

清华经济学系列英文版教材

# 计量经济学导论

Introductory Econometrics

现代方法

A Modern Approach

Sixth Edition

[美] 杰弗里·M. 伍德里奇 (Jeffrey M. Wooldridge) 著

第 6 版

清华大学经济学系列英文版教材

# 计量经济学导论

Introductory Econometrics

现代方法

A Modern Approach

Sixth Edition

[美] 杰弗里·M. 伍德里奇 (Jeffrey M. Wooldridge) 著

第6版

清华大学出版社  
北京

北京市版权局著作权合同登记号 图字 01-2017-0768 号

Jeffrey M. Wooldridge

Introductory Econometrics, 6e

EISBN: 9781305270107

Copyright © 2016 Cengage Learning Asia Pte Ltd.

Original edition published by Cengage Learning. All rights reserved. 本书原版由圣智学习出版公司出版。版权所有，盗印必究。

Tsinghua University Press is authorized by Cengage Learning to publish and distribute exclusively this custom reprint edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

此客户定制影印版由圣智学习出版公司授权清华大学出版社独家出版发行。此版本仅限在中华人民共和国境内（不包括中国香港、澳门特别行政区及中国台湾）销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可，不得以任何方式复制或发行本书的任何部分。

Cengage Learning Asia Pte. Ltd.

5 Shenton Way, # 01-01 UIC Building, Singapore 068808

本书封面贴有 Cengage Learning 防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

#### 图书在版编目 (CIP) 数据

计量经济学导论：现代方法：第6版 = Introductory Econometrics: A Modern Approach: 英文 / (美) 杰弗里·M. 伍德里奇著. —北京：清华大学出版社, 2017 (2018.8重印)

(清华经济学系列英文版教材)

ISBN 978-7-302-46852-3

I. ①计… II. ①杰… III. ①计量经济学—高等学校—教材—英文 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2017)第 052919 号

责任编辑：王 青

封面设计：常雪影

责任印制：宋 林

出版发行：清华大学出版社

网 址：http://www.tup.com.cn, http://www.wqbook.com

地 址：北京清华大学学研大厦 A 座

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者：清华大学印刷厂

装 订 者：三河市铭诚印务有限公司

经 销：全国新华书店

开 本：203mm×260mm

印 张：37.25

版 次：2017 年 4 月第 1 版

印 次：2018 年 8 月第 2 次印刷

定 价：76.00 元

产品编号：072622-02

世纪之交，中国与世界的发展呈现最显著的两大趋势——以网络为代表的信息技术的突飞猛进，以及经济全球化的激烈挑战。无论是无远弗届的互联网，还是日益密切的政治、经济、文化等方面的国际合作，都标示着 21 世纪的中国是一个更加开放的中国，也面临着一个更加开放的世界。

教育，特别是管理教育总是扮演着学习与合作的先行者的角色。改革开放以来，尤其是 20 世纪 90 年代之后，为了探寻中国国情与国际上一切优秀的管理教育思想、方法和手段的完美结合，为了更好地培养高层次的“面向国际市场竞争、具备国际经营头脑”的管理者，我国的教育机构与美国、欧洲、澳洲以及亚洲一些国家和地区的大量的著名管理学院和顶尖跨国企业建立了长期密切的合作关系。以清华大学经济管理学院为例，2000 年，学院顾问委员会成立，并于 10 月举行了第一次会议，2001 年 4 月又举行了第二次会议。这个顾问委员会包括了世界上最大的一些跨国公司和中国几家顶尖企业的最高领导人，其阵容之大、层次之高，超过了世界上任何一所商学院。在这样高层次、多样化、重实效的管理教育国际合作中，教师和学生与国外的交流机会大幅度增加，越来越深刻地融入到全球性的教育、文化和思想观念的时代变革中，我们的管理教育工作者和经济管理学习者，更加真切地体验到这个世界正发生着深刻的变化，也更主动地探寻和把握着世界经济发展和跨国企业运作的脉搏。

