PRINCIPLES OF

ECONOMETRICS

5th Edition

R. CARTER HILL | WILLIAM E. GRIFFITHS | GUAY C. LIM

WILEY

Principles of Econometrics

Fifth Edition

R. CARTER HILL
Louisiana State University

WILLIAM E. GRIFFITHS
University of Melbourne

GUAY C. LIM
University of Melbourne

WILEY

EDITORIAL DIRECTOR	Michael McDonald
EXECUTIVE EDITOR	Lise Johnson
CHANNEL MARKETING MANAGER	Michele Szczesniak
CONTENT ENABLEMENT SENIOR MANAGER	Leah Michael
CONTENT MANAGEMENT DIRECTOR	Lisa Wojcik
CONTENT MANAGER	Nichole Urban
SENIOR CONTENT SPECIALIST	Nicole Repasky
PRODUCTION EDITOR	Abidha Sulaiman
COVER PHOTO CREDIT	© liuzishan/iStockphoto

This book was set in STIX-Regular 10/12pt by SPi Global and printed and bound by Strategic Content Imaging.

This book is printed on acid free paper. ∞

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

ISBN: 9781118452271 (PBK)
ISBN: 9781119320951 (EVALC)

Library of Congress Cataloging-in-Publication Data

Names: Hill, R. Carter, author. | Griffiths, William E., author. | Lim, G. C. (Guay C.), author.
Title: Principles of econometrics / R. Carter Hill, Louisiana State University, William E. Griffiths, University of Melbourne, Guay C. Lim, University of Melbourn.
Description: Fifth Edition. | Hoboken : Wiley, 2017. | Revised edition of the authors' Principles of econometrics, c2011. | Includes bibliographical references and index. |
Identifiers: LCCN 2017043094 (print) | LCCN 2017056927 (ebook) | ISBN 9781119342854 (pdf) | ISBN 9781119320944 (epub) | ISBN 9781118452271 (paperback) | ISBN 9781119320951 (eval copy)
Subjects: LCSH: Econometrics. | BISAC: BUSINESS & ECONOMICS / Econometrics.
Classification: LCC HB139 (ebook) | LCC HB139 .H548 2018 (print) | DDC 330.01/5195—dc23
LC record available at https://lccn.loc.gov/2017043094

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

10 9 8 7 6 5 4 3 2 1

Carter Hill dedicates this work to his wife, Melissa Waters
Bill Griffiths dedicates this work to Jill, David, and Wendy Griffiths
Guay Lim dedicates this work to Tony Meagher

Brief Contents

Preface

Principles of Econometrics, *Fifth Edition*, is an introductory book for undergraduate students in economics and finance, as well as first-year graduate students in economics, finance, accounting, agricultural economics, marketing, public policy, sociology, law, forestry, and political science. We assume that students have taken courses in the principles of economics and elementary statistics. Matrix algebra is not used, and we introduce and develop calculus concepts in an Appendix. The title *Principles of Econometrics* emphasizes our belief that econometrics should be part of the economics curriculum, in the same way as the principles of microeconomics and the principles of macroeconomics. Those who have been studying and teaching econometrics as long as we have will remember that *Principles of Econometrics* was the title that Henri Theil used for his 1971 classic, which was also published by John Wiley & Sons. Our choice of the same title is not intended to signal that our book is similar in level and content. Theil's work was, and remains, a unique treatise on advanced graduate level econometrics. Our book is an introductory level econometrics text.

Book Objectives

Principles of Econometrics is designed to give students an understanding of why econometrics is necessary and to provide them with a working knowledge of basic econometric tools so that

i. They can apply these tools to modeling, estimation, inference, and forecasting in the context of real-world economic problems.

ii. They can evaluate critically the results and conclusions from others who use basic econometric tools.

iii. They have a foundation and understanding for further study of econometrics.

iv. They have an appreciation of the range of more advanced techniques that exist and that may be covered in later econometric courses.

The book is neither an econometrics cookbook nor is it in a theorem-proof format. It emphasizes motivation, understanding, and implementation. Motivation is achieved by introducing very simple economic models and asking economic questions that the student can answer. Understanding is aided by lucid description of techniques, clear interpretation, and appropriate applications. Learning is reinforced by doing, with clear worked examples in the text and exercises at the end of each chapter.

Overview of Contents

This fifth edition is a major revision in format and content. The chapters contain core material and exercises, while appendices contain more advanced material. Chapter examples are now identified and separated from other content so that they may be easily referenced. From the beginning, we recognize the observational nature of most economic data and modify modeling assumptions accordingly. Chapter 1 introduces econometrics and gives general guidelines for writing an empirical research paper and locating economic data sources. The Probability Primer preceding Chapter 2 summarizes essential properties of random variables and their probability distributions and reviews summation notation. The simple linear regression model is covered in Chapters 2–4, while the multiple regression model is treated in Chapters 5–7. Chapters 8 and 9 introduce econometric problems that are unique to cross-sectional data (heteroskedasticity) and time-series data (dynamic models), respectively. Chapters 10 and 11 deal with endogenous regressors, the failure of least squares when a regressor is endogenous, and instrumental variables estimation, first in the general case, and then in the simultaneous equations model. In Chapter 12, the analysis of time-series data is extended to discussions of nonstationarity and cointegration. Chapter 13 introduces econometric issues specific to two special time-series models, the vector error correction, and vector autoregressive models, while Chapter 14 considers the analysis of volatility in data and the ARCH model. In Chapters 15 and 16, we introduce microeconometric models for panel data and qualitative and limited dependent variables. In appendices A, B, and C, we introduce math, probability, and statistical inference concepts that are used in the book.

