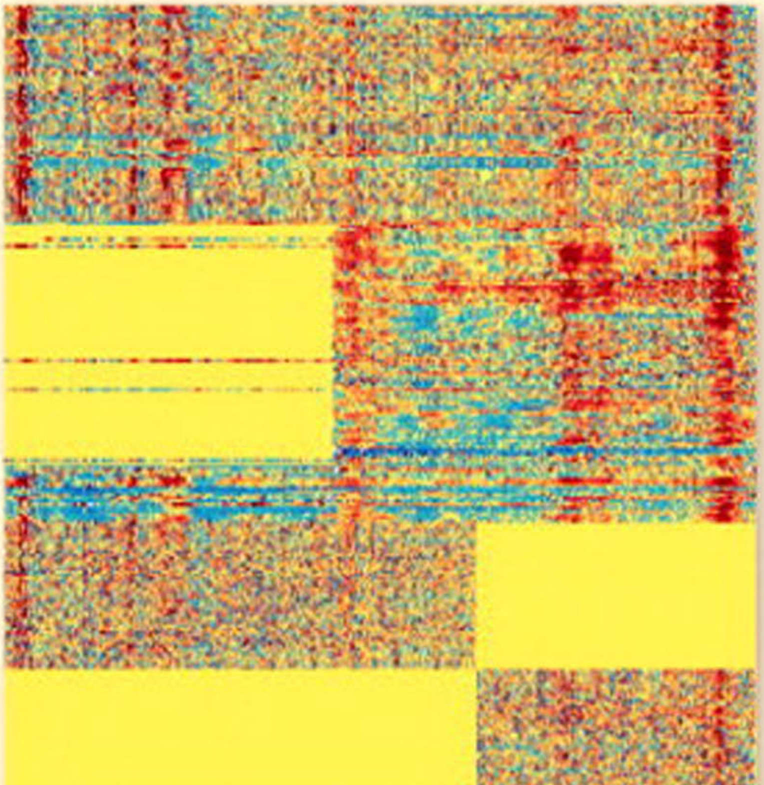
UPDATED EDITION

INTRODUCTION TO ECONOMETRICS

THIRD EDITION



James H. Stock
Mark W. Watson

Introduction to Econometrics

The Pearson Series in Economics

Abel/Bernanke/Croushore
*Macroeconomics**

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**

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.

Introduction to Econometrics

THIRD EDITION UPDATE

James H. Stock
Harvard University

Mark W. Watson
Princeton University

PEARSON

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

Vice President, Product Management: Donna Battista
Acquisitions Editor: Christina Masturzo
Editorial Assistant: Christine Mallon
Vice President, Marketing: Maggie Moylan
Director, Strategy and Marketing: Scott Dustan
Manager, Field Marketing: Leigh Ann Sims
Product Marketing Manager: Alison Haskins
Executive Field Marketing Manager: Lori DeShazo
Senior Strategic Marketing Manager: Erin Gardner
Team Lead, Program Management: Ashley Santora
Program Manager: Carolyn Philips
Team Lead, Project Management: Jeff Holcomb
Project Manager: Liz Napolitano
Operations Specialist: Carol Melville
Cover Designer: Jon Boylan
Cover Art: Courtesy of Carolin Pflueger and the authors.
Full-Service Project Management, Design, and Electronic Composition: Cenveo® Publisher Services
Printer/Binder: Edwards Brothers Malloy
Cover Printer: Lehigh-Phoenix Color/Hagerstown
Text Font: 10/14 Times Ten Roman

About the cover: The cover shows a heat chart of 270 monthly variables measuring different aspects of employment, production, income, and sales for the United States, 1974–2010. Each horizontal line depicts a different variable, and the horizontal axis is the date. Strong monthly increases in a variable are blue and sharp monthly declines are red. The simultaneous declines in many of these measures during recessions appear in the figure as vertical red bands.

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on appropriate page within text.

Photo Credits: page 410 left: Henrik Montgomery/Pressens Bild/AP Photo; page 410 right: Paul Sakuma/AP Photo; page 428 left: Courtesy of Allison Harris; page 428 right: Courtesy of Allison Harris; page 669 top left: John McCombe/AP Photo; bottom left: New York University/AFP/Newscom; top right: Denise Applewhite/Princeton University/AP Photo; bottom right: Courtesy of the University of Chicago/AP Photo.

Copyright © 2015, 2011, 2007 Pearson Education, Inc. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, 221 River Street, Hoboken, New Jersey 07030.

Many of the designations by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or call caps.

Library of Congress Cataloging-in-Publication Data
Stock, James H.
Introduction to econometrics/James H. Stock, Harvard University, Mark W. Watson, Princeton University.—Third edition update.
pages cm.—(The Pearson series in economics)
Includes bibliographical references and index.
ISBN 978-0-13-348687-2—ISBN 0-13-348687-7
1. Econometrics. I. Watson, Mark W. II. Title.
HB139.S765 2015
330.01’5195—dc23
2014018465

PEARSON

www.pearsonhighered.com

ISBN-10: 0-13-348687-7
ISBN-13: 978-0-13-348687-2

Brief Contents

This page intentionally left blank

Contents

PART THREE Further Topics in Regression Analysis

PART FIVE The Econometric Theory of Regression Analysis

Key Concepts

PART ONE Introduction and Review

PART TWO Fundamentals of Regression Analysis

PART THREE Further Topics in Regression Analysis

PART FOUR Regression Analysis of Economic Time Series Data

PART FIVE Regression Analysis of Economic Time Series Data

General Interest Boxes

This page intentionally left blank

Preface

Econometrics can be a fun course for both teacher and student. The real world of economics, business, and government is a complicated and messy place, full of competing ideas and questions that demand answers. Is it more effective to tackle drunk driving by passing tough laws or by increasing the tax on alcohol? Can you make money in the stock market by buying when prices are historically low, relative to earnings, or should you just sit tight, as the random walk theory of stock prices suggests? Can we improve elementary education by reducing class sizes, or should we simply have our children listen to Mozart for 10 minutes a day? Econometrics helps us sort out sound ideas from crazy ones and find quantitative answers to important quantitative questions. Econometrics opens a window on our complicated world that lets us see the relationships on which people, businesses, and governments base their decisions.

