields with advanced tools that enable more accurate, reliable analysis. The first sections of the chapter on economic impact models identify the severe weaknesses in many common economic impact reports, adding in some examples of particularly absurd gross exaggeration in published reports. These sections close with a plea for ethics in that subfield. As much of our work has been in public policy, we use these tools on a regular basis and hope that the analytical tools will help improve the level of analysis and that our admonitions about ethics in using them will not fall on deaf ears. Chapter 6, Chapter 7, and Chapter 8 cover the economics of a firm. Chapter 9 describes modeling a firm, focusing on the crucial question of how a firm makes money. An iterative method of matching market value and

expenses based on value is introduced. Chapter 10 discusses business valuation, noting the various methods used to estimate the value of a business and the various definitions of value. The chapter differs from most other references in this field in two ways: first, we concentrate on what we consider the most important consideration in the market value of a company, namely its prospects for future earnings; and, second, we utilize the greater analytical power of quantitative models that the MATLAB environment provides. d. Retail sales are discussed in Chapter 4. Much of the material presented here, including the derivation of the distance-sales relationship from microeconomic foundations, is unique to this book. Other portions provide craft knowledge of this interesting field that is rarely described in print.

Each of these chapters describes briefly the theory behind the models and then outlines an approach. These chapters can be read, skimmed, or skipped, depending on the interest at hand. A reader dealing with a finance problem, or one interested in fuzzy logic applications, could jump right to this chapter (after, of course, reviewing the initial chapters) without necessarily reading up on economic or fiscal impact models.

Analytics is a big buzz and a need for today's industries to solve their business problems. Analytics helps in mining the structured and unstructured data in order to withdraw the effective insights from the data, which will help to make effective business decisions. SAS and R are highly used tools in analytics across the globe by all industries for data mining and building machine learning and predictive models. This book focuses on industrial business problems and a practical analytical approach to solve those problems by implementing predictive models and machine learning techniques using SAS and R analytical languages. The primary objective of this book is to help statisticians, developers, engineers, and data analysts who are well versed in writing codes; have a basic understanding of data and statistics; and are planning to transition to a data scientist profile. The most challenging part is practical and hands-on knowledge of building predictive models and machine learning algorithms and deploying them in industries to address industrial business problems. This book will benefit the reader in solving the business problems in various industrial domains by sharpening their

analytical skills in getting practical exposure to various predictive model and machine learning algorithms.

We have provided the actual code for a number of MATLAB programs we have prepared to illustrate the techniques described in this book. Longer sections are listed as the code fragment “A Simple Startup File”. We call these code fragments because they are written in a specific software language or code, and because many of the longer code sections are shortened, the code included in this text may be only a portion or fragment of the entire file.

Keep in mind that software companies change the menus in their products frequently, and you may need to use a different set of commands if you are using a different version of the software or on a different operating system. Also, there may be a shortcut method using a combination of keystrokes or mouse actions that can accomplish the same operation in a different way.

The versions of MATLAB used during the development of this book included R11, R12, R13, and a prereleased version of R14. We believe that all of the routines will run, in some cases with minor modifications, on R12 and R13. Most will run on R11 as well. Based on our testing of the comparison business economics toolbox with the prereleased version of R14, we anticipate that all of the routines included here will also work under R14, perhaps with minor modifications. Users should expect that different versions will have some interface differences as well.

This book is about applying economic principles to practical problems in business, finance, public policy, and other fields. The principles themselves are timeless; the applications described in this book can be done in a number of software environments. Indeed, the first principle we recommend – important enough that we call it a maxim – is to think through the problem before trying to restate it in mathematical terms.

In this book, we provide detailed descriptions of analysis in three of these environments: mathematical and simulation models (MATLAB and Simulink), Geographic Information Systems (GIS), and spreadsheets (particularly Microsoft Excel). This book, however, does not provide a tutorial for these software programs. Therefore, we provide below suggestions for those readers who would benefit from additional information on the software products described in this book.

CHAPTER 1. MATHEMATICAL AND SIMULATION MODELS IN BUSINESS ECONOMICS

Economics is a social science. It seeks to describe the incentives operating on individuals as they work, save, invest, contribute, vote, and consume. Like any field of science, economics can be studied at the widest theoretical level, at a minute practical level, and at any point in between. This book looks at the practical applications of economic reasoning, using techniques that are based in rigorous theory and proven in actual practice. We entitle the book *Business Economics and Finance* because the majority of the applications focus on individuals in business dealings. However, there are many topics that focus on government organizations and their interaction with individual workers, managers, and consumers. It is worth pausing to review how this practical, applied economics differs from the theoretical economics that is taught in colleges and universities today. While there is plenty of cross-over among the subdisciplines, and most economists devote some of their time to both, it is worth considering the differences. Table 1.1, “Theoretical vs. Applied Economics,” highlights some of the important characteristics of both.

Why Practice Business Economics? As Table 1.1 indicates, academic economics normally deals in a smoother world in which data are available or easily assumed, model structures are specified often without argument, and rigorous statistical inferences are common practice. Why, then, practice applied economics with its messy or difficult-to-obtain data, lack of specification of models, and stricter budget constraints? The answer to that question is much like the answer to the question “Why climb the mountain?” Because it is there! There are a huge number of practical questions facing individuals, their employers, their governments, and the

This comparison is inspired by a similar one prepared by W.L.Martinez and A.R.Martinez in *Handbook of Computational Statistics* (Boca Raton, FL: CRC Press, 2002), in which they compare academic statistics with computational statistics. stores and service providers that sell to them. These questions normally cannot be solved with theoretical economics alone; they must be approached with applied economics. This book deals with applied economics and

shows how advanced use of mathematical and simulation software can help answer the tough, messy, and immediate challenges of today.

TABLE 1.1 Theoretical vs. Applied Economics

Characteristic	Theoretical Economics	Applied Economics
Data	Usually assumed to exist in exactly the right periodicity	Must be acquired, often with significant effort; will contain errors, breaks, and different periodicities
Data sources	Data are assumed to exist; unobservable variables may be used in equations Desired data often do not exist.	Observed data must be used from private, government, and nonprofit organizations, and experiments
Analysis	Simple to complex algorithms; often developing theory using mathematics	Simple to complex algorithms; often testing applications of existing theory using statistics
Questions	Can be posed by researcher, based on interest; abstraction allows for intense focus on specific questions	Usually posed by events beyond the control of the researcher; often multiple questions are posed
Time and budget constraints	Time constraints are largely based on academic calendar; academic budgets include time for pure research	Time constraints are largely based on events; budget constraints established by clients

TABLE 1.1 (continued)

Inference practices for parameters	Rigorous statistical inference, given strong assumptions and assumed data	Practical limitations, given potential for specification error and data problems
Assumptions about data and error process	Strong, often unverified assumptions regarding data and errors; error processes often assumed to be clean	Relatively few assumptions about data and errors; error processes often messy
Knowledge of underlying structure	Often strong assumptions about underlying structure, based on theory	Structure may be unclear, or is typically inferred from actual behavior
Key focus of specification	Optimal analysis	Robust analysis
Importance of location and geography	Often ignored or assumed away	Normally quite important and may be the focus of inquiry

The challenge of practical business economics should not dissuade one from learning in an academic setting. Many, though not all, of the techniques described in this book originated in academic settings. Even the mountain climber who climbs mountains because they are there probably went to a mountain-climbing school. This book borrows some techniques of academic economics in laying out data and sources, giving proper credit to those that have developed or written about a problem before, and in conveying information to others interested in this field. While practical problems require practical solutions, one should still remain true to the essential philosophies of social science.

Simulation Models vs. Spreadsheets Spreadsheets are probably the most common analytical program used in business economics. Spreadsheets have powerful advantages, including low costs, intuitive use, and an ability to easily print the data used in its calculations. For most accounting work and many other uses (including uses we will describe in later chapters), you cannot beat a

spreadsheet package. However, for analytical work requiring dynamic interaction among variables – meaning that one variable affects another, which affects two more, and the latter affects the results of the next-period variables – spreadsheets fall far behind. Although you can “trick” a spreadsheet into performing some advanced analysis and even limited dynamic equations, such uses are beyond their intended scope.

For example, the rainfall-runoff TOPMODEL first developed in 1979 and written in Fortran (found at <http://www.es.lancs.ac.uk/hfdg/topmodel.html>); Agricultural Non-Point Source (AGNPS), developed by the USDA Agricultural Research Service to predict soil erosion and nutrient transport within agricultural watersheds and written in C (found at: <http://www.geog.uni-hannover.de/phygeo/grass/agnps.html>): and numerous others, particularly in electronic devices and software design. MATLAB® and Simulink® are trademarks of The MathWorks, Inc., which can be found at <http://www.MathWorks.com>. For example, some years ago the author created a multiperiod model of the educational market, which was used to forecast the migration of students to private schools should a school-choice program involving those schools be approved by voters in a large industrial state. The model, implemented in Microsoft Excel, took nearly a dozen individual spreadsheets, each with many columns, to project the fiscal impact. The modeling task took at least a month. Explaining how it was done took many pages in the resulting report. More recently, using the tools and skills explained in this book, a similar project was completed in half the time with more accuracy. See Patrick L. Anderson et al., *The Universal Tuition Tax Credit*, Midland, MI: The Mackinac Center, 1998; available at <http://www.mackinac.org>. 20 Even multiplication is difficult to show adequately. For example, consider how taxes are shown on accounting statements. Unless the tax is a straightforward multiple (such as a flat-rate sales tax), the tax liability is normally shown without providing the reader the ability to review the tax rate, base, or calculations. Of course, it is possible to see the formulas in a spreadsheet, but the variables used are typically cell reference, and a special command or set of keystrokes is often required to reveal the formulas used to generate results in each cell of a spreadsheet. It is

possible to display the results of one scenario involving multiple, interactive variables over more than one time period. This is what accounting statements for a firm provide. However, if you look at a set of income statements and balance sheets for a corporation and ask “What if sales were higher, commissions lower, and personnel costs sharply reduced,” you run immediately into the limitations of a 2-D environment. Although you could display this alternate scenario, the environment itself would not keep track of the relationships among these variables or allow an efficient manner of generating a large number of alternate scenarios involving true interactivity among the variables.

Spreadsheet software has serious limitations as an analytical tool. These limitations include:

1. A spreadsheet is inherently a two-dimensional device. Many problems can be solved in a 2-D environment, including almost all accounting problems. Indeed, spreadsheet is a term borrowed from accounting. However, most problems in business are not two dimensional.

2. While spreadsheets easily display the data used for calculations, they do not easily display the formulas used to calculate that data. If all the calculations are simple additions and the proper formatting is applied, this limitation is largely overcome. However, when anything more complex than simple addition or subtraction occurs, spreadsheets often obscure the calculations.

3. Most problems involving multiple variables that interact with each other over time cannot be modeled correctly using a spreadsheet. Even if the analyst creates multiple worksheets showing different scenarios in an attempt to overcome the 2-D limitations of the spreadsheet itself, the number of potential scenarios overwhelms the ability of the software to calculate and then express the results. With two dimensions, a spreadsheet can typically handle either a small set of variables over time, or a large set of variables at one time – but not both. The sheer size of a spreadsheet model necessary to model a multiperiod, multivariable problem can be overwhelming.

4. Although not a defect in the software itself, the widespread use of spreadsheets induces an intellectual laziness regarding how variables interact together. It is much too easy to look at a column of numbers, notice that they add up correctly and then conclude that the

number at the bottom of the column must be correct.

Comparison with Simulation Models Consider how a true simulation modeling environment, such as MATLAB or Simulink, overcomes some of the limitations of spreadsheet models:

1. A simulation model is n-dimensional. You can include as many variables as you think are important and relate them in multiple ways. You are not restricted to thinking in rows and columns.

2. An enormous variety and number of operators and functions can be used in simulation models. Furthermore, these operators and functions are clearly visible when creating the model, rather than being buried underneath a cell in a spreadsheet. This allows – in fact, requires – you to specify exactly how variables interact.

3. Simulation models are designed to handle a large number of variables which interact and change over time. Such an environment can therefore model more effectively the actual workings of markets with many participants.

4. Using a simulation model environment forces the analyst to explicitly identify the variables, and describe their interaction over time. Indeed, part of the relative difficulty in using the environment is the requirement to explicitly describe all these factors in a manner that is absolutely consistent. While within a spreadsheet environment you can easily mix apples with oranges, it is more difficult to do so within MATLAB. While you can make errors using simulation models, gross specification errors are less likely.

CHAPTER 2. LIBRARY FUNCTIONS FOR BUSINESS ECONOMICS

Common Analytical Tools in Economics

Economists and all business analysts that use economic reasoning rely on a mental library of analytical tools. These tools are not well implemented in any single software environment. Therefore, economists typically use a range of tools, such as spreadsheets, statistical software, graphics utilities, and different modeling languages to complete their tasks. MATLAB does not provide a complete environment for all these tasks. However, it provides much more power to perform analysis than spreadsheets, statistical software, and modeling languages alone, and can often accomplish tasks in one software environment that might otherwise take two or three.

Library of Tools in MATLAB and Simulink In this chapter, we recommend a number of tools for use in the MATLAB environment, some of which we have programmed specifically for the field. We call these a library of tools partially because the metaphor of shelves of books is somewhat appropriate (a toolbox metaphor is also fitting) and partially because MATLAB provides for the creation of special libraries of m-files and Simulink blocks.

Microeconomic Theory and Practical Applications

We first discuss the economic theory that underpins the techniques, and then present the techniques themselves, along with notes for practical application.

Comparative Statics Many of these analytical tools are derived from comparative statics, the study of how changes in one variable affect others in a static model. Such tools have been developed by some of the great minds of the 20th century, including Alfred Marshall, John Hicks, Jon Von Neumann, and Paul Samuelson.⁶¹ We present below some of the building blocks of comparative statics in economics, first discussing theory and then presenting implementations using the MATLAB environment.

Estimating Demand Schedules Consider the ubiquitous demand curve. It is based on axioms and implications of theory that establish the following rule: assuming that all other variables stay the same, an increase in the price of a product or service will result in lower

demand for that product. In business economics, we will not derive the theories of microeconomics. Instead, we will apply them to real-world challenges. Numerous such challenges in business economics and related fields require knowledge of consumer demand. We recommend the following steps for acquiring data and for specifying, estimating, and implementing a demand function.

Step One: Acquire Data the first step is very important. The required data must not only demonstrate equilibrium prices (prices at which supply and demand intersect), but also Alfred Marshall's (1842-1924) *Principles of Economics* (London: Macmillan, 1890, 1920) introduced innovations such as the ubiquitous supply-demand charts, with price and quantity as the axes. He also defined the concept of elasticity of demand in the 1885 article "On the Graphic Method of Statistics" for the *Journal of the [London] Statistics Society*, although other economists before him had described similar concepts. Sir John Hicks helped develop the notion of demand, substitution, and expenditures, and represented the "Slutzky" decomposition of demand changes into income and substitution effects. The theory was originally stated by Russian economist Eugene Slutsky. John Von Neumann helped develop the utility theory and economic behavior under "game" conditions. Paul Samuelson's *Foundations of Economic Analysis* (Cambridge, MA: Harvard University Press, 1947) is the seminal work for establishing mathematical analysis of microeconomic questions within modern economics. He was awarded the 1970 Nobel Prize in economics for developing the "Static and Dynamic Economic Theory." See Hal Varian, *Microeconomics*, 1st ed. (New York: Norton, 1978), especially the notes to chapters on the theory of the firm, theory of the market, and theory of the consumer; Economics New School Web site, *History of Economic Thought*, at <http://cepa.newschool.edu/het/home.htm>; various entries; *The New Palgrave Dictionary of Economics* (New York and London: Stockton and Macmillan, 1987); the biographical entry for Paul Samuelson at the Nobel Web site at: <http://www.nobel.se/economics>; the biographical entry for John Von Neumann at the University of St. Andrews (Scotland) site at: <http://www.history.mcs.standrews.ac.uk/history/References/VonNeumann.html>. Meticulous microeconomists will note the theoretical

existence of Giffen goods, in which a reduction in price of sufficient quantity would result in a decrease in demand. Potatoes are often offered as an example. An autographed copy of this book awaits the first researcher that identifies, acquires data for, and estimates the demand curve for a Giffen good.

Changes in demand during times when prices change. Unless the data provide instances where demand is higher and lower, there can be no demand schedule or curve derived from the data. Similarly, there must be data that distinguishes changes in demand for a specific commodity from such changes for other commodities. This is the essence of the identification problem in econometrics.⁶³ Assuming that we acquire such data, we will go to the next step.

Step Two: Estimate Demand Equation Once the data have been acquired, we can proceed to specify a demand equation and estimate its parameters. MATLAB provides extensive statistical functions for the estimation task, including the Curve-Fitting Toolbox. Examples of demand equation specifications with good theoretical foundations include the following: Distance in a transformed demand equation: One portion of the true price of a product or service is the cost of commuting to the store and back. This commute could be as short as a quick exit off the highway and back on (for example, when stopping for gas on an automobile drive) or several round-trip visits (for example, when selecting a college or choosing an automobile dealership). Thus, a distance-sales relationship is fundamentally a demand equation, in which a portion of the price is transformed into a different variable, namely distance. We can evaluate this more rigorously by decomposing the total price of a good into two components: the transaction price and the distance price:

$$p = p^t + p^d \tag{2.1}$$

where: p – is the true price of the good and the other terms are the transaction price (the price paid in the purchase transaction for goods and services unrelated to distance) and the price related to the distance between the consumer’s location and the retail location. We insert this price into the general demand function:

$$x=f(y, p) \tag{2.2}$$

where: y – is income and p is the relative price of the good to other goods. If we differentiate the demand function with regard to price, we get:

$$dx = \frac{\partial x}{\partial y} dy + \frac{\partial x}{\partial p} dp \tag{2.3}$$

At this point we make some assumptions that simplify this relation. Assume first that the transaction price is the same at retailers located throughout the area. This will be true, or nearly true, in many cases. Second, assume that income does not change during the time period of the data, an assumption that will again be consistent with the facts in many cases. The assumptions and the resulting demand equation can be summarized as:

$$dy=0 \tag{2.4}$$

$$dp=dp^t+dp^d \tag{2.5}$$

$$dp^t=0 \tag{2.6}$$

$$dx = \frac{\partial x}{\partial p} dp^d \tag{2.7}$$

Thus, under certain assumptions that will often be fulfilled, the demand equation can be transformed into a distance-sales equation.

Indirect Addilog Demand Model: This model is based on an indirect utility function of consumers, in which their desire for goods is based on a power function of the relative price of goods and their income. A linear form of the model, in which x_i/x_j is the relative demand for goods i and j and p_i/p_j is the relative price, is:

$$\ln(x_i/x_j)=a+b \ln(y)+c \ln(p_i/p_j) \tag{2.8}$$

Note that the parameters a , b , and c in this equation represent combinations of parameters in the original consumer utility function. The decomposition of those parameters may or may not be important in applied work. However, it is important that business economists evaluating demand at least consider the foundations for specifying the equation in a certain fashion. Linear Demand Model: This is the simplest model and the one most likely represented in textbooks. It also has the weakest theoretical basis. However, under certain limited conditions it is the most practical. The most basic linear model is of familiar form:

$$x_i = a + bp_i \tag{2.9}$$

Most analysts should include relative prices and income, using this form:

$$x_i = a + by + c(p_i/p_j) \tag{2.10}$$

Semi-Log Demand Model: A variation of the linear-in-the-logs model described above is:

$$\ln(x_i) = a + by + c(p_i/p_j) \tag{2.11}$$

Other Demand Models: There are a variety of other specifications for demand functions using logs, semi-logs, powers, rational, and other combinations. The Curve-Fitting Toolbox provides a straightforward manner of specifying and estimating multiple versions. The conditions mentioned under “Practical Basis for Linear Demand Models” on page 101 are likely to hold true as useful approximations of reality. Directly Specified Demand Models: In many cases, you will have a prior knowledge, often from previous research, that suggests a specification of a demand model. Consider carefully whether the previous analyses were based on data that are actually comparable to the market participants you are modeling and, if not, make allowance for behavior to be different.

Practical Basis for Linear Demand Models The practical and theoretical basis for a linear demand equation is based on the following:

1. The observed demand is usually not of individual consumers but of a large number of aggregate consumers. Therefore restrictive utility functions used in microeconomic theory for individual consumers are not the direct underpinnings of models of aggregate consumer demand. This allows the demand functions for aggregate consumers to be modeled as if there were a large group of consumers for whom there is a continuous demand function. This is especially appropriate when studying demand for a narrow product category targeting a narrow class of consumers.

2. Given the large number of different factors that affect numerous consumers, an appeal may be made to the central limit theorem to justify an assumption of normal errors when estimating such an equation. Note that such an appeal requires a large sample size or other information on the distribution of small-sample variables.

3. Although the underlying demand function of aggregate consumers is probably not linear, the observed range of data may be approximated by a linear function which may include variables that are products, logs, or other combinations of other variables.

4. In practical applications, the relative price can be approximated by deflating the nominal price by a general price index, and the budget constraint can be relaxed when the price of the commodity is very small relative to overall income.

The central limit theorem is one of the most important in statistics and econometrics. It may be summarized as: the distribution of the mean of a very large number of independent random variables converges to a normal distribution. A more careful mathematical derivation builds from the sample of any distribution with a finite mean and variance; almost all economic and business variables of interest satisfy this quality. The theorem holds that, if X_i are the members of a random sample from a distribution with a mean μ and positive variance σ^2 , then the random variable where n is the sample size, and is the sample mean of the X_i , has a limiting distribution that is normal, with mean zero and variance one. As the name implies, the central limit theorem describes the limiting distribution of a sample of variables. It does not mean that a small sample of variables about which we know very little is distributed normally. The desirable properties of standard ordinary least squares estimators for

the parameter vector β in the standard model $Y=X\beta+\varepsilon$ will still hold when the expectation of the dependent variable Y , $E(Y)$, is linear in the unknown parameter β . This allows for cases where the explanatory variables X can be the natural logs of other variables or even products or powers of other variables. For a discussion of these and other practical applications of the linear model, see H. Theil, *Principles of Econometrics* (New York: John Wiley & Sons, 1971), Section 3.9, "Limitations of the Standard Linear Model." Be careful not to indiscriminately use the U.S. CPI, which has a well-known tendency to exaggerate underlying price inflation due to a persistent substitution bias and an overweighing for many consumers in energy prices and housing costs.

General Cautions on Estimating Equations Always consider the following cautions when specifying and estimating equations in an economic model such as a demand equation:

1. Microeconomic theory is based on well-behaved utility functions. These utility functions and other axioms typically prohibit purchases of negative quantities or negative prices, ensure consumers live within a budget constraint, prefer more to less, and like to pay less rather than more. A model should not violate those assumptions unless there is a good reason to believe that ordinary axioms do not apply.

2. Note that many models fail to follow a number of the assumptions listed above. In particular, linear models will allow negative prices and quantities; models that do not include income ignore the budget constraint; and use of nominal prices (especially with time-series data) violates the assumption that relative prices are the basis for consumer choices. You may be able to confidently use a model that, across all real numbers, violates one or more assumptions. One may do so only if the data and theory strongly support the model, the data are well outside the implausible range, and the range of the model is restricted.

3. The fact that data can be loaded into a regression model and the model is estimated with high "goodness-of-fit" statistics (such as R^2) means very little. It is much better to use a model for which there is strong theoretical and practical support than to simply run regressions until good results are obtained. It is occasionally useful to remind oneself that "correlation is not causality".

4. If demand is based on multiple variables and a model is made using only one, then the model is misspecified. There may be serious consequences to this, including the estimation of parameters that are biased. Furthermore, omitting variables from an equation when such variables are observed by the market participants and affect their behavior will typically result in biased estimates. There are econometric techniques that sometimes minimize the effect of such biases. However, it is much better to properly specify the model in the first place. Again, excellent goodness-of-fit statistics provide no benefit when the “fit” is biased.

5. Consider the measurement accuracy of your variables and whether any of them include a certain amount of random error. Any error in the dependent or independent variable results in the estimated parameter having higher variance. In general, standard econometric techniques continue to work well when the explanatory variables in a regression equation have a certain amount of randomness, as long as the random behavior is not correlated with that of the dependent variable. However, when that randomness is correlated with the behavior of the dependent variable, the resulting estimates are most likely biased.

2.1 Model Demand Equation

Using the data and estimation results from the previous two steps, we can model consumer demand for a specific commodity. MATLAB provides a number of methods, the most general being the creation of a function. Demand function: We can program a MATLAB function to calculate demand based on one or more inputs. For example, the following function produces a demand schedule for a commodity, using a single (price, quantity) pair and one additional parameter (slope of the demand curve). The linear demand curve used by this function and illustrated in the graph it produces is very simple. Most applications will require a more sophisticated demand function. However, it shows how demand can be modeled with a function within MATLAB and then illustrated with a graph.

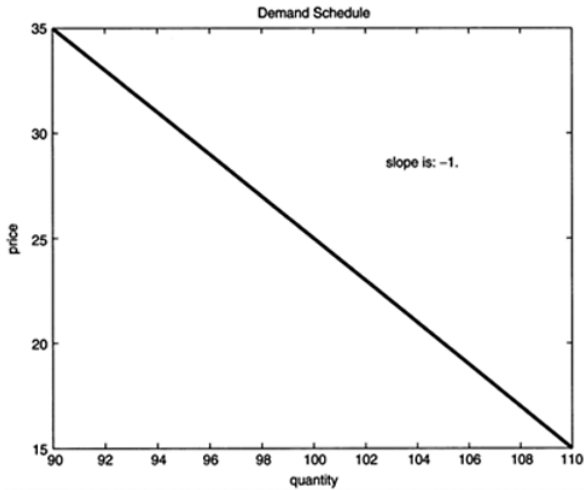


FIGURE 2.1 Linear demand function

Bias in parameter estimates means that the method systematically produces an estimate that differs from the actual value. An unbiased estimate will still usually differ from the underlying value – it is, after all, an estimate – but has the desirable property of having an expected value equal to the actual value. A related concept is consistency, which applies to estimators that, as the sample size increases, are expected to converge to the actual underlying value. Estimators that are both biased and inconsistent, therefore, must be expected to provide the wrong answer no matter how much data is available.

Code

```

if nargin <3
    disp('Using default values ...')
    price_i = 25
    quantity_i = 100
    slope = -1
end
% Check Data
if slope > 0

```

```

        warning('Price Elasticity of Demand is Positive. This
parameter is normally negative. ');
    end if price_i < 0
        warning('Price is negative. This parameter is normally
positive. ');
    end
    % Call Demand Drawing Subfunction [p, q] =
drawdemand(price_i, quantity_i, slope);
    %-----
    function [p, q] = drawdemand(price_i, quantity_i, slope);
    % create index; don't allow price below 1; rescale % if p is
entered as a number below 2
    if price_i < 2.5
        disp('Rescaling by multiplying price by 100.')
        price_i = price_i*100
    end
    inc = min(price_i-1, 10); p = [(price_i-inc):(price_i + inc)];
    % calculate line from y = mx + b linear form (p = % mq + b,
linear inverse demand; implies q = (p-b) % /m demand.)
    b = price_
    i slope*quantity_i;
    %y-intercept (price axis intercept) q_unlim = (p-b)/slope; q =
max(0, q_unlim);
    % limit quantities to nonnegative numbers
    %----- plot-----
    figure; plot(q, p); title('Demand Schedule');
    elasticity_message = {'elasticity is:' [num2str(slope), 'at starting
point.']}]; text(.7, .7, elasticity_message, 'units', 'normal');
xlabel('quantity'); ylabel('price');

```

Code Fragment 1 “Simple Linear Demand Function”

Supply Function

We can also model a supply schedule. A MATLAB function for this is printed in Code Fragment 2, “Supply Function”. The function takes as arguments a price for a supplied quantity, an elasticity of supply, and a minimum price at which producers will supply goods or services.

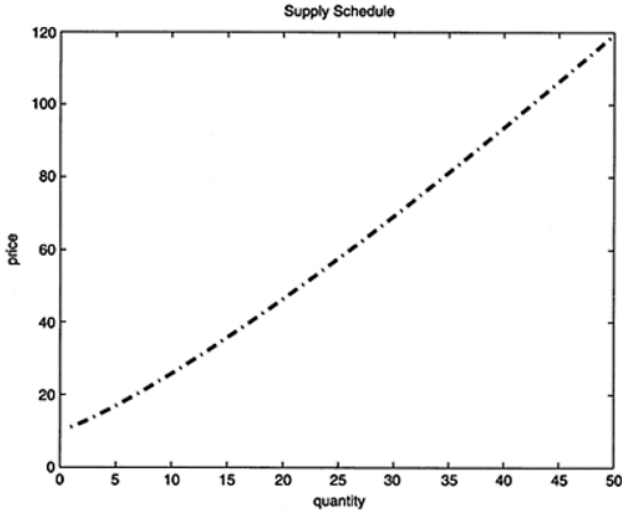


FIGURE 2.2 Supply schedule

The function produces the quantity supplied for the price (which may be zero if it is too low) and a structure that contains the input parameters, plus the complete supply schedule. The function also graphs the schedule, as illustrated in Figure 2.2, “Supply schedule.” Finally, it includes a graphical comparison of the schedule, the elasticity parameter, and the actual arc elasticities at each interval. See Figure 2.3, “Elasticity of supply schedule.”

Desirable Qualities of the Supply Function This particular supply function is interesting and can be applied directly to many real world problems. In particular:

- It has a “nearly constant” price elasticity of supply, as described in the next section, which is defined by a single parameter.
- It explicitly incorporates a minimum price (below which producers will not make new product) and makes the supply very sensitive to this initial price.
 - It allows the user to vary the price elasticity of supply.
 - The function itself is written in a modular fashion. The actual program itself is very short. The program relies on subfunctions that handle the work and graphics.
 - The function returns a data structure which records useful information about assumptions and data produced from the function.

This makes it easy to record information from a session for reviewing later.

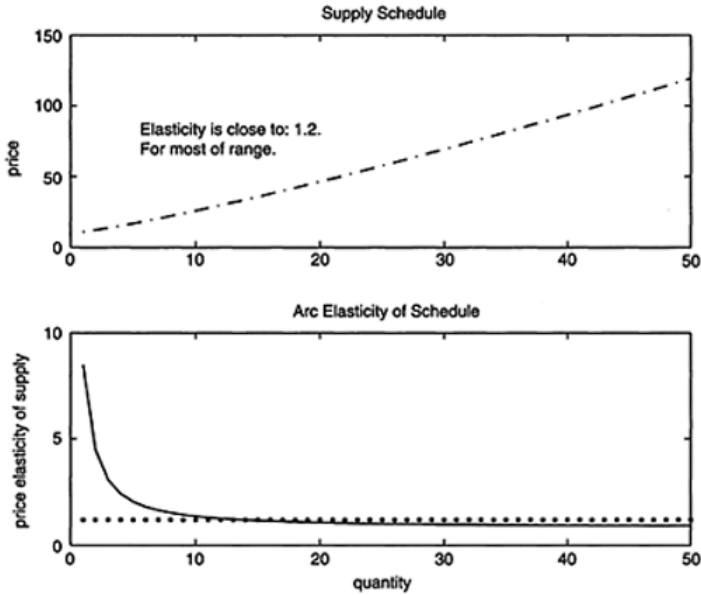


FIGURE 2.3 Elasticity of supply schedule

Elasticity of Supply

We derive the price elasticity of supply for this function below. A similar derivation can be used for demand and supply functions with similar forms. The basic supply function expresses the quantity demanded as a function of price. We include an explicit minimum-price argument to the function and transform it into an inverse function in which price is determined from quantity, an elasticity parameter, and the minimum price parameter:

$$\begin{aligned}
 p &= q^n + p^{min} \text{ from} \\
 q^k &= p - p^{min}
 \end{aligned}
 \tag{2.12}$$

Differentiating with respect to quantity:

$$\frac{dp}{dq} = kq^{k-1} \quad (2.13)$$

Multiplying by the q/p ratio to calculate the price elasticity of supply:

$$\frac{dp}{dq} \cdot \frac{q}{p} = k \left(q^{k-1} \cdot \frac{q}{p} \right) = k \cdot \left(\frac{q^k}{p} \right) \quad (2.14)$$

$$\frac{dp}{dq} \cdot \frac{q}{p} = k \cdot \left(\frac{p - p^{\min}}{p} \right) \quad (2.15)$$

Recalling from the equation above the definition of qK and substituting it.

Thus, the elasticity simplifies to the parameter k multiplied by a factor that, as prices grow, converges to one. For a good part of the usable range of the function the factor is slightly less than 1, so the elasticity of the schedule is slightly less than the elasticity parameter used in the function. This is illustrated in the lower panel of Figure 2.3, “Elasticity of supply schedule.”

Putting it Together

Once you can model supply and demand, you can attack many of the problems in typical economic thinking in a rigorous manner. For example, you can put them together as in Figure 2.4, “Demand and supply graph.”

This is sensitive to the range of plotting and the minimum price. If the minimum price is large relative to the range used for plotting, the elasticities in that range may be significantly lower than the elasticity parameter.

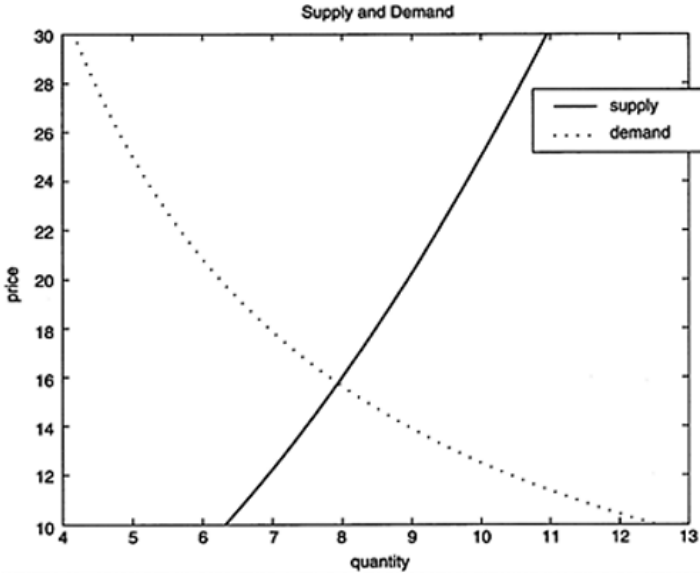


FIGURE 2.4 Demand and supply graph

The vast graphics capabilities of the MATLAB environment make it especially attractive for the business economist. MATLAB'S Handle Graphics allows the experienced user to control almost every aspect of a graphic from the figure window to the axes, lines, annotations, and the tick marks. However, that power has its price. The price in the case of MATLAB graphics is complexity.

CHAPTER 3. ECONOMIC IMPACT MODELS

Economic and Fiscal Impact Models

One of the most common questions faced by state and local governments and their taxpayers is evaluating the economic impact of a project. Many decisions about tax incentives, public financing of roads and infrastructure, changes in laws, and expensive public works and transportation expenditures are made partially on the basis of the economic impact of a proposed project. Fiscal impact models are also used widely in state and federal governments. Most tax policies are discussed largely in terms of fiscal impact. Fiscal impact includes the direct tax revenue and expenditure effects, while economic impact includes wages, expenditure, employment, and fiscal effects. Simulation models are a powerful way to address these questions, and we provide detailed instructions on their use later in this chapter. However, because there has been so much poor practice – and some outright charlatantry – in this field, we first discuss the ethics of economic impact analysis and its fundamental mathematics, before turning to the models.

Ethics in Economic Impact

There is no other area of applied economics in which more sloppy work, exaggerated results, and just plain balderdash have been produced than in “economic impact” analysis. Because the claimed economic impact of a proposed development can affect political support for a proposed project – and sometimes taxpayer funding – an incentive often exists to exaggerate the benefits. Sadly, that incentive has proven too much to resist for many practitioners in this field.

Sports Exaggerations

The debates over public funding for new sports stadia is the best documented area in which impact models have been systematically misused to exaggerate benefits. The clear beneficiary in many episodes has been those who own or operate a stadium facility built with public funds. The author of this book has seen multiple economic impact analyses for minor and major leagues’ sports stadia and major events that grossly exaggerate the benefits and understate

the costs.

One of the most pernicious trends in the cottage industry of economic impact analyses in recent years – and one that allows for weak analysis to gain currency – is the syndrome of the missing report. Our efforts to track down the reports that were cited in news stories often revealed that the only documents actually published were press releases. A prime example of this is the economic impact of the Atlanta Super Bowl in 2000. While numerous press releases and news stories cited the \$290-million-plus figure, repeated attempts in 2002 and 2003 to actually obtain a copy of the report proved fruitless. Letter from Patrick L. Anderson to NFL Commissioner Tagluabue, August 26, 2003; correspondence with Bruce Seamans, Georgia State University, August 2003. A copy of the June 2000 report was later provided by the NFL in September 2003 – 31/2 years after the event. Other examples include an analysis for a midsize city of a minor league baseball stadium completed by one of the largest accounting firms in the country; and the claims made for the economic impact of the Pontiac Silverdome, first when it was originally constructed in the 1970s and then when it was being abandoned by its only major tenant, the Detroit Lions, after the season ending in 2002. In 2003, after the Lions moved to Detroit's Ford Field, the city accepted proposals to do just that

This missing report syndrome is characterized by widely published claims, but no published study. In some of these cases, studies actually did exist, but were not made available for review when the publicity began for the result. In some cases, an understandable and commendable attempt at civic cheerleading results in an enormously attractive claim that, upon inspection, proves to be without firm basis. For example, in August 2002, a \$372 million economic impact was claimed for a Super Bowl in Detroit that would not occur until 2006 and was planned for a stadium that had not yet been completed. The claim was repeated often in the press; see, e.g., R.J.King, "Super Bowl Hotel Rooms Go Fast," *Detroit News*, August 21, 2002, p. 1. As of August, 2003, no report documenting the \$372 million economic impact was available. The figure may have been based on

The host committee, local journalists, and fans want to believe an event will be successful and appear to relax their normal impulse to

check whether the claim has a firm grounding.

A Sports Example: The Super Bowl the Super Bowl is one of the premier sporting events in the U.S. each year. Indeed, if the local economic effects of a Super Bowl were near the \$300 million or more claimed in recent years, “super” would be an understatement. Unfortunately, the claimed economic impacts of at least two recent Super Bowls suffer from the missing report syndrome and have resulted in widely published claims that are nearly impossible to document. The two Super Bowls for which we confirmed there were no published reports available at the time of publicity, but widely published claims, are the Atlanta and Detroit events; see citations above. However, an extensive Internet search found thousands of press citations for claimed economic impacts of many recent Super Bowl events – but no published reports or even executive summaries of those reports. In particular, the Houston and San Diego events had widely published claims but no evidence that any actual report was made available at the times the claims were made publicly. In some of these cities, a certain amount of attention was paid by the press to the validity of the claims. For example, the Atlanta Business Chronicle published an exchange of letters in February 2000 between Bruce Seamans, a consultant to the NFL-Atlanta Sports Council project that produced the \$290 million figure, and Robert Baade and Victor Matheson (cited below) who challenged the NFL estimate. The main topic of the exchange was revealing: whether or not the net effect was zero. Correspondence and manuscript from Bruce Seamans.

A skeptical analysis entitled “Super Bowl or Super (Hyper) Bole,” by economists Robert Baade and Victor Matheson, looked systematically at Super Bowls. They concluded that “the economic impact of the Super Bowl is likely on average to be onetenth or less the magnitude of the most recent NFL estimate.” They note that other research based on regression analysis found an impact not significantly different from zero.⁹⁰ A review article by Craig Depkin and Dennis Wilson stated that “most economists consider the forecasted impacts offered by advocates inflated,” and identified a number of sources of error – practitioner bias, measurement error, and leakage.⁹¹ We will consider the actual impact of a large sporting event further in “An Example: The Fantasy Game” on page 134.

extrapolation from previous reported economic impact figures. Letter to NFL Commissioner (cited above); telephone discussions with Detroit Super Bowl Committee and Detroit Convention and Visitors Bureau, August 2003. A later press announcement – timed to appear on the February 1, 2004, date of the Houston Super Bowl – again stated that a new study would be released showing a \$300 million-plus benefit.

Non-Sports Exaggerations

Unfortunately, misuse of economic impact analysis is not confined to sports stadia. Major investments in a state, especially ones that are being considered for tax abatements or direct state investment, routinely have economic impact analyses performed. In many cases, these are performed quite competently under the guidelines of state law or other policy.

Economic Impact of Y2K Sometimes the economic impact analyses are not performed competently. A spectacular example – hopefully not repeated for another millennium – is the incredible claims about the economic impact of the Y2K bug and the expenditures by government and businesses to control it. We issued a short series of reports in December 1999 indicating the implausibility of such exaggerated claims. During that time, the press and government officials routinely published cost estimates of \$300 billion to over \$1 trillion with the most common estimate being \$424 billion attributed to the Gartner Group. While there was some uncertainty regarding the actual effects, and government overreaction did trigger unnecessary expenditures in both the public and private sector, there was little real analysis underlying the huge claims.⁹⁶ Furthermore, there did not appear to be much journalistic fact-checking going on as the oft-quoted source for one of the estimates turned out to be ephemeral. Our own analysis (which, though not extensive, was at least in writing and based on some actual data) indicated the actual cost to the U.S. economy was about \$100 billion, and much of that was caused by overreaction. Finally, indicating in a comic fashion the underlying exaggeration, the U.S. government set up a toll-free Y2K hotline to deal with emergencies – and scheduled people to answer it only during specified business hours – which stopped just short of midnight on New Year's Eve.

Economic Impact of Broadband Another analysis was prepared recently for a major industrial state considering the imposition of a tax on telephone lines to support the creation of a development authority. This proposed authority was to help finance the development of broadband Internet technology, which would spur economic growth. The analysis prepared by the Gartner Group contained an estimate of positive economic impact of “an ubiquitous broadband network.” They estimated the positive impact of the authority – which would receive about \$420 million over 12 years – would be a Brobdingnagian \$440 billion over a 10-year period. This enormous increase in gross state product was specified in some detail, including 770,000 new jobs broken down into categories such as marketing, finance, and information. These new jobs were supposed to rise in a state with the employment of 4.8 million, meaning a 16% increase in statewide employment and \$440 billion in GSP from approximately \$31 million in additional annual financing. That is a return of over 1000 to 1.102 how could such a gargantuan economic impact be projected? The methodology section of the available report indicates that teledensity (phone lines per capita) in thirdworld countries was compared with GDP, and the same calculations were performed for the U.S. and other industrialized countries. The data show that industrialized countries had more phones per capita. Therefore, putting more broadband connections into Michigan was projected to have the same proportionate impact as industrializing and wiring third-world countries. Perhaps the authors should have left out any description of their methodology and simply included this statement: “The self-organizing techno-economic paradigm of the Internet cannot be understood by conventional arguments of economic growth.”

An Example: The Fantasy Game

How could the net economic impact of a single sporting event exceed \$300 million? As an illustration of the vagaries of economic impact analysis, we will analyze a truly fantastic event which we call “the fantasy game.” We are interested in the local economic impact, by which we mean the additional income to residents of the local area. Because this is a fantasy, we will throw out all caution in making assumptions about the economics of the event:

- Assume, for a moment, that for the fantasy game, every seat of a 65,000 seat stadium was filled.
- Assume that every single attendee was from out of town and that they all traveled there for just this one event.
- Assume that in addition to the attendees, an extra 1,000 people traveled there to work during the event.
- Assume further that all the fans and all the workers stayed two nights in a local hotel, with each visitor staying in his or her own room, paying \$150 per night.
- Assume that every single visitor spent \$200 per day on food, tickets, T-shirts, and other local expenditures.
- Since there are some costs – even in our fantasy game scenario – of food, drink, T-shirts, and hotel rooms, assume that such costs are just 50% of what the visitors spent, and that all of the remaining profit goes straight to local owners and employees.
- Now, just for the fun of it, assume further that each of the 65,000 fans and 1,000 workers drop an extra \$50 in cash on the sidewalk each day they visit. Despite the hordes of cash-rich/pocket-weak visitors wandering the streets, all of the \$3.3 million per day in lost currency is picked up by local residents.

Clearly, this is a fantasy. Whatever heroics might occur on the field would pale in comparison with the supernaturally heroic assumptions made about the fans. Such a game would never be played on earth. However, what would the local economic impact of such a fantasy be? Table 3.1, “Economic Impact of the Fantasy Game,” provides an optimistic calculation.

Sum of Benefits from the Fantasy Game

As the calculations in the table indicate, our fantasy game has a positive economic impact of \$13 million for local expenditures other than hotel rooms, another \$9.9 million for the hotel expenditures, and \$6.6 million for the cash left on the street. The local economic impact totals \$29.7 million. Such an estimate was produced with wildly optimistic assumptions. Yet, the total is still much smaller than the \$300 million estimates publicized for recent Super Bowls. It is interesting that the amount is about one-tenth of that amount – exactly the ratio of reality-to-claim found by Baade and Matheson. Indeed, even if we used a multiplier to double this amount, the

estimate would still be less than one sixth of the \$374 million claimed for the Detroit Super Bowl. If we were realistic in our assumptions – noting that some attendees are local residents, that some expenditures would have occurred anyhow, and that most visitors do not drop \$50 bills on the sidewalk – the number would go down considerably. There is simply no squaring the circle; you just can’t get to \$300+ million from 65,000 attendees at a one-day event.

TABLE 3.1 Economic Impact of the Fantasy Game

	Visitors and Days	Local Food and Other Expenditures		Hotel		Money Left on Sidewalk		Grand Total
		Per Day	Total	Per Day	Total	Per Day	Total	
Workers	1,000							
Plus: Fans	65,000							
Equals: Visitors	66,000							
Times: days	2							
Equals: Visitor-days	132,000	\$200	\$26,400,000	\$150	\$19,800,000	\$50	\$6,600,000	
Less: costs to local economy of expenditures (50%)			\$(13,200,000)		(\$9,900,000)			
			\$13,200,000		\$9,900,000		\$6,600,000	
Grand Total								\$29,700,000

There’s obviously something going on in these economic impact analyses that are not based on economics. Part of this is an ethical shortcoming among practitioners, as well as promoters eager to show “big” numbers for their projects. The other part is a lack of knowledge among practitioners. These deficiencies motivate the next sections.

A Plea for Ethics in Impact Analysis

There are reasons beside ethical lapses for the prevalence of poor impact estimates. These include:

- Many of these “economic impact” analyses are not produced by economists, but rather by accounting, “futurists,” marketing, or other

firms.

- There are now a number of commercially available models that can be used as a “black box,” in which a few numbers are placed, and out of which a much larger set of numbers comes.

- The journalists who write about these events have proven gullible or have agreed to become secret cheerleaders themselves.

However, we are left with the uncomfortable knowledge that some of these analyses are produced by people who are knowledgeable enough to know that the results are exaggerated but who simply succumb to the temptation to produce the big number. It is my hope that, after reading this book, you will know enough to avoid the mistakes that have produced erroneous estimates in the past. It is my plea that, regardless of your knowledge, you will not succumb to the temptation to exaggerate your results for an economic impact or any other report. To do so is worse than being unprofessional; it’s just plain wrong.

The Economics of Impact Models

The fundamental mathematics and economic reasoning behind impact models are quite rigorous. The conceptual reasoning behind what is known as input-output models was developed by Wassily Leontief in the early part of the 20th century.¹⁰⁶ Leontief’s innovation, along with those of the pioneers of econometrics and the creators of the national income and products accounts, paved the way for economics to become an empirical science. For his contribution, Leontief was awarded the 1973 Nobel Prize in Economics. Leontief could build on a few of the pioneers of economics, particularly Francois Quesnay whose *Tableau Économique* (1759) predated Adam Smith and became a founding document for the Physiocrats.¹⁰⁷ Quesnay graphed the interactions among sectors in “zig zag” tables, showing how much of the product of one sector went to another.

The history of input-output analysis is an interesting one. The U.S. Army used such an analysis during World War II to select bombing targets in Germany, apparently choosing ball-bearings as the most critical industrial component. The production of the input-output tables by the U.S. government and the further production of multipliers based on those tables have dramatically improved the

ability to assess the economic effects of changes in demand in any one area. Not surprisingly, there arose a cottage industry of economists and others who use these tables or products derived from them to project the economic impact of projects and events.

Widespread Lack of Knowledge The improvement in the availability of input-output tables has not, unfortunately, been matched by a growth in the number of people who know how to use them. In recent years, the study of input-output models in academia has often been relegated to a mathematical topic. This has left many economists bereft of knowledge about the underlying theory and assumptions. Meanwhile, practitioners of economic impact analysis, who may have never received any economic training, are typically not required to learn anything about the fundamental assumptions on which it is based. The RIMS II User Guide, which is the explanatory document prepared to accompany the U.S. government's multipliers, relegates the discussion of input-output analysis to an appendix. While the User Guide itself contains numerous straightforward cautions on the proper use, these are apparently quite easily ignored.

This lack of familiarity with the theory, combined with the availability of commercial software and multipliers for every region in the U.S., has made it easy to produce bogus economic impact conclusions using apparently scientific methods, including multipliers provided by a U.S. government agency. The combination of this knowledge gap and the ethical weaknesses of human beings have resulted in the poor state of the economic impact industry. We plead for ethics in economic impact analysis in the section above. In the following section, we address the knowledge gap.

Input-Output Equations

The underlying concept is that the entire economy can be viewed as a huge production function into which we pour inputs and from which comes output. Economists segment the types of inputs by type, including labor from different areas, industrial goods, agricultural goods, and other sources. They then segment the outputs by type as well, including consumption by various sectors, profits, and other outputs. The fundamental input-output relations within the economy can be described by a system of n simultaneous equations in n variables. The number of goods or services described determines

the number of equations. The household sector (which supplies labor) can be included in these equations as an exogenous variable, in which case the input-output system is “open.” If we describe all goods and services, including those provided by the household sector, as intermediate goods that are determined endogenously, then the system is “closed.” There are many references on the mathematics of input-output models.

The Technology Matrix Those familiar with matrix algebra will anticipate that these equations can be represented by simple operations among matrices. Indeed, the fundamental equation is an input matrix multiplied by an input vector to produce a vector of outputs. The relationships among industries are captured by the coefficients in the technology matrix. The technology matrix consists of coefficients a_{ij} which describe the amount of input i that is used by industry j . (The subscripts can be recalled easily by noting that the first subscript denotes the input industry and the second denotes output.) In the following three-industry example, the technology matrix would be:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{3.1}$$

The Fundamental Equation This matrix is used in a fundamental equation relating the vector of outputs x , the input matrix of coefficients A , and the final demand vector d :

$$(I-A)x=d \tag{3.2}$$

where: $(I-A)$ – is called the Leontief matrix. Intuitively, this equation means that the total outputs of all industries x , minus the amount consumed in the process of production Ax , equals the amount demanded and consumed by the rest of the economy d .

