

---

1	Basic Objectives	xxi
2	Innovations	xxii
3	Math	xxv
4	Statistics	xxvi
5	Conclusion	xxvii

---

### 1 Basic Objectives

The purpose of this book is to teach sound fundamentals to students, at any level, who are being introduced to econometrics for the first time and whose preparation or enthusiasm may be limited. This book is a teaching tool rather than a reference work. The way that a typical student needs to approach this material initially is different from the way that a successful student will return to it. This book engages novices by starting where they are comfortable beginning. It then stimulates their interests, reinforces their technical abilities, and trains their instincts.

For these purposes, the first objective of this book is the same as that of a medical doctor: Do no harm. Armed with only some vocabulary from a conventional textbook and a standard computer software package, anyone can run a regression, regardless of his or her preparation. The consequences of econometric work that lacks precision, care, and understanding can range from inept student research to bad public policy and disrepute for the econometric enterprise. These outcomes are best prevented at the source, by ensuring that potential perpetrators are exposed to the requirements, as well as the power, of econometric techniques.

Many of the students who take this course will actually run regressions in their subsequent professional life. Few will take another econometrics course. For most, this is a one-shot opportunity. It is crucial that we—students and teacher—make the best of it. At minimum, this book is designed so that those who finish it are aware that econometric tools shouldn't be used thoughtlessly.

This book's additional objectives form a hierarchy of increasing optimism. Beyond basic awareness of the power of econometrics, we want students to understand and use responsibly the techniques that this book has taught them. Better still, we want students to recognize when they have an econometric challenge that goes beyond these techniques. Best of all, we want students to have enough preparation, respect, and perhaps even affection for econometrics that they'll continue to explore its power, formally or otherwise.

---

## 2 Innovations

In pursuit of these objectives, this text makes five distinctive choices:

1. It emphasizes depth rather than breadth.
2. It is curiosity driven.
3. It discusses violations of the standard assumptions immediately after the presentation of the two-variable model, where they can be handled insightfully with ordinary algebra.
4. The tone is conversational whenever possible, in the hope of encouraging engagement with the formalities. It is precise whenever necessary, in order to eliminate confusion.
5. The text is designed to evolve as students learn.

This book engages only those subjects that it is prepared to address and that students can be expected to learn *in depth*. The formal presentations are preceded, complemented, and illuminated by extensive discussions of intuition. The formal presentations themselves are usually complete, so that there is no uncertainty regarding the required steps. End-of-chapter exercises provide students with opportunities to experiment on their own—both with derivations and their interpretations. Consequently, this book devotes more pages to the topics that it does cover than do other books. The hope is that, in compensation, lectures will be more productive, office hours less congested, and examinations more satisfying.

Equivalently, this book does not “expose” students to advanced topics, in order to avoid providing them with impressions that would be unavoidably superficial and untrustworthy. The text includes only one chapter that goes beyond the clas-

sical regression model and its conventional elaborations. This chapter introduces limited dependent variables, a topic that students who go on to practice econometrics without further training are especially likely to encounter.

A typical course should complete this entire text, with the possible exception of chapter 15, in a single semester. Students who do so will then be ready, should they choose, for one of the many texts that offer more breadth and sophistication.

Motivation is the second innovation in this book. The sequence of topics is driven by curiosity regarding the results, rather than by deduction from first principles. The empirical theme throughout is the relationship between education and earnings, which ought to be of at least some interest to most students.

The formal presentation begins in chapter 3, which confronts students with data regarding these two variables and simply asks, “What can be made of them?” Initial answers to this question naturally raise further questions: The covariance indicates the direction of the association, but not its strength. The correlation indicates its strength, but not its magnitude. The desire to know each of these successive attributes leads, inexorably, to line fitting in chapter 4. Finally, the desire to generalize beyond the observed sample invokes the concept of the population.

This contrasts with the typical presentation, which begins with the population regression. That approach is philosophically appealing. Pedagogically, it doesn’t work. First-time students are not inclined to philosophic rigor. Worse, many first-time students never really get the concept of the population. It’s too abstract. If the course starts with it, students begin with a confusion that many never successfully dispel.