Introduction to Econometrics is designed for a first course in undergraduate econometrics. It is our experience that to make econometrics relevant in an introductory course, interesting applications must motivate the theory and the theory must match the applications. This simple principle represents a significant departure from the older generation of econometrics books, in which theoretical models and assumptions do not match the applications. It is no wonder that some students question the relevance of econometrics after they spend much of their time learning assumptions that they subsequently realize are unrealistic so that they must then learn "solutions" to "problems" that arise when the applications do not match the assumptions. We believe that it is far better to motivate the need for tools with a concrete application and then to provide a few simple assumptions that match the application. Because the theory is immediately relevant to the applications, this approach can make econometrics come alive.

New to the Third Edition

- Updated treatment of standard errors for panel data regression
- Discussion of when and why missing data can present a problem for regression analysis
- The use of regression discontinuity design as a method for analyzing quasi-experiments

- Updated discussion of weak instruments
- Discussion of the use and interpretation of control variables integrated into the core development of regression analysis
- Introduction of the "potential outcomes" framework for experimental data
- Additional general interest boxes
- Additional exercises, both pencil-and-paper and empirical

This third edition builds on the philosophy of the first and second editions that applications should drive the theory, not the other way around.

One substantial change in this edition concerns inference in regression with panel data (Chapter 10). In panel data, the data within an entity typically are correlated over time. For inference to be valid, standard errors must be computed using a method that is robust to this correlation. The chapter on panel data now uses one such method, clustered standard errors, from the outset. Clustered standard errors are the natural extension to panel data of the heteroskedasticity-robust standard errors introduced in the initial treatment of regression analysis in Part II. Recent research has shown that clustered standard errors have a number of desirable properties, which are now discussed in Chapter 10 and in a revised appendix to Chapter 10.

Another substantial set of changes concerns the treatment of experiments and quasi-experiments in Chapter 13. The discussion of differences-in-differences regression has been streamlined and draws directly on the multiple regression principles introduced in Part II. Chapter 13 now discusses regression discontinuity design, which is an intuitive and important framework for the analysis of quasi-experimental data. In addition, Chapter 13 now introduces the potential outcomes framework and relates this increasingly commonplace terminology to concepts that were introduced in Parts I and II.

This edition has a number of other significant changes. One is that it incorporates a precise but accessible treatment of control variables into the initial discussion of multiple regression. Chapter 7 now discusses conditions for control variables being successful in the sense that the coefficient on the variable of interest is unbiased even though the coefficients on the control variables generally are not. Other changes include a new discussion of missing data in Chapter 9, a new optional calculus-based appendix to Chapter 8 on slopes and elasticities of nonlinear regression functions, and an updated discussion in Chapter 12 of what to do if you have weak instruments. This edition also includes new general interest boxes, updated empirical examples, and additional exercises.

The Updated Third Edition

- The time series data used in Chapters 14–16 have been extended through the beginning of 2013 and now include the Great Recession.
- The empirical analysis in Chapter 14 now focuses on forecasting the growth rate of real GDP using the term spread, replacing the Phillips curve forecasts from earlier editions.
- Several new empirical exercises have been added to each chapter. Rather than include all of the empirical exercises in the text, we have moved many of them to the Companion Website, **www.pearsonhighered.com/stock_watson**. This has two main advantages: first, we can offer more and more in-depth exercises, and second, we can add and update exercises between editions. We encourage you to browse the empirical exercises available on the Companion Website.

Features of This Book

Introduction to Econometrics differs from other textbooks in three main ways. First, we integrate real-world questions and data into the development of the theory, and we take seriously the substantive findings of the resulting empirical analysis. Second, our choice of topics reflects modern theory and practice. Third, we provide theory and assumptions that match the applications. Our aim is to teach students to become sophisticated consumers of econometrics and to do so at a level of mathematics appropriate for an introductory course.

Real-World Questions and Data

We organize each methodological topic around an important real-world question that demands a specific numerical answer. For example, we teach single-variable regression, multiple regression, and functional form analysis in the context of estimating the effect of school inputs on school outputs. (Do smaller elementary school class sizes produce higher test scores?) We teach panel data methods in the context of analyzing the effect of drunk driving laws on traffic fatalities. We use possible racial discrimination in the market for home loans as the empirical application for teaching regression with a binary dependent variable (logit and probit). We teach instrumental variable estimation in the context of estimating the demand elasticity for cigarettes. Although these examples involve economic reasoning, all can be understood with only a single introductory course in economics, and many can be understood without any previous economics coursework. Thus the instructor can focus on teaching econometrics, not microeconomics or macroeconomics.

We treat all our empirical applications seriously and in a way that shows students how they can learn from data but at the same time be self-critical and aware of the limitations of empirical analyses. Through each application, we teach students to explore alternative specifications and thereby to assess whether their substantive findings are robust. The questions asked in the empirical applications are important, and we provide serious and, we think, credible answers. We encourage students and instructors to disagree, however, and invite them to reanalyze the data, which are provided on the textbook's Companion Website (**www.pearsonhighered.com/stock_watson**).

Contemporary Choice of Topics

Econometrics has come a long way since the 1980s. The topics we cover reflect the best of contemporary applied econometrics. One can only do so much in an introductory course, so we focus on procedures and tests that are commonly used in practice. For example:

- ***Instrumental variables regression.*** We present instrumental variables regression as a general method for handling correlation between the error term and a regressor, which can arise for many reasons, including omitted variables and simultaneous causality. The two assumptions for a valid instrument—exogeneity and relevance—are given equal billing. We follow that presentation with an extended discussion of where instruments come from and with tests of overidentifying restrictions and diagnostics for weak instruments, and we explain what to do if these diagnostics suggest problems.

- ***Program evaluation.*** An increasing number of econometric studies analyze either randomized controlled experiments or quasi-experiments, also known as natural experiments. We address these topics, often collectively referred to as program evaluation, in Chapter 13. We present this research strategy as an alternative approach to the problems of omitted variables, simultaneous causality, and selection, and we assess both the strengths and the weaknesses of studies using experimental or quasi-experimental data.

