SEVENTH EDITION

β

USING ECONOMETRICS

A PRACTICAL GUIDE

A.H. STUDENMUND

Table B-1 Critical Values of the *t*-Distribution

Degrees of Freedom	Level of Significance				
	One-Sided: 10% Two-Sided: 20%	5% 10%	2.5% 5%	1% 2%	0.5% 1%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
(Normal)					
∞	1.282	1.645	1.960	2.326	2.576

Source: Reprinted from Table IV in Sir Ronald A. Fisher, *Statistical Methods for Research Workers*, 14th ed. (copyright © 1970, University of Adelaide) with permission of Hafner, a division of the Macmillan Publishing Company, Inc.

USING ECONOMETRICS

SEVENTH EDITION

β

USING ECONOMETRICS

A PRACTICAL GUIDE

A. H. Studenmund
Occidental College

with the assistance of

Bruce K. Johnson
Centre College

PEARSON

Boston Columbus Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Vice President, Business Publishing: Donna Battista
Editor-in-Chief: Adrienne D'Ambrosio
Senior Acquisitions Editor: Christina Masturzo
Acquisitions Editor/Program Manager: Neeraj Bhalla
Editorial Assistant: Diana Tetterton
Vice President, Product Marketing: Maggie Moylan
Director of Marketing, Digital Services and Products: Jeanette Koskinas
Field Marketing Manager: Ramona Elmer
Product Marketing Assistant: Jessica Quazza
Team Lead, Program Management: Ashley Santora
Team Lead, Project Management: Jeff Holcomb
Project Manager: Liz Napolitano
Operations Specialist: Carol Melville
Creative Director: Blair Brown
Art Director: Jon Boylan
Vice President, Director of Digital Strategy and Assessment: Paul Gentile
Manager of Learning Applications: Paul DeLuca
Digital Editor: Denise Clinton
Director, Digital Studio: Sacha Laustsen
Digital Studio Manager: Diane Lombardo
Digital Studio Project Manager: Melissa Honig
Digital Studio Project Manager: Robin Lazrus
Digital Content Team Lead: Noel Lotz
Digital Content Project Lead: Courtney Kamauf
Full-Service Project Management and Composition: Cenveo® Publisher Services
Interior Designer: Cenveo® Publisher Services
Cover Designer: Jon Boylan
Printer/Binder: Edwards Brothers
Cover Printer: Phoenix Color/Hagerstown

Library of Congress Cataloging-in-Publication Data
Names: Studenmund, A. H., author.
Title: Using econometrics : a practical guide / A. H. Studenmund, Occidental College.
Description: Seventh Edition. | Boston : Pearson, 2016. | Revised edition of the author's Using econometrics, 2011. | Includes index.
Identifiers: LCCN 2016002694 | ISBN 9780134182742
Subjects: LCSH: Econometrics. | Regression analysis.
Classification: LCC HB139 .S795 2016 | DDC 330.01/5195--dc23
LC record available at http://lccn.loc.gov/2016002694