Summary of Changes and New Material

This edition includes a great deal of new material, including new examples and exercises using real data and some significant reorganizations. In this edition, we number examples for easy reference and offer 25–30 new exercises in each chapter. Important new features include

- Chapter 1 includes a discussion of data types and sources of economic data on the Internet. Tips on writing a research paper are given "up front" so that students can form ideas for a paper as the course develops.
- A Probability Primer precedes Chapter 2. This Primer reviews the concepts of random variables and how probabilities are calculated from probability density functions. Mathematical expectation and rules of expected values are summarized for discrete random variables. These rules are applied to develop the concepts of variance and covariance. Calculations of probabilities using the normal distribution are illustrated. New material includes sections on conditional expectation, conditional variance, iterated expectations, and the bivariate normal distribution.
- Chapter 2 now starts with a discussion of causality. We define the population regression function and discuss exogeneity in considerable detail. The properties of the ordinary least squares (OLS) estimator are examined within the framework of the new assumptions. New appendices have been added on the independent variable, covering the various assumptions that might be made about the sampling process, derivations of the properties of the OLS estimator, and Monte Carlo experiments to numerically illustrate estimator properties.
- In Chapter 3, we note that hypothesis test mechanics remain the same under the revised assumptions because test statistics are "pivotal," meaning that their distributions under the null hypothesis do not depend on the data. In appendices, we add an extended discussion of test behavior under the alternative, introduce the noncentral t-distribution, and illustrate test power. We also include new Monte Carlo experiments illustrating test properties when the explanatory variable is random.
- Chapter 4 discusses in detail nonlinear relationships such as the log-log, log-linear, linear-log, and polynomial models. We have expanded the discussion of diagnostic residual plots and added sections on identifying influential observations. The familiar concepts of compound interest are used to motivate several log-linear models. We add an appendix on the concept of mean squared error and the minimum mean squared error predictor.
- Chapter 5 introduces multiple regression in the random-x framework. The Frisch–Waugh–Lovell (FWL) theorem is introduced as a way to help understand interpretation of the multiple regression model and used throughout the remainder of the book. Discussions of the properties of the OLS estimator, and interval estimates and t-tests, are updated. The large sample properties of the OLS estimator, and the delta method, are now introduced within the chapter rather than an appendix. Appendices provide further discussion and Monte Carlo

properties to illustrate the delta method. We provide a new appendix on bootstrapping and its uses.

- Chapter 6 adds a new section on large sample tests. We explain the use of control variables and the difference between causal and predictive models. We revise the discussion of collinearity and include a discussion of influential observations. We introduce nonlinear regression models and nonlinear least squares algorithms are discussed. Appendices are added to discuss the statistical power of F-tests and further uses of the Frisch–Waugh–Lovell theorem.
- Chapter 7 now includes an extensive section on treatment effects and causal modeling in Rubin's potential outcomes framework. We explain and illustrate the interesting regression discontinuity design. An appendix includes a discussion of the important "overlap" assumption.
- Chapter 8 has been reorganized so that the heteroskedasticity robust variance of the OLS estimator appears before testing. We add a section on how model specification can ameliorate heteroskedasticity in some applications. We add appendices to explain the properties of the OLS residuals and another to explain alternative robust sandwich variance estimators. We present Monte Carlo experiments to illustrate the differences.
- Chapter 9 has been reorganized and streamlined. The initial section introduces the different ways that dynamic elements can be added to the regression model. These include using finite lag models, infinite lag models, and autoregressive errors. We carefully discuss autocorrelations, including testing for autocorrelation and representing autocorrelations using a correlogram. After introducing the concepts of stationarity and weak dependence, we discuss the general notions of forecasting and forecast intervals in the context of autoregressive distributed lag (ARDL) models. Following these introductory concepts, there are details of estimating and using alternative models, covering such topics as choosing lag lengths, testing for Granger causality, the Lagrange multiplier test for serial correlation, and using models for policy analysis. We provide very specific sets of assumptions for time-series regression models and outline how heteroskedastic and autocorrelation consistent, robust, standard errors are used. We discuss generalized least squares estimation of a time-series regression model and its relation to nonlinear least squares regression. A detailed discussion of the infinite lag model and how to use multiplier analysis is provided. An appendix contains details of the Durbin–Watson test.
- Chapter 10 on endogeneity problems has been streamlined because the concept of random explanatory variables is now introduced much earlier in the book. We provide further analysis of weak instruments and how weak instruments adversely affect the precision of IV estimation. The details of the Hausman test are now included in the chapter.
- Chapter 11 now adds Klein's Model I as an example.
- Chapter 12 includes more details of deterministic trends and unit roots. The section on unit root testing has been restructured so that each Dickey–Fuller test is more fully explained and illustrated with an example. Numerical examples of ARDL models with nonstationary variables that are, and are not, cointegrated have been added.
- The data in Chapter 13 have been updated and new exercises added.
- Chapter 14 mentions further extensions of ARCH volatility models.
- Chapter 15 has been restructured to give priority to how panel data can be used to cope with the endogeneity caused by unobserved heterogeneity. We introduce the advantages of having panel data using the first difference estimator, and then discuss the within/fixed effects estimator. We provide an extended discussion of cluster robust standard errors in both the OLS and fixed effects model. We discuss the Mundlak version of the Hausman test for endogeneity. We give brief mention to how to extend the use of panel data in several ways.