A Robinson Crusoe Example This equation can be illustrated by the following thought experiment. Assume you are the only worker in a “Robinson Crusoe” economy located on an island of a single human inhabitant and the only job you have is to pick coconuts on

the beach. You, however, as the most diligent beachcomber on your island, tend to eat some of the coconuts. The rest you save for other projects, including late-night eating, coconut bocce ball, coconut painting, and building coconut outrigger canoes. The consumption during the production process is commonly called intermediate demand, while the consumption by the households that ultimately consume goods and services is known as final demand. On this island, however, you fulfill both roles. If you eat one out of four coconuts while you pick them, then the only coefficient in the input matrix A is $a_{11}=1/4$. If the input matrix x is the single entry $x=10$, representing 10 coconuts, then the fundamental equation for your economy are:

$$(1-[1/4]).[10 \text{ coconuts}]=7.5 \text{ coconuts} \quad (3.3)$$

Thus, the final demand for coconuts on this island is seven-and-a-half coconuts, which becomes the payment to the household sector as well as its final consumption. The input matrix explains how the total output of coconuts in the economy, 10, was allocated.

Extending the Example

We can now expand this to economies where there is one open sector that contributes a factor input and receives the resulting production. In most economies, the household sector provides labor and purchases the final product; therefore, most economies are modeled with an open system. In such systems, the sum of the factor inputs for each industry (the column sum in the input matrix A) must total less than one, or there is no incentive to produce it. The difference between the column sum (the sum of all factor inputs other than household) and 1 is the amount given to human labor, including workers and management. Thus, $(I-A)x=d$ makes sense for small and large economies.

Restrictive Assumptions

There are important, restrictive assumptions that underlie input-output analysis. The most important are:

1. Fixed input-output ratios: This is a much more restrictive assumption than is normally used with production functions. It

means that the producers cannot vary the relationship among inputs despite growing or shrinking demands, or fluctuations in prices. If we remember that the matrix A summarizes the technology of production, this assumption boils down to one of fixed technology. This assumption is clearly not fulfilled in most situations. However, it is a reasonable approximation when the overall changes in demand or supply are modest and occur within a short time frame.

2. Single-industry production of single goods: The model assumes that each industry produces a single product and that no other industry produces such goods. This, again, is a condition rarely fulfilled. However, the assumption can be somewhat relaxed by redefining the goods in certain industries as composites of multiple goods and services. This assumption becomes more visible when we consider substitution across goods or industries. As the labor force can readily substitute across industries and as consumers (and intermediate producers) can often substitute as well, we see the need to carefully consider the effect of this assumption. In practical terms, we can often conjecture that this assumption is fulfilled when the required changes being considered occur during a short period of time and when no other major substitutions are likely.

Mathematics of Multiplier Analysis

An input-output model can be used to predict the change in inputs resulting from a change in final demand and vice versa. By deriving multipliers from the input-output equation, we can directly calculate the change in one sector caused by a change in another. Any results derived from a model will only be valid when the assumptions underlying the model are valid. For the following derivation, we will assume that this is the case. Recall the fundamental equation $(I-A)x=d$. If we solve the equation for the vector of outputs from each industry x , we find by inverting the $(I-A)$ matrix. This is sometimes called the Leontief inversion. Mathematically trained readers will note that not all matrices have inverses, and we have not formally required that all the entries in the final demand vector Y will be positive. Simon and where the $(I-A)^{-1}$ matrix is sometimes called the requirements matrix as it contains the required:

$$x=(I-A)^{-1}d \tag{3.4}$$

quantities of final demand for all sectors, based on new output for each sector. The requirements matrix can be seen as containing the partial differentials of output of each sector x with respect to the final demand d : where (3.5)

$$(I-A)^{-1}=B=[b_{ij}] \tag{3.5}$$

$$b_{ij} = \frac{\partial x_i}{\partial d_j} \tag{3.6}$$

Consider the contents of the requirements matrix B for an economy with two industries plus the household sector:

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \tag{3.7}$$

In this economy, b_{11} and b_{22} are the coefficients indicating the same-industry (direct) multipliers for industry 1 and 2, respectively. The indirect effects are indicated by the sum of the other coefficients in the same column (other than the row for the household sector), meaning the effect of a change in the final demand of sector i on the output of each sector j . As this is an open economy model, the induced effect (caused by spending and respending of the household sector) is denoted by coefficients b_{3j} for sectors $j=1, 2, 3$. The last coefficient, b_{33} , would be change in output from the household sectors due to a change in demand of the same amount.

Total Output Multipliers

In this structure, the columns of matrix B summarize the impact of changing total demand for one industry's goods on the output of all industries. If we include the household sector, the induced effect caused by the respending of earnings is included in the multiplier. The total output multipliers that include induced effects are usually

greater than one because a new dollar spent in a community would be used to buy factor inputs (including wages paid to local workers), and a portion of those wages and purchase funds would be respent in the same community on additional purchases.

Uses of Input-Output Analysis

The discussion on economic input models in this chapter should not imply that this is the only use for input-output analysis. Other – often better – uses include planning for economic growth; examining the impact of strikes, work stoppages, and wars; and understanding the effects of tax policy and location changes.

Rules for the Proper Use of Multipliers Now that we have reviewed the fundamental theory and mathematics, we can present the proper use of multipliers in economic impact analysis. The key rules for doing so, with references to the theory underlying the rules and the consequences for violating them, are:

1. Only bona fide changes in final demand or factor outputs should be multiplied. Recall that the multiplier matrix B consists of partial differentials. A differential is, by definition, a change caused by a change. If there is no change, there is no differential. Therefore, there is nothing to multiply. Violating this rule is one of the most common reasons for exaggerated impact projections. If a project or event causes \$1 million in spending to shift from one entertainment venue to another, and entertainment is an industry in your economic impact model, then the change in final demand for the entertainment industry is zero. It does not matter whether the multiplier is high or low. The RIMS II User Guide contains multiple examples in which this rule is followed. It states the rule explicitly: “These multipliers measure the economic impact of a change in final demand, in earnings, or in employment on a region’s economy”. As the emphasized words indicate, a multiplier should only be applied to a change in final demand that occurs within a region’s economy.

2. Select and use the appropriate region consistently. The input-output tables for the entire U.S. economy are quite different from those for any one state or region. The U.S. economy as a whole produces just about everything. Most regions do not. The households in the U.S. can buy most goods and services from producers within the U.S. For states, especially regions or counties within the states,

this is often not true. If you are looking at the economic impact on a county, then all expenditures and income from outside the county should not be considered. This will considerably reduce the multiplier effect of expenditures inside the community, as relatively little respending will occur in the community. However, it will mean that it is more realistic to consider additional expenditures by outside entities as new expenditures. The choice of region is often determined by others. However, once it is chosen, it must be consistently used.

3. Subtract all substitution effects before using a multiplier. Many projects result in the substitution of one form of production or consumption for another. A new sports stadium will probably be associated with either the demolition, or a sharp reduction in the usage, of an existing stadium. A new shopping mall will certainly reduce the sales at nearby existing retailers. The opening of a new gaming casino will probably cannibalize existing casinos (if there are others nearby) and will certainly cannibalize other types of entertainment. All these projects can be analyzed properly, and in some cases the economic impact will be positive. We have performed economic impact analyses for all these types of projects and in each case accounted for substitution effects. In the case of sports stadiums, our analyses of Tiger Stadium (after the Tigers had left for Comerica Park) compared retail activity in a planned conversion to no activity at all, which was the alternative. For the Pontiac Silverdome, we found the alternative usage of the stadium site (rather than hosting a small number of professional football games and other stadium events) would actually result in more economic benefits to the community. For a retail development in Arizona, we accounted for displacement effects on existing retailers. Finally, for a proposed casino in western Michigan, we accounted for cannibalization of the casino and other expenditures.

In other cases, after adjusting for substitutions, there will be little or no net economic impact.

4. Use the bill of goods method when available. Often, information will be available on the likely allocation of expenditures by sector for a new project. For example, a publicly funded facility plan will normally include some accounting for the amount of money to be spent on land, construction, operations, and employment. This information should not be ignored in favor of lumping the whole

project into an industry category and hoping the averages on which the input-output tables are based are close. They are probably not. The RIMS II User Guide specifically recommends that practitioners use a bill of goods method for allocating expenditures to different industries, along with the appropriate multipliers (after deducting any substitution effects) for those industries. For example, consider the operation of a new facility. (We deal with construction in the next rule.) Assume that the new facility will be a manufacturing facility employing 1000 workers of which 250 will work in a distribution center, and that 500 of those workers are transferring from other facilities in the area. Assume further that almost all the facility's factor inputs will be shipped in from out of state and that little of the remaining production will remain in the state. This information allows you to separate out substitution effects (such as the transferring employees and operating expenditures) for about half the plant. It also allows you to use the multiplier information for a wholesale trade or similar distribution category for a portion of the plant's operations. Finally, you can also make adjustments for the shipping in and out of most of the plant's materials and output, including accounting for the significant transportation and logistics operations that it implies. This will be far more accurate than simply using a manufacturing multiplier for the plant's annual expenditures in the area. The Simulink models we present in this chapter and the next frequently rely upon separating the economic impact of multiple categories of expenditures, consistent with the bill of goods method.

5. Separate construction from operation. Most new facilities have a construction phase and then go into operation. The construction phase will normally be qualitatively and quantitatively different, in terms of its economic impact, from the operation phase. These should be handled separately for the construction phase, carefully consider how much of the materials will come from within the region. In some cases, the regional averages for construction will approximate the effects. In others, a bill of goods method will be far more accurate. For example, we once analyzed a half-billion dollar power plant construction project on the west coast of Michigan. As we knew that a large portion of the expenditure would go for giant turbines to be shipped from Europe, we deducted this from the expenditures in the local economy. We also adjusted for

transportation costs and for the share of the construction payroll that would remain in Michigan.

Construction phases may also include demolition and other activities that should not be ignored. For example, major projects often involve upgrades for transportation networks (such as roads, bridges, highway ramps, port facilities, etc.); when these are bona fide new investments, they should be considered. For operations, be sure to carefully consider employment and wage earnings. If workers are likely to be absorbed from the existing labor pool (especially if one effect of the new project will be a decline in employment at other businesses in the area), then the net economic impact of their wages will be small and may be close to zero.

6. Remember the assumption of constant technology. Recall that one of the limiting assumptions in input-output models is that of unchanging technology. The multipliers that derive from these models are based on this assumption. If your project involves a deviation from the average use of factor inputs for specific outputs, then you must adjust any multipliers or other uses of the input-output tables to account for this. This happens much more often than is expected. New facilities arise because something is new. That newness may just be new population, in which case the assumption is probably correct. However, a new type of plant, a new way of distributing goods and services, a cheaper production process, significant changes in prices, changes in law, changes in environmental regulations, and other changes that occur quite frequently would violate this assumption. Violating the assumption does not mean you cannot perform an economic impact analysis. It means you must use the standard multipliers only on the portion of the project that is standard.

Conclusion: Rules for the Use of Multipliers

The preceding rules are not terribly difficult to follow. If they are followed, economic impact analysis can be a rich tool, allowing customized, region-specific analyses of many events. In the following section, we will look at an example analysis that properly uses multipliers to determine the economic and fiscal impacts of a change in policy. In the succeeding chapter, we will look at more models that examine different phenomena.

Economic Impact Simulation Models

Below we describe an actual economic impact simulation model and explain briefly how it was used. This approach can be used as a guide for other economic impact analyses.

Economic and Fiscal Impacts: City of Detroit Model We developed a sophisticated economic model for the City of Detroit to estimate the effects of a change in a state law that would eliminate the city's long-standing residency requirement for city workers. This required a particularly complex analysis, as all of the following were direct results of the policy change:

- The number of residents in the city would change.
- The economic contribution of those residents would change.

For example, putting in enough retail service establishments to service a growing population probably does not require rethinking the “constant technology” assumption.

- The taxes paid (and city services used) by those residents would change. In addition to these direct results, we anticipated some indirect results as well.

Modeling Resident Behavior the fundamental microeconomic dynamics of the model matched that of the residents. Should the law change, those residents that were formerly constrained in their residence choices would become free to move to other locations. Some of these individuals would move outside the city. We anticipated that most residents who wanted to migrate outside the city would not do so immediately, given the large transaction costs associated with selling and buying a home, as well as the other disruptions that occur when a family moves. Therefore, we built into the simulation model a partial adjustment feature and projected the results out five years or more.

Scenarios: GUI for Model The model allowed for one change-of-residency to affect the migration of workers in and outside the city. This change, in turn, affected five different tax bases and the tax revenue. We created a custom graphical user interface (GUI) to allow us to input different scenarios. Each scenario assumed different speed of adjustment, number of migrants, and other factors. Thus, we were able to confidently project the economic and fiscal impact of the change in law after reviewing the likely impact under numerous scenarios.

Use of the Model and Output The resulting report was used in a number of arbitration proceedings on labor contracts, budget preparations for the city, as well as in testimony to the state's legislature.

Implementing the Model

We took the following approach:

1. Tax, economic, and geographic data were analyzed to provide key assumptions for the model. These assumptions included the baseline tax revenue, tax rates, tax bases, collection factors, population, income, and other variables for the model. The demographic and geographic data were analyzed in a geographic information systems (GIS) facility which allowed for current residence locations of city employees to be used as predictors of the likelihood of further migration.

2. The key behavioral assumption – the effect of the change in migration given a change in the residency law – was based on survey data, census data, and a variety of other sources.

3. A Simulink model was created. It had subsystems that calculated the key variables at each step in the cause-and-effect chain. The Simulink model allowed others, through visual inspection, to see these cause-and-effect relationships.

4. We designed the model with explicit control variables that governed the speed and magnitude of adjustment. Different control variables were used to generate multiple scenarios, allowing us to test the sensitivity of the model to changes in assumptions.

5. We created a GUI that allowed the user to directly input or change the key behavioral and economic assumptions.

6. A callback simulated the model under two scenarios: a baseline scenario (no change in law) and a change-in-law scenario. The callback then calculated the difference in the output variables over a multiple-year period.

7. Output variables were imported into Excel for presentation. The Simulink report generator was also used to describe the model. Certain figures were printed for use in the report.

8. In addition to a written report, we constructed a Web page that provided the executive summary of the report, along with selected exhibits.

Schematics of the Simulink Model

The following figures show schematic diagrams of the various subsystems within the fiscal and economic impact model, starting with the top-level view and going through the major subsystems.

1. Top-Level View: Figure 3.1, “Fiscal and economic impact model,” shows the top-level view. The key behavioral variables are shown on the left side. These include resident and nonresident population, overall population and economic growth, and a switch to turn the residency requirement off and on.

2. Population Migration: These variables flow into a population module which calculates the population (as affected by migration in and out of the city) under different scenarios. This module is illustrated in Figure 3.2, “Population migration subsystem”.

3. Economic Growth: The base variables also feed into an economic growth module which calculates income and property tax base growth factors. These are used in the remainder of the model.

4. Tax Base and Tax Revenue: From the outputs of these subsystems, the resulting tax bases for city taxes are calculated for multiple years in the tax base module. Using these tax bases, tax revenue is calculated in the tax revenue subsystem.

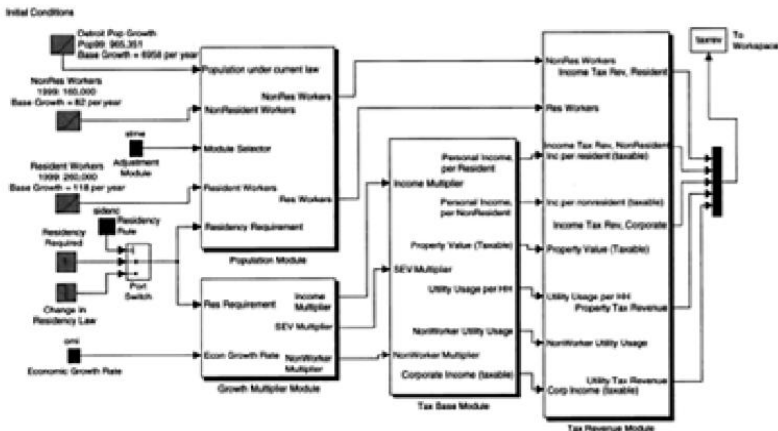


FIGURE 3.1 Fiscal and economic impact model

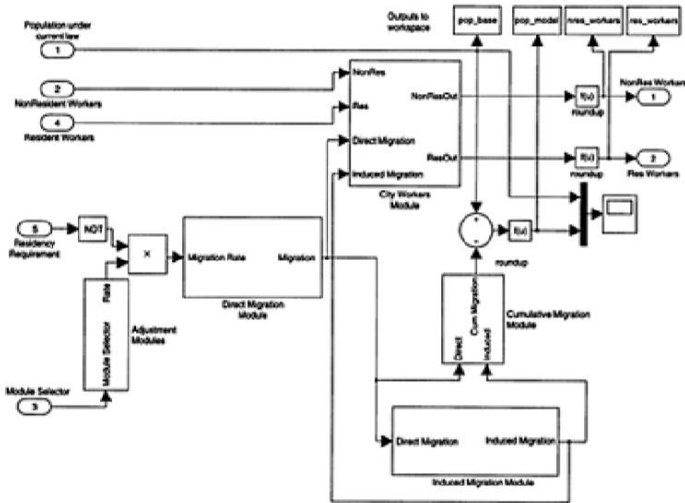


FIGURE 3.2 Population migration subsystem

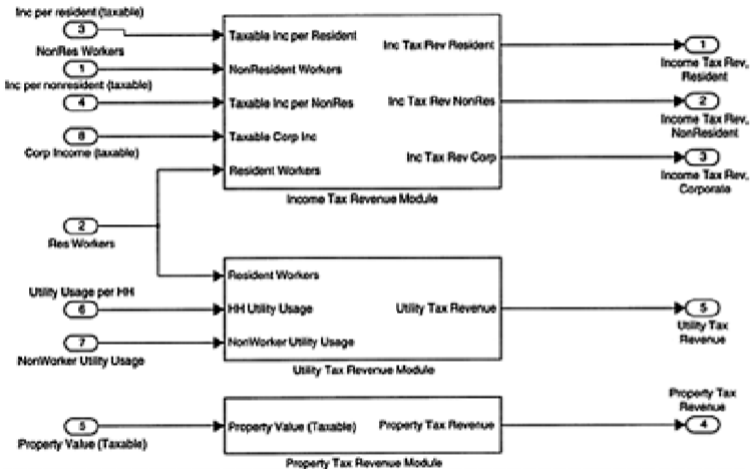


FIGURE 3.3 Tax revenue subsystem

Modeling Dynamic Behavior: Adjustment

One key behavioral question is the speed of adjustment. If we assume that 20% of the employees will move, that tells only part of

the story How many will move each year? If this is an important question in its own right, and data are available, the best approach is to directly estimate the adjustment path. However, this is often impossible. Therefore, we offer two practical approaches for cases in which the speed of adjustment is not the most critical aspect of the model.

Option I: Borrow Depreciation Algorithms One approach is to borrow functions that have been defined for depreciation. Depreciation in finance and taxation is the scheduled decline in the book value of an asset. Automobiles are probably the most well-known of depreciable assets. Numerous guidebooks and services provide estimates of their resale value. To project migration, we could use depreciation algorithms. For example, a simple method would be to assume that if 20% of the people wish to migrate over the next 5 years, one fifth would move each year. This is analogous to straight-line depreciation, as plotting the migration over time would produce a straight line. Most human behavior, however, does not follow straight lines. Thus, a more accurate adjustment path would often be closer to “double declining balance” or other methods in which earlier years have faster adjustment. Using MATLAB functions, numerous adjustment paths can be programmed, but in many cases these depreciation functions will do the trick.

Implementing this Model

To allow us to vary the adjustment method, we inserted a switch in the subsystem. This switch was controlled by a drop-down box on the GUI. Thus, scenarios differing only in the speed of adjustment could be run sequentially, revealing the sensitivity of the model to changes in this assumption.

Option II: Specify Partial Adjustment Model Another option is to specify a partial adjustment model that governs the path of adjustment toward a desired level of a variable. Such models can have strong theoretical grounding based on the costs of adjustment. For example, investment can be considered movement toward a desired level of capital stock. Such models are sometimes called stock adjustment models. Specifications for partial adjustment models include those based on polynomials, ratios, and geometric series. Such models can be readily implemented in MATLAB.

CHAPTER 4. FISCAL IMPACT MODELS

Fiscal Impact Methodology

We describe in this chapter fiscal impact analysis, which we define as the impact on revenue and expenditures of an organization from some outside event. In this chapter we assume the organization in question is a government, but the term fiscal could also apply to a private company or even a family

Definition of Fiscal Impact

As discussed in the chapter on economic impact models, we insist on a specific, conservative, and realistic definition of impact. We define fiscal impact by including only bona fide new tax revenue or reduced government expenditures. Therefore, to arrive at the net fiscal impact of a proposed event, we must subtract any lost tax revenue or increased expenditures from the increased tax revenue that would be realized from the event.

Relationship between Economic and Fiscal Impacts

Tax revenue is normally generated by economic activity. Income, sales, and value added taxes are directly related to economic activity, whereas property taxes and other taxes are indirectly related. Certain excise taxes based on usage (notably gasoline taxes in most states) are also directly related to economic activity. Therefore, most fiscal impact analyses are based on assumptions about economic activity. A common error in fiscal impact analysis is to assume the level of economic activity without considering how the change in policy affects the activity. The most obvious illustration of this is to estimate the effect of a change in tax policy on tax revenue. Elementary economic theory reveals that tax rates affect the after-tax earnings of an activity or the after-tax cost of a service or good. Therefore, tax policy must affect economic activity, taxable earnings, and taxable sales of goods and services. While economic theory indicates that tax rates affect economic activity, it does not tell us how much activity is affected. That is the realm of applied economics. In some analyses, the level of economic activity involved will be a straightforward assumption. In such cases, make this assumption explicit and support it. Remember, theory states that tax

policy will affect economic activity. If you are assuming that it does not, you must explain why you are assuming that human beings do not behave as they normally do. One approach which, though honest, is incomplete, is to acknowledge that you have ignored the impact on behavior caused by a change in tax policy. In such cases, state that your estimated tax revenue is likely to be higher (or lower) than the estimate you have made ignoring human nature. The classic hidden pairing is to support (or oppose) tax increases when you are really advocating increased (or decreased) government expenditures.

If you are constrained by institutional policy to performing static analysis, then state this prominently and note how actual human behavior is not static. For further discussion, see Chapter 8.

On the Political Economy of Taxes

The question of tax policy is connected with expenditure policy through the government's budget constraint. The two are sometimes discussed separately and sometimes artfully combined. Make sure you are explicit in limiting the explanation of your analyses to the topic you actually studied. For example, if you analyzed property taxes and their effect on business location decisions or economic development, you should ensure your report describes the effect of such taxes on business and economic development. Even if the tax is used to pay for public schools, do not use the analysis of tax policy to form a pretense for a discussion of education policy unless that policy is directly connected to tax policy. Some discussion of the uses of the tax, however, is useful in explaining the entire issue. When doing so, be sure you do not fall into the related trap of presuming that more tax revenue will invariably result in better services. In particular, if you analyze taxes, do not confuse the analysis of taxation with one about how the revenue is spent.

Fiscal Impact of Industrial Development

Methodology Summary Fiscal impact methodology includes categorizing and modeling both the costs and benefits to a local government and its taxpayers due to the subject development. Consider the analysis of a proposed industrial development. The analysis should start with describing the scale of the project in terms of employment, tax base, and other factors for each year over the

next decade or two. These variables are then used as a basis for calculating both the benefits of development (including those associated with new jobs, new employees, additional taxes, revenue sharing, payments from the state, and permit fees) and costs to the community (including additional police and fire burdens, capital expenditures, general financial administration, and other burdens). The analyst should also identify key assumptions, describe the methodology, and identify any important factors that cannot or were not quantified.

Description of Each Part of the Model

We illustrate a fiscal impact analysis performed on a proposed industrial development in a rural township in southeastern Michigan. We describe each part of a Simulink fiscal impact model using a schematic for each system and subsystem in the model. Along with the data and some model-specific parameters, the schematic describes the mathematics of the model. The following are brief descriptions of each subsystem in the model.

Overview: Fiscal Impact Model The overview of the fiscal impact model shows the big picture of the model. You will note the Project subsystem on the left (Figure 3.3, “Fiscal model schematic – I,”). Within this system we project new residences, employment, and property tax base variables. The output from this system – these variables – flows into two other subsystems: cost and benefit. This overview, the most important view of the model, illustrates how the same projections about the project facility drive analyze the related costs and benefits. The following paragraphs describe the portions of the model that are within each subsystem. In each subsystem, variables have the initial values described in the data tables.

Project Subsystem

This subsystem is composed of residential growth, employment, and property tax base subsystems.

Fiscal Residential Growth Subsystem: This uses a number of relocated workers, spin-off residences as a share of relocated workers, spin-off residences as a share of new workers, portion of new workers resident in the township, and portion of relocated workers resident in the township to calculate the number of new

residences from employment, spin-off residences, and total new residences. Employment Subsystem: This separates construction and operations workers. The look-up tables for construction, transfer, and new workers specify the schedule of employment during the time period for each category. Property Tax Base, Development Subsystem: This summarizes the property tax base, including change in personal property value, initial personal property value, annual growth rate of personal property, change in real property value, initial real property value, and annual growth rate of real property. A compound growth module is used to calculate the values of these variables as they grow over time.

Cost Module the cost module includes calculations for the costs of new residences, which include fire, police, and government administrative costs per new residence, plus other direct costs per new residence; and employment cost which includes fire, police, and government administrative costs per job, plus other direct costs per job.

Benefit Module the benefit module calculates the benefits to the community. These benefits include state revenue sharing, permit fees, and property taxes. A portion of these benefits is estimated within the property tax revenue and permit fee subsystems. Other benefits are estimated from variables for average residents per new household, state revenue sharing and other revenue per capita, and portion of average state revenue per capita to the township.

Property Tax Revenue Subsystem: This uses variables of average house value, property growth rate, assessment ratio, and property tax rate to generate property tax revenue. A zero property tax rate for the township means that property tax revenue will be zero. Permit Fee Subsystem: This summarizes permit fee revenue flow to the local government.

Notes on Model Structure

The following are additional notes on the model structure:

1. This is a fiscal impact analysis. It does not define or estimate the larger economic impact, economic substitution effects, or other variables, including concerns about amenities, character, quality of life, access to jobs, tax base, and other issues.

2. There are generally four sources of revenue generated by this

kind of development: permit fees (usually a one-time fee), property tax revenue (which normally grows and arises from both direct investment and any indirect or induced development), revenue sharing (provided on a per capita basis by the state to the township), and fee-based revenue for specific services. We have modeled each effect.

3. Other costs are largely offset by fees paid for services, such as utility costs, as explained in the report. For this category of services, we have excluded both the fees and associated costs.

4. Fiscal costs are allocated separately for residential and employer taxpayers. These include fire, police, and administrative costs. U.S. Census data were used to estimate the per-capita costs of this category of municipal costs for small and medium-sized municipalities in the state of Michigan. These total per-capita costs were further divided into per-resident and per-employee costs.

5. We also projected separately the increase in employment and residents created by the new plant.

Source Data

The spreadsheets summarizing input data we used in our model and data sources were included in the data tables in our report.

The following exhibits also summarize the methodology and results of the analysis. 1. The fiscal impact projection graphs are in Figure 4.1, “Municipal costs,” and Figure 4.2, “Cost and benefit.” 2. The fiscal impact model is illustrated by schematic diagrams in Figure 4.3, “Fiscal model schematic – I,” Figure 4.4, “Fiscal model schematic – II,” and Figure 4.5, “Fiscal model schematic – III.” These diagrams illustrate, in flowchart fashion, how variables describing the project itself (including employment, tax base, and additional residences) are used to estimate costs as well as benefits to the community.

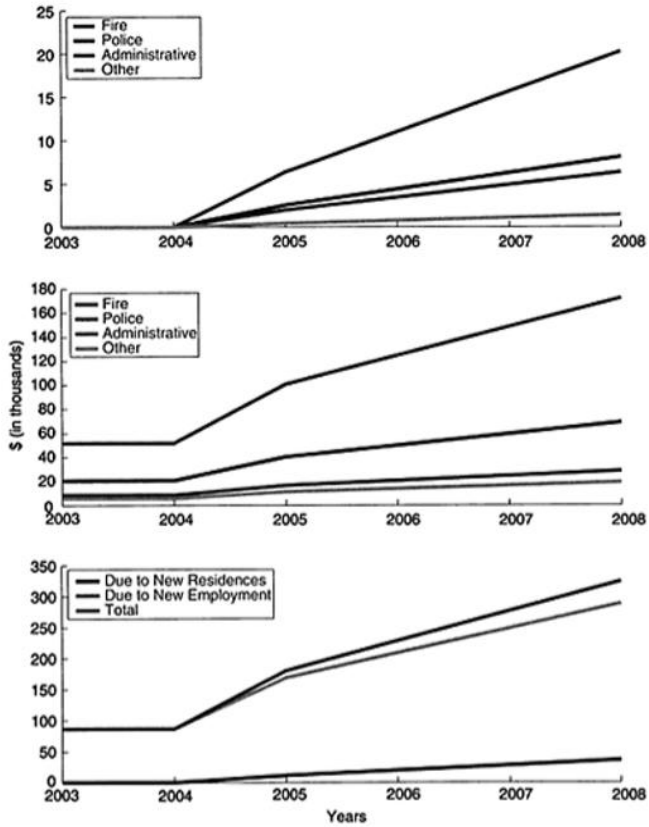


FIGURE 4.1 Municipal costs

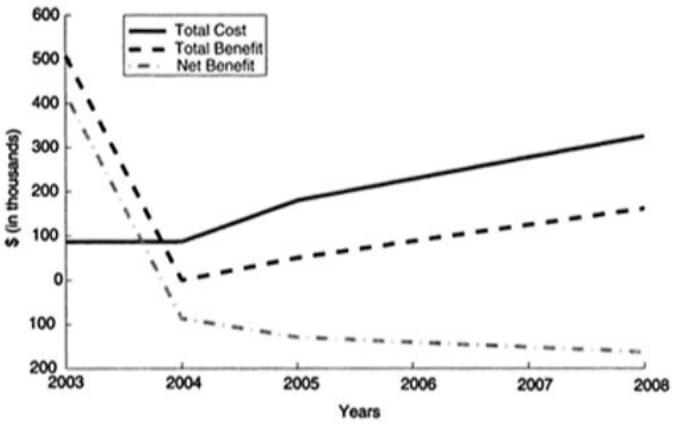
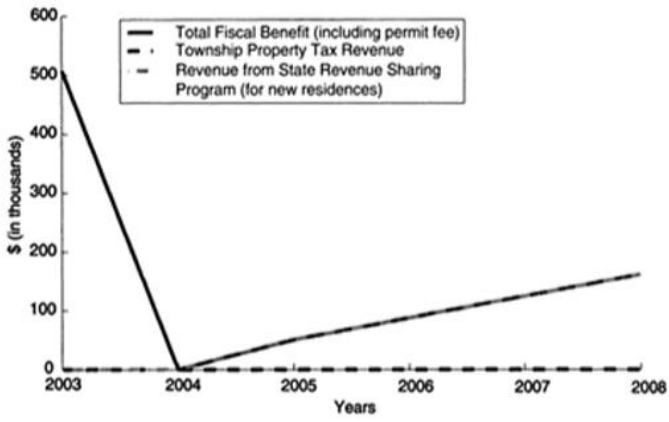


FIGURE 4.2 Cost and benefit

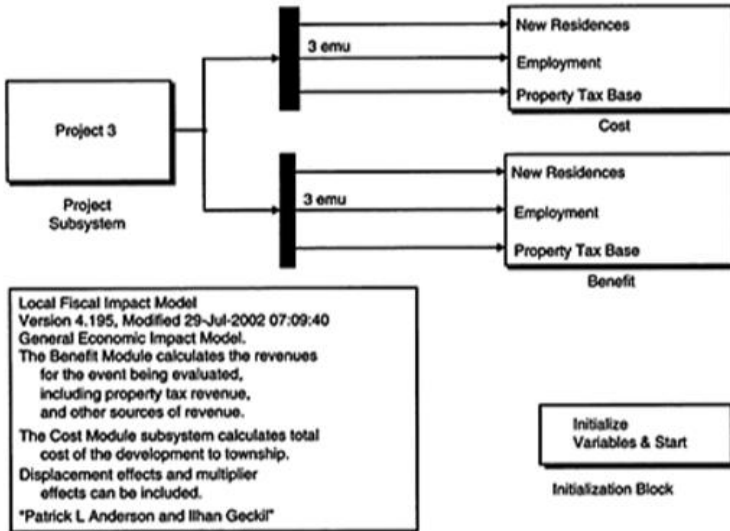


FIGURE 4.3 Fiscal model schematic – I

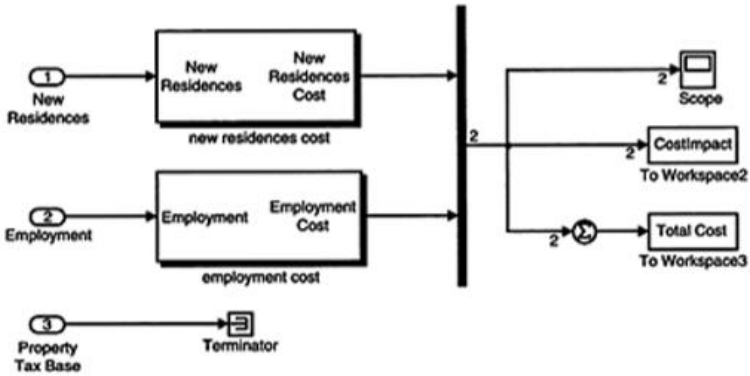


FIGURE 4.4 Fiscal model schematic – II

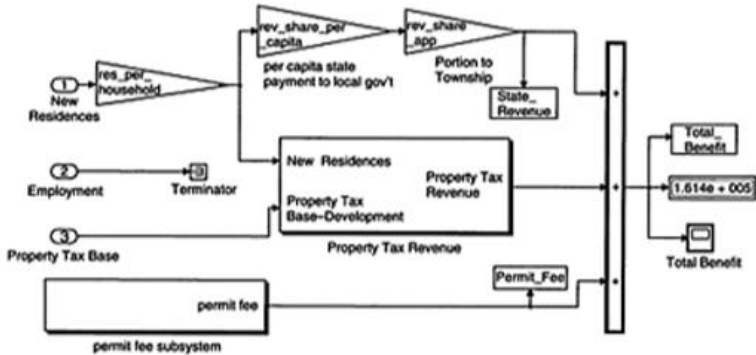


FIGURE 4.5 Fiscal model schematic – III

```

% Load workspace (.mat) file, if necessary
% (make sure file is located in reachable % directory.)
load X_AutoSupplier_input.mat;
% Get data from GUI (workstation) or from HTML (Web %
server) or from .mat % file.
%-----
% Time Parameters
Tstart = 2003; Tstop = 2008; Tstep = 1; Years =
[Tstart:Tstep:Tstop]';
sset = simset('SrcWorkspace', 'base');
%-----
% Sim command takes model, timespan, options, and
% external input arguments.
% Options structure and Timespan should be set
% above. sim('X_AutoSupplier', [Tstart:Tstep:Tstop] ,sset, []) ;
display('Model Simulated.') % Show output variables
display('Echoing Key Variables and Parameters.');
```

```

echo on; Years;
initial_real_property_value;      change_real_property_value;
initial_personal_property_value;  change_personal_property_value;
real_growth; personal_growth; construction_workers; new_workers;
transfer_workers; prty_tax_rate; permit_fee; rev_share_per_capita;
rev_share_app;                    resident_worker_app_new;
resident_worker_app__relocated;    spin_off_residences_new;
spin_off_residences_relocated;     change_property_value;

```

```

avg_hous_value;      fire_cost_per_job;      police_cost_per_job;
govt_adm_cost_per_j  ob;      other_direct_cost_per_job;      f
ire_cost_jper_residence;      police_cost_per_residence;
govt_adm_cost_per_residence;      other_direct_cost_per_residence;
echo off % -----
% Graph Results display('Graphing Results. ');
% Define variables used for graphing

cost_labels = {'Fire', 'Police', 'Administrative', 'Other'};
revenues_labels = {'Payroll Revenue', 'Local Purchases
Revenue',... 'Property Tax Revenue', 'Imports Revenue', 'Tourism
Revenue'};
% Y = [Var1./millions, Var2./millions] ;
% Call specific graphing commands or call graphing
% callback here. figure(1), subplot(3, 1, 1); plot(Years,
Res_Cost./1000,'linewidth', 2); legend([cost_labels], 2);
subplot(3, 1, 2); plot(Years, Emp_Cost./1000, 'linewidth', 2);
legend([cost_labels], 2); subplot(3, 1, 3);
plot(Years, Cost_Impact./1000, 'linewidth', 2);
hold; plot (Years, (Cost_Impact ( : , 1) +Cost_Impact ( : , 2)) . /
1000, 'r', 'linewidth', 2);
legend('Due to New Residences', 'Due to New Employment',
'Total', 2); orient tall; figure(2), subplot(2, 1, 1); plot(Years,
Total_Benefit./1000, 'r-' , 'linewidth', 2); hold; plot(Years,
Property_Tax_Revenue./1000,'b--', 'linewidth', 2); plot(Years,
State_Revenue./1000,'g-', 'linewidth', 2); legend('Total Fiscal
Benefit (including permit fee)', 'Township Property Tax... Revenue',
'Revenue from State Revenue Sharing Program (for new...
residences)'); subplot(2, 1, 2); plot(Years, TotalCost./1000, 'r-',
'linewidth', 2); hold; plot(Years, Total_Benefit./1000,'b-' ,
'linewidth', 2) ; plot(Years, (Total_Benefit-TotalCost)./1000,'g-' ,
'linewidth', 2) ; legend('Total Cost', 'Total Benefit', 'Net Benefit');
orient tall;

```

Code Fragment 2 Fiscal models 1

CHAPTER 5. APPLICATIONS FOR BUSINESS

This chapter begins the focus on the economics of the firm. There is substantial accounting literature on the proper way to account for the past actions of a firm. There are also numerous finance books covering specific topics in the capitalization of a firm. In this book, we touch on relevant accounting concepts and analyze briefly the capitalization of a firm. However, our focus is something often treated as an assumption in many of these texts: how a firm earns money. The fundamental economics of a firm revolve around earning a profit. Accounting and capitalization are important, but there can be well-capitalized, properly accounted firms that go out of business and sloppily managed firms that stay in business for years. Therefore, an understanding and modeling of the process of earning a profit should be the core of business economics.

We discuss in this chapter:

- Methods for projecting the future values of business variables (such as earnings or revenue), first with certainty, and then allowing for uncertainty
 - How to model different types of uncertainty, including random deviation around a well-determined path; Brownian motion (or “random walk”) processes; and “jump” processes, which simulate the effects of unexpected defaults, terminations, and other adverse events.
- The derivation of net present value formulas for various types of cash flows
 - How earnings from individual lines of business produce the gross earnings affect the firm
 - How earnings – or losses – affect the firm’s equity investors and lenders
 - An introduction to the use of simulation models to represent the revenue, losses, and earnings of a firm over time
 - An introduction to iterative methods, and an example method to estimate the value of a real estate enterprise when its earnings are partially based on the market value of property
 - Advanced topics such as the expected present value of an enterprise (or security) subject to business termination (or default) risk. The material in this chapter is essential for understanding the

following chapter on business valuation. The methods for projecting future values are used in other applications in this book.

Projecting Trend Growth: Simple Uncertainty

One of the most elemental tasks in projecting the future earnings of a company is projecting other variables into the future. Typically, these projections are phrased as statements such as, “We project 6% growth in GDP next year,” or “Experts anticipate sales of fancy new technology gadgets will grow 10% next year.” These statements imply a kind of certainty that simply does not exist. Whenever we project the future, we project with uncertainty. While we cannot avoid uncertainty, we can anticipate it. Therefore, even when we forecast trends that seem well established, it is best to anticipate some type of uncertainty. Therefore, we deal first with forecasting trend growth with simple uncertainty. Later, we will introduce more complex forms of uncertainty.

Simple Uncertainty: Baseline, High, and Low Scenarios Most knowledgeable people routinely forecast important variables with some confidence. By “some confidence,” we mean approaching the statistical definition of confidence intervals, although not reaching them.

For example, the “experts” cited in the statement above would probably be able to recommend high- and low-bound estimates to accompany their baseline estimates of sales growth. The spread (interval) between those high- and low-bound estimates implies a type of confidence interval, even though some elements of a statistical confidence interval will be lacking. A true statistical confidence interval will have both a “confidence level” (say, 90%) and an interval.

A Trend Projection Function for most tasks in business economics, important variables should be forecast using some allowance for uncertainty. We describe a utility program `TrendProject.m` that takes a base level of activity and forecasts it into the future incorporating the simple uncertainty of a high, medium, and low forecast trend. This function, a part of the Business Economics Toolbox, uses the following steps:

Wise men and women would do best to remember the following aphorisms: “Prediction is very difficult, especially if it’s about the

future.” (Niels Bohr, Nobel laureate in Physics); “To expect the unexpected shows a thoroughly modern intellect.” (Oscar Wilde); “A good forecaster is not smarter than everyone else, he merely has his ignorance better organised.” (Anonymous). These and other remarks on forecasting were collected by D.B. Stephenson and found at <http://www.met.rdg.ac.uk/cag/forecasting7quotes.html>.

Despite the ubiquitous statements in the general press about the confidence intervals from survey research (“Voters favor Smith by 10 points with a confidence interval of plus or minus 3 points”), such statements are meaningless without an accompanying confidence.

Applications for business

- Get the key assumptions about growth.
- Get the starting period and amount.
- Run the growth calculations three times, creating a high- and low-range estimate, in addition to a best-guess estimate. This is a true compounded growth calculation.
- Plot the results.

Trend Projection Code, Graphics, and Output the graphical output from this function is in Figure, “Trend projection graphic.” The program code is reprinted in Code Fragment, “Trend Projection Function”. The command window output shows the high, low, and middle projections for the default input assumptions.

5.1 Business Valuation and Damages Estimation

Economics, Finance, and Valuation

The previous chapter introduced the economics of the firm. Using a firm's earnings and a discount rate, we introduced a model for the value of a financial asset such as stock in a company. In this chapter we examine a subset of finance Trend projection graphic – the practice of valuing a particular asset, namely a business. There are numerous books on valuation, most of which are written from the point of view of an accountant or a professional appraiser. Some of these are quite detailed in the topics they cover. However, their focus is often on standard valuation equations and discount factors rather than on the fundamental economics of the firm. Conversely, for a reader with a strong economics background, the emphasis here on understanding the underlying industry and business prospects of the firm will seem obvious. Here, we approach valuation as an application of economics, and therefore treat carefully and systematically the actual generation of value rather than the evaluation of accounting statements. With such grounding, we then present quantitative methods to estimate the value of a firm.

Use of This Chapter by Readers of Different Backgrounds For those using this book as an introduction to valuation, the material presented in this and the previous chapter will provide a firm grounding in the methodology of valuing a financial asset. For a practitioner, however, other references will be needed to properly complete a business appraisal or assess damages. For practitioners and other readers familiar with the finance and accounting references in valuation, this book will introduce certain rigorous quantitative methods to forecast earnings and estimate business value.

In business economics, the asset to be valued is often a business enterprise, a line of business within a larger firm, or a security such as a stock, bond, or option. However, valuation is hardly restricted to economists. Indeed, it is so common that virtually all adults practice it and most become quite good at it

The first statement of accounting is generally credited to Luca Pacioli (1447–1517), a Franciscan monk and mathematician in what is now Italy. He published *Summa de Arithmetica, Geometrica, Proportioni et Proportionalite* in 1494, summarizing mathematical

knowledge of the time. The book contained a section on “Details of Accounting and Recording” that described bookkeeping as used in Venice. Venice, the Adriatic city which was the center of the Republic of Venice in the 15th century, was a leading trading center along with Florence (also the center of a republic and a city on the Arno River), Genoa (whose republic stretched across the Tyrrhenian Sea on the West Coast of the Italian peninsula, to include the island of Corsica), and other Italian cities. See, e.g., Gary Giroux, *A Short History of Accounting and Business*, a “book in progress,”; John Garrity and Peter Gay, Eds., *The Columbia History of the World* (New York: Harper & Row, 1972), Chapter 15. Giroux goes as far as crediting accountants with developing writing itself. While this is truly a heroic stretch – the index to *The Columbia History* fails to mention accounting but otherwise describes writing in ancient Sumer, Babylon, and Mesopotamia – it is clear that a good system of accounting was necessary for the industrial revolution, and that the ability to account for business transactions was a prerequisite for trade well before then.

The history of market value is as old as humans themselves, or at least dates from the time man was vanquished from the Garden of Eden. Man had to work to feed and take care of the family, and this required a careful evaluation of the amount of work needed to acquire food, shelter, and other needs. The *Columbia History* notes that the term capital from which capitalized and capitalism derive stems from the Latin *caput*, or head, coming in turn from a Babylonian term of the same meaning and sound. Debts and wages in ancient Mesopotamia were paid in silver, and the farmer was able to borrow during the seasons of sowing and cultivation and repay the debt (with a customary 33% interest rate) at harvest time. Depending on the abundance of the harvest and other economic conditions, the amounts of silver necessary to purchase a measure of grain would fluctuate.

An accountant prepares a balance sheet for a business, including schedules showing the assets (such as cash, accounts receivable, tangible and intangible assets, and any depreciation or amortization), liabilities (such as short- and long-term loans), and the equity investments and retained earnings. Is this a valuation exercise like the other examples above? The answer is, generally, no. Accounting

concepts are very important in business; indeed, they are vital for both trade and the organization of all businesses. However, accounting is based on the principle of historical cost, not current market value. Doubleentry accounting was documented as far back as Renaissance Italy and helped Italian traders ply the seas, with investors, buyers, and sellers relying on a set of accounts. Accounting statements account for the expenditures and receipts of an enterprise. Historical cost is the elemental concept in accounting for past expenditures, since you cannot provide a set of accounts today without basing them on the expenditures and receipts that produced them. Historical cost is determined by the transactions of the firm itself. Summarizing those transactions over a set of periods produces the balance sheet, income statement, and other fundamental accounting records. To the extent a value of an asset is implied by these records, it is commonly called “book” value. Market value is based on an entirely different principle, namely the capitalized sum of expected future benefits. No buyer determines the price he or she is willing to pay by adding up the historical costs of others. Buyers are interested in the benefits that will be returned to them should they make the purchase. Therefore, the market value of an asset is determined by views about the future and, in particular, the views of other participants in the market. The demonstrations of such views – actual market transactions – are the fundamental statements of value in economics.

Book Value and Market Value an accounting-based valuation of an asset – its “book” value – can be identical to its market value. However, such an event is more a coincidence than anything else. Theoretically, on the actual date of a transaction, the book value and market value are the same. Even this observation requires some simplifying assumptions, such as assuming that the decision to buy and the purchase transaction are made at the same moment; that there are no distorting effects of sales or other taxes; that the purchase price was the same as the market price on that day; and there was no compulsion to buy nor were there undisclosed material facts about the asset.

At all other times, however, the market value is likely to be different. Indeed, because market value is based on the expectations of future events, the market value of assets will typically fluctuate.

The most visible example of this is the stock market, where security prices will change minute to minute. Indeed, one of the advantages of historical-cost accounting is that it creates some stability in the financial marketplace because companies can prepare fairly stable balance sheets periodically without “marking to market” all their assets. For large companies with significant pension obligations, real estate, and foreign-exchange risks, the swing in apparent value from marking some of these assets to market every quarter – but not others – can mask more important business conditions. Certainly some parts of the balance sheet benefit from such treatment (such as financial derivatives), but a frequent mark-to-market of the entire balance sheet would be prohibitively expensive and confusing in many instances. Thus, accounting statements can never be relied upon as the indicator of value in the marketplace. That is not to say that book value has no usefulness. Without it, it would probably be impossible to account for a firm. Furthermore, among very similar assets, the ratio between book and market value can provide an insight into how the market is pricing certain assets.

Business Accounting statements have an essential, important role, but that role is not to predict market value.

TABLE 5.1 Concepts of Value in Accounting and Economics

Concept	Accounting	Economics
Statements of value	Book accounting records, including balance sheets and income statements	Market prices, including records of recent transactions of similar assets
Origin of value	Historical transaction records	Expectations of future benefits; supply and demand
Determinate actors	The firm itself	Others in the market
Determinate action	Transactions of the firm	Transactions in the market

5.2 The Role of Economist and Accountant in Business Valuation

The above discussion highlights the difference in approach between economics and accounting. Of course, a business valuation is not performed by an entire school of thought; it is performed by individuals. Most such individuals will be primarily trained in either economics or accounting. Which training is better suited for the task? We concluded that accounting was never intended to provide the basis for market value and that standard accounting statements can never be relied upon as an indicator of market value. Furthermore, we argue that understanding the economic conditions in the industry is essential to understanding the growth prospects of a business and therefore its value. Economists are certainly trained for this task. However, the first place to start in valuing a firm is its accounting statements. The evaluation of costs, the calculation of profit margins, etc., will normally be done more knowledgeably and effectively by an accountant. To properly value a firm, you must apply both intrinsic economic skills and accounting skills. Some individuals have both; many do not. In such cases, we recommend that a person with complementary skills be included on the valuation team. This suggestion is also made by Patrick Gaughan, who describes independently and in very similar terms the advantages and deficiencies of accountants and economists. In our experience, economists natively grasp the underlying growth prospects of the firm and more quickly see how a change in conditions (or interventions by other parties) affects the value of firms. However, the inclusion of an accountant skilled in evaluating accounting statements greatly reduces the chances of misreading the fundamental condition of the business and improves the accuracy of cash flow models.

Example Valuation Tasks

Consider the following common tasks, all of which require the economic thinking that is the kernel of valuation:

1. A housewife goes to the grocery store to purchase goods for the home. She considers the available food items and supplies in multiple stores, deciding to buy some goods in advance, shift some

purchases from store to store, and from one commodity to another substitute on the basis of her own assessment of relative values, the time value of money, and the cost of storing items for future use.