我国管理教育的发展，闭关锁国、闭门造车是绝对不行的，必须同国际接轨，按照国际一流的水准来要求自己。正如朱镕基同志在清华大学经济管理学院成立十周年时所发的贺信中指出的那样：“建设有中国特色的社会主义，需要一大批掌握市场经济的一般规律，熟悉其运行规则，而又了解中国企业实情的经济管理人才。清华大学经济管理学院就要敢于借鉴、引进世界上一切优秀的经济管理学院的教学内容、方法和手段，结合中国的国情，办成世界第一流的经管学院。”作为达到世界一流的一个重要基础，朱镕基同志多次建议清华的 MBA 教育要加强英语教学。我体会，这不仅因为英语是当今世界交往中重要的语言工具，是连接中国与世界的重要桥梁和媒介，而且更是中国经济管理人才参与国际竞争，加强国际合作，实现中国企业的国际战略的基石。推动和实行英文教学并不是目的，真正的目的在于培养学生——这些未来的企业家——能够具备同国际竞争对手、合作伙伴沟通和对抗的能力。按照这一要求，清华大学经济管理学院正在不断推动英语教学的步伐，使得英语不仅是一门需要学习的核心课程，而且渗透到各门专业课程的学习当中。

课堂讲授之外, 课前课后的大量英文原版著作、案例的阅读对于提高学生的英文水平也是非常关键的。这不仅是积累相当的专业词汇的重要手段, 而且是对学习者思维方式的有效训练。

我们知道, 就阅读而言, 学习和借鉴国外先进的管理经验和掌握经济理论动态, 或是阅读翻译作品, 或是阅读原著。前者属于间接阅读, 后者属于直接阅读。直接阅读取决于读者的外文阅读能力, 有较高外语水平的读者当然喜欢直接阅读原著, 这样不仅可以避免因译者的疏忽或水平所限而造成的纰漏, 同时也可以尽享原作者思想的真实表达。而对于那些有一定外语基础, 但又不能完全独立阅读国外原著的读者来说, 外文的阅读能力是需要加强培养和训练的, 尤其是专业外语的阅读能力更是如此。如果一个人永远不接触专业外版图书, 他在获得国外学术信息方面就永远会比别人差半年甚至一年的时间, 他就会在无形中减弱自己的竞争能力。因此, 我们认为, 有一定外语基础的读者, 都应该尝试一下阅读外文原版, 只要努力并坚持, 就一定能过了这道关, 到那时就能体验到直接阅读的妙处了。

在掌握大量术语的同时, 我们更看重读者在阅读英文原版著作时对于西方管理者或研究者的思维方式的学习和体会。我认为, 原汁原味的世界级大师富有特色的表达方式背后, 反映了思维习惯, 反映了思想精髓, 反映了文化特征, 也反映了战略偏好。知己知彼, 对于跨文化的管理思想、方法的学习, 一定要熟悉这些思想、方法所孕育、成长的文化土壤, 这样, 有朝一日才能真正“具备国际战略头脑”。

以往, 普通读者购买和阅读英文原版还有一个书价的障碍。一本外版书少则几十美元, 多则上百美元, 一般读者只能望书兴叹。随着全球经济合作步伐的加快, 目前在出版行业有了一种新的合作出版的方式, 即外文影印版, 其价格几乎与国内同类图书持平。这样一来, 读者可以不必再为书价发愁。清华大学出版社这些年在这方面一直以独特的优势领先于同行。早在1997年, 清华大学出版社敢为人先, 在国内最早推出一批优秀商学英文版教材, 规模宏大, 在企业界和管理教育界引起不小的轰动, 更使国内莘莘学子受益良多。

为了配合清华大学经济管理学院推动英文授课的急需, 也为了向全国更多的 MBA 试点院校和更多的经济管理学院的教师和学生提供学习上的支持, 清华大学出版社再次隆重推出与世界著名出版集团合作的英文原版影印商学教科书, 也使广大工商界人士、经济管理类学生享用到最新最好质优价廉的国际教材。

