Using Econometrics

A PRACTICAL GUIDE

THIRD EDITION

A. H. STUDENMUND

Using Econometrics: A Practical Guide

Preface

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

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

The material covered by this book is traditional, but there are five specific features that we feel distinguish *Using Econometrics:*

1. Our approach to the learning of econometrics is simple, intuitive, and easy to understand. We do not use matrix algebra, and we relegate proofs and calculus to the footnotes.
2. We include numerous examples and example-based exercises. We feel that the best way to get a solid grasp of applied econometrics is through an example-oriented approach.
3. Although most of this book is at a simpler level than other econometrics texts, Chapters 6 and 7 on specification choice are among the most complete in the field. We think that an understanding of specification issues is vital for regression users.
4. We use a new kind of learning tool, called an interactive regression learning exercise, to help students simulate econometric analysis by giving them feedback on various kinds of decisions without relying on computer time or much instructor supervision.
5. We include ECSTAT, a basic econometric software package, with each copy of the text at virtually no additional cost. ECSTAT was used to produce the regression results in the book, and students tend to appreciate the program's accuracy and simplicity. *Using*

Econometrics is not tied to ECSTAT in any way, however, so the text fits in well with all standard regression programs.

The formal prerequisites for using this book are few. Readers are assumed to have been exposed to some microeconomic and macroeconomic theory, basic mathematical functions, and elementary statistics (even if they have forgotten most of it). All the statistical concepts necessary for econometric study are covered in the text, but they are covered only to the extent needed for an understanding of regression analysis.

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

What's New in the Third Edition?

We believe that *Using Econometrics* is the best-selling text in the field. As a result, we've tried hard to retain the clarity and practicality that made the book a success while adding new material, more examples, more exercises, and more data sets. We surveyed all the professors who adopted the text, and their suggestions were instrumental in deciding what material to add (and, in some cases, drop).

Specific new topics include:

1. heteroskedasticity-corrected standard errors,
2. nonstationarity, the Dickey-Fuller test, and cointegration,
3. advice to students on running their own regressions,
4. the meaning of and use of logged variables, and
5. a number of detailed data-intensive literature-based exercises drawn from *Applied Econometrics: Problems with Data Sets*, by William Lott and Subhash Ray.

In addition, we've improved ECSTAT by adding to the disk all the text's data sets in four different formats (ECSTAT, ASCII, Lotus 1-2-3, and MICROTSP). This means that the ECSTAT disk can be used as a data source disk even for classes that utilize econometric software other than ECSTAT.

Finally, we've expanded the instructor's manual, which includes answers to odd-numbered questions, lecture notes, sample examinations, and an additional interactive exercise.

Acknowledgements

If this book has a spiritual father, it's Henry Cassidy of the Federal Home Loan Mortgage Corporation. It was Henry who saw the need for a follow-on to P. Rao and R. L. Miller's legendary *Applied Econometrics* and who coauthored the first edition of *Using Econometrics* as an expansion of his own work of the same name. I will always be in debt to Henry for his expertise, inspiration, and sense of humor.

For the third straight edition, Carolyn Summers of the National Education Association has been a superb econometric adviser, editorial consultant, proofreader, and indexer. Carolyn's influence on this book cannot be overestimated; she truly has been the Rock of Gibralter of the project. In fact, it's fair to say that without Carolyn, *Using Econometrics* never would have become the success that it is today.

As mentioned, we have added a number of data-intensive exercises that originally were published by William F. Lott and Subhash C. Ray in their excellent text, *Applied Econometrics: Problems with Data Sets* (1992). We appreciate the permission of both the authors and Harcourt Brace to reproduce their data sets and, in some cases, questions.

The quality of outside reviewers was once again quite high. In particular, I owe a real debt of gratitude to Doug Steigerwald of the University of California at Santa Barbara for taking the time to give me in-depth feedback on the new material. Others in this excellent group of reviewers were Ragui Assaad (University of Minnesota/Humphrey Institute), Rob Engle (University of California at San Diego), M. A. Grove (University of Oregon), William Nebesky (Grinnell College), Wim Vijverberg (University of Texas/Dallas) and Phanindra Wunnava (Middlebury College).

Particularly helpful in the editorial and production process were John Greenman, Amy Muller, Sheryl Nelson, Diane Williams, and Robert Dohner (the author of ECSTAT). Others who provided timely and useful assistance to this edition were Humberto Barreto (Wabash College), Larry Iannaccone (University of Santa Clara), Amber McClain, Justin Meyer (Silver Oak), Barbara Passarelle, and, of course, my family. Finally, but perhaps most importantly, I'd like to thank my superb Occidental College colleagues and students for their feedback and encouragement. These especially include Jeff Carchidi, Aaron Cheris, Cindy Deng, Mary Hirschfeld, Kevin Holliday, Phil Jerome, Andy Lewis, Robby Moore, John Pike, Stephanie Ream, Mike Tamada, and Betty Tracy.

A. H. Studenmund

Contents

PART

I

The Basic Regression Model

1

An Overview of Regression Analysis

1.1 What Is Econometrics?

"Econometrics is too mathematical; it's the reason my best friend isn't majoring in economics."

"There are two things you don't want to see in the making—sausage and econometric research."[1]

"Econometrics may be defined as the quantitative analysis of actual economic phenomena."[2]

"It's my experience that 'economy-tricks' is usually nothing more than a justification of what the author believed before the research was begun."