10 9 8 7 6 5 4 3 2 1

PEARSON

www.pearsonhighered.com

ISBN 10: 0-13-418274-X
ISBN 13: 978-0-13-418274-2

Dedicated to the memory of

Green Beret

Staff Sergeant

Scott Studenmund

Killed in action in Afghanistan on June 9, 2014

The Pearson Series in Economics

Abel/Bernanke/Croushore
*Macroeconomics**

Acemoglu/Laibson/List
*Economics**

Bade/Parkin
*Foundations of Economics**

Berck/Helfand
The Economics of the Environment

Bierman/Fernandez
Game Theory with Economic Applications

Blanchard
*Macroeconomics**

Blau/Ferber/Winkler
The Economics of Women, Men, and Work

Boardman/Greenberg/Vining/Weimer
Cost-Benefit Analysis

Boyer
Principles of Transportation Economics

Branson
Macroeconomic Theory and Policy

Bruce
Public Finance and the American Economy

Carlton/Perloff
Modern Industrial Organization

Case/Fair/Oster
*Principles of Economics**

Chapman
Environmental Economics: Theory, Application, and Policy

Cooter/Ulen
Law & Economics

Daniels/VanHoose
International Monetary & Financial Economics

Downs
An Economic Theory of Democracy

Ehrenberg/Smith
Modern Labor Economics

Farnham
Economics for Managers

Folland/Goodman/Stano
The Economics of Health and Health Care

Fort
Sports Economics

Froyen
Macroeconomics

Fusfeld
The Age of the Economist

Gerber
*International Economics**

González-Rivera
Forecasting for Economics and Business

Gordon
*Macroeconomics**

Grant
*Economic Analysis of Social Issues**

Greene
Econometric Analysis

Gregory
Essentials of Economics

Gregory/Stuart
Russian and Soviet Economic Performance and Structure

Hartwick/Olewiler
The Economics of Natural Resource Use

Heilbroner/Milberg
The Making of the Economic Society

Heyne/Boettke/Prychitko
The Economic Way of Thinking

Holt
Markets, Games, and Strategic Behavior

Hubbard/O'Brien
*Economics**

*Money, Banking, and the Financial System**

Hubbard/O'Brien/Rafferty
*Macroeconomics**

Hughes/Cain
American Economic History

Husted/Melvin
International Economics

Jehle/Reny
Advanced Microeconomic Theory

Johnson-Lans
A Health Economics Primer

Keat/Young/Erfle
Managerial Economics

Klein
Mathematical Methods for Economics

Krugman/Obstfeld/Melitz
*International Economics: Theory & Policy**

Laidler
The Demand for Money

Leeds/von Allmen
The Economics of Sports

Leeds/von Allmen/Schiming
*Economics**

Lynn
Economic Development: Theory and Practice for a Divided World

Miller
*Economics Today**

Understanding Modern Economics

Miller/Benjamin
The Economics of Macro Issues

Miller/Benjamin/North
The Economics of Public Issues

Mills/Hamilton
Urban Economics

Mishkin
*The Economics of Money, Banking, and Financial Markets**

*The Economics of Money, Banking, and Financial Markets, Business School Edition**

*Macroeconomics: Policy and Practice**

Murray
Econometrics: A Modern Introduction

O'Sullivan/Sheffrin/Perez
*Economics: Principles, Applications and Tools**

Parkin
*Economics**

Perloff
*Microeconomics**

*Microeconomics: Theory and Applications with Calculus**

Perloff/Brander
*Managerial Economics and Strategy**

Phelps
Health Economics

Pindyck/Rubinfeld
*Microeconomics**

Riddell/Shackelford/Stamos/Schneider
Economics: A Tool for Critically Understanding Society

Roberts
The Choice: A Fable of Free Trade and Protection

Rohlf
Introduction to Economic Reasoning

Roland
Development Economics

Scherer
Industry Structure, Strategy, and Public Policy

Schiller
The Economics of Poverty and Discrimination

Sherman
Market Regulation

Stock/Watson
Introduction to Econometrics

Studenmund
Using Econometrics: A Practical Guide

Tietenberg/Lewis
Environmental and Natural Resource Economics
Environmental Economics and Policy

Todaro/Smith
Economic Development

Waldman/Jensen
Industrial Organization: Theory and Practice

Walters/Walters/Appel/Callahan/Centanni/ Maex/O'Neill
Econversations: Today's Students Discuss Today's Issues

Weil
Economic Growth

Williamson
Macroeconomics

*denotes MyEconLab titles   Visit www.myeconlab.com to learn more.

This page intentionally left blank

CONTENTS

PREFACE

Econometric education is a lot like learning to fly a plane; you learn more from actually doing it than you learn from reading about it.

Using Econometrics represents an innovative approach to the understanding of elementary econometrics. It covers the topic of single-equation linear regression analysis in an easily understandable format that emphasizes real-world examples and exercises. As the subtitle *A Practical Guide* implies, the book is aimed not only at beginning econometrics students but also at regression users looking for a refresher and at experienced practitioners who want a convenient reference.

What's New in the Seventh Edition?

Using Econometrics has been praised as "one of the most important new texts of the last 30 years," so we've retained the clarity and practicality of previous editions. However, we're delighted to have made a number of substantial improvements in the text.