- The Chapter 16 discussion of binary choice models is reorganized and expanded. It now includes brief discussions of advanced topics such as binary choice models with endogenous explanatory variables and binary choice models with panel data. We add new appendices on random utility models and latent variable models.
- Appendix A includes new sections on second derivatives and finding maxima and minima of univariate and bivariate functions.
- Appendix B includes new material on conditional expectations and conditional variances, including several useful decompositions. We include new sections on truncated random variables, including the truncated normal and Poisson distributions. To facilitate discussions of test power, we have new sections on the noncentral t-distribution, the noncentral Chi-square distribution, and the noncentral F-distribution. We have included an expanded new section on the log-normal distribution.
- Appendix C content does not change a great deal, but 20 new exercises are included.
- Statistical Tables for the Standard Normal cumulative distribution function, the t-distribution and Chi-square distribution critical values for selected percentiles, the F-distribution critical values for the 95th and 99th percentiles, and the Standard Normal density function values appear in Appendix D.
- A useful "cheat sheet" of essential formulas is provided at the authors' website, www.principlesofeconometrics.com, rather than inside the covers as in the previous edition.

For Instructors: Suggested Course Plans

Principles of Econometrics, Fifth Edition is suitable for one or two semester courses at the undergraduate or first year graduate level. Some suitable plans for alternative courses are as follows:

- One-semester survey course: Sections P.1–P.6.2 and P.7; Sections 2.1–2.9; Chapters 3 and 4; Sections 5.1–5.6; Sections 6.1–6.5; Sections 7.1–7.3; Sections 8.1–8.4 and 8.6; Sections 9.1–9.4.2 and 9.5–9.5.1.
- One-semester survey course enhancements for Master's or Ph.D.: Include Appendices for Chapters 2–9.
- Two-semester survey second course, cross-section emphasis: Section P.6; Section 2.10; Section 5.7; Section 6.6; Sections 7.4–7.6; Sections 8.5 and 8.6.3–8.6.5; Sections 10.1–10.4; Sections 15.1–15.4; Sections 16.1–16.2 and 16.6;
- Two-semester survey second course, time series emphasis: Section P.6; Section 2.10; Section 5.7; Section 6.6; Sections 7.4–7.6; Sections 8.5 and 8.6.3–8.6.5; Section 9.5; Sections 10.1–10.4; Sections 12.1–12.5; Sections 13.1–13.5; Sections 14.1–14.4;
- Two-semester survey course enhancements for Master's or Ph.D.: Include Appendices from Chapters 10, Chapter 11, Appendices 15A–15B, Sections 16.3–16.5 and 16.7, Appendices 16A–16D, Book Appendices B and C.

Computer Supplement Books

There are several computer supplements to *Principles of Econometrics, Fifth Edition*. The supplements are not versions of the text and cannot substitute for the text. They use the examples in the text as a vehicle for learning the software. We show how to use the software to get the answers for each example in the text.

- *Using EViews for the Principles of Econometrics, Fifth Edition,* by William E. Griffiths, R. Carter Hill, and Guay C. Lim [ISBN 9781118469842]. This supplementary book presents the EViews 10 [www.eviews.com] software commands required for the examples in *Principles of Econometrics* in a clear and concise way. It includes many illustrations that are student friendly. It is useful not only for students and instructors who will be using this software as part of their econometrics course but also for those who wish to learn how to use EViews.
- *Using Stata for the Principles of Econometrics, Fifth Edition*, by Lee C. Adkins and R. Carter Hill [ISBN 9781118469873]. This supplementary book presents the Stata 15.0 [www.stata.com] software commands required for the examples in *Principles of Econometrics*. It is useful not only for students and instructors who will be using this software as part of their econometrics course but also for those who wish to learn how to use Stata. Screen shots illustrate the use of Stata's drop-down menus. Stata commands are explained and the use of "do-files" illustrated.
- *Using SAS for the Principles of Econometrics, Fifth Edition*, by Randall C. Campbell and R. Carter Hill [ISBN 9781118469880]. This supplementary book gives SAS 9.4 [www.sas.com] software commands for econometric tasks, following the general outline of *Principles of Econometrics, Fifth Edition*. It includes enough background material on econometrics so that instructors using any textbook can easily use this book as a supplement. The volume spans several levels of econometrics. It is suitable for undergraduate students who will use "canned" SAS statistical procedures, and for graduate students who will use advanced procedures as well as direct programming in SAS's matrix language; the latter is discussed in chapter appendices.
- *Using Excel for Principles of Econometrics, Fifth Edition,* by Genevieve Briand and R. Carter Hill [ISBN 9781118469835]. This supplement explains how to use Excel to reproduce most of the examples in *Principles of Econometrics*. Detailed instructions and screen shots are provided explaining both the computations and clarifying the operations of Excel. Templates are developed for common tasks.
- *Using GRETL for Principles of Econometrics, Fifth Edition*, by Lee C. Adkins. This free supplement, readable using Adobe Acrobat, explains how to use the freely available statistical software GRETL (download from http://gretl.sourceforge.net). Professor Adkins explains in detail, and using screen shots, how to use GRETL to replicate the examples in *Principles of Econometrics*. The manual is freely available at www.learneconometrics.com/gretl.html.
- *Using R for Principles of Econometrics, Fifth Edition*, by Constantin Colonescu and R. Carter Hill. This free supplement, readable using Adobe Acrobat, explains how to use the freely available statistical software *R* (download from https://www.r-project.org/). The supplement explains in detail, and using screen shots, how to use *R* to replicate the examples in *Principles of Econometrics, Fifth Edition*. The manual is freely available at https://bookdown.org/ccolonescu/RPOE5/.