- ***Forecasting.*** The chapter on forecasting (Chapter 14) considers univariate (autoregressive) and multivariate forecasts using time series regression, not large simultaneous equation structural models. We focus on simple and reliable tools, such as autoregressions and model selection via an information criterion, that work well in practice. This chapter also features a practically oriented treatment of stochastic trends (unit roots), unit root tests, tests for structural breaks (at known and unknown dates), and pseudo out-of-sample forecasting, all in the context of developing stable and reliable time series forecasting models.

- ***Time series regression.*** We make a clear distinction between two very different applications of time series regression: forecasting and estimation of dynamic causal effects. The chapter on causal inference using time series data (Chapter 15) pays careful attention to when different estimation methods, including generalized least squares, will or will not lead to valid causal inferences and when it is advisable to estimate dynamic regressions using OLS with heteroskedasticity- and autocorrelation-consistent standard errors.

Theory That Matches Applications

Although econometric tools are best motivated by empirical applications, students need to learn enough econometric theory to understand the strengths and limitations of those tools. We provide a modern treatment in which the fit between theory and applications is as tight as possible, while keeping the mathematics at a level that requires only algebra.

Modern empirical applications share some common characteristics: The data sets typically are large (hundreds of observations, often more); regressors are not fixed over repeated samples but rather are collected by random sampling (or some other mechanism that makes them random); the data are not normally distributed; and there is no *a priori* reason to think that the errors are homoskedastic (although often there are reasons to think that they are heteroskedastic).

These observations lead to important differences between the theoretical development in this textbook and other textbooks:

- ***Large-sample approach.*** Because data sets are large, from the outset we use large-sample normal approximations to sampling distributions for hypothesis testing and confidence intervals. In our experience, it takes less time to teach the rudiments of large-sample approximations than to teach the Student t and exact F distributions, degrees-of-freedom corrections, and so forth. This large-sample approach also saves students the frustration of discovering that, because of nonnormal errors, the exact distribution theory they just mastered is irrelevant. Once taught in the context of the sample mean, the large-sample approach to hypothesis testing and confidence intervals carries directly through multiple regression analysis, logit and probit, instrumental variables estimation, and time series methods.

- ***Random sampling.*** Because regressors are rarely fixed in econometric applications, from the outset we treat data on all variables (dependent and independent) as the result of random sampling. This assumption matches our initial applications to cross-sectional data, it extends readily to panel and time series data, and because of our large-sample approach, it poses no additional conceptual or mathematical difficulties.
- ***Heteroskedasticity.*** Applied econometricians routinely use heteroskedasticity-robust standard errors to eliminate worries about whether heteroskedasticity is present or not. In this book, we move beyond treating heteroskedasticity as an exception or a "problem" to be "solved"; instead, we allow for heteroskedasticity from the outset and simply use heteroskedasticity-robust standard errors. We present homoskedasticity as a special case that provides a theoretical motivation for OLS.

Skilled Producers, Sophisticated Consumers

We hope that students using this book will become sophisticated consumers of empirical analysis. To do so, they must learn not only how to use the tools of regression analysis but also how to assess the validity of empirical analyses presented to them.

Our approach to teaching how to assess an empirical study is threefold. First, immediately after introducing the main tools of regression analysis, we devote Chapter 9 to the threats to internal and external validity of an empirical study. This chapter discusses data problems and issues of generalizing findings to other settings. It also examines the main threats to regression analysis, including omitted variables, functional form misspecification, errors-in-variables, selection, and simultaneity—and ways to recognize these threats in practice.

Second, we apply these methods for assessing empirical studies to the empirical analysis of the ongoing examples in the book. We do so by considering alternative specifications and by systematically addressing the various threats to validity of the analyses presented in the book.

Third, to become sophisticated consumers, students need firsthand experience as producers. Active learning beats passive learning, and econometrics is an ideal course for active learning. For this reason, the textbook website features data sets, software, and suggestions for empirical exercises of different scopes.

Approach to Mathematics and Level of Rigor

Our aim is for students to develop a sophisticated understanding of the tools of modern regression analysis, whether the course is taught at a "high" or a "low" level of mathematics. Parts I through IV of the text (which cover the substantive material) are accessible to students with only precalculus mathematics. Parts I through IV have fewer equations and more applications than many introductory econometrics books and far fewer equations than books aimed at mathematical sections of undergraduate courses. But more equations do not imply a more sophisticated treatment. In our experience, a more mathematical treatment does not lead to a deeper understanding for most students.

That said, different students learn differently, and for mathematically well-prepared students, learning can be enhanced by a more explicitly mathematical treatment. Part V therefore contains an introduction to econometric theory that is appropriate for students with a stronger mathematical background. When the mathematical chapters in Part V are used in conjunction with the material in Parts I through IV, this book is suitable for advanced undergraduate or master's level econometrics courses.

Contents and Organization

There are five parts to *Introduction to Econometrics*. This textbook assumes that the student has had a course in probability and statistics, although we review that material in Part I. We cover the core material of regression analysis in Part II. Parts III, IV, and V present additional topics that build on the core treatment in Part II.

Part I

Chapter 1 introduces econometrics and stresses the importance of providing quantitative answers to quantitative questions. It discusses the concept of causality in statistical studies and surveys the different types of data encountered in econometrics. Material from probability and statistics is reviewed in Chapters 2 and 3, respectively; whether these chapters are taught in a given course or are simply provided as a reference depends on the background of the students.

Part II

Chapter 4 introduces regression with a single regressor and ordinary least squares (OLS) estimation, and Chapter 5 discusses hypothesis tests and confidence intervals in the regression model with a single regressor. In Chapter 6, students learn how they can address omitted variable bias using multiple regression, thereby estimating the effect of one independent variable while holding other independent variables constant. Chapter 7 covers hypothesis tests, including *F*-tests, and confidence intervals in multiple regression. In Chapter 8, the linear regression model is extended to models with nonlinear population regression functions, with a focus on regression functions that are linear in the parameters (so that the parameters can be estimated by OLS). In Chapter 9, students step back and learn how to identify the strengths and limitations of regression studies, seeing in the process how to apply the concepts of internal and external validity.