For this reason, this book refers only obliquely to the contemporary approach to regression analysis in terms of conditional expectations. This approach is deeply insightful to those with a thorough understanding of statistical fundamentals. The many undergraduates who struggle with the summations of chapter 2 do not truly understand expectations, much less those that are conditional. We cannot expect them to appreciate any approach that is based first on population relationships.

The sample is a much more accessible concept. It’s concrete. In several places in this book, it actually appears on the page. The question of how to deal with the sample provokes curiosity regarding its relationship with the underlying population. This gives the student both foundation and motivation to confront the population concept in chapter 5.

Subsequent chapters repeat this pattern. Each begins by raising questions about the conclusion of a previous chapter. In each, the answers raise the level of econometric sophistication. The text aspires to engage students to the point that they are actually eager to pursue this development.

The third innovation places the discussions of inference, heteroscedasticity, correlated disturbances, and endogeneity in chapters 6 through 10, immediately after the presentation of the model with one explanatory variable. In the case of inference, students may have seen the basic results in the univariate context before. If they did, they probably didn't understand them very well. Bivariate regression provides the most convenient and accessible context for review. Moreover, it is all that is necessary for most of the results.

Heteroscedasticity, correlated disturbances, and endogeneity all concern violations regarding the ordinary assumptions about the disturbance terms. The methods of diagnosis and treatment do not vary in any important way with the number of explanatory variables. The conventional formulas are often especially intuitive if only one explanatory variable is present.

This arrangement relieves the presentation of the multivariate model of the burden of conveying these ancillary issues. Accordingly, chapter 11 concentrates on what is truly novel in the multivariate model: the effects of omitted variables, the implications of including an irrelevant variable, the consequences of correlations among the explanatory variables, and statistical tests of joint significance.

In addition, this arrangement ensures that shorter courses, such as those that take place within a quarter rather than a semester, or courses in which progress is slower can still address these alternative assumptions. If the students in such courses must omit some basic material, they are better served by exposure to the problems to which regression may be subject than to an expansion of the basic results regarding the estimation of slopes.

This innovation, like the second, is a departure from the ordinary order of presentation. These discussions usually appear after the multivariate model. However, the presence of additional explanatory variables adds nothing, apart from additional notation. With an audience that is wary of notation to begin with, this is counterproductive.

As its fourth innovation, this book adopts a conversational tone wherever possible. Appropriate respect for the power of statistical analysis requires some familiarity with formal derivations. However, formal discussions reinforce the prejudice that this material is incompatible with the typical student sensibility. This prejudice defeats the purpose of the derivations.

Their real point is that they are almost always intuitive, usually insightful, and occasionally surprising. The real objective in their presentation is to develop the intuitive faculty, in the same way that repeated weight-lifting develops the muscles. The book pursues this objective by placing even greater emphasis on the revelations in the formulas than on their derivations.

At the same time, the book is meticulous about formal terminology. The names of well-defined concepts are consistent throughout. In particular, the text is rigorous in distinguishing between population and sample concepts.

This rigor is distinctive. The word “mean” provides an egregious example of ambiguity in common usage. As a noun, this term appears as a synonym for both the average in the sample and the expected value in the population. With an audience who can never confidently distinguish between the sample and the population in the first place, this ambiguity can be lethal. Here, “mean” appears only as a verb.

This ambiguity is also rampant in discussions of variances, standard deviations, covariances, and correlations. The same nouns are typically employed, regardless of whether samples or populations are at issue. This text distinguishes carefully between the two. Chapter 3 qualifies all variances, standard deviations, covariances, and correlations as sample statistics. The common Greek symbols for variances, standard deviations, and correlations appear only after chapter 5 explains the distinction between sample statistics and population parameters.

Finally, this book is designed to evolve with the students’ understanding. Each chapter begins with the section “What We Need to Know When We Finish This Chapter,” which sets attainable and appropriate goals as the material is first introduced. The text of each chapter elaborates on these goals so as to make them accessible to all. Once mastery is attained, the “What We Need to Know . . .” sections, the numbered equations, and the tables and figures serve as a concise but complete summary of the material and a convenient source for review. A companion Web site—[www.sup.org/econometrics](http://www.sup.org/econometrics)—provides these review materials independent of the book. It also provides instructors with password-protected solutions to the end-of-chapter exercises. Instructors can release these to students as appropriate so that they can explore the material further on their own.