2. A father brings his daughter to a sporting event for which he expects to buy tickets. As they walk to the stadium, other fans (and professional scalpers) offer to sell tickets. Using his knowledge of past events of this type, he waits until the right moment before the start of the game and goes to the right spot outside the stadium to get the best seats for the money.

3. A judge determines that one business unlawfully deprived another of the ability to earn income in their new location. The judge must then establish a remedy, which by law and custom is normally the lost profit of the firm.

4. A business executive wishes to purchase another firm and consolidate the businesses into one large enterprise. What is the fair market value of these various firms? Most people would identify the last example as a task requiring economic reasoning, and some would also identify the previous example. Yet all these examples involve the same essential economic reasoning that goes into the last example and is the basis for valuation. Professional economists or other experts may be called upon to handle the last two example tasks, but the accuracy of the valuation estimates in the first two examples are excellent goals for the most rigorous analysis required of professionals.

Fair Market Value

When we say “value”, we generally mean fair market value, defined as follows:

The fair market value is the price at which the property would change hands between a willing buyer and a willing seller, neither being under any compulsion to buy or to sell and both having reasonable knowledge of relevant facts. It is interesting to note that, at least among taxing, finance, and valuation authorities, there is very little theoretical debate about the standard of value. The market of buyers and sellers – not the labor inputs in a commodity, its social worth, or its cost of production to the seller – defines value. On this point, the economics literature is the more obtuse. Ideas such as mercantilism and the labor theory of value floated about for centuries, which grounded value in nearly meta-physical qualities that may never

meet the market. By comparison, the basis of accounting is historical cost – a specific fact with little metaphysical about it.

The Gold Standard of Value: Actual Market Transactions We present below some innovative ways to estimate value in the context of tried-and-true advice from other practitioners in the field. In doing so, we will always define value as fair market value, the price a willing buyer would pay to a willing seller with both having adequate information. An actual market transaction of this type is the gold standard of valuation.

Estimation Methods vs. Actual Value

The chapter could stop here if there was a ready market for all assets and they all traded frequently in standardized commodities. However, closely held businesses, various securities of small- and medium-sized companies, most investments in real estate, and many other assets do not. Therefore, techniques of estimating the value of these assets have been developed. In this chapter, we describe three approaches to estimating value and provide quantitative models for the income approach. When actual market transactions are not available, projecting and then discounting the future income from an asset is often the best method of estimating its value in the open market.

However, do not confuse an estimate using a specific method with an actual transaction price. All valuation methods are attempts to estimate the price to which a willing buyer and seller would agree. This distinction between an estimate of a parameter and the underlying parameter is normally quite explicit in disciplines such as statistics, econometrics, and economics. In statistics and econometrics, even the symbols for the variables have evolved to indicate whether a certain quantity is an (often unknown) parameter or an estimate for one. For example, the known sample mean of a random variable X is often expressed with a bar over the variable, and the Greek letter mu is often used to describe the unknown population mean. The relationship between the two is summarized as $E = \mu$ which means that the mathematical expectation of the sample mean is the population mean. Similarly, a common practice in economics (especially since the advent of “rational expectations” models) is to incorporate expectations of consumers or workers in modeling their current behavior. The difference between those

expectations (which are normally unknown) and the actual variables is quite explicit.

Such a distinction is vital in financial valuation as well but is sometimes missing in valuation reports and the discussion of such reports. To be clear, when we use the term value as a verb in this book, we mean estimating the sale price upon which a willing buyer and seller would agree.

Other Value Standards While we will discuss fair market value, defined by willing buyer-willing seller arm's length transactions, there are other variations of the term "value" which are occasionally used. In particular:

1. *Intrinsic value*: This is the value of the securities of a company based on the underlying (or intrinsic) value of the company's assets. For example, an investor may look at the underlying real estate investments owned by a REIT (a publicly traded Real Estate Investment Trust) and believe that the intrinsic value of those assets is higher or lower than the market capitalization of the firm. If the intrinsic value is higher, he or she may decide to buy stock in the hope that the rest of the market eventually feels the same way. Intrinsic value will often vary from the market price for a variety of reasons, including the fact that securities in a company are not the same as an actual fractional interest in all the company's assets. Businesses are owned through securities, but those securities are not the same as a fraction of all the firm's assets. There are further, specialized meanings for the term intrinsic value in the investment community. The search for intrinsic value in stocks is a well established technique that is sometimes called value research. There are many other investment styles, including market-timing, asset-allocation, market indexing, technical or "chartist" and the tried-and-occasionally-true "dartboard" approach. In addition, a portion of the market value of certain options is its intrinsic value.

2. *Investment value*: This is the value a particular investor would place on an asset. Note that the investment value depends on the investor and, therefore, will not be the same for all investors. This term is sometimes confused with intrinsic value. However, there is a difference; an investor may not wish to purchase stock in a company that he recognizes has substantial intrinsic value. The investment value to that investor of the asset is much lower than its intrinsic

value. The market value, however, will be the same.

3. *Fair value*: This term is inherently vague and, therefore, must be defined by a statement of “fairness”. Indeed, the term is often used in laws and regulations in which the fairness criteria should be spelled out. There are some typical uses. One involves business combinations such as mergers or acquisitions, or transactions among equity holders in a business. The fair value here may not be the value set in an open market, where there are many buyers and sellers. Instead, the fair value is the price to which these parties would agree to buy and sell an asset without any compulsion. Such prices may deviate from the fair market value because the asset owners or acquirers would be few in number, and their willingness to pay for a particular asset may not match the prices set for similar assets in the open market. However, fair value should be reasonably close to fair market value because arbitrage possibilities should keep the value to even an unmotivated buyer reasonably close to the market price. Furthermore, allowing fair value to be markedly different from market value would allow for abusive tax evasion. For example, the break-up of a firm may require the various investors and lenders to place a fair value on the underlying assets. If the fair values are different from the market values, the taxes paid on gains or losses on these assets will be significantly different from what should be paid. In some cases, this difference is merely a deferral. In others, it would amount to an abusive tax shelter. In addition, the disclosure of fair values of assets and liabilities, based on market values, helps prevent investors from being misled about the prospects or performance of a company. Note that the Financial Accounting Standards Board, in line with the effort of regulatory bodies throughout the financial industry, is currently revising its statement of fair value to bring it closer in line with fair market value.

4. *Liquidation value*: The value of a firm that is being liquidated is almost always lower than that of a going concern, and is clearly different from its fair market value as an ongoing business. The “willing buyer, willing seller” test is violated if one party is compelled to sell, which is normally the case in liquidation. When assets are being liquidated, the conditions of the sale will often affect the market price significantly. For this reason, various types of liquidation value (such as “forced” and “orderly”) are used to

distinguish such a premise of value from the fair market value.

Valuation Methods. The market value of any asset is defined as the price a willing buyer would pay to a willing seller in an arm's-length transaction, where both parties have adequate information. In most cases, there is no purchase transaction or active market that establishes such a value, so methods of estimating the value of assets have been developed and refined.

Approaches to Valuation There are three generally recognized methods of business valuation described in the fields of economics, finance, and accounting.

1. Sales or market approach: The most direct method, if data are available, is to review recent transactions for similar assets. The closer the asset being valued is to assets that have recently been sold on the open market, the better this method works.

2. Income approach: The economic income or discounted cash flow method is the basis for most financial analyses. It converts an expected future stream of benefits into a present-value amount.

3. Asset-based approach: This approach is based on evaluating the underlying assets being sold, looking for direct information on market values for these assets.

When better data on underlying assets are not available, accounting information based on the historical cost of assets and their depreciation over time may be considered. For valuation purposes, these accounting figures must normally be adjusted significantly. All these approaches – when done properly – are consistent with the fundamental economics principle that the value of an asset is based on the capitalized stream of future benefits, adjusted for risk and other factors. These practical approaches are recognized by the Internal Revenue Service, appraisal authorities, and business valuation authorities. Similarly, in the government activity in which values are most important – taxation – state laws and court decisions typically codify these approaches as acceptable methods of estimating the value of property so that it can be assessed for property tax purposes. Taxation authorities, however, should not be considered unbiased interpreters of valuation theory; government entities have a vested interest in enhancing their own sources of revenue and therefore tend to adopt procedures, rules, and laws that

increase the basis on which taxes are paid.

Fundamental Equivalence of Methods If done correctly and with good information, the values estimated using the market or economic income approach should be very close to each other and should represent a good estimate of the amount a willing buyer would pay a willing seller. In the absence of the necessary information regarding other approaches, asset-based approaches can be used, but only with substantial adjustments.

Deficiencies in the Asset Approach The deficiencies of certain approaches to valuation, notably the use of historical-cost accounting records to suggest business values, has not gone unnoticed. The IRS, for example, abandoned valuation by the “excess earnings” method, which relies on historical-cost asset values, unless no other method is available. Unfortunately, it is still widely used, resulting in a “plethora of misapplication.

New Approaches to Valuation In recent decades, economists have developed much more sophisticated models for the value of assets. Much of this development has been in the sub-field of finance and investment, and has been focused on publicly traded assets such as stocks, bonds, and derivative securities. These models explicitly incorporate risk, and in some cases specifically include management policies. The most developed of this new approach are “option value” methods, in which the risk of certain events is explicitly included in the value calculation. These methods have been widely used for financial assets, and more narrowly in business valuation.

Recommended Order of Methods

We suggest below a recommended order of valuation methods. In suggesting this order, we attempt to maximize the use of information that represents actual transactions and the motivations of likely buyers and sellers, and minimize the amount of judgment, estimation, and use of unreliable historical-cost information. This order is defined for operating business valuation challenges and may not be applicable for other valuation tasks. By “operating” we mean generating income for investors by creating products or providing services. Estimating the value of an operating business often requires more analysis than estimating the value of a passive investment, or of a financial security like a stock or bond. After a brief general

discussion of each method, we provide economic analyses based on these approaches, including MATLAB routines, as well as examples.

Best Evidence: Market Price In cases where information on actual market prices are available and the assets in question are genuinely comparable, actual transaction prices should be given the highest weight in an analysis. When using market transactions as the basis for valuing an asset, be certain that the transactions contain all the elements needed for a true market value determination:

- Willing buyer
- Willing seller
- Arm's-length transaction
- Good information available to both buyer and seller
- Known sale price for the underlying asset

To truly understand the market values implied by asset sales, look carefully at the actual transactions. Adjust for any creative financing, seller financing, partial sales, or other unusual terms that can obscure the true cost of the underlying asset to the buyer. Also be careful to disentangle the price assigned to the business enterprise (which is normally the wealth-producing asset, complete with important intangible assets) from hard assets like real estate, inventory, rolling stock, and other items that can be resold.

Second Best Method: Income Method If a market transaction price is not available, an economic income or capitalized income method is usually the next best choice for estimating the value of an operating business. Financial theory is based on the principle that value can be estimated by summing the present value of future cash flows to be derived from the asset and adjusting for risk and other factors. A business enterprise, or a separable line of business within an enterprise, is an asset that can be valued with this approach. Cash flows in such a valuation model (often called a “discounted cash flow” or “capitalized income” model) are those derived from forecasted revenue and expenses. Standard texts in business valuation rely on such an approach and describe various methods of identifying the proper cash flows, choosing the right discount rate, and adjusting for risk and other factors. For example:

From a financial point of view, the value of a business or business interest is the sum of future economic benefits to its owner, each discounted back to a present value at an appropriate discount rate.

The discount rate used to convert future cash flows into present value terms depends on current rates of return for similar assets and the risk associated with the specific asset.

Critical Assumptions of the Income Method

The income method relies on certain critical assumptions, including:

1. Going-concern premise. Generally, both accounting and valuation exercises are based on the assumption that the business is a “going concern”. This point is glossed over in some valuation texts but made explicit in others. For example, Pratt, et al., cited previously, discuss the going-concern premise repeatedly. On the other hand, Damodaran, also cited earlier, does not even list it in his index, and discusses it only indirectly. It appears that valuation references with authors having accounting backgrounds are much more likely to state the point explicitly, while authors from a finance background assume that a knowledgeable practitioner would adjust the cash flow projections accordingly. As we consider the first principle of valuation to understand how a firm makes money, we favor the accountant’s rule of making this check an explicit part of every valuation.

In some cases – most notably a bankruptcy – this principle is violated, and the assumption should not be used. Furthermore, there are other circumstances in which the going-concern assumption may not hold even for a firm that operates without bankruptcy. Because of this, we believe the goingconcern assumption should be questioned in every valuation assignment. If it is not a going concern, then a valuation method based on past performance will be unreliable, and probably would produce an estimate far in excess of the actual market value.

2. Historical results or other objective data predict future performance. In most cases, the valuation of a business starts with the examination of the past year’s financial performance. From that base year, growth assumptions and other assumptions about the future are used to project out the results. The examination of a base period is an excellent place to start. Clearly, past performance is the strongest indicator of future performance for a wide number of variables. However, if there has been a significant change in the

management of the firm, the marketplace for the firm's products or services, material change in the firm's accounting or other important policies, or a record of past failures to make required payments or report accurate financial and sales information, then reported past performance is especially suspect.

In some cases, the historical results of an entity are not the best basis for projecting future performance, but a separate set of variables are. For example, consider retailers active in multiple markets, many of which have recently had declining population and declining sales. If retailers can identify their target market near a new store location, and the demographic and geographic information suggest that the target market is growing rapidly there, then a faster-than-historical-average growth assumption would be reasonable. In any case, the grounds for projecting future revenue growth in excess of recent periods must be substantial, and stated explicitly in the report.

3. Market and product positioning are predictable. Another critical assumption is that the product or service provided by the subject company remains consistent with its position in the past. If this is not the case, the effect of such a change must be explicitly considered, and the effect on revenue must be supported by analysis and other data.

4. Management quality and focus will be maintained. The income method typically assumes that management quality remains consistent and that the management focus will remain. If this is not the case, such a contrary assumption must be supported by other analysis.

5. The regulatory environment will change relatively little. If an industry is heavily regulated and most industries are regulated in some fashion – then such regulations have a strong bearing on the market value of firms in that industry.

6. The overall economic environment is predictable. One does not need to predict small movements in interest rates or quarterly GDP to provide a solid economic base forecast for use in a valuation. However, be attuned to fundamental assumptions about the economy and the market that are embedded in any sales forecast.

7. All other material, relevant conditions have been explored. The items above apply to nearly every valuation project. However, most valuations require special attention to one or more significant factors that are not universally present in businesses. Remember, think first

about what creates value for the business. Anything that affects the earnings of the business must be considered in your value estimate.

Third Best Method: Asset-Based Methods an asset-based approach using historic accounting data normally cannot be expected to produce an adequate estimate of the value of a business. It often takes far more time to individually value each asset than to value an entire entity. Furthermore, since the entity as a whole is often the reporting unit, the best data on past performance are often available for the entity. Finally, if an acquirer is purchasing an entity as a whole, that acquirer has demonstrated that the entity as whole, not differentiated assets is the target. While in some cases the acquirer will wish to keep certain operating units and sell others, the appropriate valuation method in those cases will still be an income or market approach to valuing the individual operating units

For operating companies, the asset approach is less desirable than an income approach in terms of both time and accuracy considerations. It often takes far more time to individually value each asset than to value an entire entity. Furthermore, since the entity as a whole is often the reporting unit, the best data on past performance are often available for the entity. Finally, if an acquirer is purchasing an entity as a whole, that acquirer has demonstrated that the entity as whole, not differentiated assets is the target. While in some cases the acquirer will wish to keep certain operating units and sell others, the appropriate valuation method in those cases will still be an income or market approach to valuing the individual operating units

However, asset-based approaches can be reasonable in the following situations:

1. The overall asset to be valued can be broken into discrete units, and the sum of their individual assets is reasonably close to the values of the entire asset.

2. Good market values, or values that can be estimated with an income approach, are available for each underlying asset.

3. There are no significant intangible, going-concern, or other value-affecting attributes to the assets that change because of a transaction.

4. The assets are commodities with a ready resale market.

5. The assets consist of substantial contingent claims or highly risky enterprises, in which an option-based approach best represents

the actual market for the asset.

Note on the Equivalence of Asset-Based Methods The recommended order of preference in methods is based on an assumption that the entire business has a value that exceeds that of certain parts, and that the true market value of a firm depends on the combination of all the essential parts. In cases where the value of the business largely consists of the sum of separate, severable operating units, one can view each separable part as an individual valuation task and apply the best method for each. Even in such cases, though, the value of the entity as a whole should be considered. In some cases, the individual business units may be valued higher on their own than as part of a larger business entity. The history of mergers and acquisitions among publicly traded companies provides ample examples of mergers that resulted in a reduction in market capitalization, indicating that investors viewed the sum as being worth less than the parts.

Note on Liquidation Value A fundamental accounting principle is that businesses are evaluated with the assumption that they are going concerns. In cases where such a going-concern assumption is violated, all or some of the firm's assets will often be liquidated. The liquidation valuation of a firm will normally depend on its "hard" assets, which include tangible assets that can be resold. Intangible assets of a firm – such as the goodwill it has created over the years of business, the firm's reputation, the network of contacts, pipeline of future orders, and other vital aspects of an ongoing business – are often dissipated in a bankruptcy.

Alternative: Real-Option or Dynamic Programming Methods although much less developed, in some cases newer methods may be superior to the traditional approaches. Two newer methods are discussed under the section "Dynamic Programming and Real Options Methods" that follows.

Dynamic Programming and Real Options Methods

In this section we describe a newer method for business valuation, the "real option" approach, which has become a practical tool in certain business valuations situations. We also introduce the dynamic programming method which, until now, has been largely confined to academic settings.

The Real Options Approach Human beings have always valued risk, and fairly sophisticated mathematical and statistical treatises on the subject of games of chance appeared centuries ago. However, good tools for quantifying risk in financial and business terms have been developed over the past few decades. The seminal work by Black and Scholes that produced a workable option pricing model began a rush of new tools for pricing financial options. The insights gained can be applied to numerous other entities as well.

For some subset of business economics problems, the best approach to valuing an asset will involve explicit use of option pricing theory. This is sometimes called “contingent claims valuation” or “real options.” Here “real” is used to contrast the underlying real asset from the financial assets that are the basis for most derivative securities that are actively traded. This newer class of methods does not fit neatly into the standard categorization of three approaches (market, income, and asset), but we tentatively place it in the asset category for the time being, recognizing that a new categorization scheme may be necessary.

Valuing Equity as an Option One of the most intriguing insights emerging from the original 1972 article by Black and Scholes was sparsely discussed in it. It was the observation that the equity in a corporation could be viewed as a call option on the value of the firm, with the strike price being the value of the debt. To motivate this, note that equity holders have the residual claim on the firm’s assets. If the firm is liquidated, the proceeds will first pay off all bondholders. Anything left will go to the equity holders. Of course, most shareholders do not want to liquidate the firm because they perceive its value as a going concern to be higher than its liquidation value. In such cases, their stocks are equivalent to “in the money” call options on the value of the firm. These securities can be sold and resold indefinitely as long as the market perceives the firm’s value as higher than its debt.

Option Methods Produce Similar Value Estimates Like all other valid methods, a properly completed option-method valuation, given good information, will provide a similar estimate of fair market value. Consider a simple example of this: the valuation of a well-established, profitable, publicly traded firm. Standard methods would estimate the fair market value of the firm as follows:

1. Using a market approach, the market capitalization of the stock and the market value of the debt would be summed up to produce an estimate of the enterprise value of the firm. In many cases, a simplifying assumption that the market value of the debt is its face value is employed. However, for long-term fixed rate borrowing, preferred stock, and other securities, adjustments may be required. See also, for this approach and the income approach

2. Using an income approach, and then the earnings of the firm would be projected forward, and discounted by the weighted average cost of capital. The weights on the debt and equity factors would be set by the market value of debt and equity, which would be similar to the analysis performed for the market approach. The results should be close, though probably not identical, to that from the market approach. If the results were not close, we should review all the assumptions and calculations, since the market value should not stray far from the intrinsic value (for a publicly traded firm). The prevailing discount rates, premiums, and discounts implied by market prices, however, do fluctuate. For this reason, we would not expect the two methods to produce exactly the same result.

3. For the reasons described earlier in this chapter, an asset approach would be the least reliable, and it probably would be dispensed with.

4. An option method would involve the evaluation of the option-adjusted value of the equity and debt. The option-adjusted value of the equity, given that the firm is profitable and has strong prospects of remaining so, would be equal to its regular market value. The value of the debt would be similarly estimated. The sum of the two would be an estimate of the enterprise value of the firm. Thus, for this company, the market approach and option approach would yield exactly the same result, and the income approach should yield a similar estimate.

Options Approach: Troubled Firms the real options approach becomes especially useful when the going-concern assumption becomes untenable or uncertain. Consider the sub-group of firms that are near bankruptcy, or near the termination of important business arrangements that underlie their ability to earn profits in the future. For such firms, a standard discounted earnings model is a weak

approach because the cash flows and resulting earnings in the future are so uncertain. An option-pricing model, on the other hand, will often provide a better value estimate.

Options Approach: Unusually Promising Firms the flip side of using an options approach for troubled firms is their application to especially promising firms. In this case, the “real options” method calls for assessing the underlying assets for potential surprising increases in value, such as the introduction of new technology or a potential to dramatically improve productivity in new markets. In this case, the “volatility” in earnings will be greater than in most firms in the same industry; only much of the volatility will be generated by potential spikes upward in earnings. This approach does explain some of the deviations between underlying current earnings and stock market prices for firms considered to have unusually promising futures. However, it must be used with caution in valuing firms that will be sold, especially if the “unusually promising” management, technology, or other attribute will no longer be present after the sale.

Applying an Options Approach While the Black-Scholes model was created for financial options, its insights and the extensive development in the field since then – can be applied quantitatively to businesses today. MATLAB has extensive tools for use with financial options and derivatives of all types, including both extensions of the Black-Scholes model and binomial tree models that simulate risks over a discrete set of periods. In order to use an option pricing model, you must identify the various assets and any contingent claims on those assets. The contingent claims (options) must be identified by strike price and the events that could trigger the claim. At this point, the key variables are not detailed cash flows (as in an income model) but historical variation (volatility) of the value of the underlying assets, and time. For companies in troubled circumstances, the analyst can sometimes come up with more reliable estimates of these variables than of cash flows. Keep in mind that much of option theory (as in portfolio theory) was developed with the assumption of symmetric variation, so be cautious in using standard formulas if you think the risks are one-sided. The binomial tree method, discussed next, below may be a better method in such instances.

Binomial Trees While the Black-Scholes model deserves its place in the pantheon of finance, it is not the only, or the best, model to use in many situations. For business valuations in cases where a firm is in or near bankruptcy, a simpler binomial tree model will often work quite well. For starters, such a tree can be hand-drawn, with the branches signifying values, given the results of future decisions or other events. If probabilities can be assigned to each branch, then the combination of the probabilities associated with each event and the values that would result from each series of events form the basis for estimating the worth of securities in the firm. Even if a full option-pricing approach is not employed, such analysis can provide vital insight into the actual market values investors will place on a risky asset.

Valuing Financial Options Given the importance of options in the financial markets, there are many references to their use and widely available software to calculate option values as well as related sensitivity and other parameters. MATLAB contains extensive option capabilities and is featured in excellent references on the subject. As this topic is extensively covered in other texts, we will not address financial options further.

Dynamic Programming Methods a method of solving complicated, multistage optimization problems called dynamic programming was originated by the American mathematician Richard Bellman in 1957. Like most brilliant insights, the method promised a radical simplification of some intransigent problems. However, the method was difficult or even impossible to implement until quite recently. Two parallel developments now make it possible to use this method in business valuation:

1. The increasing use of dynamic programming in theoretical work, primarily in academic settings, and the associated advancements in technique.

2. The increasing speed in computing power, which makes possible the enormous number of calculations that are necessary to implement certain algorithms for solving dynamic programming problems. The technique is almost unknown in business valuation. We introduce it here as a method that has great potential to improve valuation, but at present is not well developed.

The Dynamic Programming Method we present below a

simplified derivation of the technique. In this derivation, we model the management of a business as a multiperiod optimization problem.

1. A business is an organization which will live through multiple periods and for which a mixture of both reasonably predictable and unknowable events will occur. These events will present the management of the company with information, which can be summarized as data in a state vector. At each time period, holding the information available, the management takes certain actions, such as hiring, firing, purchasing, pricing, advertising, and selling. The variables can be represented as Markov processes, as the information available in one period summarizes all the useful information for the prediction of these variables in the next period.

2. The challenge (the “optimization problem”) presented to the managers of the company is to take actions in a manner that maximizes the value of the firm. If we take the value of the firm to be the expected future profits, discounted for time and risk, we can express this optimization problem in the following functional equation:

$$V(s, t) = \max_x \{f(s, x) + \delta E(V(s_{t+1}, x_{t+1}))\} \quad (5.1)$$

Here, $V(s, t)$ is the value of the firm given the state s at the time period t . This value consists of two parts: the current profit of the firm $f(s, x)$ and the expected value of the firm in the next period after discounting by the factor δ . The maximization problem involves the control variables or actions variable x , so the maximization operator references this variable or vector of variables. Because both the current profit and the future profits of the firm depend on the actions of the firm’s management, the action variable is an argument to the profit and value functions. Equation (5.1) is known as a functional equation because the expression $V(s, t)$ is not, strictly speaking, a function of just the variables s and t , but instead the maximization of a family of functions. We will also refer to it as the Bellman equation for this optimization problem.

3. The existence of a solution to the optimization problem under the conditions we will normally encounter in business settings has been rigorously established.

4. Actually solving the problem we have now stated has been the greatest difficulty in using this technique. Bellman himself called the problem the “curse of dimensionality” because the size of the problem is magnified exponentially by the number of variables and their possible paths over time. It is here that the advancements over the past decade in both analytical methods and numerical computing allow for the emergence of a practical dynamic programming method.

5. There are a number of methods for solving dynamic programming problems, each of which has certain advantages. These include: a. Recursively solving the problem backwards from a known terminal value; this is known as “backward recursion.” b. Iterating on the values created at each step with variations in the policy created by the application of the control variables; this is known as “policy iteration.” The more traditional approach of specifying an initial set of values for all variables, calculating the value function at these points, and then iteratively searching for higher values until such searches yield no further improvement. This is known as “function iteration.” The advances in numeric computing techniques in economics has produced a small number of usable computer algorithms which perform the enormous number of calculations that are often required to solve dynamic programming problems. Many of these numerical algorithms are, at the current time, either experimental or newly developed and therefore cannot easily be applied to business economics problems. However, we expect progress in this field to continue at a good pace, and the recent.

Introduction of usable dynamic programming utilities in MATLAB will enable determined economists to apply this technique to a variety of problems without further wait. In particular, the toolbox and related text authored by Miranda and Fackler provide an excellent introduction to the use of dynamic programming methods in economics.

Distinguishing the Dynamic Programming Approach Bellman’s key insight began with segmenting an optimization problem into twoparts: the benefits this period, and the value of the expected discounted benefits in the future. For our purposes, we will separate the valuation problem into two parts: the return on an investment in the current period, and the eventual returns of both income and capital gains. While we are considering the market value of a firm,

the technique was developed for use in optimizing a wide array of functions. One of the advantages of the dynamic programming method in business valuation is its proper assertion of the primary importance of management policy. While traditional methods of valuation assume a very passive role for managers, dynamic programming puts management policy front-and-center.

Although we are limited in our ability to quantitatively model management policies, the dynamic programming method explicitly considered policy actions that are largely ignored in traditional cash flow models. A second advantage is that it better incorporates the complex uncertainty that exists in almost all business settings. A real, human manager will not set all expenses to grow smoothly, in the face of changing revenue. The dynamic programming method incorporates this explicitly compare this to the simple uncertainty that underlies most cash flow models.

Mathematical Similarity of Methods

Both the real options and dynamic programming methods can be used to value business assets. If both produce a market value estimate that theoretically should be the same, then we would hope the two methods would be equivalent under a set of assumptions appropriate to the specific problem. The methods are similar mathematically, but are not identical. In particular, the source of discount rates and the availability of other traded assets are critical differences in assumptions. These differences notwithstanding, given complete information, and some limiting assumptions, the two approaches produce the same answer.

Conclusion: Real Options and Dynamic Programming Methods
This section presents approaches that are either relatively new to the field of business valuation or almost unknown. Like most newer developments, these methods require more work and care to use in practical settings than other, more established methods. However, they both offer powerful analytical and practical advantages that, in some situations, should not be ignored.

Information Necessary to Value a Firm

We discussed above various methods of valuing a firm. In this section, we discuss a strongly related topic: what information is

needed.

We organize the information necessary to value a firm into three groups: accounting information, economic assumptions, and management policies.

Accounting Information.

1. Historical sales and revenue: This includes volume information as well as revenue. If there are different areas in which the products or services are delivered, or different lines of business, a segregation of those revenue streams and related costs would be helpful. If the valuation exercise is focused on one line of business, properly apportioned historical financial information is vital.

2. Cost and profitability information: The costs of the firm – both variable and fixed – as well as capital costs, interest and taxes, and noncash costs such as amortization and depreciation are required for an analysis using the income approach. If the firm is to be valued using a market or asset approach, then this information is still important but may not be decisive. Together with the other information collected, this must answer clearly the question “How does this firm make money?”

3. Capital structure and discount rate information: All companies must be capitalized in some manner, and the cost of that capital is the basis for the discount applied to future earnings. For large companies acquiring relatively small entities, this step may be straightforward. For closely held companies, it can be quite challenging. Note that the term “cost of capital” is often used loosely; so be certain you have independently estimated this critical variable and documented your assumptions. Of course, any recent changes in capital structure (including issuance of new stock, repurchases, options, substantial price changes, etc.) should be considered, as should unusual financing arrangements or the use of an ESOP or similar plan.

4. History and form of business operation: The type of business, its ownership, its management, and broad history are also essential information for valuing a firm. In firms with franchises or special licenses, a review of history will often reveal agreements carrying with them valuable rights.

5. Any material, peculiar conditions: Material events nondisclosed are a serious risk in the business of valuation. All the pencil lead in the world will not salvage an analysis of a firm that has

its officers fleeing the ship; has had accounting policies abruptly changed; is losing money and does not know why; or is nearing bankruptcy or failing to pay its creditors.

6. Growth assumptions for the company: As value is a forward-looking concept, the assumptions about the rate of growth in the future (even if it is negative) are critical. In our view, the forecasting of income growth is generally more important than the selection of a discount rate.

7. Economic conditions in the industry: Always consider the industry in which the firm operates its position in the market, and the size of its trade area. Good market research, even if it is rudimentary, will provide helpful limits on growth assumptions later on.

8. Conditions in the geographic trade area: For those companies operating in a specific trade area, the conditions in that trade area should be considered. For a retailer or wholesaler, is the population growing or declining? For firms operating on a nationwide basis, the trade area may indeed be the entire country, although even then certain areas of the country will often have significantly higher growth prospects than others.

9. Value of franchises, licenses, and intangible assets: The firm may have special franchise rights, licenses, trademarks, patents, or other intangibles that are essential to their future earnings capacity. These must be considered explicitly.

10. Market transactions for similar businesses: If market transactions involving similar enterprises can be identified, these should be given heavy weight.

Management Policies

11. Management and management policies: Who manages the firm is an essential aspect of the value of that firm. It may be that new management is expected to perform better than the previous management, or fix problems, or simply be a caretaker until a new owner is found. Management policies, such as reinvestment practices, pricing strategy, marketing expenditures, and management appointments, are important. These vital assumptions should be considered explicitly.

12. Important business plan assumptions: If a company is planning a major expansion or has just completed one, will be losing

one division or gaining another, is shifting production or changing processes, or is similarly changing the fundamental processes of its business, the valuation must recognize this.

Checklists for Information The above list can form the basis for a checklist for requesting information. Checklists for documents are presented in a number of valuation references. We emphasize in our list the fundamental drivers of value – including management, business plans, economic conditions, and franchises or other intangible assets – rather than a list of financial statements. This is, not coincidentally, also the approach followed by the IRS.

Note that this list is one of categories and subcategories, not specific documents. In keeping with the first maxim of business economics – “think first, calculate later” – we suggest thinking carefully about the business before digging into a cashflow schedule. Recall that a number of now-defunct firms reported balance sheets healthy right up to the moment the market knew they were doomed. Alternatively, other firms limp along for years with significant problems. In either case, simply reviewing accounting statements would have given a valuation economist a false sense of the predicament of the firm.

Steps To Value a Business or Estimate Damages

Each valuation exercise should include the following steps:

1. Consider the economic role of the enterprise. This inquiry must answer the question “How does this firm make money?” If the firm can be considered a bundle of assets, the intrinsic value of those assets should be considered. This is asking the question “How does this firm make money?” for each individual asset. The sum of the parts – and any advantages or disadvantages of their being combined – should also be considered.

Note that the valuation exercise is separate from ascertaining blame or deciding who are the “good” or “bad” parties in a dispute. That task may be more important than the valuation exercise, but it cannot unduly color the results of the valuation exercise.

2. Specify clearly the asset to be valued or the cause of the lost profits or other reductions in benefits for which damages are to be assessed. If only a portion of a business is to be sold, it is essential that the dividing line be clear in the valuation exercise. If you are

estimating the value of the lost profits or other reduction in benefits to a business, you must similarly be quite explicit about what streams of benefits are being considered.

3. Identify any trade restrictions, legal covenants, laws governing or restricting trade or competition, contracts or arrangements with suppliers and customers, and other factors that affect the market and potential profitability of the firm or line of business.

4. Select the best approach for valuing the business or estimating damages.

5. Value the business using the available data and the best approach.

6. Check the resulting value estimate against other sources, even if they are flawed.

7. Summarize the results, including the relevant assumptions and reasons for selecting certain methods.

Consider the value of a beer, wine, or liquor distributorship. Such a business is similar in type to other franchised businesses, as well as many retailers and wholesalers.

How Does It Make Money? Such a distributor typically has an agreement with a supplier of alcoholic beverages, which allows the distributor to distribute specific brands of products in specific areas. From the difference between the wholesale price to the distributor and the price to the retailer, the distributor earns revenue. From that gross margin must be paid operating costs, capital costs, and taxes. The remainder is profits.

What Factors Govern the Future Stream of Income? In the case of the distribution rights to a line of beer, the relevant stream of future benefits is defined principally by the following: 1. The market area in which the distributor has the rights to sell a specific brand of beer. 2. The demand for that type of beer in that market area 3. The supply and price of substitute brands of beer and other beverages 4. The fixed and variable costs of distributing beer in that area, including the costs of promotion, marketing, and other tasks required of distributors by most suppliers 5. The selling price of that beer in that market area 6. The limitations in the state law and the franchise agreement on competition within the market area and on terminating the distribution agreement.

What Is the Best Approach? If market information (such as recent

sales of similar distributors) is available, it should be given the strongest weight. If not, the income method is the best and the only appropriate method. Asset-based methods would severely underestimate the business value of a profitable distributor, given that many of the key factors that ensure future income are intangible.

Shorthand Manner as a Comparison, Not a Method as a check on our results, we could use imperfect but available “rule of thumb” information. In most industries, there is a shorthand manner of describing the value of assets. This is true in the alcoholic beverage industry as well; many industry participants discuss sales in terms of “dollars per case,” meaning the sales price of the business divided by the volume of sales (in cases) for the trailing 12-month period. A shorthand manner of describing a transaction, however, should not be confused with a method of valuing a business.

Valuing the Distributor: Assumptions in conducting this valuation exercise, we make a number of important assumptions which include:

1. The distributor is an ongoing enterprise.
2. The market structure in the area is not substantially disrupted by the change of the distribution rights for one line of beer, or from one distributor to another.
3. The consumer demand for specific lines of beer and the overall preference for beer relative to other competing beverages will not change substantially in the future.
4. No other legal, regulatory, geographic, income, or other disruptions of a substantial magnitude will occur in the near future. Similar assumptions are often required in other industries as well.

Valuing the Distributor: Collecting the Data We follow the strategy introduced in “Getting Data into a Model” on page 66, in which we collect data in a spreadsheet, import it into MATLAB for analysis and graphing, and then export a portion of it back for reporting. The following code fragment lists the variables that were collected for each business.

Table 5.2 and Table 5.3 illustrate the worksheets that contain the input data for the business. Table 5.2, “Cost-of-Capital Estimates for Small Beer Distributors, 2003” illustrates a cost-of-capital analysis for a specific industry – in this case, small, privately owned beer distributors, for the year 2003. Table 5.3, “Input Sheets for Individual Distributors” illustrates the input variables used for a

sample company within this industry.

```
{sales_rev_base, sales_case_base, ptr, sales_trend,  
...  
SalableAssets, discount_rate_liq_sales,  
sales_decay_rate,  
discount_rate_close, ...  
earnings_pretax_share, earnings_pretax, ...  
growth_low, growth_high, growth_mid, numperiods, ...  
case_mult_low, case_mult_high, ...  
income_tax_rate, eps_sub_ratio, sub_rate, ptr_sub, ...  
wacc, wacc_low, wacc_high, lost_profit_alt};
```

Code Fragment 3 Input List for Valuation Function

We now collect the information for a distributor, which includes a large number of variables, for use in our calculations. The method we recommend is to: 1. Collect the data on the input sheets with the relevant source and other notes. 2. Transfer that data into the MATLAB workspace using the Excel Link command or other methods. 3. Pack the information about each business into a data structure.

Alternative Methods

As an alternative to a custom-programmed script such as this, you could also:

1. Keep the data in the workspace and perform any needed calculations on the variables. One drawback to this approach is the lack of saved information, unless a careful diarystyle record is kept of both the data and the commands.

2. Save the data in standard numeric arrays (matrices), which might be saved all together in a .mat data file. The drawback to this approach is the confusion that results when you have more than one subject business.

3. Manually create a structure using the Struct command or other methods. We also describe these methods in Chapter 4.

Valuation Estimates by Multiple Methods Once the data are assembled into a data structure, you can use multiple valuation formulas to provide plausible value estimates.

TABLE 5.2 Cost-of-Capital Estimates for Small Beer Distributors, 2003

	Concept	Rates	
Equity	Equity cost for industry, simple CAPMa	11.1%	
	Less: discount for state laws or other factors	-0.5%	
	Plus: micro-capitalization and franchise risk premiumc	2.0%	12.6%
Debt	Prime rated	4.7%	
	Premium for small businesses	1.0%	
	Other adjustment	0.0%	
	Equals: pre-tax debt cost	5.7%	
	Times: 1-tax ratef	68.0%	
	Equals: after-tax debt cost		3.8%
Leverage	Share of equityg	70.0%	
	Share of debt	30.0%	100%
Adjusted Discount Rate	Weighted cost of capitalh as adjusted	9.9%	
Memo	Likely high discount rate	11.9%	
	Likely low discount rate	8.0%	
	Risk-free rate	5.0%	
	Plus: equity premium	5.5%	
	Times: industry	1.10%	
	Equals: industry cost-of-equity	11.1%	

We say “plausible” because no formula alone, even with good data, will always provide a good value estimate. The multiple-method approach provides the analyst with a range of estimates, some quite plausible and some that should be rejected as improper. In particular, if the business fails the going-concern test, the liquidation value estimate will become very important.

TABLE 5.3 Input Sheets for Individual Distributors

Variable Name	Variables	Values
name_of_case	name of firm	distA
<i>Business Sales, Costs, and Profitability</i>		
sales_case_base	Sales (cases)—base year	48,000
ptr	Price of case to retailer, average for year	\$21.00
sales_rev_base	Calculated: sales revenue on case sales, base year	\$1,008,000
earnings_pretax_share	Earnings pretax (after operating expenses, capital expenses)	4.95%
earnings_pretax	Calculated: pretax earnings	\$49,896
<i>Sales Growth and Projections</i>		
sales_trend	Actual trend in annual sales (4-Year average rate of annual change)	-5.00%
growth_low	Expected annual revenue growth, low	-7.00%
growth_high	Expected annual growth, high	-3.75%
growth_mid	Expected annual growth, middle	-5.00%
numperiods	Number of periods to forecast sales at current profitability	7%
Discount Rates		
wacc	Weighted average cost of capital, best estimate	10%
wacc_low	Weighted average cost of capital, low estimate	8%
wacc_high	Weighted average cost of capital, high estimate	12%
<i>5-Year Lost Profits, Assuming Partial Substitution</i>		
discount_rate_close	Discount rate (closing business) on profit	10%

TABLE 5.3 (continued)

sales_decay_rate	Sales decay rate	7%
eps_sub_ratio	Earnings pretax share; ratio of margin of substitute beer to current brands	0.675
sub_rate	Substitution ratio (portion of lost current brand substituted)	0.9
ptr_sub	Price to retailer, average, for substitute brands	17
Value-Per-Case Estimation Methods, Assumptions for Low and High Estimates		
case_mult_low Case multiplier (low) \$5.50		
case_mult_high Case multiplier (high) \$11.00		

For a true going concern, it is not important. However, for a firm that is struggling and may have to close its doors or restructure significantly, an income approach based on the past few years' results (especially if one of those years was a profitable one) will almost always lead to an exaggerated value estimate.

The calculation of all these does not imply they are equally likely guesses; only that they are a set of calculations that could be performed on these data. A handful of these become, because the assumptions and methods are most proper for the case at hand, the estimates actually used. Each method is only briefly described below; the general approaches are described in "Approaches to Valuation" on page 202, as well as in "Steps to Value a Business or Estimate Damages" on page 167. In addition, a practitioner in this field should have one or more references besides this book. Typically, the inventory of goods should be considered separately from the hard assets of the firm since the manufacturer or another buyer will often purchase or repurchase undamaged, fresh inventory at a modest discount. The other assets (such as equipment, vehicles,

furniture and fixtures, warehouses, etc.) frequently have market values that are substantially discounted from their historic costs

Furthermore, the calculations done under each method are first-cut analyses, and would often require additional thought and adjustment.

Method 1: Liquidation The liquidation method is appropriate when the firm is not a going concern or when the assets themselves will be sold quickly. Liquidation value should always be considered conceptually to be certain that the going-concern assumption is properly made or not made. In addition, the liquidation values of the hard assets of a firm often become a portion of the price of a going concern. In such cases, the buyer pays some amount for the potential earning power of the business itself plus some amount for the inventory and other assets in active use. The liquidation values of the firm's assets are often a first approximation for this latter category

Method 2: 5-Year Lost Profits, with Adjustment Using this method, the profits for a line of business are forecasted over a certain period, assuming that there is some substitution of sales for the discontinued (or sold) lines of business. The income from the substituted product line offsets some of the value placed on the line of business that is discontinued or sold. This approach is appropriate to estimate damages for distributors who lose a certain product and therefore the profits that would result from it, but have the opportunity to earn profits on other products using the same facilities and business model. Note that the rate of substitution and the profits on the substituted lines are critical assumptions. Assuming zero substitution – an implausible assumption for most companies that remain in business – has the effect of valuing the entire lost profits stream without any offset.

Methods 3 and 4: Per-Case Multiples The first step in our valuation methodology was to consider the economic value or intrinsic business value of the enterprise. This must answer the question “How does this firm make money?” clearly. Businesses do not make money on volume. They make money by receiving more revenue than they pay out in expenses. Therefore, while discussing the value of a firm in dollars-per-unit-volume terms may be convenient, you should not forget that past volume is not the same as future profits.

This axiom notwithstanding, it is common in numerous industries to discuss the market value of a firm by dividing it by some measure of volume. In the stock market, analysts may talk about “earnings per share.” That is a convenient measure, but it does not mean that issuing more shares will result in more earnings. In many retail and wholesale trade industries, market value may be summarized on a per-square-foot or per-store ratio. Again, simply adding stores does not mean profits will increase proportionately. These metrics are, in reality, combinations of indicators for efficiency, current profitability, industry profit margins and competitive factors, all rolled up into one handy multiple. In the beer distribution business, the common metric is “dollars per case,” meaning the market value of a distribution business divided by the trailing 12-months sales on a case basis. Using some judgment about past sales and a healthy skepticism about unverified claims of sale prices, this method relies upon directly estimating the market value of a distributor by using a multiple of the past year’s case sales.

High and low comparison values: because this metric is more reliable as a comparison than a direct estimator of value, we use a high- and low-multiple calculation to provide a plausible range.

Methods 5, 6, and 7: Capitalized Income the standard methods within the income approach are based on discounting the future cash flows to a company. We present three variations here. The first is known as the free cash flow to the firm (FCFF) method and is often the best method when reliable financial statements and earnings projections are available, but no direct market value information on comparable firms can be obtained. The FCFF method is also commonly used to estimate damages that occur when a portion of a business is interrupted, preventing the firm from earning the profits on that portion of the business for a certain time. Multiple assumptions: Because the assumptions used in an FCFF method are so critical, in this routine we automatically use low, middle, and high estimates for the key variables of earnings growth and weighted average cost of capital (WACC). These result in three different estimates, which are listed as different “methods” but are actually generated with the same algorithm, but with different assumptions. Note that the final estimate must have debt and equity weighted by market value, which is discussed below. Related Free Cash Flow to

Equity (FCFE) method: There is a related method known as free cash flow to equity, or FCFE. The key distinction is whether cash flows measured go to the firm as a whole or to the owners of the firm itself. These methods will produce equivalent estimates, provided consistent assumptions are used for both. Using proper discount rate: Beware of using assumptions geared for one method in the other. In particular, note that the WACC is an estimate of the implicit discount rate for the firm as a whole, not the equity holders. The use of a firmwide cost of capital must be reserved for firm wide cash flow discounting. For FCFE calculations, the appropriate discount rate is that for equity holders only. Using market value to weight capital structure: Another common error in valuation is to calculate a WACC, with the weights being the book value of debt and equity. The actual weights should be the market value of debt and equity. For firms with substantial intangible value, the market value of equity will typically be a much higher share of the total capitalization than the book value. Reading a balance sheet for a firm, based on historical cost accounting, and using those figures to weight the cost of capital will often result in a substantial overestimate of the market value of a firm. Recall that the value of equity will be based on investor's views about the capitalized value of future earnings of the firm. Since you are trying to estimate the results of those views (a market price), be aware that cost of capital and market value are determined simultaneously in the market as a whole. Many finance and valuation texts imply that one variable (usually WACC) can be obtained from a reference source and then simply plugged into a formula to produce – Voila! – the market value. If true market values are available, then from those market values we can calculate the discount rates implied or the earnings forecast implied, given certain other assumptions. However, if market value is not known, which is usually the case with private firms, then it cannot be used to simply calculate the weightings. Pratt provides an example in which the use of book values to weight the capital structure of a private firm – with all other variables being known – produces an overestimate of 30%. Iterative methods to weight capital: There is a practical solution to this quandary. First, use the book value to prepare a first guess of market value and use this to prepare a first guess of WACC. Use this, in turn, to prepare a second guess of market value. Use the second

guess of market value to prepare the second guess of WACC and continue until the results are close enough to warrant no further iterations. These iterations are not difficult to perform manually on a standard spreadsheet, and the entire process can be programmed in MATLAB. We present a similar application using a simulation model in Chapter 10, “Applications for Business,” under “Iterative Market Value and Tax Model” on page 226. The Business Economics Toolbox contains the function `Iterate Value` for this specific purpose.

Other Methods Because every case is different, even a routine designed to estimate the value of a single type of firm in a single industry makes explicit allowance for some alternate measure of value or damages. You will note the `alt` (alternative) variable in the routine, which is there for this purpose. Such an alternative could be produced by another expert, by another measure, or through some yet undefined means. Its appearance in the calculation and comparison of plausible values allows it to be compared directly with the results of different methods.

Graphical Output from Valuation Calculations The valuation function creates a set of figures. One of these is a bar graph showing all the plausible values calculated by the function.

Small- and Medium-Size Businesses

A common mistake is to assume that investments in privately held businesses are like investments in the stock market, and therefore an index like the S&P 500 or the Dow Jones Industrials is an appropriate indicator of a small firm’s cost of equity capital. In fact, unless you are discounting another portfolio of investments in large-cap, New York Stock Exchange-listed companies, it is probably the wrong discount rate. Most companies are small businesses that do not issue publicly traded stocks. They are often closely held and the owners may have some degree of direct or indirect personal responsibility for the business. These companies are often quite profitable and employ the majority of private sector workers in America. However, equity investors in these firms must face significant additional risk and restrictions when compared with owners of publicly traded securities. These include:

- Limitations on marketability.

- Higher risk in general.
- Other covenants, such as partnership agreements or loan covenants that restrict the ability to sell interests. Why then does anyone invest in small businesses? There are obviously benefits, which include:
 - The ability to work in a protected or niche market.
 - The ability to exploit particular skills.
 - The ability to earn higher returns which are sufficient to compensate for the risk and restrictions.
 - The hard-to-quantify desire to be your own boss. Furthermore, we have noted that some owners of “small” business keep them very strongly capitalized, borrowing far less than the leverage bank lenders would allow. Does this mean these individuals are naive investors? Be careful before using assumptions that may seem warranted for publicly traded firms when analyzing small firms. For all these reasons, investments in most companies should not be discounted using the returns from an index of large-cap, publicly traded companies.

Capital Asset Pricing Model (CAPM) and Small Businesses
 CAPM is a powerful organizing principle in the field of finance. There are many data sources that explicitly or implicitly assume the CAPM model. While CAPM and its modifications are excellent starting points for understanding the market value of a large, publicly traded company, it often fails as an approach for smaller businesses. This failure is largely due to both a lack of data on these companies and a theoretical framework that assumes a large number of potential investments with a range of risk and reward characteristics. CAPM is a good starting point for a discussion of institutional investor behavior and the behavior of investors with significant funds and a long time horizon. However, CAPM assumptions generally fail in the case of small business investors who also work at the firm in which they have invested.

Consider a prototypical investor in this sector. He or she has two clear options:

1. Decide to continue as an employee and invest additional funds in the stock market.
2. Invest those funds to become a part-owner of a small business and accept the restrictions that are normally attached to owner-operators of small firms. Those are choices that evidence a range of

risk and reward; however, they are not choices that define a continuum of investments. This does not mean that the CAPM model does not apply to small businesses; it does mean that substantial adjustment and a search for relevant data will often be required.