Part III

Part III presents extensions of regression methods. In Chapter 10, students learn how to use panel data to control for unobserved variables that are constant over time. Chapter 11 covers regression with a binary dependent variable. Chapter 12 shows how instrumental variables regression can be used to address a variety of problems that produce correlation between the error term and the regressor, and examines how one might find and evaluate valid instruments. Chapter 13 introduces students to the analysis of data from experiments and quasi-, or natural, experiments, topics often referred to as "program evaluation."

Part IV

Part IV takes up regression with time series data. Chapter 14 focuses on forecasting and introduces various modern tools for analyzing time series regressions, such as unit root tests and tests for stability. Chapter 15 discusses the use of time series data to estimate causal relations. Chapter 16 presents some more advanced tools for time series analysis, including models of conditional heteroskedasticity.

Part V

Part V is an introduction to econometric theory. This part is more than an appendix that fills in mathematical details omitted from the text. Rather, it is a self-contained treatment of the econometric theory of estimation and inference in the linear regression model. Chapter 17 develops the theory of regression analysis for a single regressor; the exposition does not use matrix algebra, although it does demand a higher level of mathematical sophistication than the rest of the text. Chapter 18 presents and studies the multiple regression model, instrumental variables regression, and generalized method of moments estimation of the linear model, all in matrix form.

Prerequisites Within the Book

Because different instructors like to emphasize different material, we wrote this book with diverse teaching preferences in mind. To the maximum extent possible, the chapters in Parts III, IV, and V are "stand-alone" in the sense that they do not require first teaching all the preceding chapters. The specific prerequisites for each chapter are described in Table I. Although we have found that the sequence of topics adopted in the textbook works well in our own courses, the chapters are written in a way that allows instructors to present topics in a different order if they so desire.

Sample Courses

This book accommodates several different course structures.

TABLE I Guide to Prerequisites for Special-Topic Chapters in Parts III, IV, and V

	Prerequisite parts or chapters								
	Part I	Part II		Part III		Part IV			Part V
Chapter	1–3	4–7, 9	8	10.1, 10.2	12.1, 12.2	14.1–14.4	14.5–14.8	15	17
10	X[a]	X[a]	X						
11	X[a]	X[a]	X						
12.1, 12.2	X[a]	X[a]	X						
12.3–12.6	X[a]	X[a]	X	X	X				
13	X[a]	X[a]	X	X	X				
14	X[a]	X[a]	b						
15	X[a]	X[a]	b			X			
16	X[a]	X[a]	b			X	X	X	
17	X	X	X						
18	X	X	X		X				X

This table shows the minimum prerequisites needed to cover the material in a given chapter. For example, estimation of dynamic causal effects with time series data (Chapter 15) first requires Part I (as needed, depending on student preparation, and except as noted in footnote a), Part II (except for Chapter 8; see footnote b), and Sections 14.1 through 14.4.

[a]Chapters 10 through 16 use exclusively large-sample approximations to sampling distributions, so the optional Sections 3.6 (the Student *t* distribution for testing means) and 5.6 (the Student *t* distribution for testing regression coefficients) can be skipped.
[b]Chapters 14 through 16 (the time series chapters) can be taught without first teaching Chapter 8 (nonlinear regression functions) if the instructor pauses to explain the use of logarithmic transformations to approximate percentage changes.

Standard Introductory Econometrics

This course introduces econometrics (Chapter 1) and reviews probability and statistics as needed (Chapters 2 and 3). It then moves on to regression with a single regressor, multiple regression, the basics of functional form analysis, and the evaluation of regression studies (all Part II). The course proceeds to cover regression with panel data (Chapter 10), regression with a limited dependent variable (Chapter 11), and instrumental variables regression (Chapter 12), as time permits. The course concludes with experiments and quasi-experiments in Chapter 13, topics that provide an opportunity to return to the questions of estimating causal effects raised at the beginning of the semester and to recapitulate core regression methods. *Prerequisites: Algebra II and introductory statistics*.

Introductory Econometrics with Time Series and Forecasting Applications

Like a standard introductory course, this course covers all of Part I (as needed) and Part II. Optionally, the course next provides a brief introduction to panel data (Sections 10.1 and 10.2) and takes up instrumental variables regression (Chapter 12, or just Sections 12.1 and 12.2). The course then proceeds to Part IV, covering forecasting (Chapter 14) and estimation of dynamic causal effects (Chapter 15). If time permits, the course can include some advanced topics in time series analysis such as volatility clustering and conditional heteroskedasticity (Section 16.5). *Prerequisites: Algebra II and introductory statistics*.

Applied Time Series Analysis and Forecasting

This book also can be used for a short course on applied time series and forecasting, for which a course on regression analysis is a prerequisite. Some time is spent reviewing the tools of basic regression analysis in Part II, depending on student preparation. The course then moves directly to Part IV and works through forecasting (Chapter 14), estimation of dynamic causal effects (Chapter 15), and advanced topics in time series analysis (Chapter 16), including vector autoregressions and conditional heteroskedasticity. An important component of this course is hands-on forecasting exercises, available to instructors on the book's accompanying website. *Prerequisites: Algebra II and basic introductory econometrics or the equivalent*.

Introduction to Econometric Theory

This book is also suitable for an advanced undergraduate course in which the students have a strong mathematical preparation or for a master's level course in econometrics. The course briefly reviews the theory of statistics and probability as necessary (Part I). The course introduces regression analysis using the nonmathematical, applications-based treatment of Part II. This introduction is followed by the theoretical development in Chapters 17 and 18 (through Section 18.5). The course then takes up regression with a limited dependent variable (Chapter 11) and maximum likelihood estimation (Appendix 11.2). Next, the course optionally turns to instrumental variables regression and generalized method of moments (Chapter 12 and Section 18.7), time series methods (Chapter 14), and the estimation of causal effects using time series data and generalized least squares (Chapter 15 and Section 18.6). *Prerequisites: Calculus and introductory statistics. Chapter 18 assumes previous exposure to matrix algebra.*

Pedagogical Features

This textbook has a variety of pedagogical features aimed at helping students understand, retain, and apply the essential ideas. *Chapter introductions* provide real-world grounding and motivation, as well as brief road maps highlighting the sequence of the discussion. *Key terms* are boldfaced and defined in context throughout each chapter, and *Key Concept boxes* at regular intervals recap the central ideas. *General interest boxes* provide interesting excursions into related topics and highlight real-world studies that use the methods or concepts being discussed in the text. A *Summary* concluding each chapter serves as a helpful framework for reviewing the main points of coverage. The questions in the *Review the Concepts* section check students' understanding of the core content, *Exercises* give more intensive practice working with the concepts and techniques introduced in the chapter, and *Empirical Exercises* allow students to apply what they have learned to answer real-world empirical questions. At the end of the textbook, the *Appendix* provides statistical tables, the *References* section lists sources for further reading, and a *Glossary* conveniently defines many key terms in the book.