祝愿我国的管理教育事业在社会各界的大力支持和关心下不断发展、日进日新; 祝愿我国的经济建设在不断涌现的大批高层次的面向国际市场竞争、具备国际经营头脑的管理者的勉力经营下早日中兴。

赵纯均 教授

清华大学经济管理学院



# 出版说明

为了适应经济全球化的发展趋势,满足国内广大读者了解、学习和借鉴国外先进的管理经验和掌握经济理论的前沿动态,清华大学出版社与国外著名出版公司合作影印出版一系列英文版经济管理方面的图书。我们所选择的图书,基本上是已再版多次、在国外深受欢迎、并被广泛采用的优秀教材,绝大部分是该领域中较具权威性的经典之作。在选书的过程中,我们得到了很多专家、学者的支持、帮助和鼓励,在此表示谢意!

我们根据国内的教学实际,删去了原版图书第 14 章 (Advanced Panel Data Methods)、第 17 章 (Limited Dependent Variable Models and Sample Selection Corrections) 和第 18 章 (Advanced Time Series Topics) 以及附录的部分内容。我们在对原版图书进行删节的同时保留了原书的页码,采用双页码标识。由此可能给读者带来的诸多不便,我们深表歉意。

由于原作者所处国家的政治、经济和文化背景等与我国不同,对书中所持观点,敬请广大读者在阅读过程中注意加以分析和鉴别。

我们期望这套影印书的出版对我国经济科学的发展能有所帮助,对我国经济管理专业的教学能有所促进。

欢迎广大读者给我们提出宝贵的意见和建议;同时也欢迎有关的专业人士向我们推荐您所接触到的国外优秀图书。

清华大学出版社

2017.2

# Preface

My motivation for writing the first edition of *Introductory Econometrics: A Modern Approach* was that I saw a fairly wide gap between how econometrics is taught to undergraduates and how empirical researchers think about and apply econometric methods. I became convinced that teaching introductory econometrics from the perspective of professional users of econometrics would actually simplify the presentation, in addition to making the subject much more interesting.

Based on the positive reactions to earlier editions, it appears that my hunch was correct. Many instructors, having a variety of backgrounds and interests and teaching students with different levels of preparation, have embraced the modern approach to econometrics espoused in this text. The emphasis in this edition is still on applying econometrics to real-world problems. Each econometric method is motivated by a particular issue facing researchers analyzing nonexperimental data. The focus in the main text is on understanding and interpreting the assumptions in light of actual empirical applications: the mathematics required is no more than college algebra and basic probability and statistics.

## Organized for Today's Econometrics Instructor

---

The sixth edition preserves the overall organization of the fifth. The most noticeable feature that distinguishes this text from most others is the separation of topics by the kind of data being analyzed. This is a clear departure from the traditional approach, which presents a linear model, lists all assumptions that may be needed at some future point in the analysis, and then proves or asserts results without clearly connecting them to the assumptions. My approach is first to treat, in Part 1, multiple regression analysis with cross-sectional data, under the assumption of random sampling. This setting is natural to students because they are familiar with random sampling from a population in their introductory statistics courses. Importantly, it allows us to distinguish assumptions made about the underlying population regression model—assumptions that can be given economic or behavioral content—from assumptions about how the data were sampled. Discussions about the consequences of nonrandom sampling can be treated in an intuitive fashion after the students have a good grasp of the multiple regression model estimated using random samples.

An important feature of a modern approach is that the explanatory variables—along with the dependent variable—are treated as outcomes of random variables. For the social sciences, allowing random explanatory variables is much more realistic than the traditional assumption of nonrandom explanatory variables. As a nontrivial benefit, the population model/random sampling approach reduces the number of assumptions that students must absorb and understand. Ironically, the classical approach to regression analysis, which treats the explanatory variables as fixed in repeated samples and is still pervasive in introductory texts, literally applies to data collected in an experimental setting. In addition, the contortions required to state and explain assumptions can be confusing to students.