Data Files

Data files for the book are provided in a variety of formats at the book website www.wiley.com/college/hill. These include

- ASCII format (*.dat). These are text files containing only data.
- Definition files (*.def). These are text files describing the data file contents, with a listing of variable names, variable definitions, and summary statistics.
- EViews (*.wf1) workfiles for each data file.
- Excel (*.xls) workbooks for each data file, including variable names in the first row.

- Comma separated values (*.csv) files that can be read into almost all software.
- Stata (*.dta) data files.
- SAS (*.sas7bdat) data files.
- GRETL (*.gdt) data files.
- R (*.rdata) data files.

The author website www.principlesofeconometrics.com includes a complete list of the data files and where they are used in the book.

Additional Resources

The book website www.principlesofeconometrics.com includes

- Individual data files in each format as well as ZIP files containing data in compressed format.
- Book errata.
- Brief answers to odd number problems. These answers are also provided on the book website at www.wiley.com/college/hill.
- Additional examples with solutions. Some extra examples come with complete solutions so that students will know what a good answer looks like.
- Tips on writing research papers.

Resources for Instructors

For instructors, also available at the website www.wiley.com/college/hill are

- Complete solutions, in both Microsoft Word and *.pdf formats, to *all* exercises in the text.
- PowerPoint slides and PowerPoint Viewer.

Acknowledgments

Authors Hill and Griffiths want to acknowledge the gifts given to them over the past 40 years by mentor, friend, and colleague George Judge. Neither this book nor any of the other books we have shared in the writing of would have ever seen the light of day without his vision and inspiration.

We also wish to give thanks to the many students and faculty who have commented on the fourth edition of the text and contributed to the fifth edition. This list includes Alejandra Breve Ferrari, Alex James, Alyssa Wans, August Saibeni, Barry Rafferty, Bill Rising, Bob Martin, Brad Lewis, Bronson Fong, David Harris, David Iseral, Deborah Williams, Deokrye Baek, Diana Whistler, Emma Powers, Ercan Saridogan, Erdogan Cevher, Erika Haguette, Ethan Luedecke, Gareth Thomas, Gawon Yoon, Genevieve Briand, German Altgelt, Glenn Sueyoshi, Henry McCool, James Railton, Jana Ruimerman, Jeffery Parker, Joe Goss, John Jackson, Julie Leiby, Katharina Hauck, Katherine Ramirez, Kelley Pace, Lee Adkins, Matias Cattaneo, Max O'Krepki, Meagan McCollum, Micah West, Michelle Savolainen, Oystein Myrland, Patrick Scholten, Randy Campbell, Regina Riphahn, Sandamali Kankanamge, Sergio Pastorello, Shahrokh Towfighi, Tom Fomby, Tong Zeng, Victoria Pryor, Yann Nicolas, and Yuanbo Zhang. In the book errata we acknowledge those who have pointed out our errors.

R. Carter Hill
William E. Griffiths
Guay C. Lim

Table of Contents

5 The Multiple Regression Model 196

6 Further Inference in the Multiple Regression Model 260

A Mathematical Tools 748

B Probability Concepts 768

C Review of Statistical Inference 812

List of Examples

CHAPTER 1

An Introduction to Econometrics

1.1 Why Study Econometrics?

Econometrics is fundamental for economic measurement. However, its importance extends far beyond the discipline of economics. Econometrics is a set of research tools also employed in the business disciplines of accounting, finance, marketing, and management. It is used by social scientists, specifically researchers in history, political science, and sociology. Econometrics plays an important role in such diverse fields as forestry and agricultural economics. This breadth of interest in econometrics arises in part because economics is the foundation of business analysis and is the core social science. Thus, research methods employed by economists, which includes the field of econometrics, are useful to a broad spectrum of individuals.

Econometrics plays a special role in the training of economists. As a student of economics, you are learning to "think like an economist." You are learning economic concepts such as opportunity cost, scarcity, and comparative advantage. You are working with economic models of supply and demand, macroeconomic behavior, and international trade. Through this training you become a person who better understands the world in which we live; you become someone who understands how markets work, and the way in which government policies affect the marketplace.

If economics is your major or minor field of study, a wide range of opportunities is open to you upon graduation. If you wish to enter the business world, your employer will want to know the answer to the question, "What can you do for me?" Students taking a traditional economics curriculum answer, "I can think like an economist." While we may view such a response to be powerful, it is not very specific and may not be very satisfying to an employer who does not understand economics.

The problem is that a gap exists between what you have learned as an economics student and what economists actually do. Very few economists make their livings by studying economic theory alone, and those who do are usually employed by universities. Most economists, whether they work in the business world or for the government, or teach in universities, engage in economic analysis that is in part "empirical." By this we mean that they use economic data to estimate economic relationships, test economic hypotheses, and predict economic outcomes.

Studying econometrics fills the gap between being "a student of economics" and being "a practicing economist." With the econometric skills you will learn from this book, including how to work with econometric software, you will be able to elaborate on your answer to the employer's question above by saying "I can predict the sales of your product." "I can estimate the effect on your sales if your competition lowers its price by $1 per unit." "I can test whether your new ad campaign is actually increasing your sales." These answers are music to an employer's ears, because they reflect your ability to think like an economist and to analyze economic data.