A Note on Discount and Capitalization Rates the economics literature uses the term discount rate in a general sense as the key parameter converting a stream of future benefits to a lump sum today. One of the uses of a discount rate is to capitalize a stream of earnings into a present-value equivalent. In such cases, the discount rate used is also called a capitalization rate or “cap rate.” When using a discount rate to capitalize a future stream of benefits, the economics literature makes little distinction between the terms “capitalization rate” and “discount rate.” However, the terms are not always used interchangeably in the finance world, particularly in those areas where the literature is dominated by terms originating in accounting rather than economics. For example, some articles state that capitalization is a specialized form of discounting, which implies a single discount rate and growth assumption in perpetuity. A variation on this – embedded in a definition by the American Society of Appraisers – is a definition of capitalization that confines its usage to one shorthand mathematical formula. This is much more restrictive than separate definitions in the same document, which indicate that discounting and capitalization are both methods that convert future streams of income to a present value. Some authors confuse the use of a discount rate with the use of a particular theory or approach for estimating the discount rate, such as the CAPM. It is not uncommon to see the capitalization of income method described as a totally different method from discounted cash flow, even when the cash flows in question and the discount rates are quite similar.

One influential authority in this field uses the convention that a capitalization of income model means that a single rate was used to discount future earnings and to predict future changes in income, while a discounted income approach means that variations in earnings or discount rates over time could be involved. Under this convention there is no mathematical difference, but the nomenclature of “capitalization” implies a restrictive set of assumptions.

CHAPTER 6. APPLICATIONS FOR FINANCE

Finance is a field within economics, which is devoted to the study of the capitalization of enterprises and households. By capitalization we mean the various forms of financing and operating, including borrowing (debt) and direct investment (equity). How an enterprise is financed affects its growth prospects, its current profits, and its fair market value. Interest in finance has grown rapidly in the past few decades, and today there are many times more workers employed in occupations more related to finance than economics. Unfortunately, that growth in interest has not coincided with a growth in the understanding of the economic underpinnings of financial transactions. In this chapter, we will focus on finance applications within business economics and, in particular, look at the economic underpinnings of business transactions. The applications we will consider include:

1. Consumer lending, using home mortgages as an example
2. Depreciation of assets, using automobiles as an example
3. Investment portfolios and, in particular, the application of the mean-variance portfolio analysis techniques
4. Other applications such as growth models and credit risk

Mortgage Analysis Mortgages are a common consumer lending product. The dollar amounts involved are substantial enough to justify detailed analysis for most consumers, and for all lenders. As an example consumer finance application, we discuss in this section an analytical routine that examines the financing of a typical house purchase. The `mort_calc` function described below accomplishes a number of tasks that provide consumers with useful information about potential mortgages. Some of these (such as the calculation of the monthly payment) are fixtures of many applications. Others, such as graphing of amortization schedules and the estimate of total monthly PITI (principal, interest, taxes, and insurance), are not. This function does the following tasks:

1. Checks to ensure that all variables are present and that the interest rate is entered correctly. This is particularly important with interest rates. Interest rates are often quoted in percentage terms (a mortgage rate of 9.5 typically means a 9.5% rate on the loan), while most computer programs (MATLAB included) look for interest rates

in decimal form. Thus, it is a good idea to check any interest rate entered to ensure that it is in the decimal form. The “if...error...end” block in the function accomplishes this.

2. Calculates the principal from the home value and the down payment. A more sophisticated application might include other fees, points, etc. Note that we approach the loan amount as a secondary consideration; the primary consideration is the price of the property.

3. Calculate the periodic payment on the mortgage loan. We have used the MATLAB `payer` function here.

4. Calculates the amortization, using a different MATLAB function other than that used for calculating the payments. We have used the MATLAB `amort` function here.

5. Performs an audit check to ensure that the two different calculations are identical. Such checks are essential in any model in which amounts that should be the same are calculated more than once. Always check, even if you are sure that you designed the model properly.

6. Creates labeling information to use in the graphic. You will note that, throughout this book, we recommend creating the labeling information separately from the plot command when possible. This allows the information to be identified easily and used consistently. Note that the information created here is used twice, once in the plot and once in the structure of key information (see the next step).

7. Creates a structure of values and information to provide to the user and then assign the structure into the base workspace. MATLAB structures are a powerful way to organize disparate types of data and include descriptive labels along with the data.

8. Creates the plot. Finally, using the variables created previously and the labeling information, it creates the graphic.

The function produces a graphical description of the mortgage balance and the amortization (principal and interest portion) of the loan payments over time. To ensure that the viewers of the graphs understand the terms of the loan being analyzed, it also includes text describing those terms, as well as notes on the date printed. See Figure 6.1, “Mortgage analysis.”

```

function [piti, schedule] = mort_calc (value, down,
rate, term);
% mortgage, taxes, and insurance calculator
% Produces monthly PITI (principal, interest, taxes,
% and insurance) payment,
% and schedule in structure form. Also plots
% amortization schedules.
% Enter value of home, down payment, annual interest
% rate (decimal format),
% and term in years.  Example: mort_calc(250000,
% 50000, .07, 30)
% PIA 1 Dec 00; 20 Dec 00 ver 3; 23 Sept 2001; 27
% Sept 2001; 10 Nov 01 ver 5.3
% copyright 2001 Patrick L. Anderson; Limited rights
% granted under license
    echo off;
format bank;
if nargin == 0
    action = 'default_values'
elseif nargin == 4
    action = 'custom values'
else error('Must enter correct values, or use
defaults.');
```

end

```

% -----
switch lower(action)
case 'default_values';
    value = 250000;
    down = 50000;
    rate = 0.0725;
    term = 30;
end;
% -----
if rate>1
    error('Please enter interest rate as a decimal;
for example, .08 for 8% per annum.');
```

end

```

p = value - down;
if p<0
```



```

    error('Value of house must be greater than down
payment.')
```

```

end
% Calculations of Payments
% -----
    disp (action);
    pmt = payper(rate/12, term*12, p);
    taxes = .037 * value * 1/12 * 1/2 * .90;
    ins = (1470/600000) * value/12;
    piti = pmt + taxes + ins;
% -----
% whos;
% -----
% amortization calculations
[PRINP, INTP, BAL, P] = amortize(rate/12, term*12, p,
0);
payment = PRINP(1) + INTP(1);
month_index = [1:1:length(PRINP)];
% Report on audit check; Do payment and
% amortization calculations agree?
% -----
if pmt == payment
    disp('Amortization and loan payment calculations
agree.')
```

```

else disp('Check to ensure payment and amortization
agree.')
```

```

end
% Labeling Information
printdate2 = ['printed on: ', datestr(now, 26),
'.'];
format bank;
loandatal = ['Loan: $', num2str(round(p)), '; rate:
', num2str(rate*100), '%; term:', num2str(term), '
months.'];
loandata2 = ['Monthly payment: $',
num2str(round(pmt)), '.'];
loandatacell = {loandatal, loandata2};
format short g;
% Create structure of values and vector of
% amortization information; put in workspace
% -----
schedule = struct('loan', p, 'payment', payment,
'rate', rate, 'piti', piti, 'term', term,
'principal', PRINP, 'interest', INTP, 'balance',
BAL, 'month index', month_index);
assignin('base', 'schedule', schedule);
disp(loandatal);
disp(loandata2);
disp('The following values are in the structure
"schedule" in the workspace:');
disp(schedule);
```

```

also available in workspace as "X");
X = [schedule.principal schedule.interest
schedule.balance] ;
assignin('base', 'X', X);
% Plotting-----
figure(10),
% set figure window title (visible) and figure
% property tag (invisible)
figuretag10 = ['Projection for Mortgage'];
%Insert name on this line
set(gcf, 'Name', figuretag10, 'Tag', figuretag10);
% first subplot
subplot(2, 1, 1),
plot(schedule.principal, 'r--'), hold on;
plot(schedule.interest, 'b'), hold off;
legend('Principal', 'Interest');
title('Principal and Interest Payments on Loan',
'fontweight', 'bold');
% second subplot
subplot(2, 1, 2),
plot(schedule.balance/1000, 'g');
Title('Outstanding Balance', 'fontweight', 'bold');
xlabel('Months');
ylabel('$thousands');
% insert key text into current axis
text(0, -.15, printdate2, 'units', 'normalized',
'fontangle', 'italic');
text(0.05, .30, loandatacell, 'units', 'normalized',
'fontangle', 'italic');

```

Code Fragment 4 Mortgage Calculator

For simple analytics, including the calculation of loan payments, there will often be more robust alternatives to using a MATLAB Web server. However, for advanced analytics and in environments where the users are more limited, a MATLAB Web server application could extend the reach of customized applications to a selected audience around the globe.

One of the more common analyses performed for companies with large tangible assets are a depreciation analysis. This often takes one of two forms: depreciation for tax purposes and depreciation for book purposes. The fact that most companies perform at least two depreciation calculations is an indication of its arbitrary imposition in the tax code.

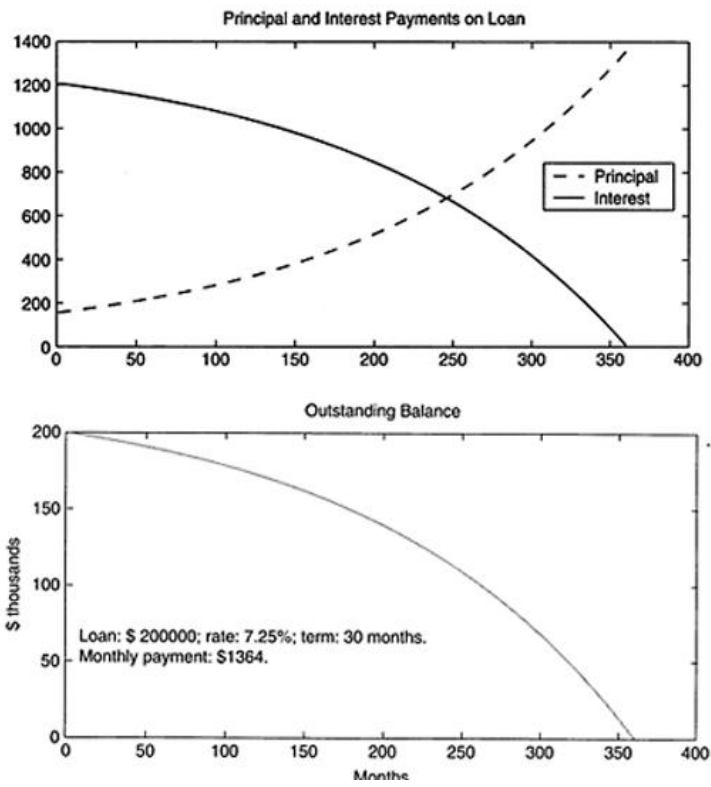


FIGURE 6.1 Mortgage analysis

Of course, the market value of those assets is not likely to coincide with a depreciation schedule imposed by a tax authority, or by an accounting convention. As market value drives the economics of an enterprise (and the wealth of a household), their managers often want to know the actual (market) value of assets, and compare them with the tax or accounting value.

6.1 Auto Depreciation

The Internal Revenue Code sets a harsh depreciation schedule for most vehicles, a straight-line depreciation over 5 years, and then normally imposes a half-year convention that effectively adds another year. For many users, the actual depreciation allowed is even less. There are “luxury car” limitations on the allowable amount of an automobile’s cost that can serve as the basis for depreciation; there are limitations on the use of itemized deductions; and there are other limitations based on the business usage share. In addition, the tax laws change from time to time. Of course, we are not providing tax advice to readers of this book. However, we are suggesting tools that will estimate the market value, tax depreciation, and other variables, given a set of assumptions that the user will choose. The auto depreciation function we created for the Business Economics Toolbox calculates depreciation in multiple ways:

The double-declining balance (DDB) method, using both a full-year and half-year convention:

- The straight-line method using a half-year convention
- A slow depreciation schedule for long-lived vehicles the function also produces three graphs and summarizes

```
function [dep_sched] = auto_dep (cost) ;
% Auto depreciation as function of original cost
% of vehicle;
% assuming DDB(10 years), with 20% salvage;
% also produces SL(5) with half-year convention
% and 10% salvage;
% [Based on standard Internal Revenue Code MACRS
% depreciation for cars; does
% not incorporate "luxury" vehicle limitations or
% expensing.]
% Also produces 150DB 10 year "slow" depreciation
% to 22% salvage.
% Example: [a, b] = auto_dep(40000)
% ...produces three graphs, depreciation series,
```

```

% and a structure called
% "dep_sched" with additional data.
%
% PLA, 21 May 2002
% (c) 2003 Anderson Economic Group. Part of
% Business Economics Toolbox.
% Use by license only.
% key assumptions
if nargin<1
cost = 40000
disp('-----Missing original value. Using
$40,000 as a default.--- ');
end
salvage = .2 * cost;
salvage_irc = .1 * cost;
salvage_b = .22 * cost;
life = 10;
life_irc = 5;
years = [1:1: (1+life)] ;
factor = 2; %use to create declining balance
factor; 2 means Double Declining Balance
factor_b = 1.5; %no factor for IRC; use straight
line; 1.5 for "slow" depreciation
% -----calculate depreciation 4 ways :-----
% first way
% note: add zeros at end of series to make sure
% same length when plotting
dep_series_a = depgendb(cost, salvage, life,
factor);
dep_series = [dep_series_a, 0];
cumdep = cumsum(dep_series)
value_series = repmat(cost, 1, length(years))
cumdep;
% second way
dep_series_hyc = [(.5 * dep_series_a) , 0] + [0,
(.5* dep_series_a)];
% third way
dep_irc = depstln (cost, salvage_irc, life_irc);
dep_series_irc_a = [repmat (dep_irc, [1, life_irc]),
zeros(1, (life-life_irc))];
dep_series_irc_hyc = [(.5 * dep_series_irc_a) , 0] +
[0, (.5* dep_series_irc_a)];
% fourth way--good resale vehicles ("slow")
dep_series_b1 = depgendb(cost, salvage_b, life,
factor_b); %trouble here?
dep_series_b = [dep_series_b1, 0];
cumdep_b = cumsum(dep_series_b)
value_series_b = repmat(cost, 1, length(years))
cumdep_b;
whos;
% Note: MATLAB description: Depreciation =
% depgendb(Cost, Salvage, Life, Factor)

```

```

% Cost: Cost of the asset.
% Salvage: Estimated salvage value of the asset.
% Life: Number of periods over which the asset is
% depreciated.
% Factor: Depreciation factor. Factor = 2 uses the
% double-declining-balance
% method.
%
% SL = DEPSTLN(COST, SALVAGE, LIFE) calculates
% straight-line depreciation
%-----
% Plots
figure(1),
subplot(2, 1, 1) ,
plot(years, dep_series);
title('Depreciation schedule, 200DB, 10 years, 20%
salvage', 'fontweight', 'bold');
% xlabel('Years');
ylabel('$Depreciation');
subplot(2, 1, 2),
plot(years, value_series);
title('Depreciated Value', 'fontweight', 'bold');
xlabel('Years');
ylabel('$Value at End of Year');
text (0.0, -0.2, ['Cost is: $' num2str(cost)] ,
'units', 'normalized', 'fontangle', 'italic');
figure(2),
subplot(2, 1, 1), plot(years, dep_series_hyc, 'mx');
title('Half-year convention schedules, 10 years DDB
and SL(5)', 'fontweight', 'bold');
legend('DDB(10), HYC');
% xlabel 'Years' );
ylabel('$Depreciation');
subplot(2, 1, 2), plot(dep_series_irc_hyc, 'ms');
legend('IRC SL(5), HYC');
xlabel('Years');
ylabel('$Depreciation');
text (0.0, -0.2, ['Cost is: $' num2str(cost)] ,
'units', 'normalized', 'fontangle', 'italic');
figure(3),
subplot(2, 1, 1), plot(years, dep_series_b, 'g');
title('Good Resale Vehicles, 1.5 DB depreciation
with 22% salvage, 10 Years', 'fontweight', 'bold');
legend('150DB, 10Yrs, 22% salvage');
% xlabel('Years');
ylabel('$Depreciation');
subplot(2, 1, 2), plot (value_series_b, 'g');
legend('Depreciated Value');
xlabel('Years');
ylabel('$Value at End of Year');
text (0.0, -0.2, ['Cost is: $' num2str(cost) ], slow

```

```

depreciation'], 'units', 'normalized', 'fontangle',
'italic');
%-----
% Create structure to report (transpose into
% column vectors)
dep_sched = struct('cost', cost, 'years', years,
'full_year', dep_series, 'half_year',
dep_series_hyc, 'irc', dep_series_irc_hyc, 'slow',
dep_series_b');
dep_sched;
% record fieldnames within schedules
names = fieldnames(dep_sched);
% create cell array of same data
cell_sched = struct2cell(dep_sched);
% % assign data to base workspace, so it can be
% accessed
% assignin ('base', 'schedules', schedules);
assignin ('base', 'names', names);
assignin ('base', 'cell_sched', cell_sched);
% % need to have all rows of same length for
% following array:
% printing_sched = [schedules.years
% schedules.full_year schedules.half_year
% schedules.slow];
% disp('-----Years, full-year, half-year
% convention depreciation.----- ');
% disp(printing_sched);

```

Code Fragment 5 Auto Depreciation

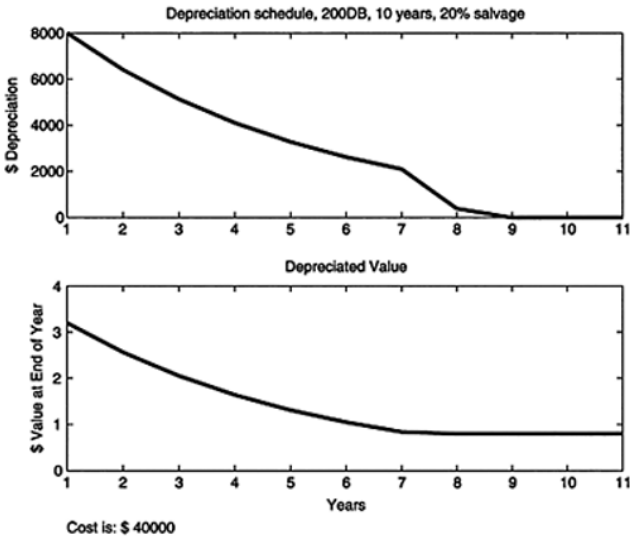


FIGURE 6.2 Auto depreciation: double declining balance

The function also produces three graphs and summarizes the data in a structure. See Figure 6.2, “Auto depreciation: double declining balance,” Figure 6.3, “Auto depreciation: half-year convention, Modified Accelerated Cost Recovery System (MACRS),” and Figure 6.4, “Auto depreciation: long-lived vehicles.” See also Code Fragment 5.

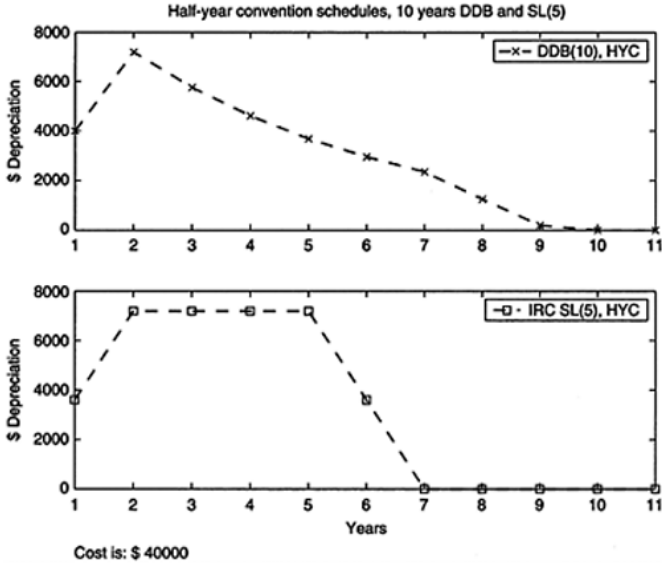


FIGURE 6.3 Auto depreciation: half-year convention, Modified Accelerated Cost Recovery System (MACRS)

The analysis of investments, particularly the analysis of portfolios of different equity and bond investments, is a field that is particularly well-suited for the tools described in this book.

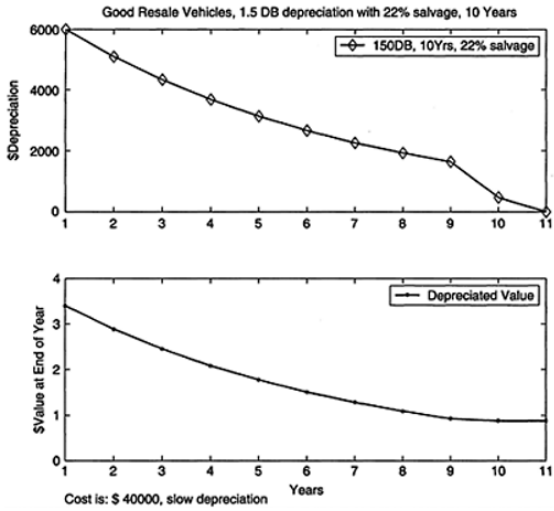


FIGURE 6.4 Auto depreciation: long-lived vehicles

CHAPTER 7. MODELING LOCATION AND RETAIL SALES

From this point, we follow the development of gravity models of spatial interaction from the time of Isaac Newton to today and then apply them to an important challenge in business economics – the analysis of retail sales. We discuss the fundamental tools of the analysis of retail sales, including the definition of trade areas, penetration rates, and other measures, and the use of Geographic Information Systems (GIS). Finally, using examples drawn from actual work in competitive retail markets, we present specifications and estimation methods for models of retail sales.

The Location Poverty of Modern Economics.

Most academic economics texts talk very little about location. Indeed, to the extent location is mentioned at all, it is in the context of being within a national or state border, or between urban and rural areas. For example, most macroeconomic texts start with a closed model of the U.S. economy and later move to an open economy. In such studies, the only location factor is whether you are on one side of the border or the other. Furthermore, prices, tax rates, and other variables are often assumed to be constant through out a country. Rarely is the vital importance of proximity to customers or competitors discussed. An academic discussion of the supply curve within a market or company often assumes a constant cost of acquiring or delivering goods.

Indeed, it is not hard to find subsectors that are almost entirely dependent on differences in location or delivery costs. There are many obvious “location, location, location” subsectors such as real estate, convenience stores, and resorts. However, the more instructive examples involve day-to-day tasks in which location is not the most obvious factor. For instance, consider the following two examples:

(1) The benefits of faster delivery have created the entire express mail industry, in which customers pay five or ten times the standard postal rate to receive a small package within one day. The success of the industry can be attributed primarily to the costs of doing business with distant parties and, therefore, to the value of a new technology that makes these costs smaller. (Of course, there are other supporting causes, including the negative effects of the original government

monopoly on the postal system's efficiency and incentive to innovate and the technological and managerial efficiencies of the leading express mail companies.)

(2) Bookstores continue to flourish, even as the Internet age provides a low-cost alternative for those merely interested in purchasing books. Indeed, if cost-per-book were the only factor in consumer decisions, traditional bookstores should (in line with the exuberant expectations of the denizens of the Internet investment bubble) have disappeared by now. They must be providing some experience for which customers gladly pay a premium, in terms of time for travel, search costs, and prices of books. As these bookstores congregate in urban and suburban strongholds, their proximity to their customers' working and living locations is clearly an essential factor in the demand for their product.

Similarly, the typical academic discussion of demand may make mention of how customer demand is affected by the total costs involved in the purchase of a product or service. However, the effect of location on the overall cost involved in a purchase is often ignored. One recent effort to reconsider the role of location in economic thought put it succinctly: "How do economists routinely deal with the question of how the economy organizes its use of space? The short answer is that mostly they do not deal with the question at all.

The expectations prove the rule. Harold Hotelling's "Stability in competition," published in 1929 in the *Econ. J.*, and Steven Salop's "Monopolistic competition with outside goods," *Bell J. Econ.*, Spring 1979, are benchmark studies of competition when consumers have different transportation costs to individual retailers. Spatial distribution of consumers and retailers is, however, usually ignored even in hedonic models of demand. labor is limited by the extent of the market." Smith's discussion of 200 years ago contains the essential insight we will use in the discussion of retail sales – how the size of the market and the costs of servicing that market limit commerce and wealth. Smith's discussion of those ancient countries that had developed significant wealth – India, China, and some in the Mediterranean region – notes how the ability to expand their markets through transportation was critical. Developments after Smith similarly focused on the production side of the economy. In a classic

work, the German economist J.H.von Thunen proposed in 1826 a model that explained how agricultural production was efficiently organized when farmers cultivated and harvested crops in areas surrounding a city. The city provided the mechanical and specialized functions that were too expensive for individual farms to purchase. The costs of transit became the critical variable, and the cost differentials caused by the differences in travel led to a pattern of different crops being cultivated at different distances from the city.

By “retail sales” we mean transactions with a consumer which result in a service or goods being provided at a specific site. Thus, sales of household goods, automobiles, and food would be considered retail sales. The same approach may also be useful for consumer banking, services such as dry cleaning or medical doctors, and entertainment venues such as casinos, racetracks, theatres, and stadia. On the other hand, the manufacture of such goods, and the wholesale trade that got them to the point of sale, would be outside the definition of retail sales.

7.1 Applying Economic Theory to Sales Location

The foundations of microeconomics provide ample support for the notion that retail activity will decrease, given increased distance from a retail outlet. This arises primarily from the fact that covering distance takes time, and time is valued by consumers and producers alike. The economics of time is based in utility theory, which is at the basis of microeconomics. Once you recognize time as an input into the utility function of a consumer, a model of microeconomic behavior will suggest that consumers shop closer to home or work in order to minimize the welfare loss due to travel time. Pioneers such as Adam Smith and von Thunen were able to explain location patterns and their relationship to productivity and wealth in the 18th and 19th centuries. While most derivations of consumer demand from utility theory omit location as a factor, location can and should be directly included in applied work.

The economics of time has been developed in other areas of applied economics, such as finance and econometrics. However, the economics of time we discuss in this chapter focuses on the value consumers place on their time. We can distinguish this from the

common uses in other fields as follows:

- The time value of money in finance arises from consumers' time preference for consumption. Fundamental microeconomics indicates that consumers must be compensated for delaying consumption, and from this originates a time value of money.

- In econometrics and other applications of statistics, the seasonal adjustment of many data series reflects the intrayear cycles of human behavior in the market, as well as natural phenomena.

- Changes in behavior that occur because of unexpected, one-time changes in income, tax policy, or demand are normally distinct from the behaviour of consumers who value time.

The careful economist will avoid assuming that the cost of increasing the distance from a consumer to a producer is merely a wage rate multiplied by the time it takes to make a one-way, or even two-way, trip. When consumers choose a producer, they are of ten estimating more than the costs for a single transaction. For example, consumers choosing a dry cleaner may wish to learn about a provider's prices and services, find a convenient location along their driving routes, and maximize nearby shopping opportunities. Their selection will be based on all these variables, and the time variable considered will often be the cost of multiple trips over a long period of time. Consider also a consumer choice of an automobile dealer. We might conjecture that the amount of money involved in a purchase of a new automobile is substantial enough that even a small discount would induce a consumer to travel a good distance. In fact, analyses of sales in quite competitive markets – even when dealers compete for sales by lowering prices – show the majority of sales go to the closest dealer, not the “lowest price” dealer. How do we reconcile the conjecture with the observation? Once you consider the number of times a consumer will visit a dealer, the apparent cost penalty per mile becomes much larger. If a consumer kicks tires on one trip, makes a purchase agreement on a second, picks up the vehicle on a third, and returns for warranty service for four additional times during the next year, he will have made a total of seven round trips to the dealership.

Sources of Customer Location Data Individual customer location data should include individual customer purchases and the customer's distance from the retail center. Many companies regularly

collect this data and may, or may not, realize it. For example, auto dealers have access to the addresses of past customers through registration information; banks know the addresses of account holders; sports teams know the addresses of season ticket holders; and travel agents know the addresses to which to send plane tickets. The more extensive the data source, the better. If you are using customer location information for multiple retail locations within a market area, you should record the distance between the retailer and the customer, even if that retailer is not the closest to the customer's location. Although these data are the most accurate to use in the analysis, their use is not always feasible. If the data is not already collected, it will likely be very difficult to compile for a short-term project. Furthermore, many establishments are unable to gather this information on their customers. For example, many coffee-shop patrons would be taken aback if they ordered a latte and the coffee shop employee asked for their address. The same difficulty holds true for most retail stores.

The location data are frequently compiled as the number or percentage of customers by geographic region. Often, the location unit is a zip-code or phone-exchange, as such data are often collected simply by asking customers for it at checkout.

There are other methods as well. Special "club" cards that connect customer purchases to their location are sometimes used, although customers must normally be given some incentive to provide this information and shoulder the burden of carrying it and presenting it at checkout. Surveys, including "intercept" surveys of actual shoppers, can also be used. However, be cautious when using survey data that is contaminated by self-selection bias or poorly designated questions.

Many major retail establishments have used this approach to collect customer location information. Such data usually need to be aggregated to show sales by geographic region and distance. Normally, the regional sales data should be collected into a matrix. One column is the distance of the region from the point of sale. The second and succeeding columns consist of sales variables, such as:

- Total sales revenue for that region
- Number of unit sales for each region
- Percentage of a retailer's sales that can be attributed to each

region

- A retailer's penetration rate in each region, if data are known for competing retailers

Once geographic data have been acquired, the first analytical step is to explore that data along the geographic dimension. Until recent years, this spatial analysis was prohibitively expensive for most applications. Proper spatial analysis is still more difficult, and the software and data requirements more extensive, than analysis that assumes away the geographic component. However, today it is possible and cost-effective in many applications. One focus of this book is to directly incorporate such analysis. In this section we make use of GIS software and data that have a spatial component.

We illustrate this with standard tools of retail sales analysis and one innovation of our own. For the following analyses, we have used data adapted from actual consumer purchases of luxury automobiles in a large metropolitan area of the U.S. These analyses involved a large dataset which included the location of competing retail outlets (automobile dealers), the sales of competing brands by zip code area, and the residence location of a subset of the customers.

7.2 Drive-Time Analysis

Most consumers in industrialized countries like the U.S. drive automobiles when shopping. Even in urban areas served by mass transit, and in countries with low vehicle ownership, buses are typically the backbone of the mass transit system. Thus, the cost in time to get to a retail outlet is best modeled by drive time – the time it takes to drive a vehicle to the location. Drive time is often the most important indicator of likelihood to shop at a certain retailer. Determining Drive Time: How does one determine drive time? The only method of doing so is to analyze specific geographic components of the area, notably the road network. If there are only one or two points to consider, this can often be done by hand, with a map and a knowledgeable local resident. By “points,” we typically mean the retail outlet. However, the term is sometimes used to include customers as well.

If there are more than a few points to consider, then a GIS

facility with adequate local data and a drive-time estimation routine is the best approach. Example Drive-Time Analysis: An example drive-time analysis is shown in Figure 7.1, “Drive-time analysis.” The irregular polygons around each point are a “drive-time area” for a specific time interval.

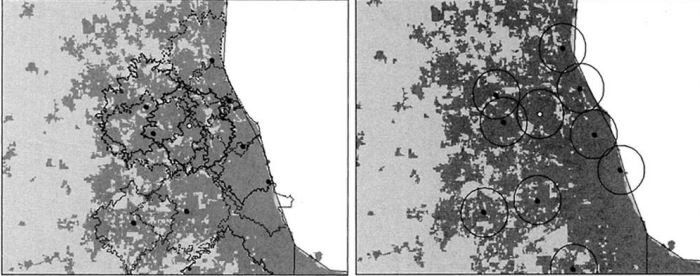


FIGURE 7.1 Equal-distance rings

A less sophisticated analysis, which can be done by hand on a paper map, is to draw concentric circles around a retail outlet (See Figure 7.1, “Equal-distance rings”). Using simple distance as a proxy for time introduces some distortion in the analysis for those customers that drive, walk, or take a bus using roads and streets. However, the traditional “as the crow flies” method is a rough guide, and is often the method used in legal contracts and state laws that govern franchised retailers. Many states have laws that allow franchised retailers in certain industries, such as automobile retailing, to protest the appointment of a same-line retailer within a certain distance from the location of a current retailer.

“As the crow flies” rings can be an acceptable method. In circumstances where distance accurately reflects the likelihood to shop, where there are no major geographic boundaries (such as rivers) in the area, no major political boundaries, or where the distances are either quite short or quite long. In fields other than retail sales, it has broad application.

When data on individual purchasers are available, a much richer understanding of the sales-distance pattern can be found by plotting the customer locations on a map. One approach is a “dot map” showing dots where customers reside. This illustrates quite

effectively the actual pattern of sales. See Figure 7.2, “Sales patterns and distance.” A second approach is to collect the number of purchasers in each surrounding area, using boundaries such as zip code areas, census tracts, or municipal boundaries. This method is often used when data deficiencies allow only a partial identification of customer residences. This thematic map shows by different shades the total number of purchasers in each area.

Retailers operate in a competitive environment. We expect that a profitable retailer will attract new entrants to a market and that customers will travel farther to find goods they cannot find nearby. This results in competition. Price is one dimension of that competition, and location is another.

One innovation we have used in modeling competition among retailers in the same region is the identification of “daylight” areas. We define daylight areas as areas in which consumers have one, and only one, convenient retailer.



FIGURE 7.2 Sales patterns and distance: alternative perspective

The method for producing a daylight map is:

1. Identify all competing retailers in a region.
2. Create a primary trade area for each retailer. This requires

analysis of the salesdistance pattern and other data, and is an important task for the analyst. These trade areas should normally be determined using the same parameters (such as drive time or distance) for each retailer. Plotting the penetration rate is an excellent visual tool for this purpose.

3. Show the trade areas for each retailer on a map. In many cases, these trade areas will overlap.

4. By darkening the areas that are in the primary trade areas of competing retailers, we reveal the daylight areas where an additional retailer could locate and serve a primary market area with limited competition.

5. The size of the daylight area and the number of consumers in that area provide a strong indication of the likely sales potential of the additional retailer.

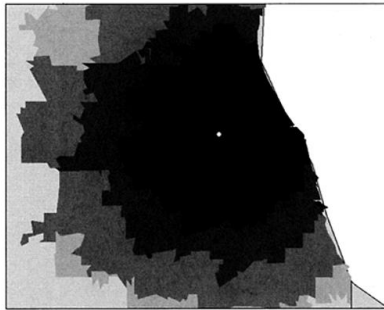


FIGURE 7.3 Market penetration rate

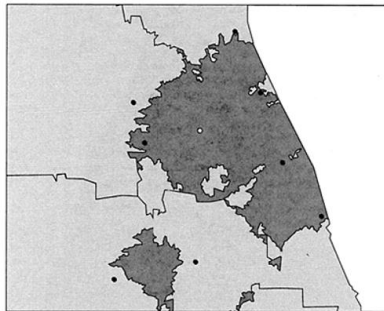


FIGURE 7.4 Daylight areas

The next step in analyzing retail sales data involves measuring the distances between the retail center and the customer locations. Ideally, drive time from each customer to each retailer would be used, but functions for easily obtaining this information for large number of locations are not yet developed in GIS facilities. In this section, we explain the best ways of determining distances. We break the processes out by the level of GIS capabilities needed to conduct the analyses and start with a discussion of the differences between individual data and data grouped by region.

Determining the distances is typically done in one of two ways, depending on the available data:

1. With full individual customer addresses available, you should measure the distance between the address and the retail location. This involves the use of geocoding software or an outside geocoding consultant.

2. With regional sales data only, the process involves finding the geographic centroid of the regions and measuring the distances between the centroids and the retail center. This, of course, loses some information. Therefore, the results will be less precise than if a larger number of actual consumer addresses were known. The simple centroid is, of course, not the most accurate point to use to represent buyers when sales in the region are distributed asymmetrically. (This is especially true when a centroid actually lies outside the region.) The center of concentration of consumers is the ideal measure. With many sufficiently small regions, however, errors in the distance measure due to the use of centroids may be assumed to be nonsystematic. The consequent analysis may be unbiased, or have negligible bias.

The better GIS facilities allow analysts to incorporate SQL (structured query language) into the assessment of distances. After geocoding the customer locations or regional centroids and determining the geographic coordinates of the retail center, SQL capabilities can be used to add distances to the database using a single query command. The actual implementation of this process varies between programs. However, those with knowledge of SQL and related GIS software should be able to adapt the commands to work with their system.³⁴⁰ This approach works for both regional sales and individual customer location data, resulting in a GIS

database that includes distance variables. These data can then be exported into an analytical program such as MATLAB or a spreadsheet program.

If you have access to a GIS or mapping software program that does not provide SQL capabilities, you are still able to gather the distance information. This involves using the ruler tool that is available within all GIS platforms to manually measure the distances between the point of sale and the customer locations. By plotting each of the sale locations on the map and, if using regional sale data, determining the centroid of each region you can obtain much of the information that can be provided with the more sophisticated analysis described above. As you manually measure each distance, record the information in the original database. Caution should be exercised when measuring distance. While some imprecision is inevitable, users should ensure that distance from each consumer is measured in the same manner. With this in mind, when only basic GIS is available, we recommend reporting simple distances rather than driving distances or drive times where greater errors are likely to occur.

Without any kind of GIS platform, the process of measuring distances will take longer. However, you can often still accomplish it using one of the following methods.

Internet mapping applications: Using mapping capabilities provided by popular Internet-based mapping engines, you can compute driving directions between two points. When the online program presents the quickest driving directions between two points, it provides a drive-time estimate. Using this tool to measure drive time between each customer location and the retail center, the analysis can proceed without the use of a GIS facility. However, this will require manually entering each of the sales into the mapping engine and individually processing each record. This can be extremely time consuming. Furthermore, the accuracy of free online systems cannot be expected to be at the same level as professional systems used by trained analysis. Paper maps: The most basic approach involves a paper map of the sales region and the use of a ruler. This process is similar to that employed in the use of basic GIS capabilities but involves manually plotting sales points on the map and manually measuring the distances. As our goal here is to increase

the efficiency and precision of your analysis, we recommend using this approach only if you are unable to complete the analysis using either a GIS program or Internet-based mapping products. Outside consultants or data vendors: If you are prepared to undertake all the analytical tasks and have the raw data on addresses, you can purchase the geocoding services from a number of outside consultants as well as data vendors. Keep in mind that this is a relatively new area of applied economics, and both the quality of the data and the knowledge of the consultants vary significantly. In particular, be certain that any outside consultant or data vendor can furnish you with the specific source of their base data and identify the results of any geocoding. The best data vendors will have data updated from a variety of sources on an annual basis and will be able to explicitly identify those sources. The best consultants will have some knowledge of the industry and the available data.

7.3 Applying a Gravity Model to the Distance-Sales Analysis

Equation (7.1) restates this model, where sales (S) equals a constant (β) multiplied by the inverse of distance (D) to a constant power (k).

$$S = \beta \cdot \frac{1}{D^k} \tag{7.1}$$

We describe below how to use past sales data to estimate the β and k constants. Once these constants are determined, the equation can describe the relationship between customer sales and distance from the retail location. We describe how to apply this equation to measure future sales, the impact of competition, or the success of a new location . The sales-distance relationship is fundamentally an implicit demand function. The one argument is a price variable proxied by distance. For this reason, one should consider the same admonitions and cautions that apply when estimating regular demand functions.

If the distance-sales relationship is an implicit demand equation, we should be able to evaluate the price elasticity of demand at least for the portion of the price that is transformed into distance. In a

well-specified model, this elasticity should illustrate rational consumer behavior. Below, we describe how to determine the elasticity of the gravity model equation. By using standard mathematical economics and by substituting a distance variable for the price variable, we define the elasticity of sales with respect to distance using the equation:

$$\epsilon = \frac{dS}{dD} \cdot \frac{D}{S} \quad (7.2)$$

where: elasticity of sales with respect to distance equals the derivative of sales with respect to distance, multiplied by distance (D) divided by sales (S). We apply this by differentiating Equation (7.1):

$$\epsilon = \frac{dS}{dD} \cdot \left(\frac{D}{S} \right) = -k \cdot \beta \cdot \frac{1}{D^{k+1}} \cdot \left(\frac{D}{S} \right) \quad (7.3)$$

Reducing the equation yields:

$$\epsilon = -k \cdot \beta \cdot \frac{1}{D^k} \cdot \frac{1}{S} \quad (7.4)$$

Substituting in this equation the original equation [Equation (7.4)] for sales (S) produces:

$$\epsilon = -k \cdot \beta \cdot \frac{1}{D^k} \cdot \frac{1}{\left(\beta \cdot \frac{1}{D^k} \right)} \quad (7.5)$$

We find that elasticity of demand with respect to price in a gravity model equals the negative k coefficient. The variable k has, in our empirical work, always been a positive number, meaning that $-k$ is a negative number. Desirable qualities of the gravity model in practice:

The following are the desirable characteristics of the gravity model:

1. The elasticity is, as expected by economic theory, negative.
2. The elasticity is constant. This should also be expected as there is no reason to expect that consumers living closer to a retail outlet would have a larger or smaller elasticity than those living farther away.³⁴⁴
3. The demand function (with respect to the distance portion of the price) can be largely parameterized by a single number, the k coefficient. If the retail price is assumed to be the same for residents of any area, then this captures quite well the variation in sales given the implicit change in price caused by varying the distance of the consumer's residence to the retail outlet. In certain cases, this assumption will not hold. An example of this is where state or national borders are within driving distance of the retail outlet, and tax rates, duties, or price schedules are significantly different in the two territories. In, say, duty-free shops, the demand will be much higher from consumers who cross a national boundary. We call this the decay factor as it indicates how rapidly sales "decay" as distance increases.
4. The parameter β captures the differences in scale and the differences in units in one constant. We call it the scale factor.

This equation thus has very desirable practical qualities and allows us to separate the portion of the relationship that is based on arbitrary factors (such as units and population) from the part of the relationship that is the distilled distance-sales relationship.

This analysis requires that the data be in one of the following two forms:

- Individual customer location data, in the form of a single-column list of distances, with each distance depicting one customer or sale.
- Regional sales data, in the form of a two-column matrix, with each row representing one sales region. One column depicts the distance of the sales region from the retail center and the second column lists the number of sales, penetration rate, or other relevant sales variables.

In order to analyze the data using the distance-sales model described in this chapter, the data can be compiled into nonoverlapping "bins" for analysis. In this case, a bin is defined by distance from the retail center and is associated with a sales figure

representing the sales at that distance. Because regional sales data should already be in this format (with one column showing distances and the other the number of sales at that distance), the process of classifying the data into bins is only necessary for individual customer location data. A bin's distance variable is determined by the average maximum and minimum distances of a bin. The sales number is an aggregate figure of all sales that fall within the maximum and minimum distances. For example, if your data set included sales at the distances of 1.0, 1.5, 2.3, 3.1, 5.5, 6.3, 7.5, 8.2, and 10.0 mi, dividing the data into three equidistant bins would result in a bin from 1.0 to 4.0mi, one from 4.0 to 7.0mi, and one from 7.0 to 10.0 mi. Four sales would be recorded in the first bin at an average distance of 2.5 mi; two sales would be recorded for the second bin at an average distance of 5.5 mi; and three sales would be recorded for the third bin at an average distance of 8.5 mi. The bins may not overlap. Although the analysis does not require that the bins are of equal distance in width, MATLAB and other software programs will often assume equaldistance bins by default.

To create bins using MATLAB, do the following:

1. Open the database of distance and sales data and import it into MATLAB. Using the procedure discussed above, you may open a dbf file in Excel and then using Excel link import the distances vector into MATLAB.

2. Once the data is in MATLAB, use the hist function to place the data in bins

```
[Sales , R ]=hist( A , BINS)
```

where: A is the distance vector, and BINS is the number of bins.

3. The resulting variable Sales is the number of sales per bin and R is the centroid of each bin (the average bin distance). Combine the sales and centroid information into a single matrix:

```
SalesMatrix=[ Sales, R ].
```

4. If there are bins with zero sales, you may wish to manually erase those bins (rows) from the matrix.

The preceding sections of this chapter demonstrated the theoretical reasons for using a social gravity model as the basis for modeling the sales-distance relationship. In most cases, the classic gravity model equation is the simplest and best choice. In other cases, a different specification may be selected that is better-suited

for the data. We describe the estimation of the parameters of such models in this section; we also discuss an alternative specification, a polynomial model.

We recommend that the following steps be done after the acquisition of the data:

1. Initial exploratory data analysis: A good approach to this and other problems is to first explore the data. Exploratory Data Analysis (EDA) is a fruitful source of analytical techniques for this purpose. We describe techniques of exploring the data in “Geographic Analysis of Sales”. Before attempting to specify and estimate a sales-distance model, explore the spatial data patterns. A second, simple EDA technique is to simply plot the distance-sales relationship on an X-Y chart, with dots indicating each (sale and distance) pair. With the data already analyzed in the manner described above, a histogram of the data, broken down by distance, would provide excellent summary information.

2. Identifying analog stores: A second step that should be taken in many cases is to look for existing retail outlets with similar characteristics and examine their sales-distance relationships. This is sometimes called the analog method after William Applebaum popularized it in the 1960, p. 346.

3. Primary model specification – gravity model: With the data prepared in bins as described above, specify and estimate the distancesales relationship. We provide examples below.

7.4 Monte Carlo Data on Retail Sales

We first provide an example of this technique using Monte Carlo data:

1. To create this test data and then estimate it, we used the `retail_test` function which is part of the Business Economics Toolbox. It first takes in assumptions about the underlying relationship and generates the random numbers to include in the data. Second, using the assumptions and the random numbers, it creates test data. Third, using the special functions we have developed for modeling retail sales, it estimates those parameters.

2. Figure 7.5, “Underlying sales-distance relationship: gravity model,” shows the test data generated to test our ability to properly

identify a gravity model. The underlying relationship is shown in the top pane. The data are based on this relationship, plus some random noise, and are shown in the bottom pane.

3. Using only the test data, we estimated the relationship as shown in Figure 7.6, “Estimated sales-distance relationship.” Note that the actual parameters were somewhat different from our estimated parameters, indicating that (as should be expected in all such cases) the random noise in the data obscured the precise relationship.

4. The command line output from the function is shown in the following code fragment.

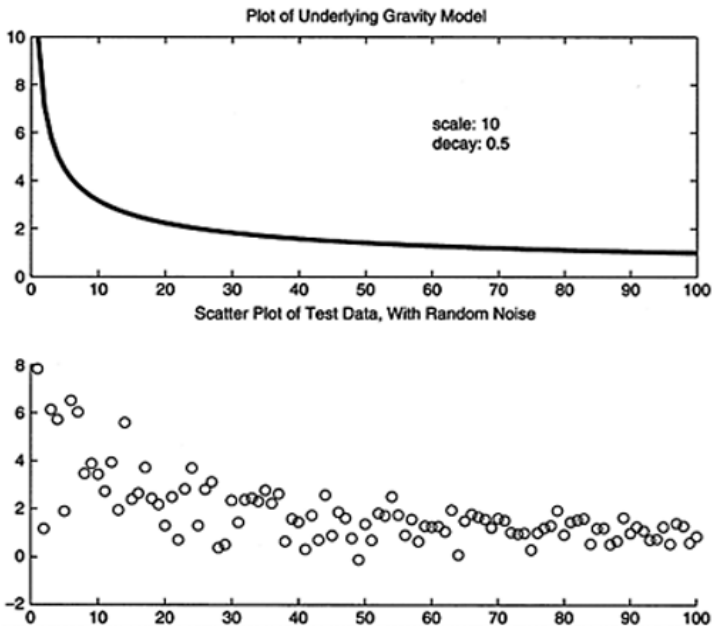


FIGURE 7.5 Underlying salesdistance relationship: gravity model

We modeled the distance-sales relationship for retail automobile purchasers in one of the largest metropolitan areas of the U.S. The results for two of these dealers, over multiple years, are shown in Figure 7.7, “Distance-Sales relationships for automobile dealers.”

This is a different geographic model, and a different segment, from the examples presented earlier. The segment in which these dealers competed was the low-price and mid-price imports. The data were collected by zip-code area, so there is some aggregation bias due to grouping all consumers within a zip-code area into a single distance class.

```
Coefficients Used in Gravity.m:  
coeff_mess =  
'Estimated Parameters'  
'Scale: 7.61'  
'Decay: 0.4051'  
'using gmeval.m function'
```

Code Fragment 6 Test Gravity Model Output

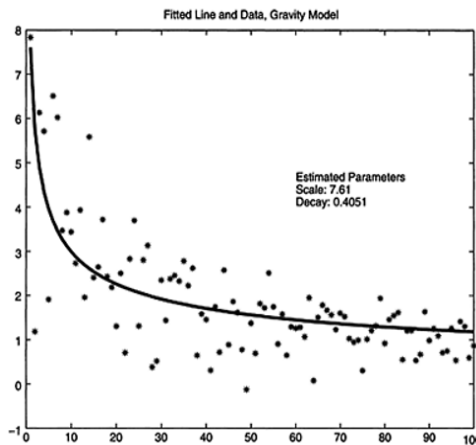


FIGURE 7.6 Estimated salesdistance relationship

Distance-Sales Relationships for Automobile Dealers We modeled the distance-sales relationship for retail automobile purchasers in one of the largest metropolitan areas of the U.S. The results for two of these dealers, over multiple years, are shown in Figure 7.7, “Distance-Sales relationships for automobile dealers.” Note that the relationship appears relatively stable over the years. Indeed, although total sales fluctuated during this time period, the

relationship between distance and sales did not. The decay and scale parameters estimated differed between the two dealers but were fairly consistent over time for each dealer.