Supplements to Accompany the Textbook

The online supplements accompanying the third edition update of *Introduction to Econometrics* include the Instructor's Resource Manual, Test Bank, and PowerPoint® slides with text figures, tables, and Key Concepts. The Instructor's Resource Manual includes solutions to all the end-of-chapter exercises, while the Test Bank, offered in Testgen, provides a rich supply of easily edited test problems and questions of various types to meet specific course needs. These resources are available for download from the Instructor's Resource Center at **www.pearsonhighered.com/stock_watson**.

Companion Website

The Companion Website, found at **www.pearsonhighered.com/stock_watson**, provides a wide range of additional resources for students and faculty. These resources include more and more in depth empirical exercises, data sets for the empirical exercises, replication files for empirical results reported in the text, practice quizzes, answers to end-of-chapter Review the Concepts questions and Exercises, and EViews tutorials.

MyEconLab

The third edition update is accompanied by a robust MyEconLab course. The MyEconLab course includes all the Review the Concepts questions as well as some Exercises and Empirical Exercises. In addition, the enhanced eText available in MyEconLab for the third edition update includes URL links from the Exercises and Empirical Exercises to questions in the MyEconLab course and to the data that accompanies them. To register for MyEconLab and to learn more, log on to **www.myeconlab.com**.

Acknowledgments

A great many people contributed to the first edition of this book. Our biggest debts of gratitude are to our colleagues at Harvard and Princeton who used early drafts of this book in their classrooms. At Harvard's Kennedy School of Government, Suzanne Cooper provided invaluable suggestions and detailed comments on multiple drafts. As a coteacher with one of the authors (Stock), she also helped vet much of the material in this book while it was being developed for a required course for master's students at the Kennedy School. We are also indebted to two other Kennedy School colleagues, Alberto Abadie and Sue Dynarski, for their patient explanations of quasi-experiments and the field of program evaluation and for their detailed comments on early drafts of the text. At Princeton, Eli Tamer taught from an early draft and also provided helpful comments on the penultimate draft of the book.

We also owe much to many of our friends and colleagues in econometrics who spent time talking with us about the substance of this book and who collectively made so many helpful suggestions. Bruce Hansen (University of Wisconsin–Madison) and Bo Honore (Princeton) provided helpful feedback on very early outlines and preliminary versions of the core material in Part II. Joshua Angrist (MIT) and Guido Imbens (University of California, Berkeley) provided thoughtful suggestions about our treatment of materials on program evaluation. Our presentation of the material on time series has benefited from discussions with Yacine Ait-Sahalia (Princeton), Graham Elliott (University of California, San Diego), Andrew Harvey (Cambridge University), and Christopher Sims (Princeton). Finally, many people made helpful suggestions on parts of the manuscript close to their area of expertise: Don Andrews (Yale), John Bound (University of Michigan), Gregory Chow (Princeton), Thomas Downes (Tufts), David Drukker (StataCorp.), Jean Baldwin Grossman (Princeton), Eric Hanushek (Hoover Institution), James Heckman (University of Chicago), Han Hong (Princeton), Caroline Hoxby (Harvard), Alan Krueger (Princeton), Steven Levitt (University of Chicago), Richard Light (Harvard), David Neumark (Michigan State University), Joseph Newhouse (Harvard), Pierre Perron (Boston University), Kenneth Warner (University of Michigan), and Richard Zeckhauser (Harvard).

Many people were very generous in providing us with data. The California test score data were constructed with the assistance of Les Axelrod of the Standards and Assessments Division, California Department of Education. We are grateful to Charlie DePascale, Student Assessment Services, Massachusetts Department of Education, for his help with aspects of the Massachusetts test score data set. Christopher Ruhm (University of North Carolina, Greensboro) graciously provided us with his data set on drunk driving laws and traffic fatalities. The research department at the Federal Reserve Bank of Boston deserves thanks for putting together its data on racial discrimination in mortgage lending; we particularly thank Geoffrey Tootell for providing us with the updated version of the data set we use in Chapter 9 and Lynn Browne for explaining its policy context. We thank Jonathan Gruber (MIT) for sharing his data on cigarette sales, which we analyze in Chapter 12, and Alan Krueger (Princeton) for his help with the Tennessee STAR data that we analyze in Chapter 13.

We thank several people for carefully checking the page proof for errors. Kerry Griffin and Yair Listokin read the entire manuscript, and Andrew Fraker, Ori Heffetz, Amber Henry, Hong Li, Alessandro Tarozzi, and Matt Watson worked through several chapters.

In the first edition, we benefited from the help of an exceptional development editor, Jane Tufts, whose creativity, hard work, and attention to detail improved

the book in many ways, large and small. Pearson provided us with first-rate support, starting with our excellent editor, Sylvia Mallory, and extending through the entire publishing team. Jane and Sylvia patiently taught us a lot about writing, organization, and presentation, and their efforts are evident on every page of this book. We extend our thanks to the superb Pearson team, who worked with us on the second edition: Adrienne D'Ambrosio (senior acquisitions editor), Bridget Page (associate media producer), Charles Spaulding (senior designer), Nancy Fenton (managing editor) and her selection of Nancy Freihofer and Thompson Steele Inc. who handled the entire production process, Heather McNally (supplements coordinator), and Denise Clinton (editor-in-chief). Finally, we had the benefit of Kay Ueno's skilled editing in the second edition. We are also grateful to the excellent third edition Pearson team of Adrienne D'Ambrosio, Nancy Fenton, and Jill Kolongowski, as well as Mary Sanger, the project manager with Nesbitt Graphics. We also wish to thank the Pearson team who worked on the third edition update: Christina Masturzo, Carolyn Philips, Liz Napolitano, and Heidi Allgair, project manager with Cenveo® Publisher Services.