Such pieces of information are keys to good business decisions. Being able to provide your employer with useful information will make you a valuable employee and increase your odds of getting a desirable job.

On the other hand, if you plan to continue your education by enrolling in graduate school or law school, you will find that this introduction to econometrics is invaluable. If your goal is to earn a master's or Ph.D. degree in economics, finance, data analytics, data science, accounting, marketing, agricultural economics, sociology, political science, or forestry, you will encounter more econometrics in your future. The graduate courses tend to be quite technical and mathematical, and the forest often gets lost in studying the trees. By taking this introduction to econometrics you will gain an overview of what econometrics is about and develop some "intuition" about how things work before entering a technically oriented course.

1.2 What Is Econometrics About?

At this point we need to describe the nature of econometrics. It all begins with a theory from your field of study—whether it is accounting, sociology, or economics—about how important variables are related to one another. In economics we express our ideas about relationships between economic variables using the mathematical concept of a function. For example, to express a relationship between income and consumption, we may write

$$CONSUMPTION = f(INCOME)$$

which says that the level of consumption is *some* function, $f(\bullet)$, of income.

The demand for an individual commodity—say, the Honda Accord—might be expressed as

$$Q^d = f(P, P^s, P^c, INC)$$

which says that the quantity of Honda Accords demanded, Q^d, is a function $f(P, P^s, P^c, INC)$ of the price of Honda Accords P, the price of cars that are substitutes P^s, the price of items that are complements P^c (like gasoline), and the level of income INC.

The supply of an agricultural commodity such as beef might be written as

$$Q^s = f\left(P, P^c, P^f\right)$$

where Q^s is the quantity supplied, P is the price of beef, P^c is the price of competitive products in production (e.g., the price of hogs), and P^f is the price of factors or inputs (e.g., the price of corn) used in the production process.

Each of the above equations is a general economic model that describes how we visualize the way in which economic variables are interrelated. Economic models of this type *guide our economic analysis*.

Econometrics allows us to go further than knowing that certain economic variables are interrelated, or even the direction of a relationship. Econometrics allows us to assign magnitudes to questions about the interrelationships between variables. One aspect of econometrics is **prediction** or **forecasting**. If we know the value of *INCOME*, what will be the magnitude of *CONSUMPTION*? If we have values for the prices of Honda Accords, their substitutes and complements, and income, how many Honda Accords will be sold? Similarly, we could ask how much beef would be supplied given values of the variables on which its supply depends.

A second contribution of econometrics is to enable us to say **how much** a change in one variable affects another. If the price for Honda Accords is increased, by how much will quantity demanded decline? If the price of beef goes up, by how much will quantity supplied increase? Finally, econometrics contributes to our understanding of the interrelationships between variables by giving us the ability to **test** the validity of hypothesized relationships.

Econometrics is about how we can use theory and data from economics, business, and the social sciences, along with tools from statistics, to predict outcomes, answer "how much" type questions, and test hypotheses.

1.2.1 Some Examples

Consider the problem faced by decision makers in a central bank. In the United States, the Federal Reserve System and, in particular, the Chair of the Board of Governors of the FRB must make decisions about interest rates. When prices are observed to rise, suggesting an increase in the inflation rate, the FRB must make a decision about whether to dampen the rate of growth of the economy. It can do so by raising the interest rate it charges its member banks when they borrow money (the discount rate) or the rate on overnight loans between banks (the federal funds rate). Increasing these rates sends a ripple effect through the economy, causing increases in other interest rates, such as those faced by would-be investors, who may be firms seeking funds for capital expansion or individuals who wish to buy consumer durables like automobiles and refrigerators. This has the economic effect of increasing costs, and consumers react by reducing the quantity of the durable goods demanded. Overall, aggregate demand falls, which slows the rate of inflation. These relationships are suggested by economic theory.

The real question facing the Chair is "*How much* should we increase the discount rate to slow inflation and yet maintain a stable and growing economy?" The answer will depend on the responsiveness of firms and individuals to increases in the interest rates and to the effects of reduced investment on gross national product (GNP). The key elasticities and multipliers are called **parameters**. The values of economic parameters are unknown and must be estimated using a sample of economic data when formulating economic policies.

Econometrics is about how to best estimate economic parameters given the data we have. "Good" econometrics is important since errors in the estimates used by policymakers such as the FRB may lead to interest rate corrections that are too large or too small, which has consequences for all of us.

Every day, decision-makers face "how much" questions similar to those facing the FRB Chair:

- A city council ponders the question of how much violent crime will be reduced if an additional million dollars is spent putting uniformed police on the street.
- The owner of a local Pizza Hut must decide how much advertising space to purchase in the local newspaper and thus must estimate the relationship between advertising and sales.
- Louisiana State University must estimate how much enrollment will fall if tuition is raised by $300 per semester and thus whether its revenue from tuition will rise or fall.
- The CEO of Proctor & Gamble must predict how much demand there will be in 10 years for the detergent Tide and how much to invest in new plant and equipment.
- A real estate developer must predict by how much population and income will increase to the south of Baton Rouge, Louisiana, over the next few years and whether it will be profitable to begin construction of a gambling casino and golf course.
- You must decide how much of your savings will go into a stock fund and how much into the money market. This requires you to make predictions of the level of economic activity, the rate of inflation, and interest rates over your planning horizon.
- A public transportation council in Melbourne, Australia, must decide how an increase in fares for public transportation (trams, trains, and buses) will affect the number of travelers who switch to car or bike and the effect of this switch on revenue going to public transportation.