The most exciting upgrades are:

1. **Econometric Labs**: These new and innovative learning tools are optional appendices that give students hands-on opportunities to better understand the econometric principles that they're reading about in the chapters. The labs originally were designed to be assigned in a classroom setting, but they also have turned out to be extremely valuable for readers who are not in a class or for individual students in classes where the labs aren't assigned. Hints on how best to use these econometric labs and answers to the lab questions are available in the instructor's manual on the *Using Econometrics* Web site.

2. **The Use of Stata throughout the Text:** In our opinion, Stata has become the econometric software package of choice among economic researchers. As a result, we have estimated all the text examples and exercises with Stata and have included a short appendix to help students get started with Stata. Beyond this, we have added a complete guide to *Using Stata* to our Web site. This guide, written by John Perry of Centre College, explains in detail all the Stata commands needed to replicate the text's equations and answer the text's exercises. However, even though we use Stata extensively, *Using Econometrics* is not tied to

Stata or any other econometric software, so the text works well with all standard regression packages.

3. **Expanded Econometric Content:** We have added coverage of a number of econometric tests and procedures, for example the Breusch-Pagan test and the Prais–Winsten approach to Generalized Least Squares. In addition, we have expanded the coverage of even more topics, for example the *F*-test, confidence intervals, the Lagrange Multiplier test, and the Dickey–Fuller test. Finally, we have simplified the notation and improved the clarity of the explanations in Chapters 12–16, particularly in topics like dynamic equations, dummy dependent variables, instrumental variables, and panel data.

4. **Answers to Many More Exercises:** In response to requests from instructors and students, we have more than tripled the number of exercises that are answered in the text's appendix. These answers will allow students to learn on their own, because students will be able to attempt an exercise and then check their answers against those in the back of the book without having to involve their professors. In order to continue to provide good exercises for professors to include in problem sets and exams, we have expanded the number of exercises contained in the text's Web site.

5. **Dramatically Improved PowerPoint Slides:** We recognize the importance of PowerPoint slides to instructors with large classes, so we have dramatically improved the quality of the text's PowerPoints. The slides replicate each chapter's main equations and examples, and also provide chapter summaries and lists of the key concepts in each chapter. The PowerPoint slides can be downloaded from the text's Web site, and they're designed to be easily edited and individualized.

6. **An Expanded and Improved Web Site:** We believe that this edition's Web site is the best we've produced. As you'd expect, the Web site includes all the text's data sets, in easily downloadable Stata, EViews, Excel, and ASCII formats, but we have gone far beyond that. We have added *Using Stata*, a complete guide to the Stata commands needed to estimate the book's equations; we have dramatically improved the PowerPoint slides; and we have added answers to the new econometric labs and instructions on how best to use these labs in a classroom setting. In addition, the Web site also includes an instructor's manual, additional exercises, extra interactive regression learning exercises, and additional data sets. But why take our word for it? Take a look for yourself at http://www.pearsonhighered.com/studenmund

Features

1. Our approach to the learning of econometrics is simple, intuitive, and easy to understand. We do not use matrix algebra, and we relegate proofs and calculus to the footnotes or exercises.

2. We include numerous examples and example-based exercises. We feel that the best way to get a solid grasp of applied econometrics is through an example-oriented approach.

3. Although most of this book is at a simpler level than other econometrics texts, Chapters 6 and 7 on specification choice are among the most complete in the field. We think that an understanding of specification issues is vital for regression users.

4. We use a unique kind of learning tool called an *interactive regression learning exercise* to help students simulate econometric analysis by giving them feedback on various kinds of decisions without relying on computer time or much instructor supervision.

5. We're delighted to introduce a new innovative learning tool called an *econometric lab*. These econometric labs, developed by Bruce Johnson of Centre College and tested successfully at two other institutions, are optional appendices aimed at giving students hands-on experience with the econometric procedures they're reading about. Students who complete these econometric labs will be much better prepared to undertake econometric research on their own.

The formal prerequisites for using this book are few. Readers are assumed to have been exposed to some microeconomic and macroeconomic theory, basic mathematical functions, and elementary statistics (even if they have forgotten most if it). Students with little statistical background are encouraged to begin their study of econometrics by reading Chapter 17, "Statistical Principles," on the text's Web site.