We also received a great deal of help and suggestions from faculty, students, and researchers as we prepared the third edition and its update. The changes made in the third edition incorporate or reflect suggestions, corrections, comments, data, and help provided by a number of researchers and instructors: Donald Andrews (Yale University), Jushan Bai (Columbia), James Cobbe (Florida State University), Susan Dynarski (University of Michigan), Nicole Eichelberger (Texas Tech University), Boyd Fjeldsted (University of Utah), Martina Grunow, Daniel Hamermesh (University of Texas–Austin), Keisuke Hirano (University of Arizona), Bo Honore (Princeton University), Guido Imbens (Harvard University), Manfred Keil (Claremont McKenna College), David Laibson (Harvard University), David Lee (Princeton University), Brigitte Madrian (Harvard University), Jorge Marquez (University of Maryland), Karen Bennett Mathis (Florida Department of Citrus), Alan Mehlenbacher (University of Victoria), Ulrich Müller (Princeton University), Serena Ng (Columbia University), Harry Patrinos (World Bank), Zhuan Pei (Brandeis University), Peter Summers (Texas Tech University), Andrey Vasnov (University of Sydney), and Douglas Young (Montana State University). We also benefited from student input from F. Hoces dela Guardia and Carrie Wilson.

Thoughtful reviews for the third edition were prepared for Addison-Wesley by Steve DeLoach (Elon University), Jeffrey DeSimone (University of Texas at Arlington), Gary V. Engelhardt (Syracuse University), Luca Flabbi (Georgetown University), Steffen Habermalz (Northwestern University), Carolyn J. Heinrich (University of Wisconsin–Madison), Emma M. Iglesias-Vazquez (Michigan State

University), Carlos Lamarche (University of Oklahoma), Vicki A. McCracken (Washington State University), Claudiney M. Pereira (Tulane University), and John T. Warner (Clemson University). We also received very helpful input on draft revisions of Chapters 7 and 10 from John Berdell (DePaul University), Janet Kohlhase (University of Houston), Aprajit Mahajan (Stanford University), Xia Meng (Brandeis University), and Chan Shen (Georgetown University).

Above all, we are indebted to our families for their endurance throughout this project. Writing this book took a long time, and for them, the project must have seemed endless. They, more than anyone else, bore the burden of this commitment, and for their help and support we are deeply grateful.

Introduction to Econometrics

CHAPTER

1 Economic Questions and Data

Ask a half dozen econometricians what econometrics is, and you could get a half dozen different answers. One might tell you that econometrics is the science of testing economic theories. A second might tell you that econometrics is the set of tools used for forecasting future values of economic variables, such as a firm's sales, the overall growth of the economy, or stock prices. Another might say that econometrics is the process of fitting mathematical economic models to real-world data. A fourth might tell you that it is the science and art of using historical data to make numerical, or quantitative, policy recommendations in government and business.

In fact, all these answers are right. At a broad level, econometrics is the science and art of using economic theory and statistical techniques to analyze economic data. Econometric methods are used in many branches of economics, including finance, labor economics, macroeconomics, microeconomics, marketing, and economic policy. Econometric methods are also commonly used in other social sciences, including political science and sociology.

This book introduces you to the core set of methods used by econometricians. We will use these methods to answer a variety of specific, quantitative questions from the worlds of business and government policy. This chapter poses four of those questions and discusses, in general terms, the econometric approach to answering them. The chapter concludes with a survey of the main types of data available to econometricians for answering these and other quantitative economic questions.

1.1 Economic Questions We Examine

Many decisions in economics, business, and government hinge on understanding relationships among variables in the world around us. These decisions require quantitative answers to quantitative questions.

This book examines several quantitative questions taken from current issues in economics. Four of these questions concern education policy, racial bias in mortgage lending, cigarette consumption, and macroeconomic forecasting.

Question #1: Does Reducing Class Size Improve Elementary School Education?

Proposals for reform of the U.S. public education system generate heated debate. Many of the proposals concern the youngest students, those in elementary schools. Elementary school education has various objectives, such as developing social skills, but for many parents and educators, the most important objective is basic academic learning: reading, writing, and basic mathematics. One prominent proposal for improving basic learning is to reduce class sizes at elementary schools. With fewer students in the classroom, the argument goes, each student gets more of the teacher's attention, there are fewer class disruptions, learning is enhanced, and grades improve.

But what, precisely, is the effect on elementary school education of reducing class size? Reducing class size costs money: It requires hiring more teachers and, if the school is already at capacity, building more classrooms. A decision maker contemplating hiring more teachers must weigh these costs against the benefits. To weigh costs and benefits, however, the decision maker must have a precise quantitative understanding of the likely benefits. Is the beneficial effect on basic learning of smaller classes large or small? Is it possible that smaller class size actually has no effect on basic learning?

Although common sense and everyday experience may suggest that more learning occurs when there are fewer students, common sense cannot provide a quantitative answer to the question of what exactly is the effect on basic learning of reducing class size. To provide such an answer, we must examine empirical evidence—that is, evidence based on data—relating class size to basic learning in elementary schools.

In this book, we examine the relationship between class size and basic learning, using data gathered from 420 California school districts in 1999. In the California data, students in districts with small class sizes tend to perform better on standardized tests than students in districts with larger classes. While this fact is consistent with the idea that smaller classes produce better test scores, it might simply reflect many other advantages that students in districts with small classes have over their counterparts in districts with large classes. For example, districts with small class sizes tend to have wealthier residents than districts with large classes, so students in small-class districts could have more opportunities for learning outside the classroom. It could be these extra learning opportunities that lead to higher test scores, not smaller class sizes. In Part II, we use multiple regression analysis to isolate the effect of changes in class size from changes in other factors, such as the economic background of the students.

Question #2: Is There Racial Discrimination in the Market for Home Loans?