To answer these questions of "how much," decision-makers rely on information provided by empirical economic research. In such research, an economist uses economic theory and reasoning to construct relationships between the variables in question. Data on these variables are collected and econometric methods are used to estimate the key underlying parameters and to make predictions. The decision-makers in the above examples obtain their "estimates" and "predictions" in different ways. The FRB has a large staff of economists to carry out econometric analyses. The CEO of Proctor & Gamble may hire econometric consultants to provide the firm with projections of sales. You may get advice about investing from a stock broker, who in turn is provided with econometric projections made by economists working for the parent company. Whatever the source of your information about "how much" questions, it is a good bet that there is an economist involved who is using econometric methods to analyze data that yield the answers.

In the next section, we show how to introduce parameters into an economic model and how to convert an economic model into an econometric model.

1.3 The Econometric Model

What is an econometric model, and where does it come from? We will give you a general overview, and we may use terms that are unfamiliar to you. Be assured that before you are too far into this book, all the terminology will be clearly defined. In an econometric model we must first realize that economic relations are not exact. Economic theory does not claim to be able to predict the specific behavior of any individual or firm, but rather describes the *average* or *systematic* behavior of *many* individuals or firms. When studying car sales we recognize that the *actual* number of Hondas sold is the sum of this systematic part and a random and unpredictable component e that we will call a **random error**. Thus, an **econometric model** representing the sales of Honda Accords is

$$Q^d = f(P, P^s, P^c, INC) + e$$

The random error e accounts for the many factors that affect sales that we have omitted from this simple model, and it also reflects the intrinsic uncertainty in economic activity.

To complete the specification of the econometric model, we must also say something about the form of the algebraic relationship among our economic variables. For example, in your first economics courses quantity demanded was depicted as a *linear* function of price. We extend that assumption to the other variables as well, making the systematic part of the demand relation

$$f(P, P^s, P^c, INC) = \beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 INC$$

The corresponding econometric model is

$$Q^d = \beta_1 + \beta_2 P + \beta_3 P^s + \beta_4 P^c + \beta_5 INC + e$$

The coefficients $\beta_1, \beta_2, \ldots, \beta_5$ are unknown **parameters** of the model that we estimate using economic data and an econometric technique. The functional form represents a hypothesis about the relationship between the variables. In any particular problem, one challenge is to determine a functional form that is compatible with economic theory and the data.

In every econometric model, whether it is a demand equation, a supply equation, or a production function, there is a systematic portion and an unobservable random component. The systematic portion is the part we obtain from economic theory, and includes an assumption about the functional form. The random component represents a "noise" component, which obscures our understanding of the relationship among variables, and which we represent using the random variable e.

We use the econometric model as a basis for **statistical inference**. Using the econometric model and a sample of data, we make inferences concerning the real world, learning something in the process. The ways in which statistical inference are carried out include the following:

- **Estimating** economic parameters, such as elasticities, using econometric methods

- **Predicting** economic outcomes, such as the enrollment in two-year colleges in the United States for the next 10 years
- **Testing** economic hypotheses, such as the question of whether newspaper advertising is better than store displays for increasing sales

Econometrics includes all of these aspects of statistical inference. As we proceed through this book, you will learn how to properly estimate, predict, and test, given the characteristics of the data at hand.

1.3.1 Causality and Prediction

A question that often arises when specifying an econometric model is whether a relationship can be viewed as both causal and predictive or only predictive. To appreciate the difference, consider an equation where a student's grade in Econometrics *GRADE* is related to the proportion of class lectures that are skipped *SKIP*.

$$GRADE = \beta_1 + \beta_2 SKIP + e$$

We would expect β_2 to be negative: the greater the proportion of lectures that are skipped, the lower the grade. But, can we say that skipping lectures **causes** grades to be lower? If lectures are captured by video, they could be viewed at another time. Perhaps a student is skipping lectures because he or she has a demanding job, and the demanding job does not leave enough time for study, and this is the underlying cause of a poor grade. Or, it might be that skipping lectures comes from a general lack of commitment or motivation, and this is the cause of a poor grade. Under these circumstances, what can we say about the equation that relates *GRADE* to *SKIP*? We can still call it a predictive equation. *GRADE* and *SKIP* are (negatively) correlated and so information about *SKIP* can be used to help predict *GRADE*. However, we cannot call it a causal relationship. Skipping lectures does not cause a low grade. The parameter β_2 does not convey the direct causal effect of skipping lectures on grade. It also includes the effect of other variables that are omitted from the equation and correlated with *SKIP*, such as hours spent studying or student motivation.

Economists are frequently interested in parameters that can be interpreted as causal. Honda would like to know the direct effect of a price change on the sales of their Accords. When there is technological improvement in the beef industry, the price elasticities of demand and supply have important implications for changes in consumer and producer welfare. One of our tasks will be to see what assumptions are necessary for an econometric model to be interpreted as causal and to assess whether those assumptions hold.

An area where predictive relationships are important is in the use of "big data." Advances in computer technology have led to storage of massive amounts of information. Travel sites on the Internet keep track of destinations you have been looking at. Google targets you with advertisements based on sites that you have been surfing. Through their loyalty cards, supermarkets keep data on your purchases and identify sale items relevant for you. Data analysts use big data to identify predictive relationships that help predict our behavior.