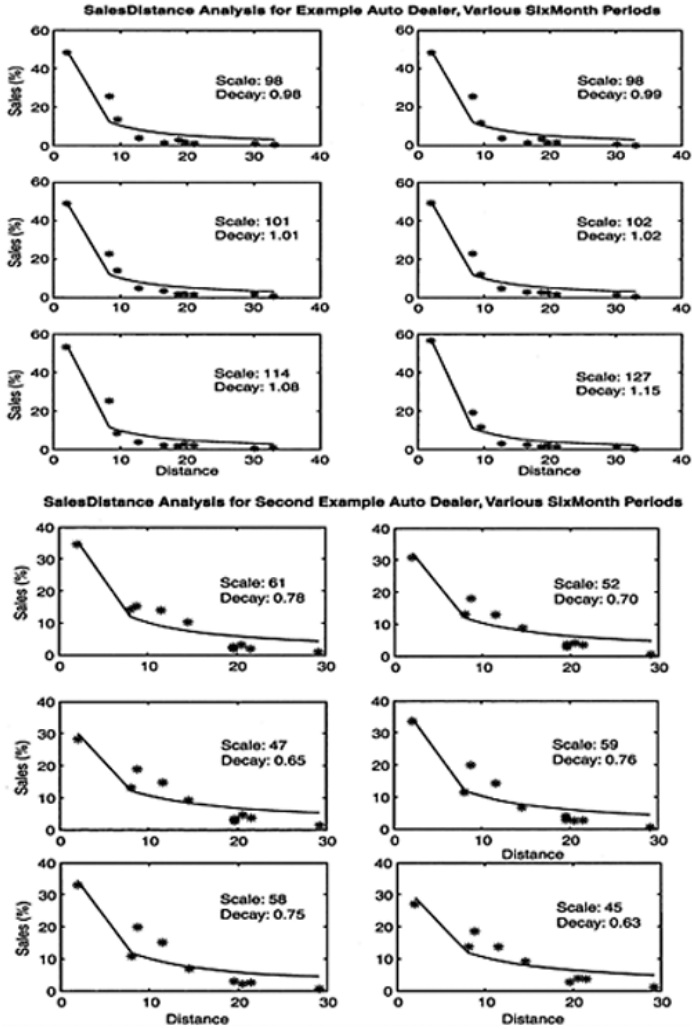


FIGURE 7.7 Distance-sales relationships for two automobile dealers

7.5 Nonlinear Optimization and Gravity Model Estimation

The specification and estimation of a gravity model requires sophisticated software. As the model is nonlinear, a simple OLS (ordinary least squares) regression will not estimate the parameters. In this case, a software environment like MATLAB is needed. In the examples presented above, we used the following methods:

1. We custom-programmed a gravity function, `gravity`, which implements the gravity model equation specified above. This function takes the data and coefficients and produces the distance-sales relationship. Note that this function will not estimate the parameters; it takes parameters as an argument and produces a distance-sales curve, given the appropriate data.

2. To estimate the parameters, we use a nonlinear optimization algorithm available within MATLAB. In particular, we make use of the `gravity.m` function and the nonlinear least squares estimation command `nlinfit`. Linear models can be directly estimated using a well-known formula that minimizes the sum of the squared errors. Nonlinear models typically require an initial guess at the correct parameters and then, iteratively, a search to find the parameters that minimize the sum of the squared residuals. The MATLAB `nlinfit` command uses the Gauss-Newton method to find these parameter estimates. Other optimization and numerical methods are also available within MATLAB. For typical sales-distance paired data, the standard Gauss-Newton algorithm (which has been modified with Levenberg-Marquardt routines to “learn” faster) in the `nlinfit` function has worked well in our applied work

3. To package the use of these multiple commands, we custom-programmed the `gmeval` function, which calls both the nonlinear least squares and gravity functions.

Best-Fit Polynomials

An alternative to the classic gravity model is a polynomial equation. Such an equation allows for the use of different factors and powers of those factors. Using such equations, one can often find a specification that offers the best fit for the data among a number of alternatives, including the gravity model. After the sales data are plotted by distance – done during the exploratory data analysis or

while examining a gravity model specification – recent versions of MATLAB make fitting polynomial equations of multiple degrees relatively simple, at least with equations involving only two variables. Once the data are plotted in MATLAB, a figure window appears showing the analysis. As with other references to specific software features, these instructions may vary from version to version. In particular, the curve-fitting features within the figure window were not available in versions before R13. If you are using an older version or a different software package, you will need to directly specify and estimate the polynomial model. Other mathematical and statistical software are capable of similar procedures. Within the Tools menu of this window, select Basic Fitting to open the basic fitting tool window. Using this tool, you can fit equations of different polynomials to the data very quickly. For example, if you select the quadratic equation box, MATLAB will fit an equation in the form:

$$y=p_1 \times x^2+p_2 \times x+p_3 \tag{7.6}$$

to the data; if you select the cubic equation box, MATLAB will fit an equation in the form:

$$y=P_1 \times x^3+p_2 \times x^2+p_3 \times x+p_4 \tag{7.7}$$

to the data (where p1, p2, p3, and p4 are constant numbers estimated by the toolbox, y is the sales variable, and x is the distance variable). If you use MATLAB to determine the best-fit equations for linear, quadratic, cubic, fourth degree, and possibly other degree polynomials, you can determine the equation that best estimates the actual sales based on distance data. To do this, return to your original data source and compare sales figures projected for each distance by the best-fit polynomials to the actual sales figures at those distances.

7.6 Gravity and Polynomial Models

Using the same method that produced the test data and gravity model described above, we generated a new set of Monte Carlo data and then estimated the sales-distance relationship using the gravity model. Although the gravity model worked well in this case, it (of course) did not explain all the variations in the data. Could a polynomial model work better? Figure 7.8, “Gravity vs. polynomial specifications,” shows the result of this exercise. Using the MATLAB basic curve-fitting tool, we were able to quickly overlay a cubic (third degree) polynomial equation onto the data. Eyeballing the data indicates that the cubic polynomial does a better job of explaining the variations for almost all of the ranges. The little hump upward and the tail downward in the cubic curve bring it closer to the distribution of the test data.

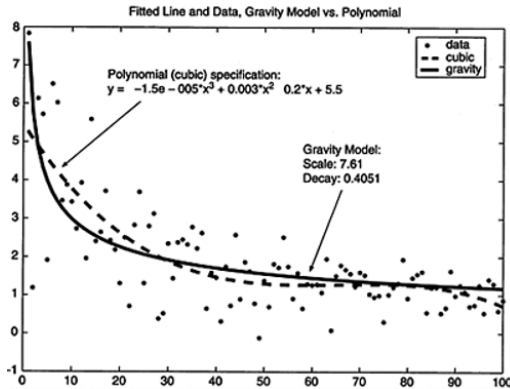


FIGURE 7.8 Gravity vs. polynomial specifications

This raises a fundamental question about using statistical methods to estimate a behavioral relationship. Is the objective of such an exercise to get the best fit, or to get the best model? Going back to our first maxim, we clearly believe it is more important to understand the market and to properly model the underlying behavior than to get the best fit with any single data set. Getting the model right means understanding not just one situation, but an entire market. Furthermore, keep in mind the well-known result from econometrics:

the goodness of fit of any equation can be improved simply by adding more explanatory variables, even if these variables are meaningless. To be precise, the goodness of fit measured by the R^2 statistic (the ratio of the explained variation to the total variation) will go up when an explanatory variable is added. For this very reason, adjusted R^2 and other measures are used.

Weaknesses of Polynomial Equations

The main attraction for using polynomials to model the distance-sales relationship is not theoretical support; neither is it consistent empirical validation. The attractions boil down to two: (1) better fits for the data at hand and (2) the software available today easily lets you try several specifications. Of course, good fit is a good thing, so why shouldn't we abandon a rigorously derived, empirically proven model that works well for one that "fits better"? There are problems with using polynomial equations to model sales-distance relationships. Among them are:

1. Weak theoretical support: There is little reason to believe, a priori, that consumers think about polynomial combinations of distances when they decide where to shop. Remember that most consumers – and most sales clerks – use calculators to determine a simple percentage sales tax on their purchases. Will the same consumers attempt to weigh a polynomial combination of distance?

2. Allowance of negative sales: Consumers cannot purchase less than zero items. However, polynomial specifications can easily produce negative predicted sales for certain distances. This problem could be ameliorated by using a method that constrains the result to nonnegative territory, but the use of such constraints then accentuates the other problems we identify here.

3. Non-monotone behavior: Consumers would not know what monotone behavior is, but they all practice it. Monotone functions have an output that constantly increases (or decreases) as the inputs increase or decrease. If consumers purchase less when they get farther away, they will likely purchase even less when still farther away. A polynomial specification – even one with a "best fit" – often results in predicting inexplicable behavior. The higher the degree of polynomial, the more the curves appear in the plot of the resulting function. This not only means that more data are necessary to

properly estimate the additional parameters but also that the estimated equation is likely to have unusual and hard-to-explain detours from a smooth path. If those detours fit the data better, then it may appear that the equation does a better job of explaining the data – but does it really explain the underlying market behavior?

4. Unclear elasticity: The gravity model presented in Equation (4) assumes a constant elasticity of sales with respect to distance. Therefore, the model assumes that the percentage change in sales, given a percentage change in distance, remains constant for all consumers and at all distances. The assumption of a constant elasticity is one which is made frequently in demand and supply behavior, and it has been thoroughly studied. Though not correct in the extreme, it has broad application in the range of circumstances faced by both buyers and sellers. On the other hand, a high-degree polynomial will often have a complicated elasticity which will vary significantly. This makes reviewing the results of an analysis difficult and makes it nearly impossible to compare demand behavior from one location to the next.

5. Intractable structural form: The structural form of a gravity model allows for simplified output and relatively simple incorporation into other economic analyses. A polynomial form, especially a third degree or higher, could be intractable.

We believe that the strong theoretical and empirical grounding of the gravity model supports its position as the primary tool of distance-sales analysis. However, we present polynomial specifications and acknowledge that within limited data ranges they can, in some cases, work better. Therefore, all the same, before you use a polynomial equation, especially a cubic specification or higher, be sure you know the reason why and have considered the weaknesses described above. A more comprehensive model of detail attraction would make use of the information in a gravity model and consider other attributes of store attractiveness. For retail markets in which all the stores are roughly the same in size and amenities, it may be possible to ignore these other variables. However, when store attributes differ significantly (such as one offering more product choices, more services, staying open later, having more expensive goods, or other factors), the location factor may still be the most important factor affecting sales, but it will not be the only factor.

Franchised retailers, who are normally contracted to meet certain uniform standards, are a subset of retailers where this assumption is more tenable. However, even in these cases, you will see variations due to the performance of the store personnel, management practices, and other individual factors.

In such cases, if data are available for sales from multiple stores, and data are also available for the factors that distinguish the retail outlets, a general predictive model can be used. Such a model can take the functional form:

$$S_i = f(d, m, y) \tag{7.8}$$

where: S_i – denotes sales of store i in the region; d is a vector of distance variables, indicating the distance between store i and other stores; m is a vector of amenity variables such as size, hours, quality or price of merchandise, and atmosphere of the area surrounding the retail outlet; and y is a vector of data on the disposable income available to consumers in the area. We include the income variable because the data may come from retail outlets from different areas. With cross-sectional data on stores of a certain type, this equation could be specified and estimated in any number of ways.

One specification for retail demand was created by David Huff in the 1960s. As will be shown below, it is very similar to a gravity model in its treatment of distance, and explicitly incorporates an additional variable. The additional variable is an important amenity: the size of the retail outlet. The Huff specification defines the attractiveness a to consumer i of store j as:

$$a_{ij} = \frac{m_j^a}{d_{ij}^b} \tag{7.9}$$

where: m_j^a is the size of store j , the exponent a is a parameter; d_{ij} is the distance between consumer i and store j , and the exponent b is a parameter indicating the sensitivity of the consumer to distance.

The model as defined in the equation has an unobservable variable (“attractiveness”) equal to a formula with two unknown parameters. It is a model of individual choice and, therefore, could not be used with aggregate data. Thus, some transformation is

necessary. To make it tractable, we first substitute the probability of purchasing goods from a certain store (which can be estimated from actual sales) for the unknowable term “attractiveness.” The equation still involves a large number of parameters. In practice, the parameters a and b are often taken on the basis of past experience to be 1 and 2, respectively. With these series of simplifying assumptions, the equation becomes:

$$s_{ij} = \frac{m_j}{d_{ij}^2} \tag{7.10}$$

We have seen this equation before. It is the gravity model equation of Equation (7.9) and Equation (7.10), with the exponent on distance (the decay parameter) set to 2 and the scale parameter set to the size of the store. Thus, the practical Huff equation is a special case of a gravity model. Indeed, using the procedure described above, in which we estimate the decay parameter directly from the data, the scale parameter is an observed indication of the amenity vector of the particular retailer. In cases where size is known to be an important variable, it can be included in the specification as suggested in the general model discussion above.

CHAPTER 8. DETERMINING SALES AND MARKET AREAS

Assuming that we have data on sales, and a valid model for sales, we can sometimes estimate the parameters of the model using the data. The resulting model can be used to project sales in other markets, or the same market when the variables change. Before estimating sales, ensure that the subject sales location compares to the sales locations on which the data are based, in terms of market and product. It is pointless to project sales for a grocery store based on data from a car dealer, or for a high-end restaurant based on data from a fast-food establishment or even a mid-range restaurant. Below, we describe the steps for estimating total sales using the general gravity model, assuming it has been estimated using available data. The same methodology can be used to project total sales for any of the best-fit polynomials as well. We restate the standard gravity model equation where sales (S) equals a constant (β) multiplied by the inverse of distance (D) to a constant power (k):

$$S = \beta \cdot \frac{1}{D^k} \quad (8.1)$$

To calculate total sales, we can integrate the equation across a specified distance range:

$$\text{Sales} = \int_a^b \left(\beta \cdot \frac{1}{D^k} \right) dD \quad (8.2)$$

where: b – represents the maximum sales distance and a represents the minimum sales distance. We complete the calculations below:

$$\begin{aligned} &= \beta \cdot \left(\frac{D^{-k+1}}{-k+1} \right) \Bigg|_a^b \\ &= \beta \cdot \left[\frac{b^{-k+1}}{-k+1} - \frac{a^{-k+1}}{-k+1} \right] \end{aligned} \quad (8.3)$$

We can then replace b with the maximum sales distance and a with the minimum sales distance from our analysis, and β and k with the estimate of the values of these parameters. We can then use this

equation to project total sales across a specified region. A numerical method can also be used to project sales, especially if the distance-sales equation is different from Equation 8.3.

Edge Effects and the Impracticality of Zero and Infinity The analysis of spatial data is often hampered by edge effects which are discontinuous, or unusual behavior near the edge of a region. For distance-sales analysis, we often encounter edge effects requiring some adjustment very close to the retailer and far away from the retailer.

Basic Adjustment

The derivation of the distance-sales relationship is based on a demand function that implicitly assumes a maximum sales distance of infinity (∞) and a minimum sales distance of zero. In practice, we recommend that you truncate the data to a maximum and a minimum distance. There are three reasons for this: truncated area will almost always be in line with the actual characteristics of the market area,³⁶⁰ the mathematics of estimating models is much more tractable if we slightly restrict the range, and the data are often aggregated or otherwise deficient. We consider the rationale for the minimum- and maximum-distance truncations separately.

Minimum-Distance Truncation

If there are very few residential properties within a mile or more of the retail establishment, a minimum retail distance of zero may result in an overestimate of the total sales of a store with a primarily residential customer base, or allow a handful of observations to bias the estimated parameters of the model. Recall our derivation of the distance-sales relationship from the fundamental microeconomics of consumer demand. Distance in the demand equation is an indicator of time and other costs of traveling to and from the retailer. Part of that time cost is the same whether the trip is across the street or across the country. This de minimis cost is not important when the distances involved are all significant. However, ignoring it can cause problems when a few consumers are within 1mi and others are within 5mi, while the majority is 10mi or more away. The first group is indeed closer than the other groups, but does it really take one fifth the time-cost for a person 1mi away to purchase something at a

retailer compared to that for a person 5 mi away? This is further complicated by the imprecision of data. Often the analyst will have aggregate data sorted, for example, by zip-code area. In metropolitan areas, the centroid of the zip-code region may be a good indicator for zip-code areas whose boundaries are more than 5mi away. However, the centroids may not be a good indicator when the retailer is located in the zip-code area or near its boundary. Finally, the mathematical specifications for distance-sales relationships do not work well for distances close to zero.

Maximum-Distance Truncation

The rationale for a maximum-distance truncation is quite apparent. Any good model, including a polynomial or gravity model, should estimate sales well within the range for which data are available. However, such models will not be accurate at distances where there have been relatively few sales in the past. When only a handful of sales arises from very distant customers, these could be considered outliers that should not be given equal weight with other observations. In addition, as discussed in “Weaknesses of Polynomial Equations”, a polynomial specification can produce negative sales projections outside its intended range. With a gravity model specification and reasonably well-behaved data, the expected sales approaches zero as the distance extends outside the range of data. Hence, such models are less susceptible to distorted results because of a handful of large-distance sales observations.

8.1 Defining Market Areas

A quantitative definition of the relationship between sales and distance allows us to define trade areas based on the percentage of sales that we expect the retail center to attract from different geographic areas. Depending on the purpose of the analysis, we may decide to use a single trade area or multiple trade-area segments (e.g., primary, secondary, and tertiary market areas). If we decide to define a single trade area for a store, we may decide that this trade area should represent the region from which 70% of customers are attracted. To determine the geographic extent of this trade area, we use the above methodology to calculate total sales and then adopt it

to determine from within which distance 70% of customers are attracted. We continue to use the standard gravity model as an example, although the same approach will work for other equations.

$$(\text{Sales} \cdot 0.70) = \beta \cdot \left[\frac{b^{-k+1}}{-k+1} - \frac{a^{-k+1}}{-k+1} \right] \quad (8.4)$$

Equation (8.4) assumes that Sales, β , k , and a (the sales-distance minimum limit) are already determined or estimated. If we solve for b , we then determine the upper limit of the distance range that represents 70% of total sales. Our market area would then be defined by a distance ring of radius b (or drive time b , if distance D was expressed in drive time). If we want to define multiple trade area segments for a retail location, we follow the same approach. For example, we may decide to define a primary market area that accounts for 50% of sales, as well as a secondary market area that accounts for another 30% of sales. To do this, we would define the distance-radius of the primary market area based on 50% of sales and then the distance-radius of the secondary market area based on 80% of sales (50 plus 30%). Being able to incorporate these trade area definitions into a GIS facility is very useful for communicating the results of the analysis and for fully understanding the reach of the market area.

Cannibalization refers to the loss of existing sales which occurs due to the introduction of a new product or location. For example, if BMW introduces a new automobile, some new car sales will likely be directed away from the purchase of competitive brands, and other new car sales will likely be redirected from the purchase of other BMWs. In this latter case, BMW “cannibalizes” some of its sales from its other models. Cannibalization can also happen with the introduction of a new retail location. This is the type of sales cannibalization that we discuss in this chapter. If BMW opens a new dealership, the company will probably pull some sales away from competitive dealerships in the area. However, the new dealership will also pull sales away from other BMW dealerships. Using the sales-distance models discussed in this chapter, the amount of cannibalization can be estimated. In industries that rely on

franchising or manufacturer-owned-and-operated stores, the introduction of a new retail outlet into a market already served by existing retailers selling the same brands generally has two effects:

1. Overall brand visibility and customer convenience are increased.

2. Some existing customers change their patronage from the previous retail location to the new one. The first effect supports higher overall sales. The second cannibalizes existing sales in that the “new” sales are really sales taken from an existing retailer of the same brand. Intelligent manufacturers are concerned about cannibalization and avoid it when possible. However, given the divergent interests between the franchisor and franchisee and the trade-off between increased brand visibility and cannibalization, there are many decisions that rely heavily on an estimate of the amount of truly new sales in an area, compared with the amount of existing sales that are merely cannibalized. The techniques in this chapter provide a method to estimate these.

Because such assessments deal with shifts in sales between retail centers, as well as changes in total sales, the analyses require both total-sales and penetration-rate data. Typically, due to the requirements of the analysis, regional sales data are used instead of individual customer data. The quality and availability of data at this level differs significantly. In some industries and markets, data are available on the sales of each retailer by region. This occurs more often for major purchase items such as automobiles and in heavily regulated industries such as telecommunications. In such cases, both total sales and penetration rate data can be derived for each region. Many times, however, some data must be derived using the selected model. This approach assumes that all retailers in the analysis face the same market behavior. Therefore, the analysis is more accurate if the different retail locations compare closely in terms of product selection, accessibility, and other factors that attract customers. To derive the required data by region, use the selected model to project sales for each retailer by geographic region. The sales figures of the different retailers can be aggregated and adjusted to provide total sales by region, and then the retailer penetration rates can be estimated. We describe methods for these tasks in the following sections.

Brand visibility and customer convenience correlate with the proximity of a potential consumer to the point of sale. If a new retail location is closer to a consumer than existing retail locations, an increase in overall brand visibility and convenience for that consumer will probably increase total sales of the product in the area. We suggest the following steps to determine changes in total product sales in a region when a new competitor enters the market:

1. For each region in the analysis, measure the distance between the region's centroid and the closest point of sale before the introduction of the new retail center.

2. For each region in the analysis, measure the distance between the region's centroid and the closest point of sale after the introduction of the new retail center. The distance will change only if the new retail center becomes the closest point of sale to the region.

3. Determine the percentage change in distance after the introduction of the new retail center. Equation (8.5) calculates the percentage change in distance (PCD). D_1 is the distance to the closest point of sale before the introduction of the new retail center and D_2 is the distance to the closest point of sale after the introduction of the new retail center:

$$PCD = \frac{D_2 - D_1}{D_1} \quad (8.5)$$

4. Determine the percentage increase in sales for each region which results from the introduction of the new retail center. Equation (8.3) estimates the percentage change in sales (PCS) based on the product of the percentage change in distance (PCD) and the elasticity of sales with respect to distance (ϵ). Equation (8.4) defines elasticity as in "Elasticity of Demand in a Gravity Model":

$$PCS = PCD \cdot \epsilon$$

$$\epsilon = \frac{dS}{dD} \cdot \frac{D}{S} \quad (8.6)$$

Elasticity differs between models. In some models, it fluctuates

with different quantities or at different distances. However, the standard gravity model has an elasticity that remains constant at all levels of sale and distance. Because of this, using the standard gravity model can simplify the analysis. Equation (8.4) shows how to determine the percentage change in sales for a gravity model analysis, using the constant variable $-k$ as elasticity:

$$PCS = PCD \cdot (-k) \quad (8.7)$$

5. Calculate total change in sales for each region that results from the introduction of a new retail center. To do this, apply the percent shift in sales to the total sales figures for each region:

$$\begin{aligned} \text{NewSales} &= PCS \cdot \text{TotalSales} \\ \text{NewTotalSales} &= \text{NewSales} + \text{TotalSales} \end{aligned} \quad (8.8)$$

After calculating the increase in total sales by region, we can sum the regional figures to estimate the total new product sales that are created as a result of increased exposure and convenience.

The following describes how to estimate for each retailer in the market the changes in sales which result from the introduction of a new facility. This approach weights the future likely cannibalization of existing retail locations, based on their existing level of sales in a region.

1. Using the methodology described above, collect data for each region on (1) the sales by retailer before the introduction of the new point of sale.

2. Estimate sales by region for the new retail center using the gravity model or best-fit polynomial chosen for the analysis.

3. Estimate the unadjusted sum of all sales by adding together the estimated or actual sales information for all retailers in a region, including the estimated sales figures for the new retail center. Note that new total sales for a region are not equal to the sum of the new retail center sales and the total sales for a region before the new retail center entered the market. This is because some of the sales that the new retail center attracts will be cannibalized from the sales at other locations.

4. Determine the penetration rate for each point of sale after the introduction of the new retail center. To do this, divide the actual or projected sales of each retailer by the unadjusted sum of all sales determined in the previous step. Equation (8.5) calculates the penetration rate for a retail center (PR1) by dividing its estimated retailer sales (RS1) by the unadjusted sum of all sales (SumSales). The retail center's total sales are determined using the selected model or by actual data:

$$PR_1 = \frac{RS_1}{SumSales} \quad (8.9)$$

5. Estimate sales for each retailer after the introduction of the new retail location. Although we use the sum of all sales figures to calculate the penetration rate, the estimated sales for each retailer (NewRS1) is determined by multiplying the penetration rate (PR1) by the total sales after the introduction of the new retail center (NewTotalSales).

$$NewRS1 = PR1 \cdot NewTotalSales \quad (8.10)$$

6. Determine the changes in retailer sales by region by subtracting retailer sales prior to the introduction of the new sales location from the retailer sales after the introduction of the new location.

CHAPTER 9. APPLICATIONS FOR MANUFACTURING

Over the past three decades, the use of statistical quality control or “statistical process control” in manufacturing has increased dramatically. With the growth of international trade, objective measurements of quality – and the processes that ensure that quality is a central objective of production – have become particularly important. This book deals with the economics of business, and an essential part of that economics is the ability to consistently produce high-quality products. Statistical methods of testing for quality are growing in importance, and a basic knowledge of this field will be helpful to many readers of this book. In this chapter, we present first the classic statistical process control methods, and then introduce the newer Taguchi methods.

The core statistical applications used in statistical process control (SPC) are the same as those used in many other applications of statistics, including those in business economics and finance. MATLAB has a set of routines that accomplishes the basic calculations used in most statistical process control applications. This chapter briefly covers such SPC applications.

The purpose of this chapter is to present SPC to an audience that is intensely interested in the future earnings of a firm, but may be lacking in knowledge about an important tool used to ensure those earnings. As the tool relies on the same fundamental economic skills used in other chapters of this book, such knowledge could be rapidly acquired. While this chapter presents basic SPC routines, it is not a thorough reference to SPC. A practitioner will want to review other references. While the routines shown here are relatively straightforward, we suggest two innovations to those working in manufacturing:

- First, consider bringing the capability of customized SPC routines to facilities in different locations using a Web-server approach. Using a centralized facility to perform the complex tasks and communicating the relatively small amount of information among the facilities using the Internet are basic principles of “thin-client” computing. As the data involved in most SPC routines are small in size (in terms of bandwidth), it could be communicated with little difficulty.

- Second, consider customizing the routines for specific production processes, incorporating past data, recording new data, checking data before analysis, and other improvements. Such customization brings benefits that can outweigh the costs of using a programming environment such as MATLAB instead of a packaged SPC software product.

Much of the development of SPC originated in the work of Walter A. Shewhart at Bell Labs in the 1920s. Shewhart developed simple routines that—without the extensive calculations made possible by calculators and computers that would not be available for another half-century – provided manufacturing engineers an excellent tool to monitor the consistency of production. He called this approach economic control of the quality of manufactured products. It is a fitting testament to the man that the same techniques are in use today.

What is a Process?

A process is any series of machines, people, facilities, logistics, or other activities that take inputs and produce a certain product. We say a process is “in control” when it consistently produces the product within the predicted range of variation. This definition implies that SPC requires disciplines to define the product, measure it, and define the process that produces the product.

The term “quality” is somewhat amorphous in business. However, the SPC discipline requires firm operational definitions of product quality. An operational definition of quality generally requires one or more specific measurements or other operational checks. If “quality” cannot be defined and measured quantitatively, then the SPC routines cannot be used.

There are two principal uses of the classic SPC routines: to recognize when a process begins to go out of control, and to aid the adjustment of a process. These two purposes fit SPC neatly into the Plan-Do-Check-Act management paradigm, which was established by the American quality expert Edward Deming and popularized first in Japan after World War II. SPC is not, in and of itself, a management paradigm. However, organizations that value product quality highly can adopt an SPC protocol that fits neatly into their management paradigm.

The principles of SPC are more important than the actual statistics. For our discussion assume that a business is following those principles; meaning that there is an operational definition of quality based on actual measurements; that we have identified the process involved; and that we have taken measurements of the products. Using these measurement data, we can produce the classic Shewhart charts. Control Charts-Xbar: Using measurements from samples taken at regular intervals, calculate the mean measurement from each sample. Plot these over time. Given past data on these measurements, we know what the mean and standard deviation should be. Control lines on the plot show the UCL (upper control line) and LCL (lower control line), indicating the expected variation in this measurement. In some instances, a specification line is also plotted, indicating the range beyond which the measurement is beyond specifications. “Out of Spec” does not mean “out of control.” Control is based on predictable variation. A process that is “in control” will occasionally produce, due to statistical variation, outliers. Furthermore, specifications may be set by those unfamiliar with the variation in manufacturing processes, or may be substantially looser or tighter than the predicted variation.

Applications an example, generated by the simple program below is Figure 9.1, “Xbar chart.”

Control Charts-S Charts: An S chart plots the standard deviation of the samples taken at regular intervals. Such a chart visually indicates when a process starts to exhibit increasing variation and, thereby, threatens to go out of control. Using this chart in conjunction with the Xbar chart provides a well-balanced view of the quality of the manufacturing output. “S chart” Capability Chart: A capability analysis uses information on variation in past measurements together with proposed specifications and statistical distributions. With this information, assuming the process remains in control, we can predict the share of parts that will be outside the specification. These figures, produced on a workstation, provide a supervisor on the assembly line strong information about the quality of the parts being produced.

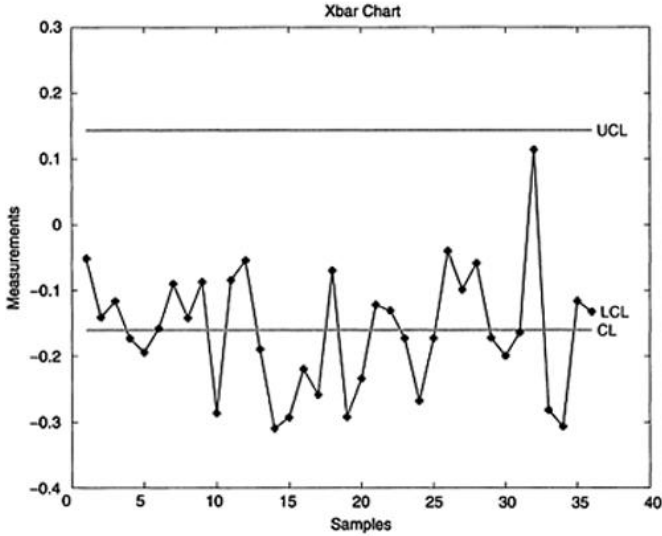


FIGURE 9.1 Xbar chart

The warnings against “garbage-in, garbage-out” are especially important when using statistical programs in which the data may be gathered, input, and analyzed by different individuals. Even when used by trained individuals, SPC routines can provide misleading information when the process analyzed is different from one that matches the underlying assumptions. For example, the capability plots rely on an assumption that the underlying sample will be random and distribution of measurements normal. If the measurement protocol does not ensure random samples, the first assumption is violated. If the process itself does not produce a normal distribution of results, then the second assumption is violated. If such data is input, the routines will produce neat, precise statistics and pretty figures – showing the wrong information.

A “Thin-Client” Approach

Consider now the first innovation we suggest: bringing the power of customized SPC routines to various locations. You could do this by sending an engineer to each location. You could also do this by having someone record data and transmit it back via telephone, mail,

or e-mail. Either approach invites delay and expense. It also creates the likelihood that the persons recording the data, analyzing the data, and receiving the analysis are all different and fail to communicate with each other about relevant factors. However, using a thin-client approach, the supervisor running the production process can receive immediate analysis of the SPC data. This achieves the central goal of SPC – enabling those running a production process to catch small changes in manufacturing output before they become quality problems. It also allows for informal or formal discussions among those who take the measurements and those who analyze the data. Indeed, it is possible that these would be the same people.

Creating a thin-client SPC program could be accomplished in the following steps:

1. Create and test SPC routines appropriate for your manufacturing facility.
2. Create and test appropriate test procedures.
3. Port the routines to a Web server running MATLAB and test the application.
4. Create a data exchange protocol for process engineers to input the data collected on the assembly line into the SPC routine. This could be simple HTML pages or a more complex XML implementation.
5. Review the graphical output and information needed, and incorporate those features into the Web pages that report the results of the analysis. Graphical output could be produced in the standard JPEG or GIF format, or in the new SVG format.

If a straightforward SPC analysis was all that was desired, specific SPC packages could probably accomplish the same goal with less expense. However, in other cases, the advantages of using MATLAB would outweigh the additional cost. You could customize this application to fit specific production processes and goals, using any of the following methods:

1. Selecting specific SPC routines, based on the control process selected for the specific production process.
2. Customizing the graphical output, for example, by adding additional text boxes, subplotting multiple plots, or providing more informative titles and annotations.
3. Specifying the tolerances allowed, sample sizes, and other

standards.

4. Catching poor data or likely measurement errors.

5. Recording the data or tests either by an automatic diary entry or through a user selected routine.

6. Providing historical data in graphical form.

7. Creating a custom graphical user interface.

The classic approach described above is the foundation of much of statistical process control. It is simple, robust, and requires relatively little advanced statistical analysis. The fact that these methods were largely developed a half-century ago, before computers could assist with the calculations, reveals the importance of the contributions of Shewhart and others. A major advance beyond the classic methods was developed by the Japanese engineer Genichi Taguchi in the 1980's.

The methods that resulted are known as "Taguchi methods," and are based on both statistics and critical insights into the role quality plays in the profitability of an enterprise. Because Taguchi methods look at the entire enterprise, rather than at the statistics of a production process, it can be considered from a management perspective that is accessible to a business economist. From this perspective, we summarize the key insights as follows:

- Customers become increasingly dissatisfied with products as they vary more frequently from the standard they expect.

- Variations arise not just from the production processes themselves, but also from other areas of the environment in which a product is produced, including human and managerial factors. Taguchi called these "noise factors."

- The variation in a process creates costs for the organization. These costs include not only the costs of returned or defective products, but also the additional managerial, production, and technical resources that must be devoted to the problem. These costs increase with variation.

- The cost of these variations can be expressed as a loss function. Such a loss function is a much richer formulation of the actual costs of imperfect processes than the classic approach, which relies largely on a discrete step function in which a process is either in control or out of control, and a product is either within specifications or outside specifications.

- In many situations, the cost of the variation (the loss function) can be modeled as a constant multiplied by the square of the deviations from the desired objective.

- This “quadratic loss” function allows an organization to estimate the value of reducing variation in a process and to make appropriate management decisions. The economic reasoning behind Taguchi methods is, in our opinion, more important than the statistical innovations.

We view Taguchi’s crowning insight as the treatment of “quality management” decisions not as solely engineering decisions, but as business management decisions. Such decisions are tractable from a classic microeconomic perspective and have implications for the profitability of the firm, the value of brands and intangible assets associated with a product, and the value of an enterprise.

CHAPTER 10. FUZZY LOGIC BUSINESS APPLICATIONS

Human beings invented logic and use it every day. However, most important decisions rely on something more than the type of rigorous logic taught in math and rhetoric classes. We are required to make decisions in a world full of partial information and imperfect data. This requires a logic that takes into account information expressed by imperfect human beings concerning events about which they have less than complete knowledge. The basics of textbook logic courses are “Boolean” logic, based on data that can take two forms: “yes” and “no.” Such logic is essential in many situations. “If the traffic light is red, stop” is an excellent rule to follow when driving. Boolean logic does quite well in such situations. However, what happens when a traffic light is green or yellow? A green light quite far in the distance may provide a trained driver with enough information to know that he should slow down, as the light is likely to change before he gets there. Another driver, facing a yellow light, may wish to consider the speed, the distance to the light, the traffic conditions, and the weather, before deciding whether to stop or go.

Decisions of this form cannot be made easily with Boolean logic. An extension of Boolean logic, known commonly as “fuzzy logic,” has been developed to deal with these situations. Such logic includes the yes/no situations for which Boolean logic is quite adequate, and extends this thinking into situations where the data are not confined to neat, yes/no information. I believe the term “fuzzy” is unfortunate, as it implies that there is something unrigorous or even illogical about such logic. In fact, there is nothing fuzzy about fuzzy logic. It requires very careful, rigorous, and logical thinking to implement correctly. When done properly, it gives exactly the same answer as Boolean logic in those situations where the required yes/no answers are available, and also gives sound answers to the many situations where the data cannot fit into yes/no categories. Many business situations boil down to probability statements rather than yes/no answers. The stock market, for example, is a market filled with participants making risk-adjusted discount decisions with imperfect information. Most decision rules of investors are best modeled as probability statements. For example, an investor will first evaluate his or her own risk profile (a probability exercise), then choose

markets that reflect that risk profile (a second probability exercise), and then pick individual securities or mutual funds within those markets (a third probability exercise). Even investors who are convinced that they cannot beat the market are making probability-based decisions when they simply hand over money to a portfolio advisor or invest in a large mutual fund.

Insurance underwriting is obviously an evaluation of probabilities. So is investing in real estate, choosing which tickets to purchase in a theatre or baseball season, choosing a restaurant, voting in an election, choosing to pursue a new business venture, closing down a money-losing subsidiary, and deciding to keep fishing when the sun goes down and the mosquitoes come out. Fuzzy logic uses rules based on probabilistic thinking for those occasions when yes/no data is not available, not complete, or represents only a fraction of the available information about an important problem.

For example, when selecting investments, one person may wish to concentrate on companies with a good record of consistent earnings. What, then, is the definition of “good?” How one could use fuzzy logic to evaluate this question, and is such a method superior to Boolean logic?

Steps in Boolean Logic Approach.

A rigorous examination of this question would involve the following steps:

1. Examine the earnings of a universe of stocks.
2. Calculate mean, standard deviation, and other descriptive statistics.
3. Using statistical inference, describe the underlying distribution of earnings based on the information you have. You might describe more than one distribution, such as earnings over the past 1- and 5-year horizons.
4. Make an assumption about whether past performance will predict future performance, with or without bias. This critical step is often overlooked or buried in a legal disclaimer. By predicting with bias, we mean that past performance will systematically understate or overstate future performance. For broad, well-established parts of investment markets, most analysts commonly assume that past performance is an unbiased predictor of future performance. It is not clear, particularly when the selection of the past period is a judgment

call, that this is the case.

5. Select, using your own preferences, the area of the distribution of returns that you consider “good.”

6. Define “good” returns with a metric that uses available data. Such a practice is standard in financial portfolio theory. Even if one had never heard of fuzzy logic, the basic elements of such logic are present in this exercise: defining conversational terms by using probabilistic and statistical measures, mixing information from multiple criteria, and returning to a straightforward quantifiable decision based on this information. Indeed, without using such logic, it would be difficult or impossible to do this exercise. In particular, the definition of “good” is inherently judgmental, even when using rigorous statistical techniques. Even a seemingly objective definition such as “more than one standard deviation above the mean” carries with it many subjective, human judgment factors.

A fuzzy logic approach would define a probability distribution of stocks, based on the same 6.2% judgment, and also incorporating other information. Suppose your overall stock-picking approach allowed certain stocks to be selected even if they had one or two bad years but had other outstanding characteristics. Suppose also that you valued stocks that had very high returns (say, better than 10% per year) more than those that just had good returns of 7%. A fuzzy logic approach would enable you to capture this information and use it.

You may observe that, for the subset of stocks that fit into the obvious categories of “good” and “bad” stocks, both the Boolean approach and fuzzy logic approach provide the same answer. This illustrates a central attraction of fuzzy logic: it is equivalent to Boolean logic, when the data can be converted into the yes/no categories essential to Boolean logic. Fuzzy logic should be considered an extension of Boolean logic, not a separate system. The same premises and measurements should produce the same answer using both systems, when the proper data are available. When only mixed data are available, the fuzzy approach should produce an answer that fits the logic used in describing the system in Boolean terms.

Human logic is as old as humanity. The classic texts of ancient Greece, as epitomized by the Socratic dialogues, were exercises in logic. In recent centuries, a rigorous mathematical logic was

established and codified as Boolean logic. However, humans exercise, as previously observed, a kind of logic that handles information which does not fit neatly into the yes/no structure of Boolean logic. The modern study of this began with the pioneering work of Lofti A. Zadeh in what he called “fuzzy sets” in 1965, which was extended to a system of fuzzy logic in the 1970s. The first fuzzy logic controller was documented by Mamdani and Assilian in 1975. Since then, the use of fuzzy logic has expanded greatly.

While fuzzy logic is used widely, the term is not. Perhaps burdened by the unfortunate “fuzzy” adjective, application developers and users of this technology often characterize it as “artificial intelligence,” “expert systems,” or “neuro-fuzzy systems.” These are all efforts to incorporate subjective, partial, or qualitative information into rigorous decisions rules, and are related or based on the basic insights of fuzzy sets.

For readers of this chapter, a well-written introduction to the field is the Fuzzy Logic Toolbox Users Guide published by The MathWorks. There are a number of other Web sites that provide overviews. Unfortunately, very few references on its use in business economics or finance exist, although there are some publications that deal with similar applications (often using different terms such as expert systems) or discuss applications that are crude versions of fuzzy logic systems, such as rudimentary credit scoring.

As a demonstration of the power of this approach consider the question of whose income tax returns will be audited in the IRS. The U.S. Internal Revenue Service, like most tax authorities, cannot audit all tax returns. It must, therefore, select a portion of those returns to audit. Generally, taxing authorities like to maximize the effect of those audits, by concentrating on those returns where an audit will be most likely to increase revenue or improve compliance. The IRS system for selecting returns is a closely guarded secret, and we do not claim to have any confidential information on their system. However, we constructed our own predictive model of the IRS system – using publicly available information, including the IRS’s own pronouncements about the focus of its enforcement and compliance efforts, as well as statistics on the characteristics of returns that are audited.

The primary purpose of this exercise is to demonstrate how the

techniques of fuzzy logic can be used to model complex business problems. An example of such a task is the decision about which tax returns to audit. We believe citizens have both a moral and legal obligation to pay their income taxes – whether they believe they will be audited or not. However, we observe that humans have resisted paying taxes for all recorded history. This creates the impetus for auditing tax returns on a selective basis.

The predictor was one of the first fuzzy logic routines ever presented on the Internet. From just before April 15, 2000, through much of 2001, it was run on a MATLAB Web server accessible to millions of taxpayers. Taxpayers across the country were able to anonymously test it with real or imagined data.

Fuzzy logic requires very rigorous rules, although they are expressed in the form of probability statements that define concepts such as “good” and “poor.” These rules are implemented by membership functions, which are probability statements. The combination of the rules and the membership functions form an inference system. A fuzzy logic inference system is sometimes abbreviated as FIS. Such an inference system is similar to Boolean logic, except that Boolean logic requires a yes/no determination. Fuzzy logic allows for probability statements that produce the identical result for strong yes/no situations but translate gray areas into probabilities.

This section states the rules which are the basis for the income tax audit predictor fuzzy inference system. The first rule is:

1. If (income is low) and (flags is no_flags) and (deductions is standard) and (accuracy is not poor) then (audit_risk is low). This incorporates the first observation about tax filers and the likelihood of audits: if your income is low, and there is nothing unusual in your return, and you make no mistakes on your return, then the likelihood of an audit is low. This rule is true, but it is inadequate by itself. We therefore summarize our other information about the factors that predict audit likelihood in addition rules.

What is “low”?

Many of the rules use terms to describe the likelihood of an audit in terms like “low” or “high.” It is the translation of concepts like “low” into mathematical statements that is the basis for the

“fuzzification” in fuzzy logic. Rather than use hard edged rules (e.g., income below \$10,000 is low), we make a probability statement, using a defined distribution. These form membership functions. Such a membership function for low income would indicate that, of course, a person with income of \$5,000 would be considered low income, but a person with income of \$11,000 would likely be low income as well. The graphical illustrations included in the next section illustrate the implementation of this research.

```

.
% Now print rules in workspace
rules_text = showrule(fismat)
rules_text =
1. If (income is low) and (flags is no_flags) and
(deductions is standard) and (accuracy is not poor)
then (audit_risk is low) (1)
2. If (income is low) and (flags is no_flags) and
(deductions is standard) and (accuracy is
compulsive) then (audit_risk is low) (1)
3. If (income is low) and (flags is no_flags) and

(accuracy is poor) then (audit_risk is medium) (1)
4. If (income is low) and (flags is not no_flags)
then (audit_risk is high) (1)
5. If (income is middle) and (flags is no_flags) and
(deductions is standard) and (accuracy is not poor)
then (audit_risk is low) (1)
6. If (income is middle) and (flags is no_flags) and
(deductions is itemized) and (accuracy is not poor)
then (audit_risk is medium) (1)
7. If (income is middle) and (flags is one_or_two)
and (deductions is itemized) then (audit_risk is
medium) (1)
8. If (income is middle) and (flags is
three_or_more) then (audit_risk is high) (1)
9. If (income is middle) and (flags is one_or_two)
and (deductions is miscellaneous) and (accuracy is
poor) then (audit_risk is high) (1)
10. If (income is high) and (flags is no_flags) and
(deductions is itemized) and (accuracy is not poor)
then (audit_risk is medium) (1)
11. If (income is high) and (deductions is
miscellaneous) and (accuracy is not compulsive) then
(audit_risk is high) (1)
12. If (income is high) and (flags is three_or_more)
and (deductions is miscellaneous) then (audit_risk
is high) (1)
% -----

```

Code Fragment 7 Audit Predictor Rules

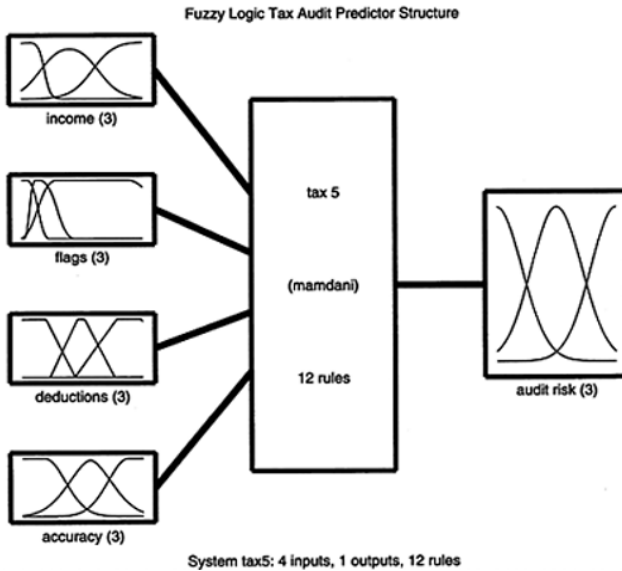


FIGURE 10.1 Graphical display of audit predictor

The following figures illustrate this inference system:

1. Figure 10.1, “Graphical display of audit predictor,” illustrates the entire structure of the inference system. Note that the number of rules is shown, as are the membership functions on the left. From the inputs and the rules, the inference system provides an output measurement which is shown on the right.

2. Figure 10.2, “Income membership functions,” shows the membership function for income. As the rules often refer to income (“if income is high,...then...”), this membership function translates a dollar figure of income into a conceptual description such as high or low. The membership function for “middle income” is centered in the range of \$40,000 to \$73,000 in annual income for the year 2000.

3. Figure 10.3, “Audit flag membership function,” shows the membership function for audit flags.

4. Figure 10.4, “Surface of rules,” provides a 3-D view of the likelihood of an audit, based on income and the number of flags. Note that the actual function is 4-D, but only 3 dimensions can be shown on this surface graph.

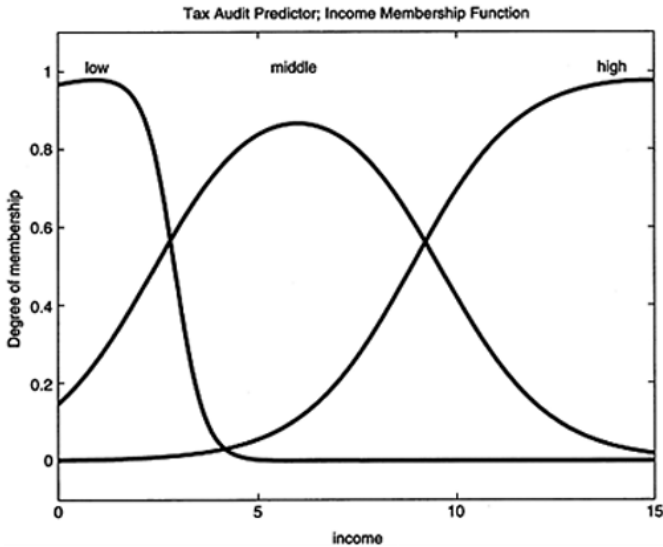


FIGURE 10.2 Income membership functions.

5. Figure 10.5, “Ruleviewer for audit risk FIS,” provides a comprehensive graphical illustration of the effect of the various inputs, membership functions, and rules on the resulting value from the FIS. In this graphic, the inputs used are an annual income of \$75,000, 5 audit flags, itemized deductions, and mediocre accuracy. These are shown in the first four columns of distribution graphics in the figure. As there are 12 rules, each input is evaluated in the 12 rows.

The audit predictor FIS evaluates such characteristics as resulting in a high probability of an audit. This is consistent with common sense; a person who makes mistakes on his return, and whose return contains a number of audit flags, should be audited more frequently than the typical taxpayer. What cannot be shown in this figure is that the ruleviewer in the MATLAB toolbox is interactive, allowing you to see how the FIS evaluates multiple scenarios quite quickly.

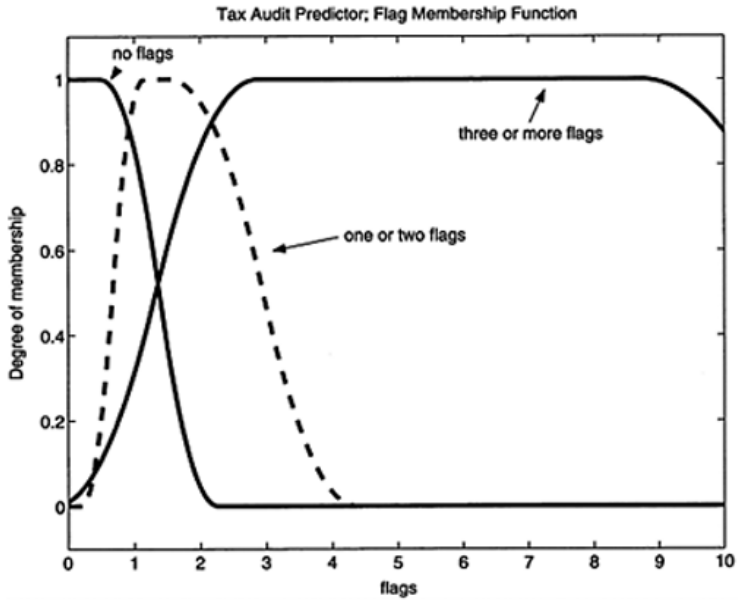


FIGURE 10.3 Audit flag membership function

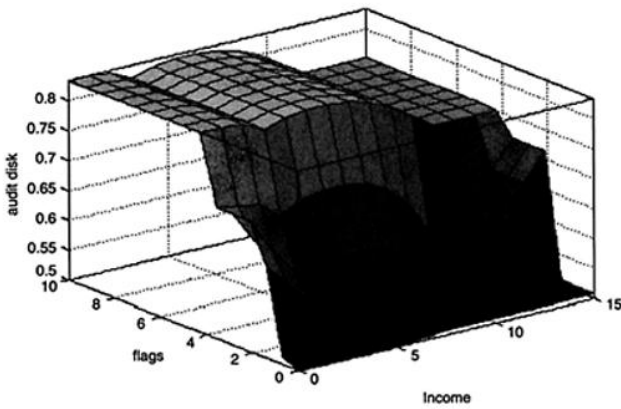


FIGURE 10.4 Surface of rules

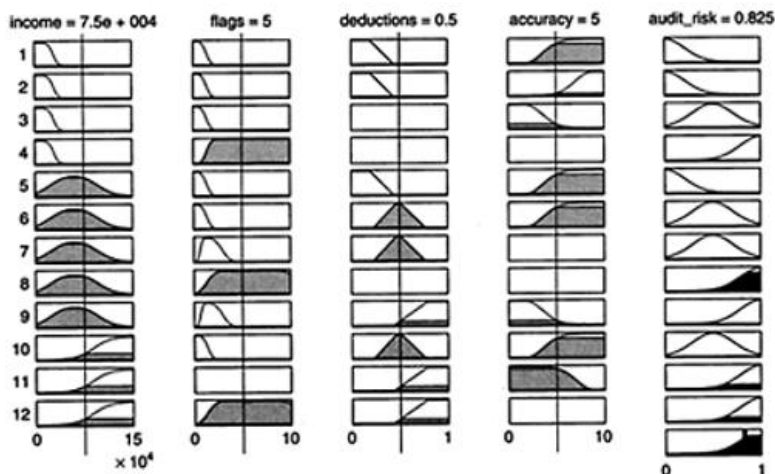


FIGURE 10.5 Ruleviewer for audit risk FIS

In particular, the program:

1. Reads in an FIS file. This FIS file must be created separately, using a tool included with MATLAB'S Fuzzy Logic Toolbox.
2. Prints the rules in the workspace.
3. Plots the membership functions for key inputs.
4. Calls an ingenious rule viewer that provides a visual explanation of the fuzzy logic at work, given different inputs. This is a particularly effective tool in the MATLAB Fuzzy Logic Toolbox.
5. Generates test inputs for the FIS. Testing the FIS is an important step, particularly since the rules and membership functions incorporate subjective information. The script we created to illustrate this approach generates random information to use in testing the system.
6. Evaluates the test data and produces an output for it, using the test input file.
7. Graphs the output and reports on the analysis in the command window. We recommend following a similar approach when implementing other fuzzy logic models.