Because the prerequisites are few and the statistics material is self-contained, *Using Econometrics* can be used not only in undergraduate courses but also in MBA-level courses in quantitative methods. We also have been told that the book is a helpful supplement for graduate-level econometrics courses.

The Stata and EViews Options

We're delighted to be able to offer our readers the chance to purchase the student version of Stata or EViews at discounted prices when bundled with the textbook. Stata and EViews are two of the best econometric software programs available, so it's a real advantage to be able to buy them at substantial savings.

We urge professors to make these options available to their students even if Stata or EViews aren't used in class. The advantages to students of owning their own regression software are many. They can run regressions when they're off-campus, they will add a marketable skill to their résumé if they learn Stata or EViews, and they'll own a software package that will allow them to run regressions after the class is over if they choose the EViews option.

Acknowledgments

This edition of *Using Econometrics* has been blessed by superb contributions from Ron Michener of the University of Virginia and Bruce Johnson of Centre College. Ron was the lead reviewer, and in that role he commented on every section and virtually every equation in the book, creating a 132-page *magnum opus* of textbook reviewing that may never be surpassed in length or quality.

Just as importantly, Ron introduced us to Bruce Johnson. Bruce wrote the first drafts of the econometric labs and three other sections, made insightful comments on the entire revision, helped increase the role of Stata in the book, and proofread the manuscript. Because of Bruce's professional expertise, clear writing style, and infectious enthusiasm for econometrics, we're happy to announce that he will be a coauthor of the 8th and subsequent editions of *Using Econometrics*.

This book's spiritual parents were Henry Cassidy and Carolyn Summers. Henry co-authored the first edition of *Using Econometrics* as an expansion of his own work of the same name, and Carolyn was the text's editorial consultant, proofreader, and indexer for four straight editions. Other important professional contributors to previous editions were the late Peter Kennedy, Nobel Prize winner Rob Engle of New York University, Gary Smith of Pomona College, Doug Steigerwald of the University of California at Santa Barbara, and Susan Averett of Lafayette College.

In addition, this edition benefitted from the evaluations of a talented group of professional reviewers:

Lesley Chiou, Occidental College
Dylan Conger, George Washington University
Leila Farivar, Ohio State University
Abbass Grammy, California State University, Bakersfield

Jason Hecht, Ramapo College
Jin Man Lee, University of Illinois at Chicago
Noelwah Netusl, Reed College
Robert Parks, Washington University in St. Louis
David Phillips, Hope College
John Perry, Centre College
Robert Shapiro, Columbia University
Phanindra Wunnava, Middlebury College

Invaluable in the editorial and production process were Jean Bermingham, Neeraj Bhalla, Adrienne D'Ambrosio, Marguerite Dessornes, Christina Masturzo, Liz Napolitano, Bill Rising, and Kathy Smith. Providing crucial emotional support during an extremely difficult time were Sarah Newhall, Barbara Passerelle, Barbara and David Studenmund, and my immediate family, Jaynie and Connell Studenmund and Brent Morse. Finally, I'd like to thank my wonderful Occidental College colleagues and students for their feedback and encouragement. These particularly included Lesley Chiou, Jack Gephart, Jorge Gonzalez, Andy Jalil, Kate Johnstone, Mary Lopez, Jessica May, Cole Moniz, Robby Moore, Kyle Yee, and, especially, Koby Deitz.

A. H. Studenmund

This page intentionally left blank

Chapter 1

An Overview of Regression Analysis

1.1 What Is Econometrics?

> *"Econometrics is too mathematical; it's the reason my best friend isn't majoring in economics."*
>
> *"There are two things you are better off not watching in the making: sausages and econometric estimates."*[1]
>
> *"Econometrics may be defined as the quantitative analysis of actual economic phenomena."*[2]
>
> *"It's my experience that 'economy-tricks' is usually nothing more than a justification of what the author believed before the research was begun."*

Obviously, econometrics means different things to different people. To beginning students, it may seem as if econometrics is an overly complex obstacle to an otherwise useful education. To skeptical observers, econometric

1. Ed Leamer, "Let's take the Con out of Econometrics," *American Economic Review*, Vol. 73, No. 1, p. 37.

2. Paul A. Samuelson, T. C. Koopmans, and J. R. Stone, "Report of the Evaluative Committee for *Econometrica*," *Econometrica*, 1954, p. 141.

results should be trusted only when the steps that produced those results are completely known. To professionals in the field, econometrics is a fascinating set of techniques that allows the measurement and analysis of economic phenomena and the prediction of future economic trends.