In general, the type of data we have impacts on the specification of an econometric model and the assumptions that we make about it. We turn now to a discussion of different types of data and where they can be found.

1.4 How Are Data Generated?

In order to carry out statistical inference we must have data. Where do data come from? What type of real processes generate data? Economists and other social scientists work in a complex world in which data on variables are "observed" and rarely obtained from a controlled experiment. This makes the task of learning about economic parameters all the more difficult. Procedures for using such data to answer questions of economic importance are the subject matter of this book.

1.4.1 Experimental Data

One way to acquire information about the unknown parameters of economic relationships is to conduct or observe the outcome of an experiment. In the physical sciences and agriculture, it is easy to imagine controlled experiments. Scientists specify the values of key control variables and then observe the outcome. We might plant similar plots of land with a particular variety of wheat, and then vary the amounts of fertilizer and pesticide applied to each plot, observing at the end of the growing season the bushels of wheat produced on each plot. Repeating the experiment on N plots of land creates a sample of N observations. Such controlled experiments are rare in business and the social sciences. A key aspect of experimental data is that the values of the explanatory variables can be fixed at specific values in repeated trials of the experiment.

One business example comes from marketing research. Suppose we are interested in the weekly sales of a particular item at a supermarket. As an item is sold it is passed over a scanning unit to record the price and the amount that will appear on your grocery bill. But at the same time, a data record is created, and at every point in time the price of the item and the prices of all its competitors are known, as well as current store displays and coupon usage. The prices and shopping environment are controlled by store management, so this "experiment" can be repeated a number of days or weeks using the same values of the "control" variables.

There are some examples of planned experiments in the social sciences, but they are rare because of the difficulties in organizing and funding them. A notable example of a planned experiment is Tennessee's Project Star.[1] This experiment followed a single cohort of elementary school children from kindergarten through the third grade, beginning in 1985 and ending in 1989. In the experiment children and teachers were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. The objective was to determine the effect of small classes on student learning, as measured by student scores on achievement tests. We will analyze the data in Chapter 7 and show that small classes significantly increase performance. This finding will influence public policy toward education for years to come.

1.4.2 Quasi-Experimental Data

It is useful to distinguish between "pure" experimental data and "quasi"-experimental data. A pure experiment is characterized by random assignment. In the example where varying amounts of fertilizer and pesticides are applied to plots of land for growing wheat, the different applications of fertilizer and pesticides are randomly assigned to different plots. In Tennessee's Project Star, students and teachers are randomly assigned to different sized classes with and without a teacher's aide. In general, if we have a control group and a treatment group, and we want to examine the effect of a policy intervention or treatment, pure experimental data are such that individuals are randomly assigned to the control and treatment groups.

Random assignment is not always possible however, particularly when dealing with human subjects. With quasi-experimental data, allocation to the control and treatment groups is not random but based on another criterion. An example is a study by Card and Krueger that is studied in more detail in Chapter 7. They examined the effect of an increase in New Jersey's minimum wage in 1992 on the number of people employed in fast-food restaurants. The treatment group was fast-food restaurants in New Jersey. The control group was fast-food restaurants in eastern Pennsylvania where there was no change in the minimum wage. Another example is the effect on spending habits of a change in the income tax rate for individuals above a threshold income. The treatment group is the group with incomes above the threshold. The control group is those with incomes below the threshold. When dealing with quasi-experimental data, one must be aware that the effect of the treatment may be confounded with the effect of the criterion for assignment.

[1] See https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/10766 for program description, public use data, and extensive literature.

1.4.3 Nonexperimental Data

An example of nonexperimental data is survey data. The Public Policy Research Lab at Louisiana State University (www.survey.lsu.edu) conducts telephone and mail surveys for clients. In a telephone survey, numbers are selected randomly and called. Responses to questions are recorded and analyzed. In such an environment, data on all variables are collected simultaneously, and the values are neither fixed nor repeatable. These are nonexperimental data.

Such surveys are carried out on a massive scale by national governments. For example, the Current Population Survey (CPS)[2] is a monthly survey of about 50,000 households conducted by the U.S. Bureau of the Census. The survey has been conducted for more than 50 years. The CPS website says "CPS data are used by government policymakers and legislators as important indicators of our nation's economic situation and for planning and evaluating many government programs. They are also used by the press, students, academics, and the general public." In Section 1.8 we describe some similar data sources.

1.5 Economic Data Types

Economic data comes in a variety of "flavors." In this section we describe and give an example of each. In each example, be aware of the different data characteristics, such as the following:

1. Data may be collected at various levels of aggregation:
   - *micro*—data collected on individual economic decision-making units such as individuals, households, and firms.
   - *macro*—data resulting from a pooling or aggregating over individuals, households, or firms at the local, state, or national levels.
2. Data may also represent a flow or a stock:
   - *flow*—outcome measures over a period of time, such as the consumption of gasoline during the last quarter of 2018.
   - *stock*—outcome measured at a particular point in time, such as the quantity of crude oil held by ExxonMobil in its U.S. storage tanks on November 1, 2018, or the asset value of the Wells Fargo Bank on July 1, 2018.
3. Data may be quantitative or qualitative:
   - *quantitative*—outcomes such as prices or income that may be expressed as numbers or some transformation of them, such as real prices or per capita income.
   - *qualitative*—outcomes that are of an "either-or" situation. For example, a consumer either did or did not make a purchase of a particular good, or a person either is or is not married.