My focus on the population model emphasizes that the fundamental assumptions underlying regression analysis, such as the zero mean assumption on the unobservable error term, are properly

stated conditional on the explanatory variables. This leads to a clear understanding of the kinds of problems, such as heteroskedasticity (nonconstant variance), that can invalidate standard inference procedures. By focusing on the population, I am also able to dispel several misconceptions that arise in econometrics texts at all levels. For example, I explain why the usual  $R$ -squared is still valid as a goodness-of-fit measure in the presence of heteroskedasticity (Chapter 8) or serially correlated errors (Chapter 12); I provide a simple demonstration that tests for functional form should not be viewed as general tests of omitted variables (Chapter 9); and I explain why one should always include in a regression model extra control variables that are uncorrelated with the explanatory variable of interest, which is often a key policy variable (Chapter 6).

Because the assumptions for cross-sectional analysis are relatively straightforward yet realistic, students can get involved early with serious cross-sectional applications without having to worry about the thorny issues of trends, seasonality, serial correlation, high persistence, and spurious regression that are ubiquitous in time series regression models. Initially, I figured that my treatment of regression with cross-sectional data followed by regression with time series data would find favor with instructors whose own research interests are in applied microeconomics, and that appears to be the case. It has been gratifying that adopters of the text with an applied time series bent have been equally enthusiastic about the structure of the text. By postponing the econometric analysis of time series data, I am able to put proper focus on the potential pitfalls in analyzing time series data that do not arise with cross-sectional data. In effect, time series econometrics finally gets the serious treatment it deserves in an introductory text.

As in the earlier editions, I have consciously chosen topics that are important for reading journal articles and for conducting basic empirical research. Within each topic, I have deliberately omitted many tests and estimation procedures that, while traditionally included in textbooks, have not withstood the empirical test of time. Likewise, I have emphasized more recent topics that have clearly demonstrated their usefulness, such as obtaining test statistics that are robust to heteroskedasticity (or serial correlation) of unknown form, using multiple years of data for policy analysis, or solving the omitted variable problem by instrumental variables methods. I appear to have made fairly good choices, as I have received only a handful of suggestions for adding or deleting material.

I take a systematic approach throughout the text, by which I mean that each topic is presented by building on the previous material in a logical fashion, and assumptions are introduced only as they are needed to obtain a conclusion. For example, empirical researchers who use econometrics in their research understand that not all of the Gauss-Markov assumptions are needed to show that the ordinary least squares (OLS) estimators are unbiased. Yet the vast majority of econometrics texts introduce a complete set of assumptions (many of which are redundant or in some cases even logically conflicting) before proving the unbiasedness of OLS. Similarly, the normality assumption is often included among the assumptions that are needed for the Gauss-Markov Theorem, even though it is fairly well known that normality plays no role in showing that the OLS estimators are the best linear unbiased estimators.

My systematic approach is illustrated by the order of assumptions that I use for multiple regression in Part 1. This structure results in a natural progression for briefly summarizing the role of each assumption:

MLR.1: Introduce the population model and interpret the population parameters (which we hope to estimate).

MLR.2: Introduce random sampling from the population and describe the data that we use to estimate the population parameters.

MLR.3: Add the assumption on the explanatory variables that allows us to compute the estimates from our sample; this is the so-called no perfect collinearity assumption.

MLR.4: Assume that, in the population, the mean of the unobservable error does not depend on the values of the explanatory variables; this is the “mean independence” assumption combined with a zero population mean for the error, and it is the key assumption that delivers unbiasedness of OLS.