You're probably thinking that such diverse points of view sound like the statements of blind people trying to describe an elephant based on which part they happen to be touching, and you're partially right. Econometrics has both a formal definition and a larger context. Although you can easily memorize the formal definition, you'll get the complete picture only by understanding the many uses of and alternative approaches to econometrics.

That said, we need a formal definition. **Econometrics**—literally, "economic measurement"—is the quantitative measurement and analysis of actual economic and business phenomena. It attempts to quantify economic reality and bridge the gap between the abstract world of economic theory and the real world of human activity. To many students, these worlds may seem far apart. On the one hand, economists theorize equilibrium prices based on carefully conceived marginal costs and marginal revenues; on the other, many firms seem to operate as though they have never heard of such concepts. Econometrics allows us to examine data and to quantify the actions of firms, consumers, and governments. Such measurements have a number of different uses, and an examination of these uses is the first step to understanding econometrics.

Uses of Econometrics

Econometrics has three major uses:

1. describing economic reality
2. testing hypotheses about economic theory and policy
3. forecasting future economic activity

The simplest use of econometrics is description. We can use econometrics to quantify economic activity and measure marginal effects because econometrics allows us to estimate numbers and put them in equations that previously contained only abstract symbols. For example, consumer demand for a particular product often can be thought of as a relationship between the quantity demanded (Q) and the product's price (P), the price of a substitute (P_s), and disposable income (Yd). For most goods, the relationship between consumption and disposable income is expected to be positive, because an increase in disposable income will be associated with an increase in the consumption of the product. Econometrics actually allows us to estimate that relationship based upon past consumption, income, and prices. In other words, a general and purely theoretical functional relationship like:

$$Q = \beta_0 + \beta_1 P + \beta_2 P_S + \beta_1 Yd \qquad (1.1)$$

can become explicit:

$$Q = 27.7 - 0.11P + 0.03P_S + 0.23Yd \qquad (1.2)$$

This technique gives a much more specific and descriptive picture of the function.[3] Let's compare Equations 1.1 and 1.2. Instead of expecting consumption merely to "increase" if there is an increase in disposable income, Equation 1.2 allows us to expect an increase of a specific amount (0.23 units for each unit of increased disposable income). The number 0.23 is called an estimated regression coefficient, and it is the ability to estimate these coefficients that makes econometrics valuable.

The second use of econometrics is hypothesis testing, the evaluation of alternative theories with quantitative evidence. Much of economics involves building theoretical models and testing them against evidence, and hypothesis testing is vital to that scientific approach. For example, you could test the hypothesis that the product in Equation 1.1 is what economists call a normal good (one for which the quantity demanded increases when disposable income increases). You could do this by applying various statistical tests to the estimated coefficient (0.23) of disposable income (Yd) in Equation 1.2. At first glance, the evidence would seem to support this hypothesis, because the coefficient's sign is positive, but the "statistical significance" of that estimate would have to be investigated before such a conclusion could be justified. Even though the estimated coefficient is positive, as expected, it may not be sufficiently different from zero to convince us that the true coefficient is indeed positive.