1.5.1 Time-Series Data

A **time-series** is data collected over discrete intervals of time. Examples include the annual price of wheat in the United States and the daily price of General Electric stock shares. Macroeconomic data are usually reported in monthly, quarterly, or annual terms. Financial data, such as stock prices, can be recorded daily, or at even higher frequencies. The key feature of time-series data is that the same economic quantity is recorded at a regular time interval.

For example, the annual real gross domestic product (GDP) for the United States is depicted in Figure 1.1. A few values are given in Table 1.1. For each year, we have the recorded value. The data are annual, or yearly, and have been "deflated" by the Bureau of Economic Analysis to billions of real 2009 dollars.

[2] www.census.gov/cps/

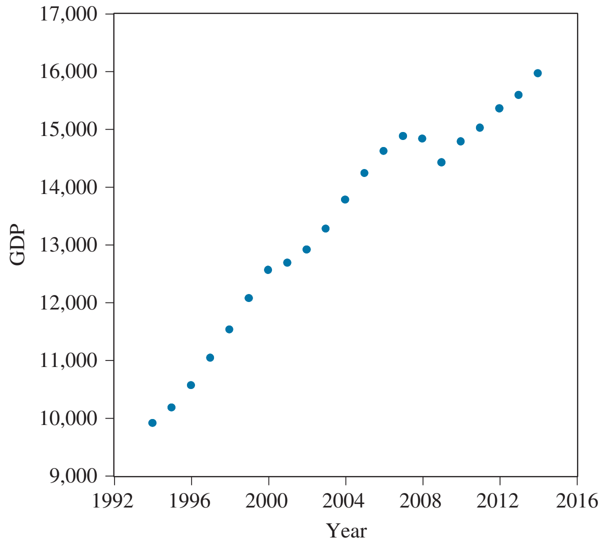

FIGURE 1.1 **Real U.S. GDP, 1994–2014.**[3]

TABLE 1.1 **U.S. Annual GDP (Billions of Real 2009 Dollars)**

Year	GDP
2006	14,613.8
2007	14,873.7
2008	14,830.4
2009	14,418.7
2010	14,783.8
2011	15,020.6
2012	15,354.6
2013	15,583.3
2014	15,961.7

1.5.2 Cross-Section Data

A cross-section of data is collected across sample units in a particular time period. Examples are income by counties in California during 2016 or high school graduation rates by state in 2015. The "sample units" are individual entities and may be firms, persons, households, states, or countries. For example, the CPS reports results of personal interviews on a monthly basis, covering items such as employment, unemployment, earnings, educational attainment, and income. In Table 1.2, we report a few observations from the March 2013 survey on the variables *RACE, EDUCATION, SEX*, and *WAGE* (hourly wage rate).[4] There are many detailed questions asked of the respondents.

[3] Source: www.bea.gov/national/index.htm

[4] In the actual raw data, the variable descriptions are coded differently to the names in Table 1.2. We have used shortened versions for convenience.

TABLE 1.2 Cross-Section Data: CPS, March 2013

Individual	Variables			
	RACE	*EDUCATION*	*SEX*	*WAGE*
1	White	Assoc Degree	Male	10.00
2	White	Master's Degree	Male	60.83
3	Black	Bachelor's Degree	Male	17.80
4	White	High School Graduate	Female	30.38
5	White	Master's Degree	Male	12.50
6	White	Master's Degree	Female	49.50
7	White	Master's Degree	Female	23.08
8	Black	Assoc Degree	Female	28.95
9	White	Some College, No Degree	Female	9.20

1.5.3 Panel or Longitudinal Data

A "panel" of data, also known as "longitudinal" data, has observations on individual micro-units that are followed over time. For example, the Panel Study of Income Dynamics (PSID)[5] describes itself as "a nationally representative longitudinal study of nearly 9000 U.S. families. Following the same families and individuals since 1969, the PSID collects data on economic, health, and social behavior." Other national panels exist, and many are described at "Resources for Economists," at www.rfe.org.

To illustrate, data from two rice farms[6] are given in Table 1.3. The data are annual observations on rice farms (or firms) over the period 1990–1997.

The key aspect of panel data is that we observe each micro-unit, here a farm, for a number of time periods. Here we have amount of rice produced, area planted, labor input, and fertilizer use. If we have the same number of time period observations for each micro-unit, which is the case here, we have a **balanced panel**. Usually the number of time-series observations is small relative to the number of micro-units, but not always. The Penn World Table[7] provides purchasing power parity and national income accounts converted to international prices for 182 countries for some or all of the years 1950–2014.

1.6 The Research Process

Econometrics is ultimately a research tool. Students of econometrics plan to do research or they plan to read and evaluate the research of others, or both. This section provides a frame of reference and guide for future work. In particular, we show you the role of econometrics in research.

Research is a process, and like many such activities, it flows according to an orderly pattern. Research is an adventure, and can be *fun!* Searching for an answer to your question, seeking new knowledge, is very addictive—for the more you seek, the more new questions you will find.

A research project is an opportunity to investigate a topic that is important to you. Choosing a good research topic is essential if you are to complete a project successfully. A starting point is the question "What are my interests?" Interest in a particular topic will add pleasure to the

[5] http://psidonline.isr.umich.edu

[6] These data were used by O'Donnell, C.J. and W.E. Griffiths (2006), Estimating State-Contingent Production Frontiers, *American Journal of Agricultural Economics*, 88(1), 249–266.

[7] www.rug.nl/ggdc/productivity/pwt