After introducing Assumptions MLR.1 to MLR.3, one can discuss the algebraic properties of ordinary least squares—that is, the properties of OLS for a particular set of data. By adding Assumption MLR.4, we can show that OLS is unbiased (and consistent). Assumption MLR.5 (homoskedasticity) is added for the Gauss-Markov Theorem and for the usual OLS variance formulas to be valid. Assumption MLR.6 (normality), which is not introduced until Chapter 4, is added to round out the classical linear model assumptions. The six assumptions are used to obtain exact statistical inference and to conclude that the OLS estimators have the smallest variances among all unbiased estimators.

I use parallel approaches when I turn to the study of large-sample properties and when I treat regression for time series data in Part 2. The careful presentation and discussion of assumptions makes it relatively easy to transition to Part 3, which covers advanced topics that include using pooled cross-sectional data, exploiting panel data structures, and applying instrumental variables methods. Generally, I have strived to provide a unified view of econometrics, where all estimators and test statistics are obtained using just a few intuitively reasonable principles of estimation and testing (which, of course, also have rigorous justification). For example, regression-based tests for heteroskedasticity and serial correlation are easy for students to grasp because they already have a solid understanding of regression. This is in contrast to treatments that give a set of disjointed recipes for outdated econometric testing procedures.

Throughout the text, I emphasize *ceteris paribus* relationships, which is why, after one chapter on the simple regression model, I move to multiple regression analysis. The multiple regression setting motivates students to think about serious applications early. I also give prominence to policy analysis with all kinds of data structures. Practical topics, such as using proxy variables to obtain *ceteris paribus* effects and interpreting partial effects in models with interaction terms, are covered in a simple fashion.

## New to This Edition

---

I have added new exercises to almost every chapter, including the appendices. Most of the new computer exercises use new data sets, including a data set on student performance and attending a Catholic high school and a time series data set on presidential approval ratings and gasoline prices. I have also added some harder problems that require derivations.

There are several changes to the text worth noting. Chapter 2 contains a more extensive discussion about the relationship between the simple regression coefficient and the correlation coefficient. Chapter 3 clarifies issues with comparing R-squareds from models when data are missing on some variables (thereby reducing sample sizes available for regressions with more explanatory variables).

Chapter 6 introduces the notion of an average partial effect (APE) for models linear in the parameters but including nonlinear functions, primarily quadratics and interaction terms. The notion of an APE, which was implicit in previous editions, has become an important concept in empirical work; understanding how to compute and interpret APEs in the context of OLS is a valuable skill. For more advanced classes, the introduction in Chapter 6 eases the way to the discussion of APEs in the non-linear models studied, which also includes an expanded discussion of APEs—including now showing APEs in tables alongside coefficients in logit, probit, and Tobit applications.

In Chapter 8, I refine some of the discussion involving the issue of heteroskedasticity, including an expanded discussion of Chow tests and a more precise description of weighted least squares when the weights must be estimated. Chapter 9, which contains some optional, slightly more advanced topics, defines terms that appear often in the large literature on missing data. A common practice in empirical work is to create indicator variables for missing data, and to include them in a multiple regression analysis. Chapter 9 discusses how this method can be implemented and when it will produce unbiased and consistent estimators.

Chapter 15 now has a more detailed discussion of the problem of weak instrumental variables so that students can access the basics without having to track down more advanced sources.

## **Targeted at Undergraduates, Adaptable for Master's Students**

---

The text is designed for undergraduate economics majors who have taken college algebra and one semester of introductory probability and statistics. (Appendices A, B, and C contain the requisite background material.) A one-semester or one-quarter econometrics course would not be expected to cover all, or even any, of the more advanced material in Part 3. A typical introductory course includes Chapters 1 through 8, which cover the basics of simple and multiple regression for cross-sectional data. Provided the emphasis is on intuition and interpreting the empirical examples, the material from the first eight chapters should be accessible to undergraduates in most economics departments. Most instructors will also want to cover at least parts of the chapters on regression analysis with time series data, Chapters 10 and 12, in varying degrees of depth. In the one-semester course that I teach at Michigan State, I cover Chapter 10 fairly carefully, give an overview of the material in Chapter 11, and cover the material on serial correlation in Chapter 12. I find that this basic one-semester course puts students on a solid footing to write empirical papers, such as a term paper, a senior seminar paper, or a senior thesis. Chapter 9 contains more specialized topics that arise in analyzing cross-sectional data, including data problems such as outliers and nonrandom sampling; for a one-semester course, it can be skipped without loss of continuity.