The third and most difficult use of econometrics is to forecast or predict what is likely to happen next quarter, next year, or further into the future, based on what has happened in the past. For example, economists use econometric models to make forecasts of variables like sales, profits, Gross Domestic Product (GDP), and the inflation rate. The accuracy of such forecasts depends in large measure on the degree to which the past is a good guide to the future. Business leaders and politicians tend to be especially interested in this use of

3. It's of course naïve to build a model of sales (demand) without taking supply into consideration. Unfortunately, it's very difficult to learn how to estimate a system of simultaneous equations until you've learned how to estimate a single equation. As a result, we will postpone our discussion of the econometrics of simultaneous equations until Chapter 14. Until then, you should be aware that we sometimes will encounter right-hand-side variables that are not truly "independent" from a theoretical point of view.

econometrics because they need to make decisions about the future, and the penalty for being wrong (bankruptcy for the entrepreneur and political defeat for the candidate) is high. To the extent that econometrics can shed light on the impact of their policies, business and government leaders will be better equipped to make decisions. For example, if the president of a company that sold the product modeled in Equation 1.1 wanted to decide whether to increase prices, forecasts of sales with and without the price increase could be calculated and compared to help make such a decision.

Alternative Econometric Approaches

There are many different approaches to quantitative work. For example, the fields of biology, psychology, and physics all face quantitative questions similar to those faced in economics and business. However, these fields tend to use somewhat different techniques for analysis because the problems they face aren't the same. For example, economics typically is an observational discipline rather than an experimental one. "We need a special field called econometrics, and textbooks about it, because it is generally accepted that economic data possess certain properties that are not considered in standard statistics texts or are not sufficiently emphasized there for use by economists."[4]

Different approaches also make sense within the field of economics. A model built solely for descriptive purposes might be different from a forecasting model, for example.

To get a better picture of these approaches, let's look at the steps used in nonexperimental quantitative research:

1. specifying the models or relationships to be studied
2. collecting the data needed to quantify the models
3. quantifying the models with the data

The specifications used in step 1 and the techniques used in step 3 differ widely between and within disciplines. Choosing the best specification for a given model is a theory-based skill that is often referred to as the "art" of econometrics. There are many alternative approaches to quantifying the same equation, and each approach may produce somewhat different results. The choice of approach is left to the individual econometrician (the researcher using econometrics), but each researcher should be able to justify that choice.

4. Clive Granger, "A Review of Some Recent Textbooks of Econometrics," *Journal of Economic Literature*, Vol. 32, No. 1, p. 117.

This book will focus primarily on one particular econometric approach: *single-equation linear regression analysis*. The majority of this book will thus concentrate on regression analysis, but it is important for every econometrician to remember that regression is only one of many approaches to econometric quantification.

The importance of critical evaluation cannot be stressed enough; a good econometrician can diagnose faults in a particular approach and figure out how to repair them. The limitations of the regression analysis approach must be fully perceived and appreciated by anyone attempting to use regression analysis or its findings. The possibility of missing or inaccurate data, incorrectly formulated relationships, poorly chosen estimating techniques, or improper statistical testing procedures implies that the results from regression analyses always should be viewed with some caution.

1.2 What Is Regression Analysis?

Econometricians use regression analysis to make quantitative estimates of economic relationships that previously have been completely theoretical in nature. After all, anybody can claim that the quantity of iPhones demanded will increase if the price of those phones decreases (holding everything else constant), but not many people can put specific numbers into an equation and estimate *by how many* iPhones the quantity demanded will increase for each dollar that price decreases. To predict the *direction* of the change, you need a knowledge of economic theory and the general characteristics of the product in question. To predict the *amount* of the change, though, you need a sample of data, and you need a way to estimate the relationship. The most frequently used method to estimate such a relationship in econometrics is regression analysis.

Dependent Variables, Independent Variables, and Causality

Regression analysis is a statistical technique that attempts to "explain" movements in one variable, the **dependent variable**, as a function of movements in a set of other variables, called the **independent** (or **explanatory**) **variables**, through the quantification of one or more equations. For example, in Equation 1.1:

$$Q = \beta_0 + \beta_1 P + \beta_2 P_S + \beta_1 Yd \tag{1.1}$$

Q is the dependent variable and P, P_S, and Yd are the independent variables. Regression analysis is a natural tool for economists because most (though not all) economic propositions can be stated in such equations. For example, the quantity demanded (dependent variable) is a function of price, the prices of substitutes, and income (independent variables).