One variation discussed in the following section is to utilize a fuzzy logic component in a simulation model. For a Simulink model of this type, the recommended procedure is:

1. Initialize the model, either with an initialization block or by manually running an initialization callback program. Such a callback may also print the rules, plot the membership function, or open a ruleviewer. This step should include the first several steps outlined above.

2. The testing tasks should be accomplished outside the general initialization routine unless you want a simple test to be part of every use.

3. Enter any custom variables.

4. Simulate the model.

5. Graph or otherwise report the output.

Another common business decision, involving subjective as well as objective information, is the evaluation of credit risk. This task is so important that “credit scoring” is now an established industry. We developed a fuzzy logic credit risk algorithm that is similar to the audit risk model presented above. However, we implemented it in a Simulink simulation model. This model demonstrates how a fuzzy logic inference system can be used even in simulation models.

Consider the role of a lender confronted with numerous business loan applications. She wishes to quickly analyze a number of applications that have already passed an initial review. We assume that, as a result of this prior review, we will only examine firms that have been historically profitable, going concerns with satisfactory debt repayment histories.

While an actual examination of business credit risk would look at more variables, for this inference system we have simplified the evaluation to two measures of a firm’s ability to pay debt: its return on assets (ROA) and its coverage ratio (earnings as a multiple of debt coverage). We take current and historic information necessary to calculate these measures and collect them for multiple applications in one data file. We also summarize the collected wisdom of experienced loan officers in a credit risk FIS. This system includes membership functions for ROA and coverage ratios, quantifying judgments such as “good” or “poor.” We then summarize our underwriting standards in a set of rules that tells us when the desirability of lending money to the applicant is high, low, or uncertain. Such rules would include statements such as “if ROA is good, and coverage is adequate, then desirability is high.” We then

create a simulation model to feed the information into the FIS and provide other graphing and reporting functions. This will allow us to take the entire data file representing multiple applicants and analyze them all using consistent standards.

The Credit Risk model in Figure 10.6, “Credit risk simulation model,” is structured as follows:

1. The model is initialized with variables for assets, earnings on those assets, and debt. For testing, random information is included to generate multiple scenarios. The same initialization file reads in a credit risk FIS which will be used in a fuzzy logic controller block in the model.

2. The earnings are calculated in earnings subsystem, using the variables created when the model was initialized. The costs of debt are calculated in another subsystem.

3. Using these variables, the model calculates ROA and coverage ratios for each scenario.

4. This information is then given to a fuzzy logic controller. The credit risk FIS, referenced by the controller, rates each case as it is generated.

5. The resulting scores are collected and reported. Although we have not shown them here, the membership functions, rules, and other attributes of this model can be illustrated as they have been in the section “An Income Tax Audit Predictor”

This simple model could be extended greatly, for example by:

1. Using the numeric score from the fuzzy logic controller to directly control the disposition of an application

2. Including more extensive underwriting processes for those applications that fail to pass the first FIS

3. Creating a second stage requiring additional information or analysis

4. Increasing the number and sophistication of the membership functions and rules

The credit risk model demonstrates the following: 1. Fuzzy logic is a powerful tool, and similar analyses are already being performed every day, although they may not be called fuzzy logic. 2. Fuzzy logic can be used directly in a simulation model.

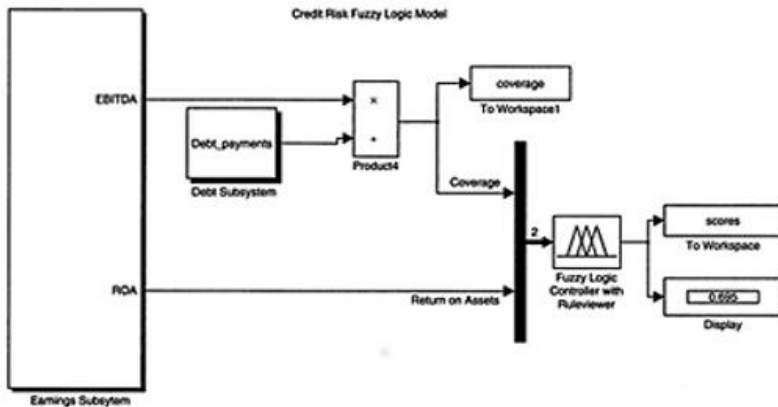


FIGURE 10.6 Credit risk simulation model

Fuzzy logic is a new field, and there are relatively few references on its use. However, based on our experience, we have identified implementation issues and troubleshooting tips.

There are a number of specific points to keep in mind when creating a model using fuzzy logic:

1. The inputs should be clearly identified in the FIS and preferably named in a way that is logical for the model as a whole.
2. The FIS can be created using a special editor which is invoked by the command `fuzzy`.
3. You must read in the FIS before it can be used. One way to accomplish this is through an initialization script given in the code fragment below.

Fuzzy logic can also be incorporated into a Simulink model, as demonstrated above. There are special blocks for membership functions and fuzzy logic controllers. The following troubleshooting tips apply to such models:

1. The FIS must be “read” into the workspace, using the command `readfis`. This is often done in an initialization routine, before the model is simulated.
2. If you receive an error message indicating that the FIS you have selected is unknown, make sure that the FIS has been read into the workspace. For example, if the `readfis` command shown in the

example Code Fragment 8 had been issued, the parameter for the fuzzy logic controller would be `fismat`, not `myfis`.

3. The inputs can be fed by a command-line script or into a Simulink model. In an m-file, the syntax is described in Code Fragment 8, “Read FIS Command” on page 400.

4. Carefully check the orientation of the inputs.

5. Look under the mask of the fuzzy logic controller as well as simply clicking it to set the parameter.

6. Check to see that the proper FIS name is included in all the blocks (the animation and the controller blocks), drilling down twice for the controller. Do not assume that the parameter will pass through to the lower levels; there may be a bug preventing this.

7. When all else fails, delete the block. Break the library link for a new block from the Simulink library and import it into the model. Go through all the above steps systematically as you hook up the new block.


```

% Fuzzy Logic Example: Risk of Audit %
% (c) 2001-2003, Anderson Economic Group
% Use with license only.
% August 2003, ver 3
% Uses: Fuzzy Logic Toolbox, generate_tax_input.m
echo on;
% -----
% Fuzzy Logic Audit Predictor
% Patrick L. Anderson

```

^{386a} If you have the block with the ruleviewer animation, you may need to look under the mask twice in order to get down to the actual FIS controller block. You enter the parameter first by clicking on the controller block, then put it into the animation block, and then look under the controller block mask again (for a total of three times) to set the parameter.

```

% AEG LLC, Lansing Michigan
% -----
% Read in Fuzzy Inference System (fis).
% Note that FIS must already be created; that the
% file must be in the path;
% that the name of the file must be in single
% quotes.
%
fismat = readfis('tax5.fis')
% Note that fuzzy inference system is a MATLAB
% structure.
% Examine the items within the structure.
fismat.name
fismat.input
fismat.input.name
fismat.rule
% -----
% Now print rules in workspace
rules_text = showrule(fismat)
% -----
% Now use specific tools to plot parts of FIS
figure(1), plotfis(fismat);
title('Fuzzy Logic Tax Audit Predictor Structure',

```

```

'fontweight', 'bold');
text (0.4, -0.2, ['Printed by AEG, 'datestr(now, 1)] ,
'units', 'normalized', 'fontangle', 'italic');
figure(2), plotmf(fismat, 'input', 1);
title('Tax Audit Predictor; Income MF',
'fontweight', 'bold');
% text (0.0, -0.2, ['Printed by AEG, 'datestr(now,
1)], 'units', 'normalized', 'fontangle', 'italic');
figure(3), plotmf(fismat, 'input', 2)
title('Tax Audit Predictor; Flag MF', 'fontweight',
'bold');
% text (0.0, -0.2, ['Printed by AEG, 'datestr(now,
1)], 'units', 'normalized', 'fontangle', 'italic');
% -----
% Now use GUI to see rules
ruleview(fismat)
echo off;
%-----
% Test script
%
% Create input vector of many cases, using random
% variables; display
generate_tax_input
% evaluate input vector
rating = evalfis(test_input, fismat);
% Display numerical and verbal assessments;
disp('Gross income, audit flags, deductions,
accuracy inputs are:');
disp(test_input);
disp('Audit risk index on a 0 to 1 scale (1 is
highest): ');
disp (rating);
% Create bar chart of ratings of sample tax returns
figure, bar (rating);
title('Ratings of test tax returns', 'fontweight',
'bold');
ylabel('Rating (higher number is higher risk)');
text (0.0, -0.1, ['Printed by AEG, 'datestr(now,
1)], 'units', 'normalized', 'fontangle', 'italic');
% -----
% Check to see what is in the workspace now.
whos
% end of script

```

Code Fragment 8 Fuzzy Logic Audit Risk

CHAPTER 11. THE NATURE OF REGRESSION ANALYSIS

11.1 Historical origin of the term regression

The term regression was introduced by Francis Galton. In a famous paper, Galton found that, although there was a tendency for tall parents to have tall children and for short parents to have short children, the average height of children born of parents of a given height tended to move or “regress” toward the average height in the population as a whole.

1. In other words, the height of the children of unusually tall or unusually short parents tends to move toward the average height of the population. Galton’s law of universal regression was confirmed by his friend Karl Pearson, who collected more than a thousand records of heights of members of family groups.

2. He found that the average height of sons of a group of tall fathers was less than their fathers’ height and the average height of sons of a group of short fathers was greater than their fathers’ height, thus “regressing” tall and short sons alike toward the average height of all men. In the words of Galton, this was “regression to mediocrity”.

The modern interpretation of regression is, however, quite different. Broadly speaking, we may say.

Regression analysis is concerned with the study of the dependence of one variable, the dependent variable, on one or more other variables, the explanatory variables, with a view to estimating and/or predicting the (population) mean or average value of the former in terms of the known or fixed (in repeated sampling) values of the latter. The full import of this view of regression analysis will become clearer as we progress, but a few simple examples will make the basic concept quite clear.

Reconsider Galton’s law of universal regression. Galton was interested in finding out why there was stability in the distribution of heights in a population. But in the modern view our concern is not with this explanation but rather with finding out how the average height of sons changes, given the fathers’ height. In other words, our concern is with predicting the average height of sons knowing the height of their fathers. To see how this can be done, consider Figure 11.1, which is a scatter diagram, or scatter gram.

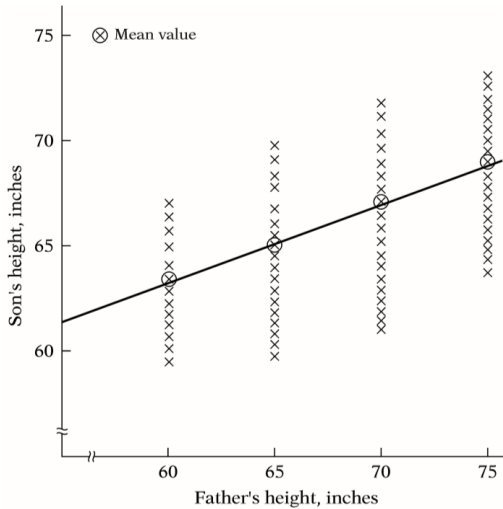


FIGURE 11.1 Hypothetical distribution of sons' heights corresponding to given heights of fathers

This figure shows the distribution of heights of sons in a hypothetical population corresponding to the given or fixed values of the father's height. Notice that corresponding to any given height of a father is a range or distribution of the heights of the sons. However, notice that despite the variability of the height of sons for a given value of father's height, the average height of sons generally increases as the height of the father increases. To show this clearly, the circled crosses in the figure indicate the average height of sons corresponding to a given height of the father. Connecting these averages, we obtain the line shown in the figure. This line, as we shall see, is known as the regression line. It shows how the average height of sons increases with the father's height.

3. Consider the scattergram in Figure 11.2, which gives the distribution in a hypothetical population of heights of boys measured at fixed ages. Corresponding to any given age, we have a range, or distribution, of heights. Obviously, not all boys of a given age are likely to have identical heights. But height on the average increases with age (of course, up to a certain age), which can be seen clearly if we draw a line (the regression line) through the circled points that

represent the average height at the given ages. Thus, knowing the age, we may be able to predict from the regression line the average height corresponding to that age.

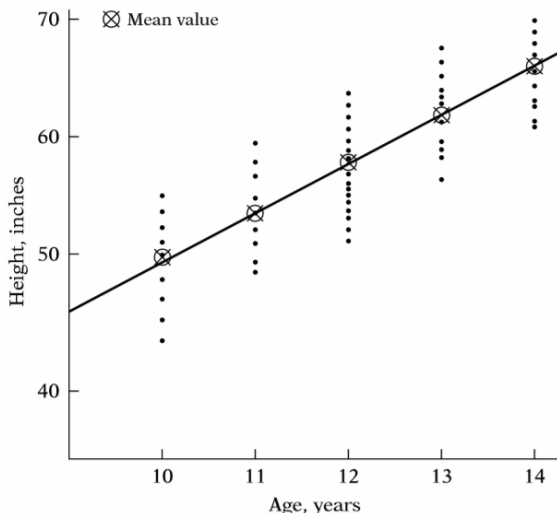


FIGURE 11.2 Hypothetical distribution of heights corresponding to selected ages

4. Turning to economic examples, an economist may be interested in studying the dependence of personal consumption expenditure on aftertax or disposable real personal income. Such an analysis may be helpful in estimating the marginal propensity to consume (MPC), that is, average change in consumption expenditure for, say, a dollar's worth of change in real income.

5. A monopolist who can fix the price or output (but not both) may want to find out the response of the demand for a product to changes in price. Such an experiment may enable the estimation of the price elasticity (i.e., price responsiveness) of the demand for the product and may help determine the most profitable price.

6. A labor economist may want to study the rate of change of money wages in relation to the unemployment rate. The historical data are shown in the scattergram given in Figure 11.3. The curve in Figure 11.3 is an example of the celebrated Phillips curve relating

changes in the money wages to the unemployment rate. Such a scattergram may enable the labor economist to predict the average change in money wages given a certain unemployment rate. Such knowledge may be helpful in stating something about the inflationary process in an economy, for increases in money wages are likely to be reflected in increased prices.

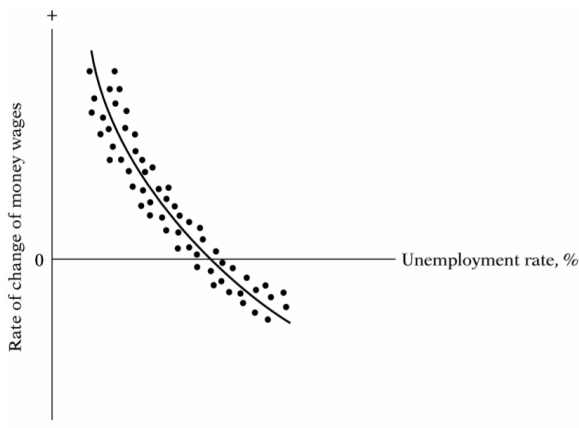


FIGURE 11.3 Hypothetical Phillips curve

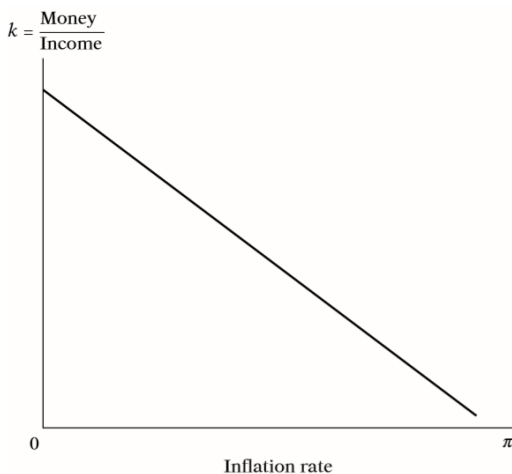


FIGURE 11.4 Money holding in relation to the inflation rate π .

7. From monetary economics it is known that, other things remaining the same, the higher the rate of inflation π , the lower the proportion k of their income that people would want to hold in the form of money, as depicted in Figure 11.4. A quantitative analysis of this relationship will enable the monetary economist to predict the amount of money, as a proportion of their income that people would want to hold at various rates of inflation.

8. The marketing director of a company may want to know how the demand for the company's product is related to, say, advertising expenditure. Such a study will be of considerable help in finding out the elasticity of demand with respect to advertising expenditure, that is, the percent change in demand in response to, say, a 1 percent change in the advertising budget. This knowledge may be helpful in determining the "optimum" advertising budget.

9. Finally, an agronomist may be interested in studying the dependence of crop yield, say, of wheat, on temperature, rainfall, amount of sunshine, and fertilizer. Such a dependence analysis may enable the prediction or forecasting of the average crop yield, given information about the explanatory variables. The reader can supply scores of such examples of the dependence of one variable on one or more other variables. The techniques of regression analysis discussed in this text are specially designed to study such dependence among variables.

11.2 Statistical versus deterministic relationships

From the examples cited in Section 1.2, the reader will notice that in regression analysis we are concerned with what is known as the statistical, not functional or deterministic, dependence among variables, such as those of classical physics. In statistical relationships among variables we essentially deal with random or stochastic variables, that is, variables that have probability distributions. In functional or deterministic dependency, on the other hand, we also deal with variables, but these variables are not random or stochastic. The dependence of crop yield on temperature, rainfall, sunshine, and fertilizer, for example, is statistical in nature in the sense that the explanatory variables, although certainly important, will not enable the agronomist to predict crop yield exactly because

of errors involved in measuring these variables as well as a host of other factors (variables) that collectively affect the yield but may be difficult to identify individually. Thus, there is bound to be some “intrinsic” or random variability in the dependent-variable crop yield that cannot be fully explained no matter how many explanatory variables we consider. In deterministic phenomena, on the other hand, we deal with relationships of the type, say, exhibited by Newton’s law of gravity, which states: Every particle in the universe attracts every other particle with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Symbolically, $F = k(m_1 m_2 / r^2)$, where F =force, m_1 and m_2 are the masses of the two particles, r =distance, and k =constant of proportionality. Another example is Ohm’s law, which states: For metallic conductors over a limited range of temperature the current C is proportional to the voltage V ; that is, $C = (1/k)V$ where $1/k$ is the constant of proportionality. Other examples of such deterministic relationships are Boyle’s gas law, Kirchhoff’s law of electricity, and Newton’s law of motion. In this text we are not concerned with such deterministic relationships. Of course, if there are errors of measurement, say, in the k of Newton’s law of gravity, the otherwise deterministic relationship becomes a statistical relationship. In this situation, force can be predicted only approximately from the given value of k (and m_1 , m_2 , and r), which contains errors. The variable F in this case becomes a random variable.

Although regression analysis deals with the dependence of one variable on other variables, it does not necessarily imply causation. In the words of Kendall and Stuart, “A statistical relationship, however strong and however suggestive, can never establish causal connection: our ideas of causation must come from outside statistics, ultimately from some theory or other.” In the crop-yield example cited previously, there is no statistical reason to assume that rainfall does not depend on crop yield. The fact that we treat crop yield as dependent on rainfall (among other things) is due to non statistical considerations: Common sense suggests that the relationship cannot be reversed, for we cannot control rainfall by varying crop yield. In all the examples cited in Section 1.2 the point to note is that a statistical relationship in itself cannot logically imply causation. To

ascribe causality, one must appeal to a priori or theoretical considerations. Thus, in the third example cited, one can invoke economic theory in saying that consumption expenditure depends on real income.

Closely related to but conceptually very much different from regression analysis is correlation analysis, where the primary objective is to measure the strength or degree of linear association between two variables. The correlation coefficient, which we shall study in detail in Chapter 3, measures this strength of (linear) association. For example, we may be interested in finding the correlation (coefficient) between smoking and lung cancer, between scores on statistics and mathematics examinations, between high school grades and college grades, and so on. In regression analysis, as already noted, we are not primarily interested in such a measure. Instead, we try to estimate or predict the average value of one variable on the basis of the fixed values of other variables. Thus, we may want to know whether we can predict the average score on a statistics examination by knowing a student's score on a mathematics examination. Regression and correlation have some fundamental differences that are worth mentioning. In regression analysis there is an asymmetry in the way the dependent and explanatory variables are treated. The dependent variable is assumed to be statistical, random, or stochastic, that is, to have a probability distribution. The explanatory variables, on the other hand, are assumed to have fixed values (in repeated sampling), which was made explicit in the definition of regression given in Section. Thus, in Figure 11.2 we assumed that the variable age was fixed at given levels and height measurements were obtained at these levels. In correlation analysis, on the other hand, we treat any (two) variables symmetrically; there are no distinction between the dependent and explanatory variables. After all, the correlation between scores on mathematics and statistics examinations is the same as that between scores on statistics and mathematics examinations. Moreover, both variables are assumed to be random. As we shall see, most of the correlation theory is based on the assumption of randomness of variables, whereas most of the regression theory to be expounded in this book is conditional upon the assumption that the dependent variable is stochastic but the explanatory variables are fixed or nonstochastic.

Although it is a matter of personal taste and tradition, in this text we will use the dependent variable/explanatory variable or the more neutral, regressand and regressor terminology. If we are studying the dependence of a variable on only a single explanatory variable, such as that of consumption expenditure on real income, such a study is known as simple, or two-variable, regression analysis. However, if we are studying the dependence of one variable on more than one explanatory variable, as in the crop-yield, rainfall, temperature, sunshine, and fertilizer examples, it is known as multiple regression analysis. In other words, in two-variable regression there is only one explanatory variable, whereas in multiple regression there is more than one explanatory variable. The term random is a synonym for the term stochastic. As noted earlier, a random or stochastic variable is a variable that can take on any set of values, positive or negative, with a given probability.

Unless stated otherwise, the letter Y will denote the dependent variable and the X 's (X_1, X_2, \dots, X_k) will denote the explanatory variables, X_k being the k th explanatory variable. The subscript i or t will denote the i th or the t th observation or value. X_{ki} (or X_{kt}) will denote the i th (or t th) observation on variable X_k . N (or T) will denote the total number of observations or values in the population, and n (or t) the total number of observations in a sample. As a matter of convention, the observation subscript i will be used for cross-sectional data (i.e., data collected at one point in time) and the subscript t will be used for time series data (i.e., data collected over a period of time). The nature of cross-sectional and time series data, as well as the important topic of the nature and sources of data for empirical analysis, is discussed in the following section.

The success of any econometric analysis ultimately depends on the availability of the appropriate data. It is therefore essential that we spend some time discussing the nature, sources, and limitations of the data that one may encounter in empirical analysis.

11.3 Types of Data

Three types of data may be available for empirical analysis: time series, cross-section, and pooled (i.e., combination of time series and cross-section) data.

Time Series Data The data shown in Table 11.1 of the Introduction are an example of time series data. A time series is a set of observations on the values that a variable takes at different times. Such data may be collected at regular time intervals, such as daily (e.g., stock prices, weather reports), weekly (e.g., money supply figures), monthly [e.g., the unemployment rate, the Consumer Price Index (CPI)], quarterly (e.g., GDP), annually (e.g., government budgets), quinquennially, that is, every 5 years (e.g., the census of manufactures), or decennially (e.g., the census of population). Sometime data are available both quarterly as well as annually, as in the case of the data on GDP and consumer expenditure. With the advent of high-speed computers, data can now be collected over an extremely short interval of time, such as the data on stock prices, which can be obtained literally continuously (the so-called real-time quote). Although time series data are used heavily in econometric studies, they present special problems for econometricians. As we will show in chapters on time series econometrics later on, most empirical work based on time series data assumes that the underlying time series is stationary. Although it is too early to introduce the precise technical meaning of stationarity at this juncture, loosely speaking a time series is stationary if its mean and variance do not vary systematically over time. To see what this means, consider Figure 11.5, which depicts the behavior of the M1 money supply in the United States from January 1, 1959, to July 31, 1999. (The actual data are given in exercise 1.4.) As you can see from this figure, the M1 money supply shows a steady upward trend as well as variability over the years, suggesting that the M1 time series is not stationary.

Cross-Section Data Cross-section data are data on one or more variables collected at the same point in time, such as the census of population conducted by the Census Bureau every 10 years (the latest being in year 2000), the surveys of consumer expenditures conducted by the University of Michigan, and, of course, the opinion polls by Gallup and umpteen other organizations. A concrete example of cross-sectional data is given in Table 11.1 This table gives data on egg production and egg prices for the 50 states in the union for 1990 and 1991. For each year the data on the 50 states are cross-sectional data. Thus, in Table 11.1 we have two cross-sectional samples.

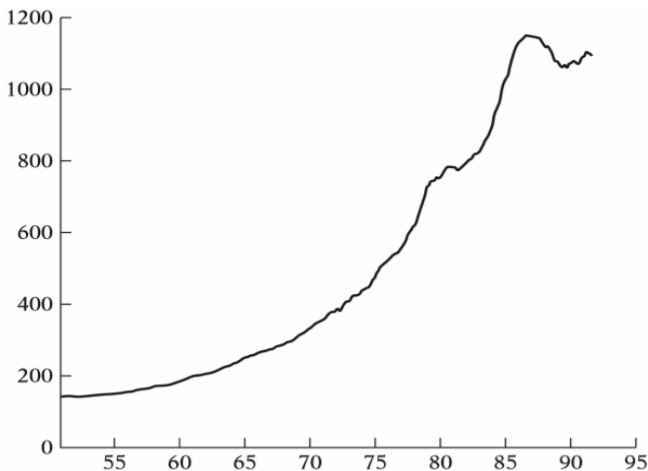


FIGURE 11.5 M1 money supply: United States, 1951:01–1999:09

Just as time series data create their own special problems (because of the stationarity issue), cross-sectional data too have their own problems, specifically the problem of heterogeneity. From the data given in Table 11.1 we see that we have some states that produce huge amounts of eggs (e.g., Pennsylvania) and some that produce very little (e.g., Alaska).

When we include such heterogeneous units in a statistical analysis, the size or scale effect must be taken into account so as not to mix apples with oranges. To see this clearly, we plot in the data on eggs produced and their prices in 50 states for the year 1990. This figure shows how widely scattered the observations are. In Chapter 11 we will see how the scale effect can be an important factor in assessing relationships among economic variables.

Pooled Data In pooled, or combined, data are elements of both time series and cross-section data. The data in Table 11.1 are an example of pooled data. For each year we have 50 cross-sectional observations and for each state we have two time series observations on prices and output of eggs, a total of 100 pooled (or combined) observations. Likewise, the data given in exercise 11.1 are pooled data in that the Consumer Price Index (CPI) for each country for

1973-1997 is time series data, whereas the data on the CPI for the seven countries for a single year are cross-sectional data. In the pooled data we have 175 observations – 25 annual observations for each of the seven countries.

TABLE 11.1 U.S. EGG PRODUCTION

State	Y ₁	Y ₂	X ₁	X ₂	State	Y ₁	Y ₂	X ₁	X ₂
AL	2,206	2,186	92.7	91.4	MT	172	164	68.0	66.0
AK	0.7	0.7	151.0	149.0	NE	1,202	1,400	50.3	48.9
AZ	73	74	61.0	56.0	NV	2.2	1.8	53.9	52.7
AR	3,620	3,737	86.3	91.8	NH	43	49	109.0	104.0
CA	7,472	7,444	63.4	58.4	NJ	442	491	85.0	83.0
CO	788	873	77.8	73.0	NM	283	302	74.0	70.0
CT	1,029	948	106.0	104.0	NY	975	987	68.1	64.0
DE	168	164	117.0	113.0	NC	3,033	3,045	82.8	78.7
FL	2,586	2,537	62.0	57.2	ND	51	45	55.2	48.0
GA	4,302	4,301	80.6	80.8	OH	4,667	4,637	59.1	54.7
HI	227.5	224.5	85.0	85.5	OK	869	830	101.0	100.0
ID	187	203	79.1	72.9	OR	652	686	77.0	74.6
IL	793	809	65.0	70.5	PA	4,976	5,130	61.0	52.0
IN	5,445	5,290	62.7	60.1	RI	53	50	102.0	99.0
IA	2,151	2,247	56.5	53.0	SC	1,422	1,420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48.0	45.8
KY	412	483	67.7	73.5	TN	277	279	71.0	80.7
LA	273	254	115.0	115.0	TX	3,317	3,356	76.7	72.6
ME	1,069	1,070	101.0	97.0	UT	456	486	64.0	59.0
MD	885	898	76.6	75.4	VT	31	30	106.0	102.0
MA	235	237	105.0	102.0	VA	943	988	86.3	81.2
MI	1,406	1,396	58.0	53.8	WA	1,287	1,313	74.1	71.5
MN	2,499	2,697	57.7	54.0	WV	136	174	104.0	109.0
MS	1,434	1,468	87.8	86.7	WI	910	873	60.1	54.0
MO	1,580	1,622	55.4	51.5	WY	1.7	1.7	83.0	83.0

Panel, Longitudinal, or Micropanel Data This is a special type of pooled data in which the same cross-sectional unit (say, a family or a firm) is surveyed over time. For example, the U.S. Department of Commerce carries out a census of housing at periodic intervals. At each periodic survey the same household (or the people living at the same address) is interviewed to find out if there has been any change in the housing and financial conditions of that household since the last survey.

The data used in empirical analysis may be collected by a governmental agency (e.g., the Department of Commerce), an international agency (e.g., the International Monetary Fund (IMF))

or the World Bank), a private organization (e.g., the Standard&Poor's Corporation), or Panel, Longitudinal, or Micropanel Data This is a special type of pooled data in which the same cross-sectional unit (say, a family or a firm) is surveyed over time. For example, the U.S. Department of Commerce carries out a census of housing at periodic intervals. At each periodic survey the same household (or the people living at the same address) is interviewed to find out if there has been any change in the housing and financial conditions of that household since the last survey an individual. Literally, there are thousands of such agencies collecting data for one purpose or another.

The Internet the Internet has literally revolutionized data gathering. If you just "surf the net" with a keyword (e.g., exchange rates), you will be swamped with all kinds of data sources. In Appendix E we provide some of the frequently visited web sites that provide economic and financial data of all sorts. Most of the data can be downloaded without much cost. You may want to bookmark the various web sites that might provide you with useful economic data. The data collected by various agencies may be experimental or nonexperimental. In experimental data, often collected in the natural sciences, the investigator may want to collect data while holding certain factors constant in order to assess the impact of some factors on a given phenomenon. For instance, in assessing the impact of obesity on blood pressure, the researcher would want to collect data while holding constant the eating, smoking, and drinking habits of the people in order to minimize the influence of these variables on blood pressure. In the social sciences, the data that one generally encounters are nonexperimental in nature, that is, not subject to the control of the researcher. For example, the data on GNP, unemployment, stock prices, etc., are not directly under the control of the investigator. As we shall see, this lack of control often creates special problems for the researcher in pinning down the exact cause or causes affecting a particular situation. For example, is it the money supply that determines the (nominal) GDP or is it the other way round?

11.4 The Accuracy of Data

Although plenty of data are available for economic research, the quality of the data is often not that good. There are several reasons for that. First, as noted, most social science data are nonexperimental in nature. Therefore, there is the possibility of observational errors, either of omission or commission. Second, even in experimentally collected data errors of measurement arise from approximation and round offs. Third, in questionnaire-type surveys, the problem of nonresponse can be serious; a researcher is lucky to get a 40 percent response to a questionnaire. Analysis based on such partial response may not truly reflect the behavior of the 60 percent who did not respond, thereby leading to what is known as (sample) selectivity bias. Then there is the further problem that those who respond to the questionnaire may not answer all the questions, especially questions of financially sensitive nature, thus leading to additional selectivity bias. Fourth, the sampling methods used in obtaining the data may vary so widely that it is often difficult to compare the results obtained from the various samples. Fifth, economic data are generally available at a highly aggregate level. For example, most macrodata (e.g., GNP, employment, inflation, unemployment) are available for the economy as a whole or at the most for some broad geographical regions. Such highly aggregated data may not tell us much about the individual or microunits that may be the ultimate object of study. Sixth, because of confidentiality, certain data can be published only in highly aggregate form. The IRS, for example, is not allowed by law to disclose data on individual tax returns; it can only release some broad summary data. Therefore, if one wants to find out how much individuals with a certain level of income spent on health care, one cannot do that analysis except at a very highly aggregate level. But such macroanalysis often fails to reveal the dynamics of the behavior of the microunits. Similarly, the Department of Commerce, which conducts the census of business every 5 years, is not allowed to disclose information on production, employment, energy consumption, research and development expenditure, etc., at the firm level. It is therefore difficult to study the interfirm differences on these items. Because of all these and many other problems, the researcher should always keep in mind that the results of research are

only as good as the quality of the data. Therefore, if in given situations researchers find that the results of the research are “unsatisfactory,” the cause may be not that they used the wrong model but that the quality of the data was poor. Unfortunately, because of the non experimental nature of the data used in most social science studies, researchers very often have no choice but to depend on the available data. But they should always keep in mind that the data used may not be the best and should try not to be too dogmatic about the results obtained from a given study, especially when the quality of the data is suspect. The variables that we will generally encounter fall into four broad categories: ratio scale, interval scale, ordinal scale, and nominal scale. It is important that we understand each.

Ratio Scale For a variable X , taking two values, X_1 and X_2 , the ratio X_1/X_2 and the distance $(X_2 - X_1)$ are meaningful quantities. Also, there is a natural ordering (ascending or descending) of the values along the scale. Therefore, comparisons such as $X_2 \leq X_1$ or $X_2 \geq X_1$ are meaningful. Most economic variables belong to this category. Thus, it is meaningful to ask how big this year’s GDP is compared with the previous year’s GDP.

Interval Scale An interval scale variable satisfies the last two properties of the ratio scale variable but not the first. Thus, the distance between two time periods, say (2000–1995) is meaningful, but not the ratio of two time periods (2000/1995).

Ordinal Scale A variable belongs to this category only if it satisfies the third property of the ratio scale (i.e., natural ordering). Examples are grading systems (A, B, C grades) or income class (upper, middle, lower). For these variables the ordering exists but the distances between the categories cannot be quantified. Students of economics will recall the indifference curves between two goods, each higher indifference curve indicating higher level of utility, but one cannot quantify by how much one indifference curve is higher than the others.

Nominal Scale Variables in this category have none of the features of the ratio scale variables. Variables such as gender (male, female) and marital status (married, unmarried, divorced, separated) simply denote categories. Question: What is the reason why such variables cannot be expressed on the ratio, interval, or ordinal scales?

As we shall see, econometric techniques that may be suitable for ratio scale variables may not be suitable for nominal scale variables. Therefore, it is important to bear in mind the distinctions among the four types of measurement scales discussed above.

EXERCISES

1.1. Table 11.2 gives data on the Consumer Price Index (CPI) for seven industrialized countries with 1982–1984=100 as the base of the index. a. From the given data, compute the inflation rate for each country. b. Plot the inflation rate for each country against time (i.e., use the horizontal axis for time and the vertical axis for the inflation rate.) c. What broad conclusions can you draw about the inflation experience in the seven countries? d. Which country's inflation rate seems to be most variable? Can you offer any explanation?

TABLE 11.2 CPI IN SEVEN INDUSTRIAL COUNTRIES, 1973–1997 (1982–1984)

Year	Canada	France	Germany	Italy	Japan	U.K.	U.S.
1973	40.80000	34.60000	62.80000	20.60000	47.90000	27.90000	44.40000
1974	45.20000	39.30000	67.10000	24.60000	59.00000	32.30000	49.30000
1975	50.10000	43.90000	71.10000	28.80000	65.90000	40.20000	53.80000
1976	53.90000	48.10000	74.20000	33.60000	72.20000	46.80000	56.90000
1977	58.10000	52.70000	76.90000	40.10000	78.10000	54.20000	60.60000
1978	63.30000	57.50000	79.00000	45.10000	81.40000	58.70000	65.20000
1979	69.20000	63.60000	82.20000	52.10000	84.40000	66.60000	72.60000
1980	76.10000	72.30000	86.70000	63.20000	90.90000	78.50000	82.40000
1981	85.60000	81.90000	92.20000	75.40000	95.30000	87.90000	90.90000
1982	94.90000	91.70000	97.10000	87.70000	98.10000	95.40000	96.50000
1983	100.4000	100.4000	100.3000	100.8000	99.80000	99.80000	99.60000
1984	104.7000	108.1000	102.7000	111.5000	102.1000	104.8000	103.9000
1985	109.0000	114.4000	104.8000	121.1000	104.1000	111.1000	107.6000
1986	113.5000	117.3000	104.7000	128.5000	104.8000	114.9000	109.6000
1987	118.4000	121.1000	104.9000	134.4000	104.8000	119.7000	113.6000
1988	123.2000	124.4000	106.3000	141.1000	105.6000	125.6000	118.3000
1989	129.3000	128.7000	109.2000	150.4000	108.1000	135.3000	124.0000
1990	135.5000	133.0000	112.2000	159.6000	111.4000	148.2000	130.7000
1991	143.1000	137.2000	116.3000	169.8000	115.0000	156.9000	136.2000
1992	145.3000	140.5000	122.1000	178.8000	116.9000	162.7000	140.3000
1993	147.9000	143.5000	127.6000	186.4000	118.4000	165.3000	144.5000
1994	148.2000	145.8000	131.1000	193.7000	119.3000	169.4000	148.2000
1995	151.4000	148.4000	133.5000	204.1000	119.1000	175.1000	152.4000
1996	153.8000	151.4000	135.5000	212.0000	119.3000	179.4000	156.9000
1997	156.3000	153.2000	137.8000	215.7000	121.3000	185.0000	160.5000

1.2. a. Plot the inflation rate of Canada, France, Germany, Italy, Japan, and the United Kingdom against the United States inflation rate. b. Comment generally about the behavior of the inflation rate in the six countries vis-à-vis the U.S. inflation rate. c. If you find that the six countries' inflation rates move in the same direction as the U.S. inflation rate, would that suggest that U.S. inflation "causes" inflation in the other countries? Why or why not?

1.3. Table 11.3 gives the foreign exchange rates for seven industrialized countries for years 1977–1998. Except for the United Kingdom, the exchange rate is defined as the units of foreign currency for one U.S. dollar; for the United Kingdom, it is defined as the number of U.S. dollars for one U.K. pound. a. Plot these exchange rates against time and comment on the general behavior of the exchange rates over the given time period. b. The dollar is said to appreciate if it can buy more units of a foreign currency. Contrarily, it is said to depreciate if it buys fewer units of a foreign currency. Over the time period 1977–1998, what has been the general behavior of the U.S. dollar? Incidentally, look up any textbook on macroeconomics or international economics to find out what factors determine the appreciation or depreciation of a currency.

TABLE 11.3 EXCHANGE RATES FOR SEVEN COUNTRIES:
1977–1998

Year	Canada	France	Germany	Japan	Sweden	Switzerland	U.K.
1977	1.063300	4.916100	2.323600	268.6200	4.480200	2.406500	1.744900
1978	1.140500	4.509100	2.009700	210.3900	4.520700	1.790700	1.918400
1979	1.171300	4.256700	1.834300	219.0200	4.289300	1.664400	2.122400
1980	1.169300	4.225100	1.817500	226.6300	4.231000	1.677200	2.324600
1981	1.199000	5.439700	2.263200	220.6300	5.066000	1.967500	2.024300
1982	1.234400	6.579400	2.428100	249.0600	6.283900	2.032700	1.748000
1983	1.232500	7.620400	2.553900	237.5500	7.671800	2.100700	1.515900
1984	1.295200	8.735600	2.845500	237.4600	8.270800	2.350000	1.336800
1985	1.365900	8.980000	2.942000	238.4700	8.603200	2.455200	1.297400
1986	1.389600	6.925700	2.170500	168.3500	7.127300	1.797900	1.467700
1987	1.325900	6.012200	1.798100	144.6000	6.346900	1.491800	1.639800
1988	1.230600	5.959500	1.757000	128.1700	6.137000	1.464300	1.781300
1989	1.184200	6.380200	1.880800	138.0700	6.455900	1.636900	1.638200
1990	1.166800	5.446700	1.616600	145.0000	5.923100	1.390100	1.784100
1991	1.146000	5.646800	1.661000	134.5900	6.052100	1.435600	1.767400
1992	1.208500	5.293500	1.561800	126.7800	5.825800	1.406400	1.766300
1993	1.290200	5.666900	1.654500	111.0800	7.795600	1.478100	1.501600
1994	1.366400	5.545900	1.621600	102.1800	7.716100	1.366700	1.531900
1995	1.372500	4.986400	1.432100	93.96000	7.140600	1.181200	1.578500
1996	1.363800	5.115800	1.504900	108.7800	6.708200	1.236100	1.560700
1997	1.384900	5.839300	1.734800	121.0600	7.644600	1.451400	1.637600
1998	1.483600	5.899500	1.759700	130.9900	7.952200	1.450600	1.657300

CHAPTER 12. TWO-VARIABLE REGRESSION ANALYSIS

Regression analysis is largely concerned with estimating and/or predicting the (population) mean value of the dependent variable on the basis of the known or fixed values of the explanatory variable(s). To understand this, consider the data given in Table 12.1. The data in the table refer to a total population of 60 families in a hypothetical community and their weekly income (X) and weekly consumption expenditure (Y), both in dollars. The 60 families are divided into 10 income groups (from \$80 to \$260) and the weekly expenditures of each family in the various groups are as shown in the table. Therefore, we have 10 fixed values of X and the corresponding Y values against each of the X values; so to speak, there are 10 Y subpopulations.

TABLE 12.1 WEEKLY *FAMILY* INCOME X , \$

$Y \downarrow$ $X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

There is considerable variation in weekly consumption expenditure in each income group, which can be seen clearly from Figure 12.1. But the general picture that one gets is that, despite the variability of weekly consumption expenditure within each income bracket, on the average, weekly consumption expenditure increases as income increases. To see this clearly, in Table 12.1 we have given the mean, or average, weekly consumption expenditure corresponding to each of the 10 levels of income.

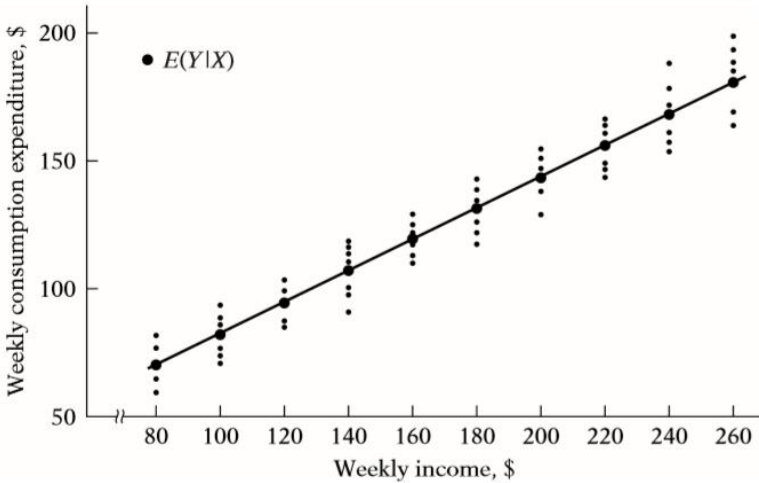


FIGURE 12.1 Conditional distribution of expenditure for various levels of income (data of Table 12.1)

Thus, corresponding to the weekly income level of \$80, the mean consumption expenditure is \$65, while corresponding to the income level of \$200, it is \$137. In all we have 10 mean values for the 10 subpopulations of Y . We call these mean values conditional expected values, as they depend on the given values of the (conditioning) variable X . Symbolically, we denote them as $E(Y|X)$, which is read as the expected value of Y given the value of X (see also Table 12.2). It is important to distinguish these conditional expected values from the unconditional expected value of weekly consumption expenditure, $E(Y)$. If we add the weekly consumption expenditures for all the 60 families in the population and divide this number by 60, we get the number \$121.20 ($\$7272/60$), which is the unconditional mean, or expected, value of weekly consumption expenditure, $E(Y)$; it is unconditional in the sense that in arriving at this number we have disregarded the income levels of the various families. Obviously, the various conditional expected values of Y given in Table 12.1 are different from the unconditional expected value of Y of \$121.20. When we ask the question, “What is the expected value of weekly consumption expenditure of a family,” we get the answer \$121.20 (the unconditional mean).

TABLE 12.2 CONDITIONAL PROBABILITIES $p(Y | X_i)$ FOR THE DATA OF TABLE 12.1

$p(Y X_i)$ ↓ $X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Conditional probabilities $p(Y X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	—	—	$\frac{1}{7}$	—	—	—	$\frac{1}{7}$	—	$\frac{1}{7}$
Conditional means of Y	65	77	89	101	113	125	137	149	161	173

But if we ask the question, “What is the expected value of weekly consumption expenditure of a family whose monthly income is say, \$140,” we get the answer \$101 (the conditional mean). To put it differently, if we ask the question, “What is the best (mean) prediction of weekly expenditure of families with a weekly income of \$140,” the answer would be \$101. Thus the knowledge of the income level may enable us to better predict the mean value of consumption expenditure than if we do not have that knowledge. This probably is the essence of regression analysis, as we shall discover throughout this text. The dark circled points in Figure 12.1 show the conditional mean values of Y against the various X values. If we join these conditional mean values, we obtain what is known as the population regression line (PRL), or more generally, the population regression curve. More simply, it is the regression of Y on X. The adjective “population” comes from the fact that we are dealing in this example with the entire population of 60 families. Of course, in reality a population may have many families. Geometrically, then, a population regression curve is simply the locus of the conditional means of the dependent variable for the fixed values of the explanatory variable(s). More simply, it is the curve connecting the means of the subpopulations of Y corresponding to

the given values of the regressor X . It can be depicted as in Figure 12.2.

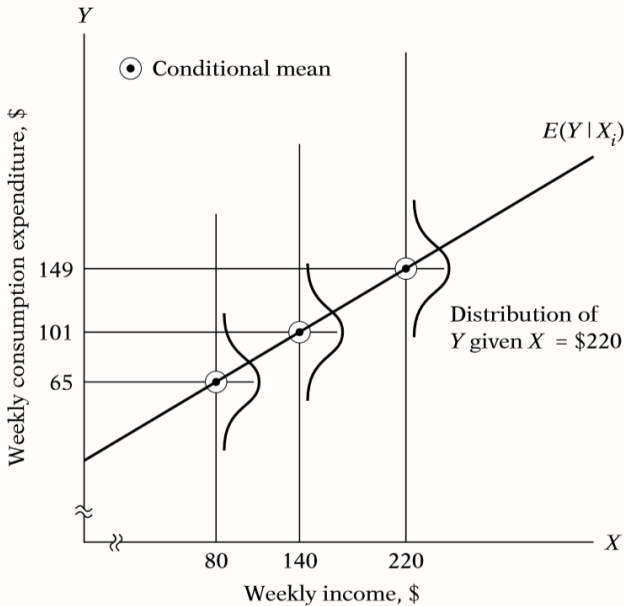


FIGURE 12.2 Population regression line (data of Table 12.1)

This figure shows that for each X (i.e., income level) there is a population of Y values (weekly consumption expenditures) that are spread around the (conditional) mean of those Y values. For simplicity, we are assuming that these Y values are distributed symmetrically around their respective (conditional) mean values. And the regression line (or curve) passes through these (conditional) mean values.

From the preceding discussion and Figures 12.1 and 12.2, it is clear that each conditional mean $E(Y|X_i)$ is a function of X_i , where X_i is a given value of X . Symbolically,

$$E(Y|X_i) = f(X_i) \quad (12.1)$$

where: $f(X_i)$ denotes some function of the explanatory variable X . In our example, $E(Y|X_i)$ is a linear function of X_i . Equation (12.1) is

known as the conditional expectation function (CEF) or population regression function (PRF) or population regression (PR) for short. It states merely that the expected value of the distribution of Y given X_i is functionally related to X_i . In simple terms, it tells how the mean or average response of Y varies with X . What form does the function $f(X_i)$ assume? This is an important question because in real situations we do not have the entire population available for examination. The functional form of the PRF is therefore an empirical question, although in specific cases theory may have something to say. For example, an economist might posit that consumption expenditure is linearly related to income. Therefore, as a first approximation or a working hypothesis, we may assume that the PRF $E(Y|X_i)$ is a linear function of X_i , say, of the type:

$$E(Y|X_i) = \beta_1 + \beta_2 X_i \quad (12.2)$$

where: β_1 and β_2 are unknown but fixed parameters known as the regression coefficients; β_1 and β_2 are also known as intercept and slope coefficients, respectively.

