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/

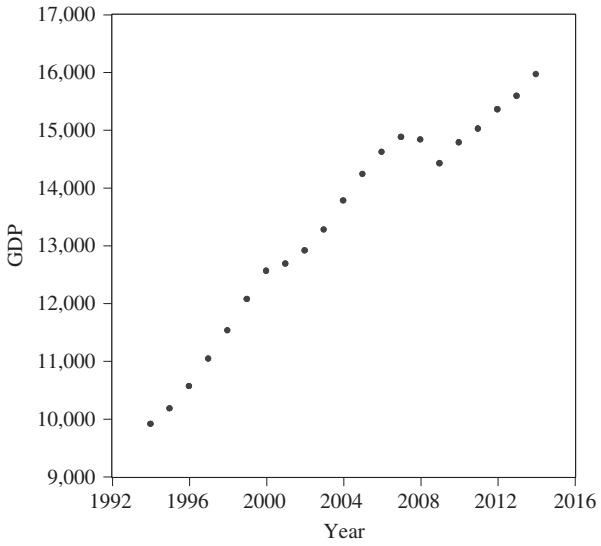

FIGURE 1.1 **Real U.S. GDP, 1994–2014.**[3]

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**

	Variables			
Individual	***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

TABLE 1.3 **Panel Data from Two Rice Farms**

FARM	*YEAR*	*PROD*	*AREA*	*LABOR*	*FERT*
1	1990	7.87	2.50	160	207.5
1	1991	7.18	2.50	138	295.5
1	1992	8.92	2.50	140	362.5
1	1993	7.31	2.50	127	338.0
1	1994	7.54	2.50	145	337.5
1	1995	4.51	2.50	123	207.2
1	1996	4.37	2.25	123	345.0
1	1997	7.27	2.15	87	222.8
2	1990	10.35	3.80	184	303.5
2	1991	10.21	3.80	151	206.0
2	1992	13.29	3.80	185	374.5
2	1993	18.58	3.80	262	421.0
2	1994	17.07	3.80	174	595.7
2	1995	16.61	4.25	244	234.8
2	1996	12.28	4.25	159	479.0
2	1997	14.20	3.75	133	170.0

research effort. Also, if you begin working on a topic, other questions will usually occur to you. These new questions may put another light on the original topic or may represent new paths to follow that are even more interesting to you. The idea may come after lengthy study of all that has been written on a particular topic. You will find that "inspiration is 99% perspiration." That means that after you dig at a topic long enough, a new and interesting question will occur to you. Alternatively, you may be led by your natural curiosity to an interesting question. Professor Hal Varian[8] suggests that you look for ideas outside academic journals—in newspapers, magazines, etc. He relates a story about a research project that developed from his shopping for a new TV set.

By the time you have completed several semesters of economics classes, you will find yourself enjoying some areas more than others. For each of us, specialized areas such as health economics, economic development, industrial organization, public finance, resource economics, monetary economics, environmental economics, and international trade hold a different appeal. If you find an area or topic in which you are interested, consult the *Journal of Economic Literature* (*JEL*) for a list of related journal articles. The *JEL* has a classification scheme that makes isolating particular areas of study an easy task. Alternatively, type a few descriptive words into your favorite search engine and see what pops up.

Once you have focused on a particular idea, begin the research process, which generally follows steps like these:

1. Economic theory gives us a way of thinking about the problem. Which economic variables are involved, and what is the possible direction of the relationship(s)? Every research project, given the initial question, begins by building an economic model and listing the questions (hypotheses) of interest. More questions will arise during the research project, but it is good to list those that motivate you at the project's beginning.
2. The working economic model leads to an econometric model. We must choose a functional form and make some assumptions about the nature of the error term.

[8] Varian, H. How to Build an Economic Model in Your Spare Time, *The American Economist*, 41(2), Fall 1997, pp. 3–10.

3. Sample data are obtained and a desirable method of statistical analysis chosen, based on initial assumptions and an understanding of how the data were collected.
4. Estimates of the unknown parameters are obtained with the help of a statistical software package, predictions are made, and hypothesis tests are performed.
5. Model diagnostics are performed to check the validity of assumptions. For example, were all of the right-hand side explanatory variables relevant? Was an adequate functional form used?
6. The economic consequences and the implications of the empirical results are analyzed and evaluated. What economic resource allocation and distribution results are implied, and what are their policy-choice implications? What remaining questions might be answered with further study or with new and better data?

These steps provide some direction for what must be done. However, research always includes some surprises that may send you back to an earlier point in your research plan or that may even cause you to revise it completely. Research requires a sense of urgency, which keeps the project moving forward, the patience not to rush beyond careful analysis, and the willingness to explore new ideas.