---

### **3 Math**

Many students would probably prefer a treatment of econometrics that is entirely math-free. This book acknowledges the mathematical hesitations and limitations of average students without indulging them. It carefully avoids any mathematical development that is not essential.

However, derivation can’t be disregarded, no matter how ill-prepared the student. Proof is how we know what it is that we know. Students have to have some comfort with the purpose and process of formal derivation in order to be well educated in introductory econometrics. Some respect for the formal properties of regression is also a prerequisite for responsible use.

This book accommodates the skills of the typical student by developing all basic results regarding the classical linear regression model and the

elaborations associated with heteroscedasticity, correlated disturbances, and endogeneity in the language of ordinary algebra. Virtually all of the mathematical operations consist entirely of addition, subtraction, multiplication, and division. There is no reference to linear algebra. The only sophistication that the book requires is a facility with summation signs. The second chapter provides a thorough review.

In general, individual steps in the algebraic derivations are simple. The text explains all, or nearly all, of these steps. This level of detail replicates what, in my experience, instructors end up saying in class when students ask for explanations of what's in the book. With these explanations, the derivations, as a whole, should be manageable. They also provide opportunities to emphasize emerging insights. Such insights are especially striking in chapter 11, where ordinary algebra informs intuitions that would be obscured in the matrix presentation.

Chapters 11 through 14 present the three-variable model entirely with ordinary algebra. By the time students reach this point, they are ready to handle the complexity. This treatment has two advantages. First, students don't need matrices as a prerequisite and faculty don't have to find time to teach, or at least review, them. Second, the algebraic formulas often reveal intuitions that are obscure or unavailable in matrix formulations.

Similarly, this book contains very few references to calculus. Only five derivatives are inescapable. Two occur when minimizing the sum of squared errors in the two-variable model of chapter 4. Three appear when executing the same task in the three-variable model of chapter 11. All five are presented in their entirety, so that the students are not responsible for their derivation. All other derivatives are optional, in the appendices to chapters 4 and 7, in chapter 15, and in several exercises.

In sum, this book challenges students where they are most fearful, by presenting all essential derivations, proofs, and results in the language with which they are most familiar. Students who have used this book uniformly acknowledge that they have an improved appreciation for proof. This will be to their lasting benefit, well beyond any regressions they might later run.

---

## 4 Statistics

This text assumes either that students have not had a prior course in introductory statistics or that they don't remember very much of it. It does not derive the essential concepts, but reviews all of them.

In contrast to other texts, this review does not appear as a discrete chapter. This conventional treatment suggests a distinction between statistics and

whatever else is going on in econometrics. It reinforces the common student suspicion, or even hope, that statistics can be ignored at some affordable cost.

Instead, the statistics review here is consistent with the curiosity-driven theme. The text presents statistical results at the moment when they are needed. As examples, it derives the sample covariance and sample correlation from first principles in chapter 3, in response to the question of how to make any sense at all out of a list of values for earnings and education. The text defines expectations and population variances in section 5.3, as it introduces the disturbances.

Results regarding the expectation of summations first appear in section 5.4, where they are necessary to derive the expected value of the dependent variable. Variances of summations are not required until the derivation of the variance of the ordinary least squares slope estimator. Accordingly, the relevant formulas do not appear until section 5.8.

Similarly, a review of confidence intervals and hypothesis tests immediately precedes the formal discussion of inference regarding regression estimates. This review is extensive both because inference is essential to the responsible interpretation of regression results and because students are rarely comfortable with this material when they begin an econometrics course. Consequently, it appears in the self-contained chapter 6. The application to the bivariate regression context appears in chapter 7.

---

## 5 Conclusion

This book tries to strike a new balance between the capacities and enthusiasms of instructors and students. The former typically have a lot more of both than the latter. This book aspires to locate a common understanding between the two. It attempts to distill the instructor's knowledge into a form that preserves its essence but is acceptable and even appealing to the student's intellectual palate. This book insists on rigor where it is essential. It emphasizes intuition wherever it is available. It seizes upon entertainment.

This book is motivated by three beliefs. First, that students are, perhaps despite themselves, interested in questions that only econometrics can answer. Second, through these answers they can come to understand, appreciate, and even enjoy the enterprise of econometrics. Third, this text, with its innovations in presentation and practice, can provoke this interest and encourage responsible and insightful application. With that, let's go!