Equation (12.4) itself is known as the linear population regression function. Some alternative expressions used in the literature are linear population regression model or simply linear population regression. In the sequel, the terms regression, regression equation, and regression model will be used synonymously. In regression analysis our interest is in estimating the PRFs like (12.2), that is estimating the values of the unknowns β_1 and β_2 on the basis of observations on Y and X .

The second interpretation of linearity is that the conditional expectation of Y , $E(Y|X_i)$, is a linear function of the parameters, the β 's; it may or may not be linear in the variable X .⁷ In this interpretation $E(Y|X_i) = \beta_1 + \beta_2 X_i$ is a linear (in the parameter) regression model. To see this, let us suppose X takes the value 3. Therefore, $E(Y|X=3) = \beta_1 + 3\beta_2$, which is obviously linear in β_1 and β_2 . All the models shown in Figure 12.3 are thus linear regression models, that is, models linear in the parameters. Now consider the model $E(Y|X_i) = \beta_1 + \beta_2 2X_i$. Now suppose $X = 3$; then we obtain $E(Y|X_i) = \beta_1 + 3\beta_2 2$, which is nonlinear in the parameter β_2 . The preceding model is an example of a nonlinear (in the parameter)

regression model.

Of the two interpretations of linearity, linearity in the parameters is relevant for the development of the regression theory to be presented shortly. Therefore, from now on the term “linear” regression will always mean a regression that is linear in the parameters; the β 's (that is, the parameters are raised to the first power only). It may or may not be linear in the explanatory variables, the X's. Schematically, we have Figure 12.3. Thus, $E(Y|X_i) = \beta_1 + \beta_2 X_i$, which is linear both in the parameters and variable, is a LRM, and so is $E(Y|X_i) = \beta_1 + \beta_2 X_i$, which is linear in the parameters but nonlinear in variable X.

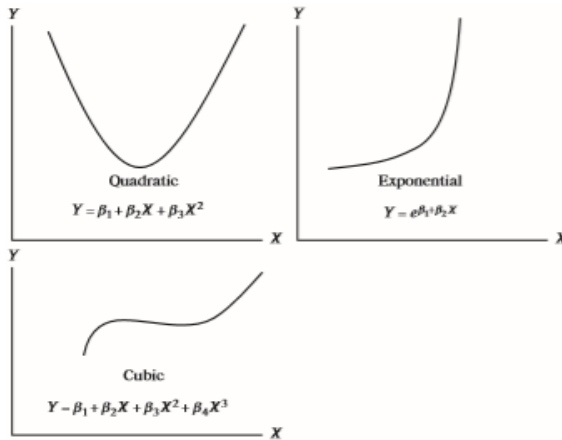


FIGURE 12.3 Linear-in-parameter functions

Therefore, we can express the deviation of an individual Y_i around its expected value as follows:

$$u_i = Y_i - E(Y|X_i) \text{ or } Y_i = E(Y|X_i) + u_i \quad (12.3)$$

where: the deviation u_i is an unobservable random variable taking positive or negative values. Technically, u_i is known as the stochastic disturbance or stochastic error term. How do we interpret? We can say that the expenditure of an individual family, given its income level, can be expressed as the sum of two components: (1) $E(Y|X_i)$,

which is simply the mean consumption expenditure of all the families with the same level of income. This component is known as the systematic, or deterministic, component, and (2) u_i , which is the random, or nonsystematic, component.

We shall examine shortly the nature of the stochastic disturbance term, but for the moment assume that it is a surrogate or proxy for all the omitted or neglected variables that may affect Y but are not (or cannot be) included in the regression model. If $E(Y|X_i)$ is assumed to be linear in X_i , as in (12.1), Eq. (12.2) may be written as:

$$Y_i = E(Y|X_i) + u_i = \beta_1 + \beta_2 X_i + u_i \quad (12.4)$$

Equation (12.3) posits that the consumption expenditure of a family is linearly related to its income plus the disturbance term. Thus, the individual consumption expenditures, given $X = \$80$ (see Table 12.1), can be expressed as:

$$\begin{aligned} Y_1 = 55 &= \beta_1 + \beta_2(80) + u_1 \\ Y_2 = 60 &= \beta_1 + \beta_2(80) + u_2 \\ Y_3 = 65 &= \beta_1 + \beta_2(80) + u_3 \\ Y_4 = 70 &= \beta_1 + \beta_2(80) + u_4 \\ Y_5 = 75 &= \beta_1 + \beta_2(80) + u_5 \end{aligned} \quad (12.5)$$

Now if we take the expected value on both sides, we obtain:

$$E(Y_i | X_i) = E[E(Y|X_i)] + E(u_i | X_i) = E(Y|X_i) + E(u_i | X_i) \quad (12.6)$$

where: use is made of the fact that the expected value of a constant is that constant itself.⁸ Notice carefully that we have taken the conditional expectation, conditional upon the given X 's.

Since $E(Y_i | X_i)$ is the same thing as $E(Y|X_i)$, Eq. (12.5) implies that:

$$E(u_i | X_i) = 0 \quad (12.7)$$

Thus, the assumption that the regression line passes through the conditional means of Y (see Figure 12.2) implies that the conditional mean values of u_i (conditional upon the given X 's) are zero. From

the previous discussion, it is clear (12.1) and (12.2) are equivalent forms if $E(u_i | X_i) = 0$. But the stochastic specification (12.2) has the advantage that it clearly shows that there are other variables besides income that affect consumption expenditure and that an individual family's consumption expenditure cannot be fully explained only by the variable(s) included in the regression model.

The disturbance term u_i is a surrogate for all those variables that are omitted from the model but that collectively affect Y . The obvious question is: Why not introduce these variables into the model explicitly? Stated otherwise, why not develop a multiple regression model with as many variables as possible? The reasons are many.

1. *Vagueness of theory*: The theory, if any, determining the behavior of Y may be, and often is, incomplete. We might know for certain that weekly income X influences weekly consumption expenditure Y , but we might be ignorant or unsure about the other variables affecting Y . Therefore, u_i may be used as a substitute for all the excluded or omitted variables from the model.

2. *Unavailability of data*: Even if we know what some of the excluded variables are and therefore consider a multiple regression rather than a simple regression, we may not have quantitative information about these variables. It is a common experience in empirical analysis that the data we would ideally like to have often are not available. For example, in principle we could introduce family wealth as an explanatory variable in addition to the income variable to explain family consumption expenditure. But unfortunately, information on family wealth generally is not available. Therefore, we may be forced to omit the wealth variable from our model despite its great theoretical relevance in explaining consumption expenditure.

3. *Corevariables versus peripheral variables*: Assume in our consumption income example that besides income X_1 , the number of children per family X_2 , sex X_3 , religion X_4 , education X_5 , and geographical region X_6 also affect consumption expenditure. But it is quite possible that the joint influence of all or some of these variables may be so small and at best nonsystematic or random that as a practical matter and for cost considerations it does not pay to introduce them into the model explicitly. One hopes that their combined effect can be treated as a random variable u_i .

4. *Intrinsic randomness in human behavior*: Even if we succeed in introducing all the relevant variables into the model, there is bound to be some “intrinsic” randomness in individual Y 's that cannot be explained no matter how hard we try. The disturbances, the u 's, may very well reflect this intrinsic randomness.

5. *Poor proxy variables*: Although the classical regression model (to be developed in Chapter 3) assumes that the variables Y and X are measured accurately, in practice the data may be plagued by errors of measurement. Consider, for example, Milton Friedman's well-known theory of the consumption function.¹¹ He regards permanent consumption (Y_p) as a function of permanent income (X_p). But since data on these variables are not directly observable, in practice we use proxy variables, such as current consumption (Y) and current income (X), which can be observable. Since the observed Y and X may not equal Y_p and X_p , there is the problem of errors of measurement. The disturbance term u may in this case then also represent the errors of measurement. As we will see in a later chapter, if there are such errors of measurement, they can have serious implications for estimating the regression coefficients, the β 's.

6. *Principle of parsimony*: Following Occam's razor, we would like to keep our regression model as simple as possible. If we can explain the behavior of Y “substantially” with two or three explanatory variables and if our theory is not strong enough to suggest what other variables might be included, why introduce more variables? Let u_i represent all other variables. Of course, we should not exclude relevant and important variables just to keep the regression model simple.

7. *Wrong functional form*: Even if we have theoretically correct variables explaining a phenomenon and even if we can obtain data on these variables, very often we do not know the form of the functional relationship between the regressand and the regressors. Is consumption expenditure a linear (invariable) function of income or a nonlinear (invariable) function? If it is the former, $Y_i = \beta_1 + \beta_2 X_i + u_i$ is the proper functional relationship between Y and X , but if it is the latter, $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$ may be the correct functional form. In two-variable models the functional form of the relationship can often be judged from the scattergram. But in a multiple regression model, it is not easy to determine the appropriate

functional form, for graphically we cannot visualize scattergrams in multiple dimensions. For all these reasons, the stochastic disturbances u_i assume an extremely critical role in regression analysis, which we will see as we progress.

12.1 The sample regression function (SRF)

By confining our discussion so far to the population of Y values corresponding to the fixed X 's, we have deliberately avoided sampling considerations (note that the data of Table 12.1 represent the population, not a sample). But it is about time to face up to the sampling problems, for in most practical situations what we have is but a sample of Y values corresponding to some fixed X 's. Therefore, our task now is to estimate the PRF on the basis of the sample information. As an illustration, pretend that the population of Table 12.1 was not known to us and the only information we had was a randomly selected sample of Y values for the fixed X 's as given in Table 12.3 unlike Table 12.1, we now have only one Y value corresponding to the given X 's; each Y (given X_i) in Table 12.3 is chosen randomly from similar Y 's corresponding to the same X_i from the population of Table 12.1. The question is: From the sample of Table 12.3 can we predict the average weekly consumption expenditure Y in the population as a whole corresponding to the chosen X 's? In other words, can we estimate the PRF from the sample data? As the reader surely suspects, we may not be able to estimate the PRF "accurately" because of sampling fluctuations. To see this, suppose we draw another random sample from the population of Table 12.1, as presented in Table 12.4. Plotting the data of Tables 12.3 and 12.4, we obtain the scattergram given in Figure 12.3. In the scattergram two sample regression lines are drawn so as to "fit" the scatters reasonably well: SRF_1 is based on the first sample, and SRF_2 is based on the second sample. Which of the two regression lines represents the "true" population regression line? If we avoid the temptation of looking at Figure 12.1, which purportedly represents the PR, there is no way we can be absolutely sure that either of the regression lines shown in Figure 12.4 represents the true population regression line (or curve).

TABLE 12.3 A RANDOM SAMPLE FROM THE POPULATION OF TABLE 12.1

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

TABLE 12.4 ANOTHER RANDOM SAMPLE FROM THE POPULATION OF TABLE 12.1

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

The regression lines in Figure 12.4 are known as the sample regression lines. Supposedly they represent the population regression line, but because of sampling fluctuations they are at best an approximation of the true PR. In general, we would get N different SRFs for N different samples, and these SRFs are not likely to be the same. Now, analogously to the PRF that underlies the population regression line, we can develop the concept of the sample regression function (SRF) to represent the sample regression line. The sample counterpart of (12.8) may be written as:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \tag{12.8}$$

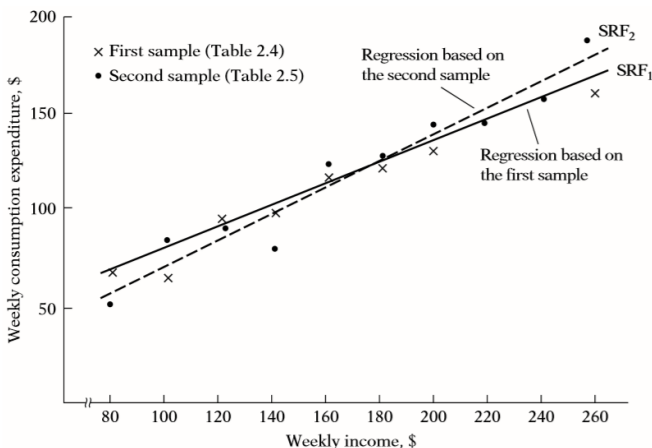


FIGURE 12.4 Regression lines based on two different samples

where \hat{Y} is read as “Y-hat” or “Y-cap”

\hat{Y}_i = estimator of $E(Y|X_i)$

$\hat{\beta}_1$ = estimator of β_1

$\hat{\beta}_2$ = estimator of β_2 Note that an estimator, also known as a (sample) statistic, is simply a rule or formula or method that tells how to estimate the population parameter from the information provided by the sample at hand. A particular numerical value obtained by the estimator in an application is known as an estimate. Now just as we expressed the PRF in two equivalent forms, (12.4) and (12.5), we can express the SRF (12.9) in its stochastic form as follows:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (12.9)$$

where, in addition to the symbols already defined, \hat{u}_i denotes the (sample) residual term. Conceptually \hat{u}_i is analogous to u_i and can be regarded as an estimate of u_i . It is introduced in the SRF for the same reasons as u_i was introduced in the PRF. To sum up, then, we find our primary objective in regression analysis is to estimate the PRF because more often than not our analysis is based upon a single sample from some population. But because of sampling fluctuations

our estimate of the PRF based on the SRF is at best an approximate one.

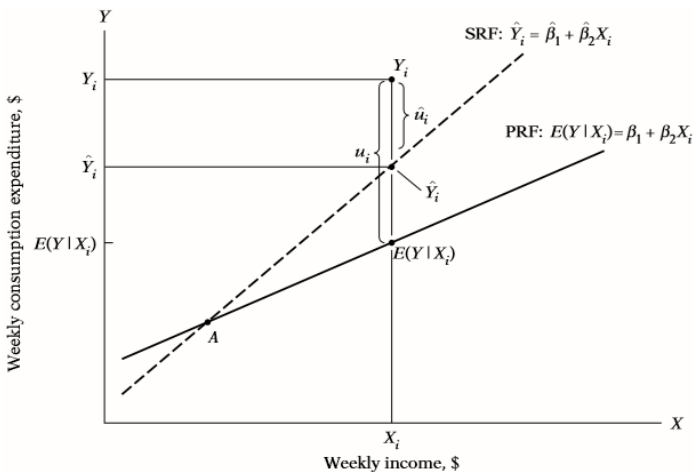


FIGURE 12.5 Sample and population regression lines

This approximation is shown diagrammatically in Figure 12.5. For $X = X_i$, we have one (sample) observation $Y = Y_i$. In terms of the SRF, the observed Y_i can be expressed as:

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{12.10}$$

and in terms of the PRF, it can be expressed as:

$$Y_i = E(Y|X_i) + u_i \tag{12.11}$$

Now obviously in Figure 12.5 \hat{Y}_i overestimates the true $E(Y|X_i)$ for the X_i shown therein. By the same token, for any X_i to the left of the point A, the SRF will underestimate the true PRF. But the reader can readily see that such over- and underestimation is inevitable because of sampling fluctuations. The critical question now is: Granted that the SRF is but an approximation of the PRF, can we devise a rule or a method that will make this approximation as “close” as possible? In other words, how should the SRF is

constructed so that $\hat{\beta}_1$ is as “close” as possible to the true β_1 and $\hat{\beta}_2$ is as “close” as possible to the true β_2 even though we will never know the true β_1 and β_2 ?

The answer to this question will occupy much of our attention in Chapter 3. We note here that we can develop procedures that tell us how to construct the SRF to mirror the PRF as faithfully as possible. It is fascinating to consider that this can be done even though we never actually determine the PRF itself.

Table 12.5 gives data on the level of education (measured by the number of years of schooling), the mean hourly wages earned by people at each level of education, and the number of people at the stated level of education. Ernst Berndt originally obtained the data presented in the table, and he derived these data from the current population survey conducted in May 1985. We will explore these data (with additional explanatory variables) in Chapter 3 (Example 3.3, p. 91). Plotting the (conditional) mean wage against education, we obtain the picture in Figure 12.6. The regression curve in the figure shows how mean wages vary with the level of education; they generally increase with the level of education, a finding one should not find surprising. We will study in a later chapter how variables besides education can also affect the mean wage.

1. The key concept underlying regression analysis is the concept of the conditional expectation function (CEF), or population regression function (PRF). Our objective in regression analysis is to find out how the average value of the dependent variable (or regressand) varies with the given value of the explanatory variable (or regressor).

2. This book largely deals with linear PRFs, that is, regressions that are linear in the parameters. They may or may not be linear in the regressand or the regressors.

3. For empirical purposes, it is the stochastic PRF that matters. The stochastic disturbance term u_i plays a critical role in estimating the PRF.

4. The PRF is an idealized concept, since in practice one rarely has access to the entire population of interest. Usually, one has a sample of observations from the population.

TABLE 12.5 MEAN HOURLY WAGE BY EDUCATION

Years of schooling	Mean wage, \$	Number of people
6	4.4567	3
7	5.7700	5
8	5.9787	15
9	7.3317	12
10	7.3182	17
11	6.5844	27
12	7.8182	218
13	7.8351	37
14	11.0223	56
15	10.6738	13
16	10.8361	70
17	13.6150	24
18	13.5310	31
Total		528

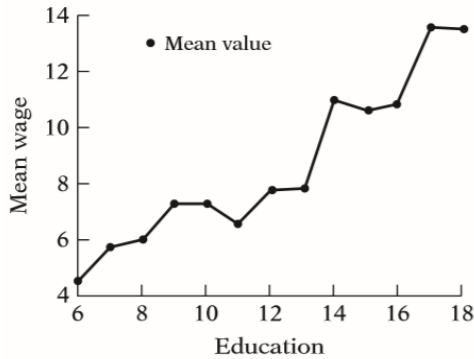


FIGURE 12.6 Relationship between mean wages and education

The method of ordinary least squares is attributed to Carl Friedrich Gauss, a German mathematician. Under certain assumptions, the method of least squares has some very attractive statistical properties that have made it one of the most powerful and popular methods of regression analysis. To understand this method, we first explain the leastsquares principle.

CHAPTER 13. APPLICATIONS OF ANALYTICS IN THE BANKING SECTOR

The Banking Industry is one of the fundamental pillars of modern day's business. The banking sector, due to the sheer nature of the industry, has always had access to vast amounts of credible consumer data. In addition to this, there are two other factors that have played a significant role in this industry being the forerunner in implementing data analytics. The first factor is that the financial sector was quick to understand the significance of computers and was an early adopter of computer technology. In 1950 the Bank of America pioneered the adoption of computers for bookkeeping and check processing. This computer, Electronics Recording Machine, Accounting (ERMA), was developed and prototyped by Stanford Research Institute (SRI) and built by the General Electric Computer Department. Development of ERMA not only revolutionized bookkeeping and check processing but also brought about key changes in bank account data handling. For example, the concept of bank account numbers as a unique identifier was first conceptualized and used under this program and is used to date for handling financial data. By October of 1993 as per the U.S. Census Bureau, the private financial sector had the highest percentage of workers using computers at work. The second factor is the regulatory environment that required the banks to have constant access to data for reporting purposes. In the early 1980s, vital changes were brought about in banking regulations and reporting to provide much needed stability to the international banking system. In 1988 the Basel Committee of Banking Regulations consisting of representatives of central banks of the Group of Ten (G10) published the "Basel Accord" or "Basel I." Basel I provided guidelines to accurately quantify the various financial risk to which the banks were vulnerable and required them to balance these risks with sufficient capital on hand at all times. Basel II⁴ and Basel III⁵ reforms followed Basel I to impact more stability to the banking sector. In terms of data this meant that financial institutions and banks had to have information from the balance sheets as well as off balance sheet entities to be available in an easily accessible and analyzable form for reporting purposes. These factors provided the necessary

foundation for implementing data analytics by providing relevant credible structured data in a digital format. In this chapter a case study of predicting bank-loan default for the banking industry is presented. R and SAS Studio have been employed in this case study. The rest of the chapter is organized as follows. In the next section, key applications of analytics that have redefined the banking industry are discussed. The following section discusses in detail about the logistic regression model, equation, assumptions, and model fit. The section after that provides the main steps and tasks for data exploration in R. In the subsequent section, the logistic regression model is performed and model output of each part is presented in R. The section after this discusses data exploration, model building, and interpreting each part in SAS Studio. In the last section, the summary of the chapter is presented. Today data analytics has become almost mainstream in the banking industry, evident from the fact that 90% of top 100 banks use SAS® and all of the major professional service networks including McKinsey & Company, Boston Consulting Group (BCG), Deloitte, PricewaterhouseCoopers (PwC), Ernst & Young(E&Y), Klynveld Peat Marwick Goerdeler (KPMG), and many more offer financial analytics services for the banking sector. A word cloud built based on a summary of applications provided by SAS® banking industry customers is shown in Figure 13.1.

Analytics is playing an important role in redefining the banking industry in a holistic sense with applications ranging from risk mitigation, fraud detection, customer acquisition and retention, forecasting, and an increase in cross-selling & up-selling. An overview of some of these key applications is provided in this section.

Cross-selling and Up-selling are sales strategies that are applied across various industries with diverse portfolios. The modern-day banking industry is a perfect example with diverse financial offerings like checking and savings accounts, credit and debit cards, short-term and long-term loans, and mortgages. Cross-selling can be defined as a sales strategy to influence existing and new customers to buy complementary products or services in addition to the original purchase while Up-selling is the sales strategy to influence existing customers to purchase a higher valued product or service of the same

product line. With the help of predictive analytics, modern banking is able to collect and consolidate the internal data from different departments like commercial loans, personal loans, term deposits, etc. The analytics team analyzes this data and builds models to derive insights. Creative approaches to gather customer data from social media like Twitter and to analyze the sentiments of customers by applying the sentiment analysis model or text analytics is also gaining significant popularity. Sentiment analysis helps banks to predict the next product offerings that a customer would be interested in purchasing. The effective cross-selling and up-selling is possible only when banks understand what their customers need, like are the customers happy from their products or services? Which type of investments is giving them high revenue? Did customers look for any customized or personalized products or services? And so on. Banks have various channels to cross-sell and up-sell their products and services by interacting with their customers by phone banking, Web, emails, SMS, walk-in, etc. Strong customer relationships and customer loyalty are always an added advantage in driving effective cross-selling and up-selling.

Customer Churn or attrition is defined as when customers leave your service or loss of customers. Attrition in any industry hampers its growth, and it is a sign of failure for the industry. Higher attrition rates signify that customers are not happy from products and service due to several reasons. Banking analytics helps in analyzing the customers' past purchasing behavior, based on their habits, need, investments, etc., and accordingly customize their product to fulfill their need and make them happy and satisfied. A broad spectrum of predictive analytics and machine learning techniques like logistic regression, decision tree, and support vector machines (SVM) are used by banks to identify the probability of customer churn and identify the factors that are responsible for customer churn. Targeted strategies are developed to address factors responsible for churn as well as provide the customers with the highest risk of attrition with promotional offers and advertisements to keep them happy and satisfied. A targeted approach can help significantly reduce costs associated with customer churn and also help banks identify and address systemic problems.

Creating or acquiring new customers is the top priority of every

industry for their growth. Predictive analytics and machine learning algorithms are helping banks in creating new customers by luring them with attractive products and conducting lots of promotional and effective advertisement activities so the customers can be aware of their products and services. Acquiring new customers always costs more than retaining old customers. Banks are deploying analytics to their complex data so they can analyze it for better insights. KYC (know your customer) helps bank to know the background of their customers in a better way, like what types of product and services customers would be interested to utilize? This will help banks to customize the products and services for the customers accordingly and keep track in managing their portfolio in order to keep them happy and satisfied as happy customers tend to purchase more products, stick for longer time, and spreads good word of mouth about their product and services. After creating new customers, improving customer experiences is very important to retain them. To survive in the cut-throat competitive world, banks are giving high priority to improving the customer experience by applying analytical and marketing strategies. Loans are the biggest source of income for the banks. Profitability for the net interest income sources like loans relies on two aspects: interest rates and low default rates. In 2014 there was a mortgage crisis that caused bad impacts on the financial market. The main reason behind this subprime mortgage crisis was heavy borrowing for home loans at lower interest rates with an assumption that house prices would inflate. Lenders used to approve loans with minimal document verification and even if the borrower lacked minimal credentials. Due to these practices, the housing market became extremely volatile and resulted in a huge number of loan defaults and foreclosures. Many U.S. investment banks went bankrupt, one famous example being Lehman Brothers. To prevent the reoccurrence of such a financial holocaust, new banking regulations were put in place and banks implemented strict measures to minimize loan default rates. These methods consisted of more stringent borrower information and document verification practices as well as developed tool sets to statistically identify a possible occurrence of default based on historic data. These tools primarily consisted of analytics and machine learning techniques like logistic regression and a decision tree to quantify the default propensity of

loan applicants.

Bank fraud is defined as the act of using illegal ways or indulging in illegal activities in order to receive money, funds, credits, or assets from a bank or any other financial institutions. The most common fraud threats are Phishing, credit/debit card scams, check scams, accounting fraud, identity theft, and money laundering.

- **Phishing:** In Phishing the customers will receive spam emails. Through these emails, those phishing will try to get access to the customer's account information by requesting customers to update their login and password information, credit card details, etc. Those spam emails seem to come from the bank by creating the exact website of the bank and trapping the customers in such spams. Whenever such type of emails is received, customers should ignore them and if necessary update their respective banks about such scam emails.

- **Credit/Debit Card Scam:** The credit/debit card fraud occurs when an individual credit card is fraudulently used by another individual or any third party to purchase goods and services, etc.

- **Check Scam:** Check fraud occurs when checks are altered and duplicated without the permission of the authorized person, for example, a 100-dollar check can be altered to 1000 dollars. In other cases, a lag in check encashment practices are exploited where the checks are used as forms of payments but no funds are available to support encashment.

- **Accounting Fraud:** Accounting fraud constitutes deliberate falsification of financial records and estimates for purposes of financial gains. Falsification of records may constitute practices like inflating assets, revenues and profit margins, and/or understating liabilities and expenses. The primary motivation of financial fraud is to lure investors or seek financial services and support from banks and other financial institutions. Various data mining, predictive analytics and machine learning technique are gaining popularity for accounting fraud detection including outlier identification, neural networks, decision tree, and regression models.

- **Identity Fraud:** Identity fraud in relation to the financial sector can be defined as the practice of illegally obtaining personal and financial details of an individual for the purposes of financial gains. All this personal information can be used to get the financial

advantages or other benefits such as getting bank account access, loans, credit cards, etc., so the person can take out the cash and loan easily and disappear from the place, which will result in huge financial loss and victims will suffer badly.

- **Money Laundering:** Money laundering occurs when a large amount of money is illegally received and the true source of the money is hidden or concealed (like from terrorist activity, smuggling drugs, etc.) so that the money seems to be received from a legal source.

In the money laundering process, the fraudulent money is deposited in one country and then transferred to other countries so they can use it safely. There are so many acts of money laundering, for example, the inactive bank accounts will become active suddenly and a huge amount of cash is deposited in that account, a person opening multiple accounts for doing their illegal bank transactions with the same name, and many more. Here are some real-time cases where fraud activities in banks were responsible for huge financial loss: for example, the 1873 Bank of England Forgeries, Moldovan bank fraud scandal, and Russian Laundromat, etc. Analytics help banks in predicting the probability of the fraudulent transactions and send the early warning signals so that preventive measures can be taken immediately and effectively, which will help both banks and their customers from huge financial loss.

In the modern-day banking industry, commercial banks have diverse sources of revenue that can be broadly classified as non-interest sources and net interest sources. Non-interest sources constitute deposit account services like ATMs, online payments, safe deposits, and cash management services like payroll processing. Net interest earnings, on the one hand, are essentially comprised of different types of loan offerings where the bank makes money by lending at a higher interest rate than what it pays on deposits. In spite of non-interest incomes on rise, over half of the money made by banks still comes from net interest earnings. A bank's success heavily relies upon how many loans it can give out while maintaining low default rates, where default means the inability of the borrowers to pay back the loan in time. In this case study, application of the logistic regression model to predict which customers have a high propensity to default on a bank loan is

demonstrated. Logistic regression or the logit model is defined as a type of regression model where the dependent or target variable is binary or dichotomous like having only two values as default / not default, fraud / not fraud, live/dead, etc., and independent or explanatory variables can be binary, continuous, ordinal, etc.

When a dependent or target variable has only two values, then then it is considered a binomial logistic regression. In other cases when the target variable has more than two values or outcome categories, then the multinomial logistic regression is applied and if multiple outcome categories are ordered, then it is considered as ordered logit model, for example, the target variable is having multiple ordered categories like excellent, good, and average. The logistic regression or logit model is used to model the data and describe the relationship between the target or dependent binary variable and independent or explanatory variable. The ogistic regression model is used in various fields like medical science, life science, actuarial science, and many more. For example, one needs to predict whether patients will be readmitted to a hospital or not? Whether a customer will default a bank loan or not? Will the customer churn or not? In all these cases, the target variable is binary or dichotomous; having only two values (yes or no). Past and historical data is used to predict the future events by applying the logistic regression model when the dependent variable is binary.

13.1 Logistic Regression Equation

In logistic regression, the probability of a binary outcome is predicted; hence in a logistic regression formula, when the probability is $y = 1$, it is represented as P and when the probability is $y = 0$, it is represented as $1 - P$.

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \tag{13.1}$$

where: $\ln\left(\frac{P}{1-P}\right) =$ Logit function;

$x_1, x_2, \dots, x_n =$ Independent variables;

β_0 : = Logistic regression model intercept;

$\beta_1, \beta_2, \dots, \beta_n$ = Logistic regression coefficients for N independent (x) variables.

In logistic regression the odd ratio is defined as the probability of the occurrence of the events (1 or Yes) to the probability of the occurrence of the nonevents (0 or No). For example, there is a sample of 60 unemployed borrowers, and 50 are defaulting on the bank loan and 10 are not defaulting on the bank loan, while in another sample of 60 employed borrowers, 40 are defaulting on a bank loan and 20 are not defaulting on a bank loan.

The Probability of defaulting on a bank loan by an unemployed borrower is calculated below.

$$P = 50/60 = 0.8$$

$$q = 1 - P = 1 - 0.8 = 0.2$$

For an unemployed borrower, the probability of defaulting on a bank loan is 0.8 and the probability of not defaulting on a bank loan is 0.2. The Probability of defaulting on a bank loan by the employed borrower is calculated below.

$$P = 40/60 = 0.6$$

$$q = 1 - P = 1 - 0.6 = 0.4$$

For employed persons the probability of defaulting on a bank loan is 0.6 and the probability of not defaulting on a bank loan is 0.4. Now in the next step these probabilities are used to compute the odds of bank-loan default for unemployed and employed borrowers.

$$\text{Odds (unemployed)} = 0.8/0.2 = 4$$

$$\text{Odds (employed)} = 0.6/0.4 = 1.5$$

ODDS Ratio: ODDS Ratio is computed as the ratio of the two odds. In this example the odds ratio is computed by dividing the odds

of the unemployed by odds of the employed.

$$OR = 4/1.5 = 2.66$$

Now the conclusion is that for an unemployed person the odds of defaulting bank loans is 2.66 times higher than the odds for defaulting on bank loans for employed persons.

Logistic Regression Curve

The logistic regression curve is an S-shaped or Sigmoidal curve as shown in Figure 13.1. In a sigmoid curve the curve starts with slow linear growth, followed by exponential growth and again slows down to become stable when dependent variable (y) is binary (0,1 or yes, no) and the independent variable (x) is numerical, the logistic regression model fits a curve to display the relationship between the dependent variable (y) and the independent variable (x).

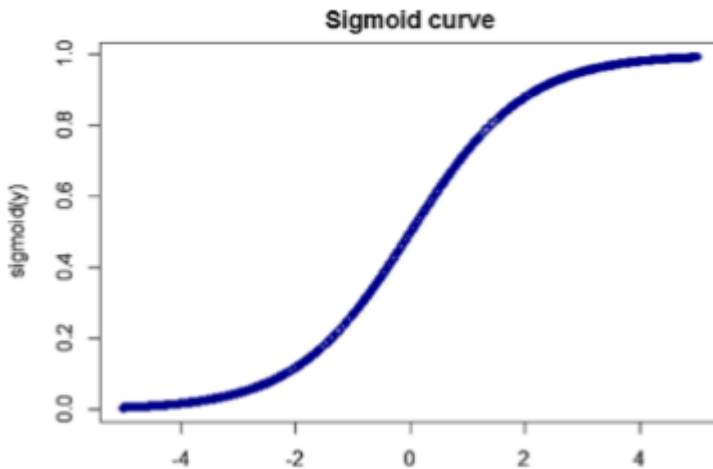


Figure 13.1 Logistic regression sigmoidal curve

All statistical models are developed keeping in mind certain assumptions about the characteristics of data. It is critical that data on which the model is being applied is aligned with these assumptions for obtaining reliable results. In addition to these assumptions, there

are also certain model-specific formats that need to be followed to ensure the data is aligned with the fundamental workings of the model. Some of these changes can be applied when the conditioning and structuring of the data is being performed prior to model implementation. Key logistic regression assumptions and format are listed below:

1. In logistic regression, it is that the target variable must be discrete and mostly binary or dichotomous. This is the key differentiator between the linear regression model and the logistic model and primary requirement for the logistic model.

2. Logistic regression estimates the probability of the occurrence of events $P (Y=1)$ so it is important that the dependent or target variable must be coded accordingly and the desired outcome must be coded to be 1.

3. Regression model should neither be overfitted nor underfitted. Overfitted data can significantly lose the predictive ability due to an erratic response to noise whereas underfitted will lack the accuracy to account for the variability in response in its entirety.

4. There should not be any multicollinearity between the independent variables in the model, and all independent variables should be independent to each other.

5. Independent or explanatory variables must be linearly related to the log odds, and in the logistic regression model, the linear relationship between the dependent or target variable and independent or explanatory variable is not required like as in the linear regression model.

In addition to these assumptions and format requirements, the logistic regression model requires quite a large sample size as maximum likelihood estimates (MLE) are less powerful than ordinary least square (OLS)

Logistic Regression Model Fitting and Evaluation

Model fit tests are done to evaluate how well the model fits the data or how accurately the model can predict the observed values or real values. Lesser is the difference between observed values and predicted values, and better is the model. These differences between the predicted and observed values are also referred to as residuals. Most commonly used methods for evaluating the model fit are the

Likelihood ratio test and Hosmer and Lemeshow test.

1. Likelihood Ratio Test: Likelihood ratio test is used to compare two nested generalized linear models. Like here, the comparison between a reduced (model with few predictors) and a full model is done. The formula which is used to measure the difference of the reduced model and full model with k degrees of freedom is the following:

$$\chi^2 = -2\ln(\text{of reduced model}) - (-2\ln(\text{likelihood of full model})) \quad (13.2)$$

If p value for model fit is less than 0.05, then signify that our full model fits significantly better than our reduced model.

2. Hosmer-Lemeshow Test: Hosmer-Lemeshow test is another approach for reflecting the goodness of fit. In this test with the help of Pearson Chi-Square, it is examined whether the observed proportion of events are same to the predicted probabilities in the model population subgroup. In this test predicted probabilities are divided into deciles, which are in 10 groups.

The formula used for calculating the Hosmer- Lemeshow test is:

$$H = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{E_g} \quad (13.3)$$

Where O_g = no of observed events in the gth group. E_g = no of expected events in the gth group. This test follows a Chi-Square distribution with 8 degrees of freedom, which is calculated as (number of group – 2). The output of this Hosmer-Lemeshow test will be the Chi-Square value and a p value.

If the p value is small, such as the p value less than 0.05, it signifies that the model is a poor fit; and if the p value is large, with the p value greater than 0.05, it signifies that the model is a good fit. The Hosmer-Lemeshow test is usually not recommended because of the selection of the number of subgroups.

13.2 Regression Model

Once the overall model fit tests are done, then the next step is to examine that how important is each independent variable or the significant contribution of the independent variables in the model. After building the logistic regression model, the logistic regression coefficient for an individual independent variable is known and from that we can conclude that if there is a change in 1 unit of n th independent variable, when all other independent variables are held constant, then the log odds of the predicted outcome also changes by B_n units. To find out the importance of an independent variable in logistic regression, there are different types of tests that can be applied like the Wald statistic test and likelihood ratio test.

1. Wald Statistic test: The Wald statistic test is used to find out the importance of the independent variables in a model or how efficient and effective are independent variables in predicting the dependent variable by looking at the logistic regression coefficient of individual independent variables.

The Wald statistic test is calculated by using this formula:

$$W_j = \left(\frac{B_j}{SE \text{ of } B_j} \right)^2 \quad (13.4)$$

where: B_j – is the regression coefficient.

$SE \text{ of } B_j$ – is the standard error of the regression coefficient.

In Wald statistics, the null hypothesis is when the coefficient of interest is equal to zero and the alternate hypothesis is when coefficients are not equal to zero. If the Wald statistic test accepts the null hypothesis, then we can conclude that removing that variable from the model will not impact the model fit.

Wald statistic test is easy to calculate but it has drawbacks, in some cases when data has large coefficients estimates, which tends to increase the standard error and hence lower the value of the Wald statistic. Due to the lower value of the Wald statistic the independent variables can be considered insignificant in the model, though they are playing significant roles in the model.

2. Likelihood ratio test: In addition to evaluation of the model fit, the likelihood ratio test is also used to find out the contribution of individual independent variables in the model. This test is calculated by using the formula:

$$G = -2(\ln(\text{Reduced model}) - \ln(\text{Full model}))$$

$$\chi^2 = -2\ln(\text{of reduced model}) - (-2\ln(\text{likelihood of full model}))$$

(13.5)

The individual independent variable is entered inside the model in an orderly manner, and then the comparison between both the models is done in order to study the contribution of each independent variable in the model. The smaller the deviance between the reduced model and the full model, the better is the correlation between the dependent or target and independent or explanatory variables.

In predictive value validation, a few measurements like the confusion matrix and receiver operating characteristic are used to find out that how accurately the model is predicting the target or dependent variable or, in other words, we can say it measures the model accuracy.

Confusion Matrix

The confusion matrix is a technique used to evaluate the predictive accuracy or the performance of the logistic regression model; it is a classification table with two dimensions consisting of 2 rows and 2 columns. Each column represents the dichotomous or binary predicted class and each row represents the dichotomous or binary actual class.

Let's take an example to explain the confusion matrix for a logistic regression model when the dependent variable is binary or dichotomous. In this example, if we were predicting customer churn, Yes means the customers would churn and No means the customers would not churn. There are a total of 185 customers that were being tested for customer churn. Out of 185 events, the classifier predicted 130 times Yes and 55 times No. In Actual, there are total 125 customers in the sample data who are churning and 60 customers that

do not churn. A sample classification table is provided in Table 13.1. Columns are representing binary or dichotomous predicted outcomes and rows are representing binary or dichotomous actual outcomes.

TABLE 13.1 Classification Table

N = 185 Actual Outcomes	Predicted Outcomes		Total Rows
	NO	YES	
NO	50 (TN)	10 (FP)	60
YES	5 (FN)	120 (TP)	125
Total Columns:	55	130	185

In the above confusion matrix table there are True Positive, True Negative, False Positive, and False Negative cases displayed.

- True Positive (TP): In this case we predicted yes (customers would churn, and in reality they do churn).
- True Negative (TN): In this case we predicted no (customers would not churn), and in reality they don't churn.

False Positive (FP): In this case we predicted yes (customers would churn) but in reality they don't churn; this is also known as a type I error.

- False Negative (FN): In this case we predicted no (customers would not churn) but in reality they do churn, and this is also known as a type II error.

Binary classifier confusion matrix is used for computing various rates like accuracy rate, error rate, true positive rate, false positive rate, specificity, precision, and prevalence. These terms are defined in the following section.

- Accuracy Rate: Accuracy rate is the total fraction of accurately predicted outcomes to the total outcomes. It is calculated by the formula:

$$\frac{TP + TN}{TP + FP + FN + TN} = \frac{120 + 50}{50 + 10 + 5 + 120} = \frac{170}{185} = 0.92 \tag{13.6}$$

where: TP = True positive, TN = True negative, FP = False positive, FN = False negative

- Error Rate: Error rate is also known as misclassification rate, and it is the total fraction of inaccurately predicted outcomes to the total outcomes. It is calculated as:

1-Accuracy rate:

$$1 - 0.91 = 0.09$$

- True Positive Rate: TPR is defined as the total fraction of accurately predicted positive outcomes to the total actual positive outcomes. It is also known as sensitivity. It is calculated by the formula:

$$\frac{TP}{TP + FN} = \frac{120}{125} = 0.96 \quad (13.7)$$

False Positive Rate: False positive rate (FPR) is defined as the fraction of inaccurately predicted positive outcomes to the actual negative outcomes, which in simpler terms means how frequently does the model predict yes, when in reality it is no. It is calculated by the formula:

$$\frac{FP}{TN + FP} = \frac{10}{60} = 0.17 \quad (13.8)$$

Specificity: Defined as the fraction of accurately predicted negative outcomes to the actual negative outcomes, which in simpler terms means how frequently does the model predict no when in reality it is no. It is calculated by the formula:

$$1 - FPR = 1 - 0.17 = 0.83 \quad (13.9)$$

Precision: Precision is defined as the fraction of accurately predicted positive outcomes to the total positive outcomes or in simpler words how frequently is the model correct, when it predicts yes. The formula for precision is:

$$\frac{TP}{TP + FP} = \frac{120}{130} = 0.92 \quad (13.10)$$

Prevalence: Prevalence is defined as the fraction of actual positive outcomes to the total outcomes. it is calculated by the formula:

$$\frac{TP + FN}{TN + TP + FP + FN} = \frac{125}{185} = 0.68 \quad (13.11)$$

Receiver Operating Characteristics (ROC) and Area Under Curve (AUC): ROC curve is a graph with the True positive rate (TPR) or Sensitivity plotted on the y-axis against the False positive rate (FPR) or 1 – Specificity on the x-axis. TPR, in simple words, is how frequently does the model predict yes when in reality it is yes, and FPR is how frequently does the model predict yes, when in reality it is no. The ROC curve is used to measure the binary or dichotomous classifier performance visually and Area under Curve (AUC) is used to quantify the model performance. Generally an AUC greater than 70% is considered as the accurate model. The more is the AUC curve toward 1, the better is the predictive accuracy of the model. Receiver Operating Characteristic (ROC) and Area under Curve is shown in Figure 13.2 and Figure 13.3. The curve above the diagonal represents the better performance, and below the diagonal it represents the worse performance and approaching toward 1, it represents the best performance. The Area under Curve (AUC) values varies from range 0.5 to 1.0 where 0.5 means the poor predictive ability and 1.0 is considered as the accurate or best predictive ability. The graph approaching toward 1 signifies that the model predictive accuracy is high. The higher is the AUC value, the better is the model predictive accuracy; and the lower is the AUC value, the worse is the model predictive accuracy. Figure 13.2 and Figure 13.3 shows the ROC and AUC curve for Model A and Model B respectively. Model A (AUC) value is 68% and Model B (AUC) value is 97%. Hence Model B is having high predictive accuracy compared to Model A.

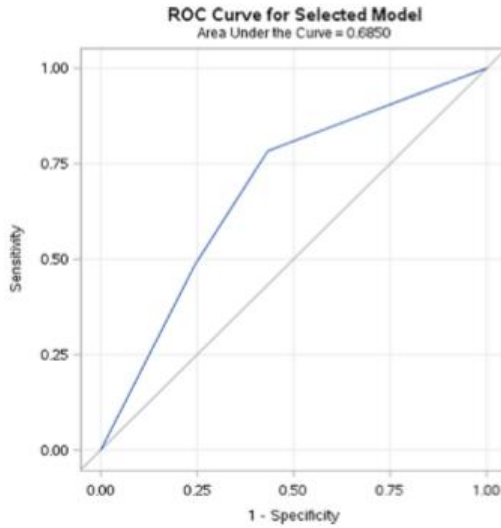


FIGURE 13.2 ROC and AUC Curve for Logistic regression: Model A

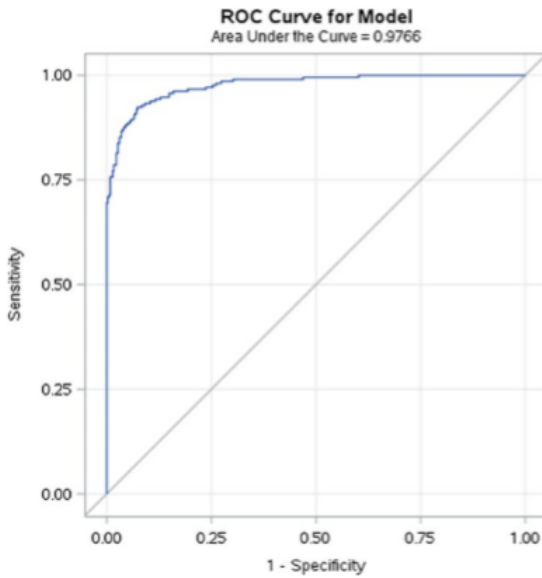


FIGURE 13.3 ROC and AUC Curve for Logistic regression: Model B

13.2 Logistic Regression Model Using R

In this banking case study, we will discuss about the data and the variables used in the data. We then discuss about the exploratory data analysis in R, which is considered as the first step in the data analysis process. Exploratory data analysis helps in taking a broad look at patterns, trends, summary, outliers, missing values, and so on in the existing data. We also discuss about building a logistic regression model and interpretation of its output in R.

Business Problem: To predict the probability of the bank-loan default.

Business Solution: To build the logistic regression model.

In this banking case study, in order to illustrate the bank-loan default, logistic regression model is created and the data is generated synthetically. In this dataset there are a total of 1,000 observations and 16 variables; 13 variables are numeric and 3 variables are categorical. The loan-default data set contains information on 1,000 loan applicants. Default is the dependent or target variable in the data, where 1 denotes presence of loan default and 0 denotes absence of loan default. There are 70% of non-defaulter applicants and 30% applicants are loan defaulters in this data. This data set is used to develop the logistic regression model in order to predict the probability of the bank-loan default. In R, create your own working directory to import the dataset.

#Read the data from the working directory, create your own working directory to read the dataset.

```
setwd("C:/Users/Deep/Desktop/data")
data1 <- read.csv ("C:/Users/Deep/Desktop/data/
loan_default.csv",header=TRUE,sep=",")
data2<-data.frame(data1)
```

Performing Data Exploration

In Exploratory data analysis, we are going to take a broad look at patterns, trends, summary, outliers, missing values, and so on in the existing data. R codes for data exploration and their output is discussed in the section below.

```
#perform exploratory data analysis to know about the data  
# display top 6 rows of dataset to see how data look like  
head (data2)
```

In this case partial output is displayed. False represents that there are no missing values present in the data, and if there are missing values present in any variables in the data, it will be represented as TRUE. If there is a large number of missing values present in the data, then the accuracy of the model is impacted. Model accuracy is impacted by missing values as it reduces statistical power of a study and can produce a biased estimate, leading to incorrect conclusions. To have an effective and accurate model, it is necessary to impute missing values; there are so many missing value imputation techniques like mean, K- nearest neighbor, fuzzy K-means, and many more that can be applied to handle the missing values present in the data.

```

$ Amount      : int  1536 947 1678 1804 1184 475 626 1224 1162 786 ...
$ Saving_amount : int  3455 3600 3093 2449 2867 3282 3398 3022 3475
                2711 ...
$ Emp_duration : int  12 25 43 0 4 12 11 12 12 0 ...
$ Age         : int  38 36 34 29 30 32 38 36 36 29 ...
$ No_of_credit_acc: int  1 1 1 1 1 2 1 1 1 1 ...

```

```
#display the column name of the data
```

```
names(data2)
```

```

[1] "Default"      "Checking_amount" "Term"
[4] "Credit_score" "Gender"          "Marital_status"
[7] "Car_loan"     "Personal_loan"   "Home_loan"
[10] "Education_loan" "Emp_status"      "Amount"
[13] "Saving_amount" "Emp_duration"    "Age"
[16] "No_of_credit_acc"

```

```
#display the summary or descriptive statistics of the data
```

```
summary(data2$Amount)
```

```

      Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
      244  1016   1226   1219   1420   2362

```

```
# Check the missing values present in the data
```

```
sum(is.na(data2))
```

```
[1] 0
```

```
is.na(data2)
```

```

      Default Checking_amount Term Credit_score Gender Marital_status
[1,] FALSE      FALSE FALSE      FALSE FALSE      FALSE
[2,] FALSE      FALSE FALSE      FALSE FALSE      FALSE
[3,] FALSE      FALSE FALSE      FALSE FALSE      FALSE
[4,] FALSE      FALSE FALSE      FALSE FALSE      FALSE
[5,] FALSE      FALSE FALSE      FALSE FALSE      FALSE
[6,] FALSE      FALSE FALSE      FALSE FALSE      FALSE

```

Code Fragment 9 INPut data

```
#To find out the correlation between the variables corr <-
cor.test(data2$Default, data2$Term, method = "pearson") corr
```

Pearson's product-moment correlation

data: data2\$Default and data2\$Term t = 11.494, df = 998, p-value <

2.2e-16

alternative hypothesis: true correlation is not equal to 0.95 percent
confidence interval: 0.2859874 0.3955278

sample estimates:

cor

0.3419185

The Pearson moment correlation method is used to find out the correlation between the two variables. There is moderate positive correlation (34%) between the default and term variable; it means both of the variables are directly proportional to each other. Default increases with the increase in term. To determine whether the correlation between variables is significant, we need to compare the p value to the significance level (0.05). In this case the p-value for the correlation between default and term is less than the significance level of 0.05, which indicates that the correlation coefficient is significant. Correlation is generally of three types: positive correlation, negative correlation, and neutral correlation. When both the variables are directly proportional to each other, then there is positive correlation between the variables; when both the variables are indirectly proportional to each other, then there is negative correlation between the variables; and when there is no relationship between the two variables, then there is neutral correlation.

Model Building and Interpretation of Full Data

In this case, initially we are using complete data for model building, and interpretation and in the next section, we are randomly splitting data into two parts: train data set and test data set. Train data set is used to build the model and test data set is used to test the performance of the model. There are variations seen in the output of both of the models due to the sampling. A detailed description is seen in the section below.

#building logistic regression model using glm on full data

```
fullmodel1 <-glm(Default~.,data = data2,family=binomial  
(link=logit ))
```

summary(fullmodel1)

Call: glm(formula = Default ~ ., family = binomial(link = logit), data = data2)

Deviance Residuals: Min 1Q Median 3Q Max -3.1820
-0.1761 -0.0439 0.0415 3.3985

Looking at the coefficients in Listing 2-1, there are variables with p values less than 0.05; these variables are the significant variables in the model. Other variables like Gender, Marital_status, Car_loan, Personal_loan, Home_loan, Education_loan, Emp_status, Amount, Emp_duration, and No_of_credit_acc where p values are greater than 0.05, are not significant. Nonsignificant variables are removed from the model and the model is rerun.