1.7 Writing an Empirical Research Paper

Research rewards you with new knowledge, but it is incomplete until a research paper or report is written. The process of writing forces the distillation of ideas. In no other way will your depth of understanding be so clearly revealed. When you have difficulty explaining a concept or thought, it may mean that your understanding is incomplete. Thus, writing is an integral part of research. We provide this section as a building block for future writing assignments. Consult it as needed. You will find other tips on writing economics papers on the book website, www.principlesofeconometrics.com.

1.7.1 Writing a Research Proposal

After you have selected a specific topic, it is a good idea to write up a brief project summary, or proposal. Writing it will help to focus your thoughts about what you really want to do. Show it to your colleagues or instructor for preliminary comments. The summary should be short, usually no longer than 500 words, and should include the following:

1. A concise statement of the problem
2. Comments on the information that is available, with one or two key references
3. A description of the research design that includes
   a. the economic model
   b. the econometric estimation and inference methods
   c. data sources
   d. estimation, hypothesis testing, and prediction procedures, including the econometric software and version used
4. The potential contribution of the research

1.7.2 A Format for Writing a Research Report

Economic research reports have a standard format in which the various steps of the research project are discussed and the results interpreted. The following outline is typical.

1. *Statement of the Problem* The place to start your report is with a summary of the questions you wish to investigate as well as why they are important and who should be interested in the results. This introductory section should be nontechnical and should motivate the reader to continue reading the paper. It is also useful to map out the contents of the following sections of the report. This is the first section to work on and also the last. In today's busy world, the reader's attention must be garnered very quickly. A clear, concise, well-written introduction is a must and is arguably the most important part of the paper.
2. *Review of the Literature* Briefly summarize the relevant literature in the research area you have chosen and clarify how your work extends our knowledge. By all means, cite the works of others who have motivated your research, but keep it brief. You do not have to survey everything that has been written on the topic.
3. *The Economic Model* Specify the economic model that you used and define the economic variables. State the model's assumptions and identify hypotheses that you wish to test. Economic models can get complicated. Your task is to explain the model clearly, but as briefly and simply as possible. Don't use unnecessary technical jargon. Use simple terms instead of complicated ones when possible. Your objective is to display the quality of your thinking, not the extent of your vocabulary.
4. *The Econometric Model* Discuss the econometric model that corresponds to the economic model. Make sure you include a discussion of the variables in the model, the functional form, the error assumptions, and any other assumptions that you make. Use notation that is as simple as possible, and do not clutter the body of the paper with long proofs or derivations; these can go into a technical appendix.
5. *The Data* Describe the data you used, as well as the source of the data and any reservations you have about their appropriateness.
6. *The Estimation and Inference Procedures* Describe the estimation methods you used and why they were chosen. Explain hypothesis testing procedures and their usage. Indicate the software used and the version, such as Stata 15 or EViews 10.
7. *The Empirical Results and Conclusions* Report the parameter estimates, their interpretation, and the values of test statistics. Comment on their statistical significance, their relation to previous estimates, and their economic implications.
8. *Possible Extensions and Limitations of the Study* Your research will raise questions about the economic model, data, and estimation techniques. What future research is suggested by your findings, and how might you go about performing it?
9. *Acknowledgments* It is appropriate to recognize those who have commented on and contributed to your research. This may include your instructor, a librarian who helped you find data, or a fellow student who read and commented on your paper.
10. *References* An alphabetical list of the literature you cite in your study, as well as references to the data sources you used.

Once you've written the first draft, use your computer's spell-check software to check for spelling errors. Have a friend read the paper, make suggestions for clarifying the prose, and check your logic and conclusions. Before you submit the paper, you should eliminate as many errors as possible. Your work should look good. Use a word processor, and be consistent with font sizes, section headings, style of footnotes, references, and so on. Often software developers provide templates for term papers and theses. A little searching for a good paper layout before beginning is a good idea. Typos, missing references, and incorrect formulas can spell doom for an otherwise excellent paper. Some do's and don'ts are summarized nicely, and with good humor, by Deidre N. McClosky in *Economical Writing*, 2nd edition (Prospect Heights, IL: Waveland Press, Inc., 1999).

While it is not a pleasant topic to discuss, you should be aware of the rules of **plagiarism**. You must not use someone else's words as if they were your own. If you are unclear about what you can and cannot use, check with the style manuals listed in the next paragraph, or consult

your instructor. Your university may provide a plagiarism-checking software, such as Turnitin or iThenticate, that will compare your paper to millions of online sources and look for problem areas. There are some free online versions as well. The paper should have clearly defined sections and subsections. The pages, equations, tables, and figures should be numbered. References and footnotes should be formatted in an acceptable fashion. A style guide is a good investment. Two classics are the following:

- *The Chicago Manual of Style*, 16th edition, is available online and in other formats.
- *A Manual for Writers of Research Papers, Theses, and Dissertations: Chicago Style for Students and Researchers*, 8th edition, by Kate L. Turabian; revised by Wayne C. Booth, Gregory G. Colomb, and Joseph M Williams (2013, University of Chicago Press).