Much of economics and business is concerned with cause-and-effect propositions. If the price of a good increases by one unit, then the quantity demanded decreases on average by a certain amount, depending on the price elasticity of demand (defined as the percentage change in the quantity demanded that is caused by a one percent increase in price). Similarly, if the quantity of capital employed increases by one unit, then output increases by a certain amount, called the marginal productivity of capital. Propositions such as these pose an if-then, or causal, relationship that logically postulates that a dependent variable's movements are determined by movements in a number of specific independent variables.

Don't be deceived by the words "dependent" and "independent," however. Although many economic relationships are causal by their very nature, a regression result, no matter how statistically significant, cannot prove causality. All regression analysis can do is test whether a significant quantitative relationship exists. Judgments as to causality must also include a healthy dose of economic theory and common sense. For example, the fact that the bell on the door of a flower shop rings just before a customer enters and purchases some flowers by no means implies that the bell causes purchases! If events A and B are related statistically, it may be that A causes B, that B causes A, that some omitted factor causes both, or that a chance correlation exists between the two.

The cause-and-effect relationship often is so subtle that it fools even the most prominent economists. For example, in the late nineteenth century, English economist Stanley Jevons hypothesized that sunspots caused an increase in economic activity. To test this theory, he collected data on national output (the dependent variable) and sunspot activity (the independent variable) and showed that a significant positive relationship existed. This result led him, and some others, to jump to the conclusion that sunspots did indeed cause output to rise. Such a conclusion was unjustified because regression analysis cannot confirm causality; it can only test the strength and direction of the quantitative relationships involved.

Single-Equation Linear Models

The simplest single-equation regression model is:

$$Y = \beta_0 + \beta_1 X \tag{1.3}$$

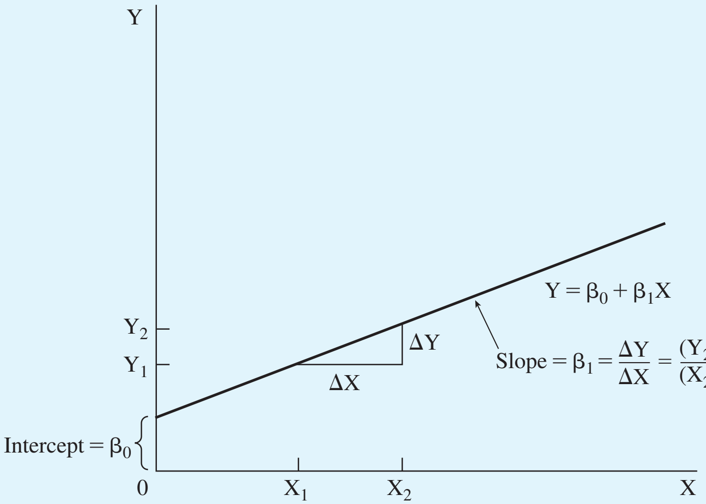

Figure 1.1 Graphical Representation of the Coefficients of the Regression Line

The graph of the equation $Y = \beta_0 + \beta_1 X$ is linear with a constant slope equal to $\beta_1 = \Delta Y/\Delta X$.

Equation 1.3 states that Y, the dependent variable, is a single-equation linear function of X, the independent variable. The model is a single-equation model because it's the only equation specified. The model is linear because if you were to plot Equation 1.3 it would be a straight line rather than a curve.

The βs are the coefficients that determine the coordinates of the straight line at any point. β_0 is the **constant** or **intercept** term; it indicates the value of Y when X equals zero. β_1 is the **slope coefficient**, and it indicates the amount that Y will change when X increases by one unit. The line in Figure 1.1 illustrates the relationship between the coefficients and the graphical meaning of the regression equation. As can be seen from the diagram, Equation 1.3 is indeed linear.

The slope coefficient, β_1, shows the response of Y to a one-unit increase in X. Much of the emphasis in regression analysis is on slope coefficients such as β_1. In Figure 1.1 for example, if X were to increase by one from X_1 to X_2 (ΔX), the value of Y in Equation 1.3 would increase from Y_1 to Y_2 (ΔY). For linear (i.e., straight-line) regression models, the response in the predicted value of Y due to a change in X is constant and equal to the slope coefficient β_1:

$$\frac{(Y_2 - Y_1)}{(X_2 - X_1)} = \frac{\Delta Y}{\Delta X} = \beta_1$$

where Δ is used to denote a change in the variables. Some readers may recognize this as the "rise" (ΔY) divided by the "run" (ΔX). For a linear model, the slope is constant over the entire function.