Most people buy their homes with the help of a mortgage, a large loan secured by the value of the home. By law, U.S. lending institutions cannot take race into account when deciding to grant or deny a request for a mortgage: Applicants who are identical in all ways except their race should be equally likely to have their mortgage applications approved. In theory, then, there should be no racial bias in mortgage lending.

In contrast to this theoretical conclusion, researchers at the Federal Reserve Bank of Boston found (using data from the early 1990s) that 28% of black applicants are denied mortgages, while only 9% of white applicants are denied. Do these data indicate that, in practice, there is racial bias in mortgage lending? If so, how large is it?

The fact that more black than white applicants are denied in the Boston Fed data does not by itself provide evidence of discrimination by mortgage lenders because the black and white applicants differ in many ways other than their race. Before concluding that there is bias in the mortgage market, these data must be examined more closely to see if there is a difference in the probability of being denied for *otherwise identical* applicants and, if so, whether this difference is large or small. To do so, in Chapter 11 we introduce econometric methods that make it possible to quantify the effect of race on the chance of obtaining a mortgage, *holding constant* other applicant characteristics, notably their ability to repay the loan.

Question #3: How Much Do Cigarette Taxes Reduce Smoking?

Cigarette smoking is a major public health concern worldwide. Many of the costs of smoking, such as the medical expenses of caring for those made sick by smoking and the less quantifiable costs to nonsmokers who prefer not to breathe secondhand cigarette smoke, are borne by other members of society. Because these costs are borne by people other than the smoker, there is a role for government intervention in reducing cigarette consumption. One of the most flexible tools for cutting consumption is to increase taxes on cigarettes.

Basic economics says that if cigarette prices go up, consumption will go down. But by how much? If the sales price goes up by 1%, by what percentage will the quantity of cigarettes sold decrease? The percentage change in the quantity demanded resulting from a 1% increase in price is the *price elasticity of demand*.

If we want to reduce smoking by a certain amount, say 20%, by raising taxes, then we need to know the price elasticity of demand to calculate the price increase necessary to achieve this reduction in consumption. But what is the price elasticity of demand for cigarettes?

Although economic theory provides us with the concepts that help us answer this question, it does not tell us the numerical value of the price elasticity of demand. To learn the elasticity, we must examine empirical evidence about the behavior of smokers and potential smokers; in other words, we need to analyze data on cigarette consumption and prices.

The data we examine are cigarette sales, prices, taxes, and personal income for U.S. states in the 1980s and 1990s. In these data, states with low taxes, and thus low cigarette prices, have high smoking rates, and states with high prices have low smoking rates. However, the analysis of these data is complicated because causality runs both ways: Low taxes lead to high demand, but if there are many smokers in the state, then local politicians might try to keep cigarette taxes low to satisfy their smoking constituents. In Chapter 12, we study methods for handling this "simultaneous causality" and use those methods to estimate the price elasticity of cigarette demand.

Question #4: By How Much Will U.S. GDP Grow Next Year?

It seems that people always want a sneak preview of the future. What will sales be next year at a firm that is considering investing in new equipment? Will the stock market go up next month, and, if it does, by how much? Will city tax receipts next year cover planned expenditures on city services? Will your microeconomics exam next week focus on externalities or monopolies? Will Saturday be a nice day to go to the beach?

One aspect of the future in which macroeconomists are particularly interested is the growth of real economic activity, as measured by real gross domestic product (GDP), during the next year. A management consulting firm might advise a manufacturing client to expand its capacity based on an upbeat forecast of economic growth. Economists at the Federal Reserve Board in Washington, D.C., are mandated to set policy to keep real GDP near its potential in order to maximize employment. If they forecast anemic GDP growth over the next year, they might expand liquidity in the economy by reducing interest rates or other measures, in an attempt to boost economic activity.

Professional economists who rely on precise numerical forecasts use econometric models to make those forecasts. A forecaster's job is to predict the future by using the past, and econometricians do this by using economic theory and statistical techniques to quantify relationships in historical data.

The data we use to forecast the growth rate of GDP are past values of GDP and the "term spread" in the United States. The *term spread* is the difference between long-term and short-term interest rates. It measures, among other things, whether investors expect short-term interest rates to rise or fall in the future. The term spread is usually positive, but it tends to fall sharply before the onset of a recession. One of the GDP growth rate forecasts we develop and evaluate in Chapter 14 is based on the term spread.

Quantitative Questions, Quantitative Answers

Each of these four questions requires a numerical answer. Economic theory provides clues about that answer—for example, cigarette consumption ought to go down when the price goes up—but the actual value of the number must be learned empirically, that is, by analyzing data. Because we use data to answer quantitative questions, our answers always have some uncertainty: A different set of data would produce a different numerical answer. Therefore, the conceptual framework for the analysis needs to provide both a numerical answer to the question and a measure of how precise the answer is.

The conceptual framework used in this book is the multiple regression model, the mainstay of econometrics. This model, introduced in Part II, provides a mathematical way to quantify how a change in one variable affects another variable, holding other things constant. For example, what effect does a change in class size have on test scores, *holding constant* or *controlling for* student characteristics (such as family income) that a school district administrator cannot control? What effect does your race have on your chances of having a mortgage application granted, *holding constant* other factors such as your ability to repay the loan? What effect does a 1% increase in the price of cigarettes have on cigarette consumption, *holding constant* the income of smokers and potential smokers? The multiple regression model and its extensions provide a framework for answering these questions using data and for quantifying the uncertainty associated with those answers.

1.2 Causal Effects and Idealized Experiments

Like many other questions encountered in econometrics, the first three questions in Section 1.1 concern causal relationships among variables. In common usage, an action is said to cause an outcome if the outcome is the direct result, or consequence, of that action. Touching a hot stove causes you to get burned; drinking water causes you to be less thirsty; putting air in your tires causes them to inflate; putting fertilizer on your tomato plants causes them to produce more tomatoes. Causality means that a specific action (applying fertilizer) leads to a specific, measurable consequence (more tomatoes).

Estimation of Causal Effects

How best might we measure the causal effect on tomato yield (measured in kilograms) of applying a certain amount of fertilizer, say 100 grams of fertilizer per square meter?