1.8 Sources of Economic Data

Economic data are much easier to obtain since the development of the World Wide Web. In this section we direct you to some places on the Internet where economic data are accessible. During your study of econometrics, browse some of the sources listed to gain some familiarity with data availability.

1.8.1 Links to Economic Data on the Internet

There are a number of fantastic sites on the World Wide Web for obtaining economic data.

Resources for Economists (RFE) www.rfe.org is a primary gateway to resources on the Internet for economists. This excellent site is the work of Bill Goffe. Here you will find links to sites for economic data and sites of general interest to economists. The **Data** link has these broad data categories:

- *U.S. Macro and Regional Data* Here you will find links to various data sources such as the Bureau of Economic Analysis, Bureau of Labor Statistics, *Economic Reports of the President*, and the Federal Reserve Banks.
- *Other U.S. Data* Here you will find links to the U.S. Census Bureau, as well as links to many panel and survey data sources. The gateway to U.S. government agencies is FedStats (fedstats.sites.usa.gov). Once there, click on *Agencies* to see a complete list of U.S. government agencies and links to their homepages.
- *World and Non-U.S. Data* Here there are links to world data, such as at the CIA World Factbook and the Penn World Tables, as well as international organizations such as the Asian Development Bank, the International Monetary Fund, the World Bank, and so on. There are also links to sites with data on specific countries and sectors of the world.
- *Finance and Financial Markets* Here are links to sources of U.S. and world financial data on variables such as exchange rates, interest rates, and share prices.
- *Journal Data and Program Archives* Some economic journals post data used in articles. Links to these journals are provided here. (Many of the articles in these journals will be beyond the scope of undergraduate economics majors.)

National Bureau of Economic Research (NBER) www.nber.org/data provides access to a great amount of data. There are headings for

- Macro Data
- Industry Productivity and Digitalization Data
- International Trade Data

- Individual Data
- Healthcare Data—Hospitals, Providers, Drugs, and Devices
- Demographic and Vital Statistics
- Patent and Scientific Papers Data
- Other Data

Economagic Some websites make extracting data relatively easy. For example, Economagic (www.economagic.com) is an excellent and easy-to-use source of macro time series (some 100,000 series available). The data series are easily viewed in a copy and paste format, or graphed.

1.8.2 Interpreting Economic Data

In many cases it is easier to obtain economic data than it is to understand the meaning of the data. It is essential when using macroeconomic or financial data that you understand the definitions of the variables. Just what is the index of leading economic indicators? What is included in personal consumption expenditures? You may find the answers to some questions like these in your textbooks. Another resource you might find useful is *A Guide to Everyday Economic Statistics*, 7th edition, by Gary E. Clayton and Martin Gerhard Giesbrecht, (Boston: Irwin/McGraw-Hill, 2009). This slender volume examines how economic statistics are constructed, and how they can be used.

1.8.3 Obtaining the Data

Finding a data source is not the same as obtaining the data. Although there are a great many easy-to-use websites, "easy-to-use" is a relative term. The data will come packaged in a variety of formats. It is also true that there are many, many variables at each of these websites. A primary challenge is identifying the specific variables that you want, and what exactly they measure. The following examples are illustrative.

The Federal Reserve Bank of St. Louis [9] has a system called **FRED** (Federal Reserve Economic Data). Under "Categories" there are links to financial variables, population and labor variables, national accounts, and many others. Data on these variables can be downloaded in a number of formats. For reading the data, you may need specific knowledge of your statistical software. Accompanying *Principles of Econometrics, 5e*, are computer manuals for Excel, EViews, Stata, SAS, R, and Gretl to aid this process. See the publisher website www.wiley.com/college/hill, or the book website at www.principlesofeconometrics.com for a description of these aids.

The CPS (www.census.gov/cps) has a tool called **DataFerrett**. This tool will help you find and download data series that are of particular interest to you. There are tutorials that guide you through the process. Variable descriptions, as well as the specific survey questions, are provided to aid in your selection. It is somewhat like an Internet shopping site. Desired series are "ticked" and added to a "Shopping Basket." Once you have filled your basket, you download the data to use with specific software. Other Web-based data sources operate in this same manner. One example is the PSID.[10] The Penn World Tables[11] offer data downloads in both Excel and Stata formats.

You can expect to find massive amounts of readily available data at the various sites we have mentioned, but there is a learning curve. You should not expect to find, download, and process the data without considerable work effort. Being skilled with Excel and statistical software is a must if you plan to regularly use these data sources.

[9] https://fred.stlouisfed.org

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

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