The structure of the text makes it ideal for a course with a cross-sectional or policy analysis focus: the time series chapters can be skipped in lieu of topics from Chapters 9 or 15. Chapter 13 is advanced only in the sense that it treats two new data structures: independently pooled cross sections and two-period panel data analysis. Such data structures are especially useful for policy analysis, and the chapter provides several examples. Students with a good grasp of Chapters 1 through 8 will have little difficulty with Chapter 13. A good way to end a course on cross-sectional methods is to cover the rudiments of instrumental variables estimation in Chapter 15.

I have used selected material in Part 3, including Chapter 13, in a senior seminar geared to producing a serious research paper. Along with the basic one-semester course, students who have been exposed to basic panel data analysis, instrumental variables estimation, and limited dependent variable models are in a position to read large segments of the applied social sciences literature.

The text is also well suited for an introductory master's level course, where the emphasis is on applications rather than on derivations using matrix algebra. Several instructors have used the text to teach policy analysis at the master's level. For instructors wanting to present the material in matrix form, Appendix E are self-contained treatments of the matrix algebra and the multiple regression model in matrix form.

At Michigan State, PhD students in many fields that require data analysis—including accounting, agricultural economics, development economics, economics of education, finance, international economics, labor economics, macroeconomics, political science, and public finance—have found the text to be a useful bridge between the empirical work that they read and the more theoretical econometrics they learn at the PhD level.

## Design Features

---

Numerous in-text questions are scattered throughout, with answers supplied in Appendix F. These questions are intended to provide students with immediate feedback. Each chapter contains many numbered examples. Several of these are case studies drawn from recently published papers, but where I have used my judgment to simplify the analysis, hopefully without sacrificing the main point. The end-of-chapter problems and computer exercises are heavily oriented toward empirical work, rather than complicated derivations. The students are asked to reason carefully based on what they have learned. The computer exercises often expand on the in-text examples. Several exercises use data sets from published works or similar data sets that are motivated by published research in economics and other fields.

A pioneering feature of this introductory econometrics text is the extensive glossary. The short definitions and descriptions are a helpful refresher for students studying for exams or reading empirical research that uses econometric methods. I have added and updated several entries for the fifth edition.

## Data Sets—Available in Six Formats

---

This edition adds R data set as an additional format for viewing and analyzing data. In response to popular demand, this edition also provides the Minitab<sup>®</sup> format. With more than 100 data sets in six different formats, including Stata<sup>®</sup>, EViews<sup>®</sup>, Minitab<sup>®</sup>, Microsoft<sup>®</sup> Excel, and R, the instructor has many options for problem sets, examples, and term projects. Because most of the data sets come from actual research, some are very large. Except for partial lists of data sets to illustrate the various data structures, the data sets are not reported in the text. This book is geared to a course where computer work plays an integral role.

## Updated Data Sets Handbook

---

An extensive data description manual is also available online. This manual contains a list of data sources along with suggestions for ways to use the data sets that are not described in the text. This unique handbook, created by author Jeffrey M. Wooldridge, lists the source of all data sets for quick reference and how each might be used. Because the data book contains page numbers, it is easy to see how the author used the data in the text. Students may want to view the descriptions of each data set and it can help guide instructors in generating new homework exercises, exam problems, or term projects. The author also provides suggestions on improving the data sets in this detailed resource that is available on the book's companion website at <http://login.cengage.com> and students can access it free at [www.cengagebrain.com](http://www.cengagebrain.com).