1. Attributed to Edward E. Leamer.

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

Obviously, econometrics means different things to different people. To beginning students, it may seem as if econometrics is an overly complex obstacle to an otherwise useful education. To skeptical observers, econometric results should be trusted only when the steps that produced those results are completely known. To professionals in the field, econometrics is a fascinating set of techniques that allows the measurement and analysis of economic phenomena and the prediction of future economic trends.

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

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

1.1.1 Uses of Econometrics

Econometrics has three major uses:

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

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

$$C = f(P, P_s, Y_d) \quad (1.1)$$

can become explicit:

$$C = -74.4 - 0.39P + 0.16P_s + 13.7Y_d \quad (1.2)$$

This technique gives a much more specific and descriptive picture of the function.[3] Let's compare Equations 1.1 and 1.2. Instead of expecting consumption merely to "increase" if there is an increase in disposable income, Equation 1.2 allows us to expect an increase of a specific amount (13.7 units for each unit of increased disposable income). The number 13.7 is called an estimated regression coefficient, and it is the ability to estimate these coefficients that makes econometrics valuable.

The second and perhaps the most common use of econometrics is **hypothesis testing,** the evaluation of alternative theories with quantitative evidence. Much of economics involves building theoretical models and testing them against evidence, and hypothesis testing is vital to that scientific approach. For example, you could test the hypothesis that the product in Equation 1.1 is what economists call a normal good (one for which the quantity demanded increases when disposable income increases). You could do this by applying various statistical tests to the estimated coefficient (13.7) of disposable income (Y_d) in Equation 1.2. At first glance, the evidence would seem to support this hypothesis because the coefficient's sign is positive, but the "statistical significance" of that estimate would have to be investigated before such a conclusion could be justified. Even though the estimated coefficient is positive, as expected, it may not be sufficiently different from zero to imply that the coefficient is indeed positive instead of zero. Unfortunately, statistical tests of such hypotheses are

3. The results in Equation 1.2 are from a model of the demand for chicken that we will examine in more detail in Section 6.1. Note that the definition of Y_d is the log of disposable income (for reasons we'll discuss in Chapter 7).

not always easy, and there are times when two researchers can look at the same set of data and come to different conclusions. Even given this possibility, the use of econometrics in testing hypotheses is probably its most important function.

The third and most difficult use of econometrics is to **forecast** or predict what is likely to happen next quarter, next year, or further into the future, based on what has happened in the past. For example, economists use econometric models to make forecasts of variables like sales, profits, Gross Domestic Product (GDP), and the inflation rate. The accuracy of such forecasts depends in large measure on the degree to which the past is a good guide to the future. Business leaders and politicians tend to be especially interested in this use of econometrics because they need to make decisions about the future, and the penalty for being wrong (bankruptcy for the entrepreneur and political defeat for the candidate) is high. To the extent that econometrics can shed light on the impact of their policies, business and government leaders will be better equipped to make decisions. For example, if the president of a company that sold the product modeled in Equation 1.1 wanted to decide whether to increase prices, forecasts of sales with and without the price increase could be calculated and compared to help make such a decision. In this way, econometrics can be used not only for forecasting but also for policy analysis.

1.1.2 Alternative Econometric Approaches

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

Different approaches also make sense within the field of economics. The kind of econometric tools used to quantify a particular function depends in part on the uses to which that equation will be put. A model

4. Clive Granger, "A Review of Some Recent Textbooks of Econometrics," *Journal of Economic Literature,* March 1994, p. 117.

built solely for descriptive purposes might be different from a forecasting model, for example.

To get a better picture of these approaches, let's look at the steps necessary for any kind of quantitative research:

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

Steps 1 and 2 are similar in all quantitative work, but the techniques used in step 3, quantifying the models, differ widely between and within disciplines. Choosing among techniques for quantifying a model given a particular set of data is often referred to as the "art" of econometrics. There are many alternative approaches to quantifying the same equation, and each approach may give somewhat different results. The choice of approach is left to the individual econometrician (the researcher using econometrics), but each researcher should be able to justify that choice.

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

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

1.2 What Is Regression Analysis?

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

1.2.1 Dependent Variables, Independent Variables, and Causality

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

$$C = f(P, P_s, Y_d) \tag{1.1}$$

C is the dependent variable and P, P_s, and Y_d are the independent variables. Regression analysis is a natural tool for economists because most (though not all) economic propositions can be stated in such single-equation functional forms. For example, the quantity demanded (dependent variable) is a function of price, the prices of substitutes, and income (independent variables).

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

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

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

1.2.2 Single-Equation Linear Models

The simplest single-equation linear regression model is:

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

Equation 1.3 states that Y, the dependent variable, is a single-equation linear function of X, the independent variable. The model is a single-equation model because no equation for X as a function of Y (or any other variable) has been specified. The model is linear because if you were to plot Equation 1.3 on graph paper, it would be a straight line rather than a curve.

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

The slope coefficient, β_1, shows the response of Y to a change in X. Since being able to explain and predict changes in the dependent variable is the essential reason for quantifying behavioral relationships, much of the emphasis in regression analysis is on slope coefficients such as β_1. In Figure 1.1 for example, if X were to increase from X_1 to X_2 (ΔX), the value of Y in Equation 1.3 would increase from Y_1 to Y_2 (ΔY). For linear (i.e.,