Call: glm(formula = Default ~ ., family = binomial(link = logit), data = data2)

Deviance Residuals: Min 1Q Median 3Q Max -3.1820
-0.1761 -0.0439 0.0415 3.3985

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	39.6415229	4.7284136	8.384	< 2e-16	***
Checking_amount	-0.0050880	0.0006759	-7.528	5.14e-14	***
Term	0.1703676	0.0520728	3.272	0.00107	**
Credit_score	-0.0109793	0.0020746	-5.292	1.21e-07	***
GenderMale	0.1950806	0.5095698	0.383	0.70184	
Marital_statusSingle	0.3351480	0.4920120	0.681	0.49576	
Car_loan	-0.6004643	2.7585197	-0.218	0.82768	
Personal_loan	-1.5540876	2.7585124	-0.563	0.57318	
Home_loan	-3.5684378	2.8457131	-1.254	0.20985	
Education_loan	0.6498873	2.7894965	0.233	0.81578	
Emp_statusunemployed	0.5872532	0.3474376	1.690	0.09098	.
Amount	0.0008026	0.0005114	1.569	0.11653	

Code Fragment 10 Coefficient Table


```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73  on 999  degrees of freedom
Residual deviance: 297.65  on 984  degrees of freedom
AIC: 329.65

Number of Fisher Scoring iterations: 7

# removing insignificant variables inorder to build final logistic model on
full data

fullmodel2<-glm(Default~Checking_amount+Term+Credit_score
+Saving_amount+Age,data = data2,family=binomial(link=logit))

summary(fullmodel2)

```

Code Fragment 11 The coefficients (labeled Estimates), their standard errors, the Z values, and the P values

The coefficients for Checking_amount, Term, Credit_score, Saving_amount, and Age are statistically significant as their P values are less than 0.05. In the logistic regression model for a one-unit increase in the predictor variables, it will display the change (increase or decrease) in the log odds of the outcome. It can be explained as:

1. For every one-unit change in Checking_amount, the log odds of bank-loan default versus bank-loan non-default decreases by (-0.004). It signifies that the probability of bank-loan default decreases with the increase in the Checking_amount. Higher is the Checking_amount, lesser is the probability of bank-loan default and lower is the Checking_amount, the higher is the probability of bank-loan default.

2. Similarly it is for Term that one-unit change in Term, the log odds of bank-loan default versus bank-loan non-default increases by (0.174). It signifies that the probability of bank-loan default increases with the increase in the Term. The greater the Term, the higher is the probability of bank-loan default; and the less is the Term, the lower is the probability of bank-loan default.

#splitting data set into training and validation dataset in 70:30

Now the data is split into two parts, train data set and test data set;

the splitting ratio is 70:30, which means 70% of data contribute to the train dataset and 30% percent of data contribute to the test dataset. Train dataset is used to build the model and test dataset is used to test the performance of the model.

```
train_obs <- floor (0.7*nrow (data2)) print(train_obs)
#Set seed in order to reproduce the sample
set.seed(2) train_ind <-
sample(seq_len(nrow(data2)),size=train_obs) test <- -train_ind
#No of observations in train dataset train_data<-data2[train_ind,]
# No of observations in test dataset
test_data<-data2[-train_ind,]
```

13.3 Model Building and Interpretation of Training and Testing Data

```
#Building logistic regression model using glm on training data
model1<-glm(Default~.,data=
train_data,family=binomial(link=logit))
summary(model1)
Call: glm(formula = Default ~ ., family = binomial(link = logit),
data = train_data)
```

```
Deviance Residuals:  Min      1Q  Median      3Q      Max
2.44674 -0.13513 -0.03275  0.02618  3.05506
```

Looking at the coefficients in Listing 2-3, there are variables whose p values are less than 0.05; hence all those variables are the significant variables in the model, but there are few variables like Gender, Marital_status, Car_loan, Personal_loan, Home_loan, Education_loan, Amount, Emp_duration, and No_of_credit_acc whose p values are not less than 0.05; hence all these variables are not significant variables in the model.

The coefficient for the categories of Emp_status interpretation is a bit different.

1. For example, defaulting on a bank loan with a Emp_statusunemployed versus with a Emp_statusemployed (omitted category) increases the log odds of the bank-loan default by (0.912). It signifies that the probability of bank-loan default increases with the increase in the Emp_statusunemployed versus Emp_statusemployed, which is considered here as the reference group,

hence omitted category. More is the Emp_statusunemployed, the higher is the probability of bank-loan default.

#Check for variance inflation factor, VIF > 5 to 10- high correlation

Variance Inflation Factor is a test for multicollinearity. VIF test helps in identifying the multicollinearity between the independent variables present in the data.

```
#install car package install.packages("car") library(car)
vif(model2)
```

Variance Inflation Factor is a test for multicollinearity, and this VIF test helps in identifying the multicollinearity between the independent variables present in the data. VIF values higher than 5 to 10 display that there is multicollinearity present and the variables having high multicollinearity values is dropped from the model building, so there is no impact in the accuracy of the model. In Listing 2-5, none of the variables is shown as having VIF values higher than 5 to 10; hence no multicollinearity present in the model.

After applying the predict function in order to predict the probability of bank-loan default and setting the probability cut-off value, like probability greater than 70%, display it as 1 (bank-loan default) else 0 (bank-loan non-default). It means all predicted as 1 or probability greater than 70 percent are considered as high likelihood of defaulting a bank loan.

```
# Predicting the model using test data
Prob <-predict(model2,test_data,type = "response")
prob1<- data.frameProb
# setting the cutoff for probability values
results <- ifelse(prob1 > 0.7,1,0)
```

Predictive Value Validation

In predictive value validation, a few measurements like the confusion matrix and receiver operating characteristic are used to find out that how accurately the model is predicting the target or dependent variable, or in other words we can say it measures the model accuracy. A detailed explanation of each concept is already explained in the above predictive value validation in the logistic regression model section.

```

#Display the confusion matrix or classification table
table(testing_high,results) #Calculating the error rate
misclassificationerror <- mean(results != testing_high)
misclassificationerror
# Calculating the accuracy rate
accuracyrate <- 1-misclassificationerror print(accuracyrate)

```

```

#Calculating the confusion matrix and statistics
install.packages("caret") library(caret)
confusionMatrix(prediction, testing_high,positive ="1")

```

In the confusionMatrix if there are only two factor levels in the data then the first level by default is used as the “positive” result. In this data that is not the case as the first level, 0 is negative so to here we override the default and correctly assign positive as “1”.

```

Accuracy : 0.9167
95% CI : (0.8794, 0.9453)
No Information Rate : 0.6833
P-Value [Acc > NIR] : <2e-16

Kappa : 0.8024
Mcnemar's Test P-Value : 0.1096

Sensitivity : 0.8211
Specificity : 0.9610
Pos Pred Value : 0.9070
Neg Pred Value : 0.9206
Prevalence : 0.3167
Detection Rate : 0.2600
Detection Prevalence : 0.2867
Balanced Accuracy : 0.8910

'Positive' Class : 1

```

Code Fragment 12 Output data

A detailed explanation of the statistic value like sensitivity, specificity, etc., is explained in detail in the predictive value validation in logistic regression model section.

```

# conducting Receiver operating characteristic (ROC) test and
Area under curve (AUC)
install.packages("ROCR")
library(ROCR)
# Compute AUC for predicting Default with the model
prob <- predict(model2, newdata=test_data, type="response")
pred <- prediction(prob, test_data$Default)
pmf <- performance(pred, measure = "tpr", x.measure =
"fpr")
plot(pmf,col= "red" )
auc <- performance(pred, measure = "auc")
auc <- auc@y.values[[1]] auc [1] 0.9688318

```

Receiver Operating Characteristic (ROC) and Area Under Curve is displayed in Figure 13.4. In the y-axis true positive rate (TPR) or sensitivity is displayed and x axis false positive rate (FPR) or 1 – specificity is displayed, the curve approaching toward 1 represents the best performance of the model.

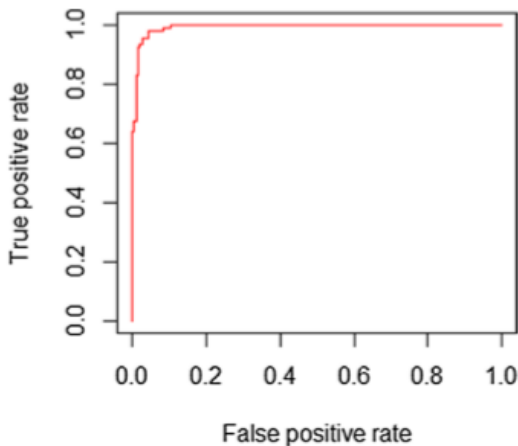


FIGURE 13.4 ROC & AUC curve

Area under Curve (AUC) values greater than 70% are considered as the model with high predictive accuracy. In this case study

illustrating how the logistic regression model, Area under curve (AUC) value is around 96%, which is considered an accurate model.

In this chapter we learned about application of data analytics in the banking sector, which is one of the first adopters of computers and data analytics. The model discussed in the chapter is the logistic regression model. Various characteristics, features, and assumptions of the model were discussed. Practical application of this model in a real-life scenario was demonstrated by help of a case study to predict bank-loan defaults, which is a big challenge for the banking sector, and it can help the bank better understand its customers in terms of lending bank loans. Model development, execution, visualization, result interpretations, and evaluation were also performed by using both R.

CONCLUSIONS

Maxims of Business Economics

Economists learn a lot through detailed academic work and then receive additional practical training. When prompted about their training, I have noted that professionals early in their career recite specific academic classes as well as opportunities to learn skills. The same question, put to successful business people later in their careers, often produces a surprisingly different set of lessons that has enabled them to succeed. Most of the remainder of this book is devoted to skills in applying business economics, including quantitative techniques, analytical techniques, and presentation techniques, as well as methodological notes and cautions. In this section, however, we summarize a set of maxims that will enable economists to be successful in their careers regardless of how much math they know. With that said, here are six maxims for business economics:

1. Remember – it is a social science.

There is a joke that “an economist is someone who looks at reality and wonders whether it would work in theory.” Like most jokes, there is a kernel of truth in that observation. Economics models human behavior. Economics applied to real-world situations, the focus of this book, models human behavior when such humans are placed in common situations. Applying economic thinking to actual human behavior forces one to confront the untidy, unexpected, and occasionally unexplainable behavior of fellow members of one’s species. Thus, the first maxim is to remember that economics is a social science, by which we mean economics only works when it models human behavior. The corollary to this is: when human beings do not behave like the model, it is probably the model that is wrong.

2. All economics is microeconomics.

All social science is an attempt to explain the behavior of individuals. Of course, the aggregate of those individual behaviors is the behavior of groups, including groups of consumers, producers, taxpayers, government officials, and other groups that are important to a business economist. However, when tempted to describe

macroeconomic behavior that is at odds with the microeconomic foundation for individual behavior, remember this maxim: if it does not work at the microeconomic level, it will not work at the macroeconomic level.

Recent pioneers in economics, including at least a few winners of the Nobel Prize, have documented failures of macroeconomics that stem directly from violation of this maxim. Robert Lucas summarized one set of such errors, in macroeconomics as follows: The prevailing strategy for macroeconomic modeling in the early 1960s held that the individual or sectoral models arising out of this intertemporal theorizing could then simply be combined to form a single model, the way Keynes, Tinbergen, and their successors assembled a consumption function, an investment function, and so on into a model of an entire economy. But models of individual decisions over time necessarily involve expected future prices. Some microeconomic analyses treated these prices as known; others imputed adaptive forecasting rules to maximizing firms and households. However it was done, the “church supper” models assembled from such individual components implied behavior of actual equilibrium prices and incomes that bore no relation to, and were in general grossly inconsistent with, the price expectations that the theory imputed to individual agents. Going back to the times before Alfred Nobel, pioneers of economics such as Adam Smith, David Ricardo, and others grounded their economic arguments on the behavior of individuals. Modern pioneers of business similarly achieve wealth and influence by providing products that individuals (persons and companies) want to buy and not by providing products that on average, or in aggregate, the economy as a whole will want to buy. While your goal may or may not be winning Many highly technical, mathematically intense, and impressive-looking economic analyses are completely worthless because the author did not think through the problem before producing his or her equations. The proliferation of software, including spreadsheets, statistical packages, and simulation software such as MATLAB, greatly expands the power of the economist to model behavior and provide detailed, multiperiod forecasts or analyses. Power, however, can be misused. Simply because one can extrapolate for another 10 years and produce a nice chart, does not mean the exercise is useful. A recurring theme

in this book is the importance of thinking first before rushing to create a complicated model. I should also note the academic corollary to this maxim: think first – differentiate later.

4. Know your data.

This is an area where business economics is quite different from theoretical economics. Analysts should always spend time understanding the source, reliability, meaning, and deficiencies in the data. Most data have some deficiencies. Some data are produced by parties with an interest that should be understood before pasting the data into an equation. Be sure that you always know your data.

5. Be ethical.

There are many opportunities in the world of academic and business economics to cut ethical corners in an attempt to produce a more sensational analysis, masquerade someone else's work as your own, or simply toot your own horn inappropriately. No book by a human author can reform this essentially human deficiency. However, if you do not recognize a standard, you will certainly not follow it. Therefore, establish an ethical standard in your work and discipline yourself and your coworkers to follow it.

6. Be humble.

The best thinkers start off by knowing what they do not know and admitting it. If you are reading this book, you are, in all likelihood, an intelligent person who either now or in the future will be considered an expert on some topics. Resist the temptation to claim expertise in an area in which you are an intellectual bystander. In your own field, be careful to cite the contributions of others and to identify the bases of your assumptions, the sources of your data, and the limitations of your knowledge and experience.

These maxims could be applied in many fields. I hope the remainder of this book imparts a great deal of practical knowledge about modeling economic behavior, and that both author and reader follow these simple maxims

REFERENCES

1. Saar-Tsechansky, M.; Provost, F., Handling Missing Values when Applying Classification Models. *J. Mach. Learn. Res.* 2007, 8, 1623–1657.
2. Sachin Gavankar, S. S., Decision Tree: Review of Techniques for Missing Values at Training, Testing and Compatibility. In *Third International Conference on Artificial Intelligence, Modelling and Simulation*, IEEE Computer Society: Sabah, Malaysia, 2015; 122–126.
3. Sebastian Land, S. F. *RapidMiner 5: RapidMiner in academic use 2012*.
4. Majeed, H.; Amir, Y., A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach. *Management & Marketing* 2017, 12 (1), 1–11.
5. De Gooijer, J. G.; Hyndman, R. J., 25 years of time series forecasting. *International Journal of Forecasting* 2006, 22 (3), 443–473.
6. Kusriani, K., Grouping of Retail Items by Using K-Means Clustering. *Procedia Computer Science* 2015, 72, 495-502.
7. Angerer, A. *The Impact of Automatic Store Replenishment Systems on Retail*. University of St. Gallen, Austria, 2005.
8. Badr Hssina, A. M., Hanane Ezzikouri and Mohammed Erritali, A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications 2014, 13–19.
9. Ball, A. *Review of Data Management Lifecycle Models: 2012*; p 14.
10. Berman, J. J., *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Elsevier Science: 2013.
11. Bondos, I., Price and service bundling- An example of the telecommunications market in Poland. In *Human capital without borders: Knowledge and learning for Quality of life*, Portoroz, Slovenia, 2014; 10.
12. Bosari, J. *What Really Goes Into Determining Your Insurance Rates?* 2013.

13. Bosari, J. What Really Goes Into Determining Your Insurance Rates? 2013.
14. Bramer, M., Avoiding Overfitting of Decision Trees. In Principles of Data Mining, Springer London: London, 2007; 119–134.
15. Braun, M. R.; Altan, H.; Beck, S. B. M., Using regression analysis to predict the future energy consumption of a supermarket in the UK. Applied Energy 2014, 130, 305-313.
16. Braun, M. R.; Altan, H.; Beck, S. B. M., Using regression analysis to predict the future energy consumption of a supermarket in the UK. Applied Energy 2014, 130, 305-313.
17. Buckinx, W.; Van den Poel, D., Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research 2005, 164 (1), 252-268.
18. Buckinx, W.; Van den Poel, D., Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. European Journal of Operational Research 2005, 164 (1), 252-268.
19. Cheshti, M.; Taher Ahmadi Shadmehri, M.; Safaye Nikoo, H., Forecasting Bank Deposits Rate: Application of ARIMA and Artificial Neural Networks. 2014; Vol. 7, 527-532.
20. Cheshti, M.; Taher Ahmadi Shadmehri, M.; Safaye Nikoo, H., Forecasting Bank Deposits Rate: Application of ARIMA and Artificial Neural Networks. 2014; Vol. 7, 527-532.
21. Cooper, D.; Schindler, P., Business Research Methods: 12th Edition. McGraw-Hill Higher Education: 2013.
22. Crone, T. P. K. a. S. F. In Demand models for the static retail price optimization problem – A Revenue Management perspective, OpenAccess Series in Informatics (OASIs), Granado, P. C. D.; Joyce-Moniz, M.; Ravizza, S., Eds. Schloss Dagstuhl--LeibnizZentrum fuer Informatik: 2014; 101–125.
23. D. Lukianov, K. Bepanskaya-Paulenko, V. Gogunskii, O. Kolesnikov. Development of the markov model of a project as a system of role communications in a team. Eastern-European Journal of Enterprise Technologies, 21-28
24. Fabian Pedregosa, G. V., Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel,

- Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay, Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011, 12, 2825-2830.
25. Fisher, E. W.; McKenney, J. L., The development of ERMA Banking system: Lessons from History. *IEEE Annals of the History of Computing* 1993, 15 (1), pp. 44–57.
 26. Gallego, G.; Wang, R., Multiproduct Price Optimization and Competition Under the Nested Logit Model with Product Differentiated Price Sensitivities. *Operations Research* 2014, 62 (2), 450–461.
 27. Gepp, A.; Kumar, K.; Bhattacharya, S., Business failure prediction using decision trees. *Journal of Forecasting* 2010, 29 (6), 536-555.
 28. Gepp, A.; Kumar, K.; Bhattacharya, S., Business failure prediction using decision trees. *Journal of Forecasting* 2010, 29 (6), 536-555.
 29. Gerbing, D. Time Series Components. (accessed Nov 15, 2017).
 30. Harvey, C. Big Data Challenges 2017.
 31. He, Y.; Han, J.-c.; Zeng, S.-h., Classification Algorithm based on Improved ID3 in Bank Loan Application. 2012; 1124-1130.
 32. He, Y.; Han, J.-c.; Zeng, S.-h., Classification Algorithm based on Improved ID3 in Bank Loan Application. 2012; 1124-1130.
 33. Hu, Y.-C.; Ansell, J., Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research* 2007, 183 (3), 1595-1606.
 34. Hu, Y.-C.; Ansell, J., Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research* 2007, 183 (3), 1595-1606.
 35. Hyndman, R. J.; Athanasopoulos, G., *Forecasting: principles and practice*. OTexts: 2014.
 36. Institute, S., *SAS/ETS 12.1 User's Guide*. Sas Inst: 2012.
 37. International Convergence of Capital Measurement and Capital Standards A Revised Framework Comprehensive Version. Supervision, B. C. o. B., Ed. Bank of international settlements: <http://www.bis.org/publ/bcbs128.pdf>, 2006.

38. International convergence of capital measurements and capital standards. Supervision, B. C. o. B., Ed. Bank of International Settlements: Basel, Switzerland, 1988.
39. K. Kolesnikova, O. Mezentseva and O. Savielieva, "Modeling of Decision Making Strategies In Management of Steelmaking Processes," 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2019, 455-460.
40. Kingsford, C.; Salzberg, S. L., What are decision trees? *Nature biotechnology* 2008, 26 (9), 1011–1013.
41. Kusrini, K., Grouping of Retail Items by Using K-Means Clustering. *Procedia Computer Science* 2015, 72, 495-502.
42. Maimon, O.; Rokach, L., *Data Mining and Knowledge Discovery Handbook*. Springer US: 2006.
43. Martz, E., Understanding Qualitative, Quantitative, Attribute, Discrete, and Continuous Data Types. *The Minitab blog*: <http://blog.minitab.com/blog/understanding-statistics/understanding-qualitative-quantitative-attributediscrete-and-continuous-data-types>, 2017.
44. Miranda, M. D.; Sassi, R. J., Using Sentiment Analysis to Assess Customer Satisfaction in an Online Job Search Company. In *Business Information Systems Workshops: BIS 2014 International Workshops*, Larnaca, Cyprus, May 22-23, 2014, Revised Papers, Abramowicz, W.; Kokkinaki, A., Eds. Springer International Publishing: Cham, 2014; 17–27.
45. Mohsen, A., RFID: an enabler of supply chain operations. *Supply Chain Management: An International Journal* 2007, 12 (4), 249–257.
46. Morisi, T. L. Commercial Banking transformed by Computer technology *Monthly Labor Review*, 1996, 30–36.
47. Morteza Arab, M. K. S., Proposing a new clustering method to detect phishing websites. *Turkish Journal of Electrical Engineering & Computer Sciences* 2017, 25, 4757-4767.
48. Morteza Arab, M. K. S., Proposing a new clustering method to detect phishing websites. *Turkish Journal of Electrical Engineering & Computer Sciences* 2017, 25, 4757-4767.
49. Nielsen, F. A., *Data Mining with Python (Working draft)*. http://www2.compute.dtu.dk/pubdb/views/edoc_download.

- php/6814/pdf/imm6814.pdf, 2017; 103.
50. Nor Mazlina Abu Bakar, I. M. T., Applying Multiple Linear Regression and Neural Network to Predict Bank Performance. *International Business research* 2009, 2 (4), 8.
 51. Nor Mazlina Abu Bakar, I. M. T., Applying Multiple Linear Regression and Neural Network to Predict Bank Performance. *International Business research* 2009, 2 (4), 8.
 52. O. Mezentseva. Intellectualization of enterprise management using business intelligence instruments / *Technology Audit and Production Reserves*. – Харків, 2019. – 4/4(48). 19-24.
 53. O.E. Kolesnikov, D.V. Lukianov, O.I. Sherstyuk, K.V. Kolesnikova. Project Manager Job Description as One of Project Management Key Success Factors. *Herald of advanced information technology*, 215-228.
 54. O.E. Kolesnikov, O.I. Sherstyuk. The range method use for forming the necessary set of project team competencies. *Bulletin of NTU “KhPI”. Series: Strategic Management, Portfolio*.
 55. Ordenes, F. V.; Theodoulidis, B.; Burton, J.; Gruber, T.; Zaki, M., Analyzing Customer Experience Feedback Using Text Mining: A Linguistics- Based Approach. *Journal of Service Research* 2014, 17 (3), 278-295.
 56. Raileanu, L. E.; Stoffel, K., Theoretical Comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence* 2004, 41 (1), 77-93.
 57. Rose, R. *Defining analytics: A conceptual framework* ORMS Today, 2016.
 58. Rossum, G. v.; Jr., F. L. D. *The Python Language Reference Release 2.6.3*. Python Software Foundation: 115.
 59. Russell, M., *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*. O'Reilly Media, Incorporated: 2011.
 60. Rutkowski, L.; Jaworski, M.; Pietruczuk, L.; Duda, P. The CART decision tree for mining data streams. *Information Sciences* 2014, 266 (Supplement C), 1-15.
 61. Seeja, K. R.; Zareapoor, M., *FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining*. *The Scientific World Journal*, 2014, 10.

62. Seeja, K. R.; Zareapoor, M., FraudMiner: A Novel Credit Card Fraud Detection Model Based on Frequent Itemset Mining. *The Scientific World Journal* 2014, 2014, 10.
63. Simon, P., *Analytics: The Agile Way*. Wiley: 2017.
64. Sokol, D. L.; Jonas, J. *Using Entity Analytics to Greatly Increase the Accuracy of Your Models Quickly and Easily*.
65. Studio, R. *Take control of your R code*.
66. V. Gogunskii, O. Kolesnikov, G. Oborska, A. Moskaliuk, K. Kolesnikova. Representation of project systems using the markov chain. *Eastern-European Journal of Enterprise Technologies*, Vol. 4, 60-65.
67. V. Morozov, O. Kalnichenko, O. Mezentseva. The method of interaction modeling on basis of deep learning of neural networks in complex it-projects. *International Journal of Computing*. 2020, T. 19, Vol. 1, 88-96.
68. Viktor Morozov, Olga Mezentseva, Maksym Proskurin. Application of Game Theory for Decisions Making on the Development of IT Products. "Lecture Notes in Computational Intelligence and Decision Making Proceedings of the International Scientific Conference" *Intellectual Systems of Decision Making and Problems of Computational Intelligence* "(ISDMCI'2020)
69. W. N. Venables, D. M. S.; Team, a. t. R. C. *An Introduction to R: Notes on R: A Programming Environment for Data Analysis and Graphics V3.4.4*, 105.
70. Wendler, T.; Grötrrup, S., *Data Mining with SPSS Modeler: Theory, Exercises and Solutions*. Springer International Publishing: 2016.
71. XSI. *Updated for 2017: The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity* 2017.
72. Zhou, Z. H., *Ensemble Methods: Foundations and Algorithms*. Taylor & Francis: 2012.
73. Бритченко И.Г. *Маркетинг менеджмент: новые решения / А.А. Шубин, Р.К. Перепелиця, Г.И. Бритченко. – Донецк: ДонНУЕТ им. М. Туган-Барановского, 2007. – 326 с.*
74. Брітченко І.Г. *Реальний капітал в сучасних умовах / Брітченко І.Г., Товстик В.А. // Схід. – № 3(27). – 1999. – С. 17-23.*

75. Брітченко І.Г. Маркетинг у банках: навч. посіб. / Брітченко І.Г., Белявцев М.І., Тягунова Н.М. – Полтава: РВЦ ПУСКУ, 2008. – 345 с.
76. Бритченко И.Г. MARKETING MANAGEMENT ОРГАНІЗАЦІЙ: потенціал и система / Брітченко І.Г., Бритченко Г.И. – Донецк: ДонНУ, 2001. – 279 с.
77. Britchenko Igor. Reputation risks, value of losses and financial sustainability of commercial banks / Kunitsyna N., Britchenko I., Kunitsyn I. // Entrepreneurship and Sustainability Issues. 5(4): 943-955. [https://doi.org/10.9770/jesi.2018.5.4\(17\)](https://doi.org/10.9770/jesi.2018.5.4(17)) ISSN 2345-0282.
78. Бритченко И.Г. Организационные основы взаимодействия спортивного бизнеса: постановка и последовательное движение к инновационному управлению в Украине / Бритченко И.Г., Саенко В.Г. // Електроний науковий журнал ВСУ ім. Черноризец Храбр (Варна, Болгарія). – № 9. – 2016. – 32 с. (ISSN 1313-7514) Режим доступу: <http://ejournal.vfu.bg>.
79. Брітченко І.Г. Підвищення конкурентоспроможності банків на ринку банківських послуг України // Науковий вісник УжНУ (Серія економіка). Випуск 1 (45). – Том. 1/ 2015. – С. 176-180. (ISSN: 2409-6857)
80. Брітченко І.Г., Момот О.М. Особливості використання скорингових систем у банківському кредитуванні фізичних осіб// Збірник наукових праць КНЕУ «Фінанси облік і аудит». – Вип. 17. – 2011 – К: КНЕУ. – С. 31-37. (ISSN 2310-9734)
81. Бритченко И.Г. Управление банковским рыноководением. – Донецк: ИЭПИ НАН Украины. – 2003. – 200 с.
82. Брітченко І.Г. Маркетинг як чинник підвищення ефективності функціонування вищого навчального закладу/Брітченко І.Г., Чайка І.П. // Наук. вісник УжНУ. Сер. Економіка. Спецвипуск. – № 33/2011. Частина 2. – 2011. – Ужгород: УжНУ. – С. 299 – 302.
83. Брітченко І.Г., Момот О.М., Саенко В.Г. Економічні проблеми розвитку підприємницької функції в секторах економіки національного господарства / під наук ред. проф. Брітченко І.Г. – Донецьк-Полтава: ООО

«Техсервіс», 2012. – 640 с.

84. Мікловда В.П., Брітченко І.Г., Кубіній Н.Ю., Дідович Ю.О. Ефективність стратегічного управління підприємствами: сучасні проблеми та перспективи їх вирішення. – Полтава: ПУЕТ, 2013. – 240 с.
85. Брітченко І.Г. Переваги застосування нових кредитних технологій // Схід. – № 3 (53). – 2003. – С. 55–58.
86. Брітченко І.Г. Виртуальные банки и их связь с реальной банковской системой // Наука, релігія, суспільство. – № 2. – 2001. – С. 3–10.
87. Брітченко І.Г., Перепелиця Р.К. Оцінка кредитоспроможності емітента корпоративних облігацій // Вісник НБУ. – № 11 (117). – 2005. – С. 32–37.
88. Брітченко І.Г., Князевич А.О. Контролінг: навч. посіб. // І.Г. Брітченко, А.О. Князевич. – Рівне: Волинські обереги, 2015. – 280 с.
89. Брітченко І.Г. Концепція виробничого менеджменту підприємства // Зб. наукових праць “Торгівля і ринок України”. – Випуск 12. – Том II. – Донецьк: ДонДУЕТ, 2001. – С. 31–36.
90. Брітченко І.Г. Формування системи державного управління інвестиційним процесом як головна передумова економічного зростання в Україні / Брітченко І.Г., Кужелев М.А. // Вестник ДонГУЭТ. – №12. – Донецьк: ДонГУЭТ, 2001. – С. 15–22.
91. Брітченко І.Г. Вплив малого бізнесу на регіональний розвиток // Науковий вісник ПУЕТ. – № 4 (43). – 2010. – Полтава: ПУЕТ, 2010. – С. 15 - 19.
92. Ножова Г.В., Брітченко І.Г. Інноваційна діяльність в контексті стратегічного управління // Науковий вісник ПУЕТ. – № 6 (51) / 2011. Частина 1 – Полтава: ПУЕТ, 2011. – С. 34–39.
93. Брітченко І.Г., Гаврилко П.П., Завадяк Р.І. Стратегічне управління інноваційною діяльністю як основа конкурентоспроможності // Науковий вісник УжНУ. – № 35/2012. Частина 1. – Ужгород: УжНУ, 2012. – С. 27 – 32.
94. Князевич А.О., Брітченко І.Г. Сутність та специфіка формування соціально-орієнтованої інноваційної

- інфраструктури // Науковий журнал "Менеджер". Вісник донецького державного університету управління. – № 3 (65)/2013. – Донецьк: ДонДУУ, 2013. – С. 68 – 73.
95. Кужелев М.О., Брітченко І.Г. Корпоративний контроль інсайдерської ренти // Науковий вісник ПУЕТ - № 1 (56) / 2014 – Полтава: ПУЕТ, 2014. – С. 281 – 285.
96. Князевич А.А. Организационно-правовые основы регулирования деятельности субъектов инновационной инфраструктуры Украины / А.А. Князевич, И.Г. Бритченко // Новая экономика. – Минск (Беларусь). – № 1(63). – 2014. – С. 137–142.
97. Князевич А.О. Кластерний підхід до створення інноваційної інфраструктури країни / А.О. Князевич, І.Г. Брітченко // Науковий вісник Мукачівського державного університету. – Серія «Економіка». – Вип. 2(4). – Ч. 2. – 2015. – С. 24-28.
98. Брітченко І.Г. Підвищення конкурентоспроможності банків на ринку банківських послуг України // Науковий вісник УжНУ (Серія економіка). Випуск 1 (45). – Том. 1/ 2015. – С. 176-180.
99. Бритченко И. Модернизация инновационной инфраструктуры Украины в процессе конвергенции с ЕС / И. Бритченко, А. Князевич // The international journal «Sustainable development». – «Технический университет», г. Варна (Болгария). – № 6 (27). – 2015. – С. 4–10.
100. Бритченко И. Модернизация инновационной инфраструктуры Украины в процессе конвергенции с ЕС / И. Бритченко, А. Князевич // The international journal «Sustainable development». – «Технический университет», г. Варна (Болгария). – № 6 (27). – 2015. – С. 4–10. (Наукометрическая база JFACTOR).
101. Брітченко І.Г. Передумови сталого розвитку національних соціально-економічних систем, що формують Європейський Союз / Брітченко І.Г., Стьопчкін А.І. // Економіка і організація управління. Збірник наукових праць ДонНУ – Вінниця. – Випуск № 3 (23). – 2016. – С. 40-49.
102. Бритченко И.Г. Бизнес в Украине: Социально-

экономические процессы ориентированные на внешнеэкономические связи / Бритченко И.Г., Саенко В.Г. // Науковий вісник УжНУ. Серія: Міжнародні економічні відносини та світове господарство – № 10 Ч. 1. – Ужгород: УжНУ, 2016. – С. 64-70.

103. Britchenko Igor. Key sources when formulating competitive advantages for hotel chains / Oleksandr P. Krupskiy, Oleksii Dzhusov, Nataliia Meshko, Igor Britchenko, Artem Prytykin // Tourism: An International Interdisciplinary Journal, Vol. 67 No. 1, 2019. – P. 34-46. ISSN 1332-7461 (Print), ISSN 1849-1545 (Online) (<https://hrcak.srce.hr/218374>)
104. Britchenko Igor. Leading methods for promoting finished product quality / Serhii Tkachenko, Igor Britchenko, Maksym Bezpartochnyi // Advances in Social Science, Education and Humanities Research. – Atlantis Press: Proceedings of the 3rd International Conference on Social, Economic and Academic Leadership (ICSEAL 2019). – Volume 318, May 2019. – P. 99-106. (<https://www.atlantis-press.com/proceedings/icseal-19/125909022>) ISBN 978-94-6252-729-4, ISSN 2352-5398
105. Britchenko Igor. Content marketing model for leading web content management / Iryna Diachuk, Igor Britchenko, Maksym Bezpartochnyi // Advances in Social Science, Education and Humanities Research. – Atlantis Press: Proceedings of the 3rd International Conference on Social, Economic and Academic Leadership (ICSEAL 2019). – Volume 318, May 2019. – P. 119-126. (<https://www.atlantis-press.com/proceedings/icseal-19/125909025>) ISBN 978-94-6252-729-4, ISSN 2352-5398
106. Britchenko Igor. Banking liquidity as a leading approach to risk management / Stanislav Arzevitin, Igor Britchenko, Anatoly Kosov // Advances in Social Science, Education and Humanities Research. – Atlantis Press: Proceedings of the 3rd International Conference on Social, Economic and Academic Leadership (ICSEAL 2019). – Volume 318, May 2019. – P. 149-157. (<https://www.atlantis-press.com/proceedings/icseal-19/125909030>) ISBN 978-94-6252-729-4, ISSN 2352-5398
107. Britchenko Igor. Central banks as leaders in ensuring financial stability / Viktoriia Biloshapka, Igor Britchenko, Iryna

- Okhrymenko // *Advances in Social Science, Education and Humanities Research*. – Atlantis Press: Proceedings of the 3rd International Conference on Social, Economic and Academic Leadership (ICSEAL 2019). - Volume 318, May 2019. – P. 173-181. (<https://www.atlantis-press.com/proceedings/icseal-19/125909033>) ISBN 978-94-6252-729-4, ISSN 2352-5398
108. Britchenko I. Blockchain Technology in the Fiscal Process of Ukraine / I. Britchenko, T. Cherniavska // *Списание «Икономически изследвания (Economic Studies)»*. – Институт за икономически изследвания при БАН, София (България). – Volume 28, Issue 5 – 2019. – P. 134-148. ISSN 02053292.
109. Britchenko I. Financial decentralization in Ukraine: prerequisites, problems, prospects / Britchenko Igor, Maksym Bezpartochnyi, Natalia Maslii // *VUZF review*. – VUZF, Sofia (Bulgaria). – № 4(4). – 2019. – P. 25-44. ISSN 2534-9228 <https://papersvuzf.net/index.php/VUZF/issue/view/13>
110. Britchenko I. Potential of Sustainable Regional Development in View of Smart Specialisation /Igor Britchenko, Tetiana Romanchenko, Oleksandr Hladkyi // *Списание «Икономически изследвания (Economic Studies)»*. – Институт за икономически изследвания при БАН, София (България). – № 6. – Volume 28, Issue 6 – 2019. – P. 88 - 110. ISSN 02053292. <https://www.iki.bas.bg/spisanie-ikonomicheski-izsledvaniia>
111. Britchenko Igor. University innovative hubs as points of growth of industrial parks of Ukraine / Britchenko I., N. Kraus, K. Kraus // *Financial and credit activity: problems of theory and practice*, Volume 4, No 31(2019), 2019. – P. 448-456. ISS (print) 2306-4994, ISSN (on-line) 2310-8770 <http://fkd.org.ua/article/view/190996>
112. Britchenko Igor. Consulting Services in Agriculture / Nadiia Serskykh, Igor Britchenko // *Modern Development Paths of Agricultural Production*. – Springer International Publishing. – 2019. – P. 217-223. (<https://www.springer.com/gp/book/>) ISBN 978-3-030-14917-8, eBook ISBN 978-3-030-14918-5 DOI https://doi.org/10.1007/978-3-030-14918-5_23
113. Britchenko I. Development of methodology of alternative

- rationale for financial ensuring of bridges building / Britchenko Igor, Maksym Bezpartochnyi, Yaroslava Levchenko // VUZF review. – VUZF, Sofia (Bulgaria). – № 5(1). – 2020. – P. 43-49. ISSN 2534-9228 <https://papersvuzf.net/index.php/VUZF/issue/view/14>
114. Britchenko Igor. University innovative hubs as points of growth of industrial parks of Ukraine / Britchenko I., N. Kraus, K. Kraus // Financial and credit activity: problems of theory and practice, Volume 4, No 31(2019), 2019. – P. 448-456. ISS (print) 2306-4994, ISSN (on-line) 2310-8770 <http://fkd.org.ua/article/view/190996>
 115. Britchenko I. Issues of shaping the students' professional and terminological competence in science area of expertise in the sustainable development era / Olena Lavrentieva, Victoria Pererva, Oleksandr Krupskyi, Igor Britchenko, Sardar Shabanov // E3S Web of Conferences. – FDP Sciences, France. – Volume 166, 10031 (2020). – 22.04.2020. – eISSN 2267-1242. – 9 pages. DOI <https://doi.org/10.1051/e3sconf/202016610031>
 116. Britchenko I. Areas and Means of Formation of Transport Regional Complexes and Mechanisms for Managing their Competitiveness in Ukraine / I gor Britchenko, Liliya Savchenko, Inna Naida, Oleksandr Tregubov // Списание «Икономически изследвания (Economic Studies)». – Институт за икономически изследвания при БАН, София (България). – № 3. – Volume 32, Issue 3 - 2020. – P. 61 - 82. ISSN 02053292. <https://www.iki.bas.bg/spisanie-ikonomicheski-izsledvaniia>
 117. Vazov R., "Inovative Approaches to Insurance Company Cash Flow Management (contemporary theoretical aspects)", VUZF Publishing House "St. Grigorii Bogoslov", 2019, ISBN: 978-954-8590-54-9
 118. Вазов Р., "Предприемачество иноваций и мениджмънт (Нови подходи и представи), издателство на ВУЗФ "Св. Григорий Богослов" 2019 г., ISBN 978-954-8590-55-6
 119. Вазов Р. Пътят на предприемача (Концепции за провала), академично издателство "Талант", годишник на ВУАРР, ISSN 1314-9113, ISSN 2535-0609

120. Вазов Р. "Предприемачество и рисков капитал", Годишник на ВУЗФ том 12 2018г., ISSN: 1312-7918
121. Вазов, Р. „Мониторинг на застрахователния риск и риска при паричните потоци на ЗАД „Булстрад Виена Иншуранс Груп” АД в условията на задълбочаваща се криза“. V международна научна конференция "Демараж или забавен каданс за икономиката и финансите?/По примера на ЕС и България"/, 2014. .
122. Вазов Р. "Парите и застраховането", списание "Ново знание" година VIII, бр. 1, 2019 г., ISSN 2367-4598. ISSN 1314-5703
123. Вазов Р. "Застраховане на катастрофични рискове посредством алтернативно трансфериране на риска", Академично издателство "Ценов" Сборник доклади Свищов 2018г., ISBN 978-954-23-1712-8
124. Вазов Р. "Dr. Consulta. Case study on social entrepreneurship and co-creation", Икономиката на България и европейският съюз колективна монография, УНСС 2018, ISBN 978-619-232-090-4
125. Radostin Vazov, Teodora Lazarova, Virginia Zhelyazkova. "Innovation leadership as a key concept in entrepreneurship", The 8th International Conference for Entrepreneurship, Innovation and Regional Development, University of Sheffield, June 2015, ISBN 978-0-9932801-0-8
126. Вазов Р., Илиева Д. "Neuro-linguistic programming techniques for perfecting presentation skills", Selected papers presented at 3rd m-sphere international conference for multidisciplinary in science and business. ISBN: 978-953-7930-06-6
127. Вазов Р., "Системата "бонус малус" предложения". "VUZF Review", Online journal for economics, Volume 3(7), Number 3, 2018 ISSN 2534-9228
128. Вазов Р., "Системата "бонус малус" коментари и тълкувания". Колективна монография "Икономиката на България и Европейският съюз в дигиталният свят", УНСС 2019 г.
129. Вазов Р. Предприемачество и глобализация, "VUZF Review", Online journal for economics, Volume 4(8), Number

1, 2019 ISSN 2534-9228

130. Вазов Р. "Трансформация на образованието. Предприемачески университети", Колективна монография "Икономиката на България и Европейският съюз в дигиталният свят", УНСС 2019 г.
131. Vazov Radostin PhD, Dr Sudhanshu Rai, Dr Maria L. Granados. Innovators intent: role of it in facilitating innovative knowledge practices in social enterprises, Technovation
132. Bezpartochnyi M. Methodological tools for assessing suppliers of trade enterprises. Formation strategy of economic structures: the tools and practices: [Collective monograph] / edited by A. Berezin, M. Bezpartochnyi. Riga: «Landmark» SIA, 2016. P. 110–120.
133. Bezpartochnyi M., Živitere M., Riashchenko V. Modern vision of effective functioning and development of tourist enterprises. Mechanisms and models of development of entities tourist business: collective monograph / edited by A. Berezin, M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2016. P. 8–14.
134. Bezpartochnyi M., Živitere M., Riashchenko V. The modern concept of operation and development of retail. Innovative approaches in the management of competitiveness of businesses: collective monograph / edited by A. Berezin, M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2016. Vol. 2. P. 8–14.
135. Bezpartochnyi M., Živitere M., Riashchenko V. Use of cluster analysis to ensure the effective functioning and development of retail. Innovative approaches in the management of competitiveness of businesses: collective monograph / edited by A. Berezin, M. Bezpartochnyi, in 2 vol. ISMA University. Riga: «Landmark» SIA, 2016. Vol. 2. P. 51–55.
136. Bezpartochnyi M. European model of consumer protection. Legal mechanisms to ensure the activities of economic entities: collective monograph. ISMA University. Riga: «Landmark» SIA, 2016. P. 50–65.
137. Bezpartochnyi M., Živitere M., Riashchenko V. Management and strategies of financial activities of trade enterprises. Organizational and economic mechanisms of development of

- the financial system: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2016. P. 101–117
138. Bezpartochnyi M., Živitere M., Riashchenko V. Specific features of personnel management system of trade enterprises. Theoretical, methodological and practical foundations of human resources management: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2016. P. 10–18.
139. Bezpartochnyi M., Živitere M., Riashchenko V. Creating a marketing strategy on agriculture market. Formation of modern social, economic and organizational mechanisms development of entities agrarian business: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2017. P. 52–68.
140. Bezpartochnyi M., Živitere M., Riashchenko V. Organizational and legal aspects of free trade zone between Ukraine and the European Union // Transformation of international economic relations: modern challenges, risks, opportunities and prospects: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2017. Vol. 1. P. 176–186 .
141. Bezpartochnyi M., Britchenko I. Development the wholesale enterprises region through business process reengineering. Transformational processes the development of economic systems in conditions of globalization: scientific bases, mechanisms, prospects: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2018. Vol. 1. P. 10–22.
142. Bezpartochnyi M., Britchenko I., Jarosz P. Reengineering business processes as a modern innovation of development wholesale enterprises region. Management of innovative development the economic entities: collective monograph / edited by M. Bezpartochnyi, I. Britchenko. Higher School of Social and Economic. Przeworsk: Wydawnictwo i Drukarnia NOVA SANDEC, 2018. Vol. 1. P. 10–24.
143. Bezpartochnyi M., Britchenko I., Jarosz P. Economic diagnostics in ensuring of competitiveness the economic

- entities. Conceptual aspects management of competitiveness the economic entities: collective monograph / edited by M. Bezpartochnyi, I. Britchenko. Higher School of Social and Economic. Przeworsk: WSSG, 2019. Vol. 1. P. 10–19.
144. Bezpartochnyi M., Britchenko I., Bezpartochna O., Mikhel V. Economic diagnostics as a tool for transformation of organizational-legal forms of economic activity in the field of agriculture. Management mechanisms and development strategies of economic entities in conditions of institutional transformations of the global environment: collective monograph / edited by M. Bezpartochnyi. ISMA University. Riga: «Landmark» SIA, 2019. Vol. 1. P. 259–269.
145. Bezpartochnyi M. Modern approaches to formation the mechanisms of management and regulation the activity of trade enterprises / Bezpartochnyi M., Britchenko I., Bezpartochna P., Radochonska-Jarosz R. // Organizational-economic mechanism of management innovative development of economic entities: collective monograph / edited by M. Bezpartochnyi, in 3 Vol. / Higher School of Social and Economic. – Przeworsk: WSSG, 2019. – Vol. 1. – pp. 300-310.
146. Bezpartochna O. Diagnostic the innovation potential of Latvia / M. Bezpartochnyi, M. Živitere, V. Riashchenko, O. Bezpartochna // Mechanisms of interaction between competitiveness and innovation in modern international economic relations: collective monograph / edited by M. Bezpartochnyi, in 4 Vol. ISMA University. – Riga: «Landmark» SIA, 2017. – Vol. 1. – pp. 62-70.
147. Bezpartochna O. Reengineering of business processes at enterprises of consumer co-operation / Rjashchenko V., Živitere M., Bezpartochna O. // Conceptual aspects management of competitiveness the economic entities: collective monograph / edited by M. Bezpartochnyi, I. Britchenko. Higher School of Social and Economic. Przeworsk: WSSG, 2019. – Vol. 2. – pp. 85–98
148. Bezpartochna O. Study of the modern production and economic relations of the enterprises of agricultural complex in Poltava region / Bezpartochna O., Britchenko I., Jarosz P.,

- Radochonska-Jarosz R. // Organizational-economic mechanism of management innovative development of economic entities: collective monograph / edited by M. Bezpartochnyi, in 3 Vol. Higher School of Social and Economic. – Przeworsk: WSSG, 2019. – Vol. 1. – pp. 142-153.
149. Bezpartochna O. The main trends in the development of tourism in Latvia / Rjashchenko V., Živitere M., Bezpartochna O. // Management mechanisms and development strategies of economic entities in conditions of institutional transformations of the global environment: collective monograph / edited by M. Bezpartochnyi, in 2 Vol. ISMA University. – Riga: “Landmark” SIA, 2019. – Vol. 2. – pp. 284-293.
150. Bezpartochnyi M. Modern management decision-making approaches / Riashchenko V., Korjuhina J., Bezpartochnyi M. // Strategies for sustainable socio-economic development and mechanisms their implementation in the global dimension: collective monograph / edited by M. Bezpartochnyi, in 3 Vol. // VUZF University of Finance, Business and Entrepreneurship. – Sofia: VUZF Publishing House “St. Grigorii Bogoslov”, 2019. – Vol. 1. – pp. 83-91.
151. Bezpartochna O. Gamification. Haven’t we overplayed? / Shamshina T., Koryuhina C., Riashchenko V., Bezpartochna O. // Organizational-economic mechanism of management innovative development of economic entities: collective monograph / edited by M. Bezpartochnyi, in 3 Vol. Higher School of Social and Economic. – Przeworsk: WSSG, 2019. – Vol. 3. – pp. 275-287.
152. Britchenko I. Pandemic economic crisis: essence, reasons, comparative characteristics, opportunities / Britchenko I., Bezpartochnyi M. // New trends in the economic systems management in the context of modern global challenges: collective monograph / scientific edited by M. Bezpartochnyi // VUZF University of Finance, Business and Entrepreneurship. – Sofia: VUZF Publishing House “St. Grigorii Bogoslov”, 2020. p. 8-19. ISBN 978-954-8590-85-3
153. Bezpartochna O. Diagnostics the economic efficiency of agricultural enterprises / Bezpartochna O., Britchenko I. //

New trends in the economic systems management in the context of modern global challenges: collective monograph / scientific edited by M. Bezpartochnyi // VUZF University of Finance, Business and Entrepreneurship. – Sofia: VUZF Publishing House “St. Grigorii Bogoslov”, 2020. p. 267-277. ISBN 978-954-8590-85-3

FORECASTING MODELING AND ANALYTICS OF ECONOMIC PROCESSES

Authors:

Olha Mezentseva – PhD, Docent of the Department of Management of Foreign Economic Activities of Enterprises, National Aviation University, Ukraine

Oksana Ilienکو – Doctor of Technical Sciences, Associate Professor of the Department of Management of Foreign Economic Activities of Enterprises, National Aviation University, Ukraine

Oleksii Kolesnikov – Doctor of Technical Sciences, Associate Professor of the Information System and Technology Department, Taras Shevchenko National University of Kyiv, Ukraine

Olena Savielieva – Doctor of Philosophy (PhD), Associate Professor of the Department of Theory and Methods of Applied and Decorative Arts and Graphics, South Ukrainian National Pedagogical University named after K.D. Ushynsky, Ukraine

Dmytro Lukianov – Candidate of Technical Sciences, Department of Construction and Operation of Buildings and Structures, Faculty of Construction and Real Estate, Branch of the Intersectoral Institute for Advanced Studies and Retraining of Personnel, Belarusian National Technical University, Republic of Belarus

Format 140 x 210
Print font 11
Circulation: 100 copies
11,35 p.s.

VUZF Publishing House “St. Grigorii Bogoslov”

Sofia, 2020

ISBN 978-954-8590-88-4