One way to measure this causal effect is to conduct an experiment. In that experiment, a horticultural researcher plants many plots of tomatoes. Each plot is tended identically, with one exception: Some plots get 100 grams of fertilizer per square meter, while the rest get none. Moreover, whether a plot is fertilized or not is determined randomly by a computer, ensuring that any other differences between the plots are unrelated to whether they receive fertilizer. At the end of the growing season, the horticulturalist weighs the harvest from each plot. The difference between the average yield per square meter of the treated and untreated plots is the effect on tomato production of the fertilizer treatment.

This is an example of a **randomized controlled experiment**. It is controlled in the sense that there are both a **control group** that receives no treatment (no fertilizer) and a **treatment group** that receives the treatment (100 g/m^2 of fertilizer). It is randomized in the sense that the treatment is assigned randomly. This random assignment eliminates the possibility of a systematic relationship between, for example, how sunny the plot is and whether it receives fertilizer so that the only systematic difference between the treatment and control groups is the treatment. If this experiment is properly implemented on a large enough scale, then it will yield an estimate of the causal effect on the outcome of interest (tomato production) of the treatment (applying 100 g/m^2 of fertilizer).

In this book, the **causal effect** is defined to be the effect on an outcome of a given action or treatment, as measured in an ideal randomized controlled experiment. In such an experiment, the only systematic reason for differences in outcomes between the treatment and control groups is the treatment itself.

It is possible to imagine an ideal randomized controlled experiment to answer each of the first three questions in Section 1.1. For example, to study class size, one can imagine randomly assigning "treatments" of different class sizes to different groups of students. If the experiment is designed and executed so that the only systematic difference between the groups of students is their class size, then in theory this experiment would estimate the effect on test scores of reducing class size, holding all else constant.

The concept of an ideal randomized controlled experiment is useful because it gives a definition of a causal effect. In practice, however, it is not possible to perform ideal experiments. In fact, experiments are relatively rare in econometrics because often they are unethical, impossible to execute satisfactorily, or prohibitively expensive. The concept of the ideal randomized controlled experiment does, however, provide a theoretical benchmark for an econometric analysis of causal effects using actual data.

Forecasting and Causality

Although the first three questions in Section 1.1 concern causal effects, the fourth—forecasting the growth rate of GDP—does not. You do not need to know a causal relationship to make a good forecast. A good way to "forecast" whether it is raining is to observe whether pedestrians are using umbrellas, but the act of using an umbrella does not cause it to rain.

Even though forecasting need not involve causal relationships, economic theory suggests patterns and relationships that might be useful for forecasting. As we see in Chapter 14, multiple regression analysis allows us to quantify historical relationships suggested by economic theory, to check whether those relationships have been stable over time, to make quantitative forecasts about the future, and to assess the accuracy of those forecasts.

1.3 Data: Sources and Types

In econometrics, data come from one of two sources: experiments or nonexperimental observations of the world. This book examines both experimental and nonexperimental data sets.

Experimental Versus Observational Data

Experimental data come from experiments designed to evaluate a treatment or policy or to investigate a causal effect. For example, the state of Tennessee financed a large randomized controlled experiment examining class size in the 1980s. In that experiment, which we examine in Chapter 13, thousands of students were randomly assigned to classes of different sizes for several years and were given standardized tests annually.

The Tennessee class size experiment cost millions of dollars and required the ongoing cooperation of many administrators, parents, and teachers over several years. Because real-world experiments with human subjects are difficult to administer and to control, they have flaws relative to ideal randomized controlled experiments. Moreover, in some circumstances, experiments are not only expensive and difficult to administer but also unethical. (Would it be ethical to offer randomly selected teenagers inexpensive cigarettes to see how many they buy?) Because of these financial, practical, and ethical problems, experiments in economics are relatively rare. Instead, most economic data are obtained by observing real-world behavior.

Data obtained by observing actual behavior outside an experimental setting are called **observational data**. Observational data are collected using surveys, such as telephone surveys of consumers, and administrative records, such as historical records on mortgage applications maintained by lending institutions.

Observational data pose major challenges to econometric attempts to estimate causal effects, and the tools of econometrics are designed to tackle these challenges. In the real world, levels of "treatment" (the amount of fertilizer in the tomato example, the student–teacher ratio in the class size example) are not assigned at random, so it is difficult to sort out the effect of the "treatment" from other relevant factors. Much of econometrics, and much of this book, is devoted to methods for meeting the challenges encountered when real-world data are used to estimate causal effects.

Whether the data are experimental or observational, data sets come in three main types: cross-sectional data, time series data, and panel data. In this book, you will encounter all three types.

Cross-Sectional Data

Data on different entities—workers, consumers, firms, governmental units, and so forth—for a single time period are called **cross-sectional data**. For example, the data on test scores in California school districts are cross sectional. Those data are for 420 entities (school districts) for a single time period (1999). In general, the number of entities on which we have observations is denoted n; so, for example, in the California data set, $n = 420$.

The California test score data set contains measurements of several different variables for each district. Some of these data are tabulated in Table 1.1. Each row lists data for a different district. For example, the average test score for the first district ("district #1") is 690.8; this is the average of the math and science test scores for all fifth graders in that district in 1999 on a standardized test (the Stanford

TABLE 1.1 Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student–Teacher Ratio	Expenditure per Pupil ($)	Percentage of Students Learning English
1	690.8	17.89	$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
⋮	⋮	⋮	⋮	⋮
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

Note: The California test score data set is described in Appendix 4.1.

Achievement Test). The average student–teacher ratio in that district is 17.89; that is, the number of students in district #1 divided by the number of classroom teachers in district #1 is 17.89. Average expenditure per pupil in district #1 is $6385. The percentage of students in that district still learning English—that is, the percentage of students for whom English is a second language and who are not yet proficient in English—is 0%.

The remaining rows present data for other districts. The order of the rows is arbitrary, and the number of the district, which is called the **observation number**, is an arbitrarily assigned number that organizes the data. As you can see in the table, all the variables listed vary considerably.

With cross-sectional data, we can learn about relationships among variables by studying differences across people, firms, or other economic entities during a single time period.

Time Series Data

Time series data are data for a single entity (person, firm, country) collected at multiple time periods. Our data set on the growth rate of GDP and the term spread in the United States is an example of a time series data set. The data set