If linear regression techniques are going to be applied to an equation, that equation *must* be linear. An equation is **linear** if plotting the function in terms of X and Y generates a straight line; for example, Equation 1.3 is linear.[5]

$$Y = \beta_0 + \beta_1 X \tag{1.3}$$

The Stochastic Error Term

Besides the variation in the dependent variable (Y) that is caused by the independent variable (X), there is almost always variation that comes from other sources as well. This additional variation comes in part from omitted explanatory variables (e.g., X_2 and X_3). However, even if these extra variables are added to the equation, there still is going to be some variation in Y that simply cannot be explained by the model.[6] This variation probably comes from sources such as omitted influences, measurement error, incorrect functional form, or purely random and totally unpredictable occurrences. By *random* we mean something that has its value determined entirely by chance.

Econometricians admit the existence of such inherent unexplained variation ("error") by explicitly including a stochastic (or random) error term in their regression models. A **stochastic error term** is a term that is added to a regression equation to introduce all of the variation in Y that cannot be explained by the included Xs. It is, in effect, a symbol of the econometrician's ignorance or inability to model all the movements of the dependent variable. The error term (sometimes called a disturbance term) usually is referred to with the symbol epsilon (ϵ), although other symbols (like u or v) sometimes are used.

5. Technically, as you will learn in Chapter 7, this equation is linear in the coefficients β_0 and β_1 and linear in the variables Y and X. The application of regression analysis to equations that are nonlinear in the variables is covered in Chapter 7. The application of regression techniques to equations that are nonlinear in the coefficients, however, is much more difficult.

6. The exception would be the extremely rare case where the data can be explained by some sort of physical law and are measured perfectly. Here, continued variation would point to an omitted independent variable. A similar kind of problem is often encountered in astronomy, where planets can be discovered by noting that the orbits of known planets exhibit variations that can be caused only by the gravitational pull of another heavenly body. Absent these kinds of physical laws, researchers in economics and business would be foolhardy to believe that *all* variation in Y can be explained by a regression model because there are always elements of error in any attempt to measure a behavioral relationship.

The addition of a stochastic error term (ϵ) to Equation 1.3 results in a typical regression equation:

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1.4}$$

Equation 1.4 can be thought of as having two components, the *deterministic* component and the *stochastic,* or random, component. The expression $\beta_0 + \beta_1 X$ is called the *deterministic* component of the regression equation because it indicates the value of Y that is determined by a given value of X, which is assumed to be nonstochastic. This deterministic component can also be thought of as the **expected value** of Y given X, the mean value of the Ys associated with a particular value of X. For example, if the average height of all 13-year-old girls is 5 feet, then 5 feet is the expected value of a girl's height given that she is 13. The deterministic part of the equation may be written:

$$E(Y|X) = \beta_0 + \beta_1 X \tag{1.5}$$

which states that the expected value of Y given X, denoted as $E(Y|X)$, is a linear function of the independent variable (or variables if there are more than one).

Unfortunately, the value of Y observed in the real world is unlikely to be exactly equal to the deterministic expected value $E(Y|X)$. After all, not all 13-year-old girls are 5 feet tall. As a result, the stochastic element (ϵ) must be added to the equation:

$$Y = E(Y|X) + \epsilon = \beta_0 + \beta_1 X + \epsilon \tag{1.6}$$

The stochastic error term must be present in a regression equation because there are at least four sources of variation in Y other than the variation in the included Xs:

1. Many minor influences on Y are *omitted* from the equation (for example, because data are unavailable).
2. It is virtually impossible to avoid some sort of *measurement error* in the dependent variable.
3. The underlying theoretical equation might have a *different functional form* (or shape) than the one chosen for the regression. For example, the underlying equation might be nonlinear.
4. All attempts to generalize human behavior must contain at least some amount of unpredictable or *purely random* variation.