## Instructor Supplements

---

### Instructor's Manual with Solutions

The *Instructor's Manual with Solutions* contains answers to all problems and exercises, as well as teaching tips on how to present the material in each chapter. The instructor's manual also contains sources for each of the data files, with many suggestions for how to use them on problem sets, exams, and term papers. This supplement is available online only to instructors at <http://login.cengage.com>.

## PowerPoint Slides

Exceptional PowerPoint® presentation slides help you create engaging, memorable lectures. You will find teaching slides for each chapter in this edition, including the advanced chapters in Part 3. You can modify or customize the slides for your specific course. PowerPoint® slides are available for convenient download on the instructor-only, password-protected portion of the book's companion website at <http://login.cengage.com>.

## Scientific Word Slides

Developed by the author, Scientific Word® slides offer an alternative format for instructors who prefer the Scientific Word® platform, the word processor created by MacKichan Software, Inc. for composing mathematical and technical documents using LaTeX typesetting. These slides are based on the author's actual lectures and are available in PDF and TeX formats for convenient download on the instructor-only, password-protected section of the book's companion website at <http://login.cengage.com>.

## Test Bank

Cengage Learning Testing, powered by Cognero® is a flexible, online system that allows you to import, edit, and manipulate content from the text's test bank or elsewhere. You have the flexibility to include your own favorite test questions, create multiple test versions in an instant, and deliver tests from your LMS, your classroom, or wherever you want. In the test bank for *INTRODUCTORY ECONOMETRICS, 6E* you will find a wealth and variety of problems, ranging from multiple-choice to questions that require simple statistical derivations to questions that require interpreting computer output.

## Suggestions for Designing Your Course

I have already commented on the contents of most of the chapters as well as possible outlines for courses. Here I provide more specific comments about material in chapters that might be covered or skipped:

Chapter 9 has some interesting examples (such as a wage regression that includes IQ score as an explanatory variable). The rubric of proxy variables does not have to be formally introduced to present these kinds of examples, and I typically do so when finishing up cross-sectional analysis. In Chapter 12, for a one-semester course, I skip the material on serial correlation robust inference for ordinary least squares as well as dynamic models of heteroskedasticity.

Even in a second course I tend to spend only a little time on Chapter 16, which covers simultaneous equations analysis. I have found that instructors differ widely in their opinions on the importance of teaching simultaneous equations models to undergraduates. Some think this material is fundamental; others think it is rarely applicable. My own view is that simultaneous equations models are overused (see Chapter 16 for a discussion). If one reads applications carefully, omitted variables and measurement error are much more likely to be the reason one adopts instrumental variables estimation, and this is why I use omitted variables to motivate instrumental variables estimation in Chapter 15. Still, simultaneous equations models are indispensable for estimating demand and supply functions, and they apply in some other important cases as well.

Chapter 19, which would be added to the syllabus for a course that requires a term paper, is much more extensive than similar chapters in other texts. It summarizes some of the methods appropriate for various kinds of problems and data structures, points out potential pitfalls, explains in some detail how to write a term paper in empirical economics, and includes suggestions for possible projects.

# About the Author

**Jeffrey M. Wooldridge** is University Distinguished Professor of Economics at Michigan State University, where he has taught since 1991. From 1986 to 1991, he was an assistant professor of economics at the Massachusetts Institute of Technology. He received his bachelor of arts, with majors in computer science and economics, from the University of California, Berkeley, in 1982, and received his doctorate in economics in 1986 from the University of California, San Diego. He has published more than 60 articles in internationally recognized journals, as well as several book chapters. He is also the author of *Econometric Analysis of Cross Section and Panel Data*, second edition. His awards include an Alfred P. Sloan Research Fellowship, the Plura Scripsit award from *Econometric Theory*, the Sir Richard Stone prize from the *Journal of Applied Econometrics*, and three graduate teacher-of-the-year awards from MIT. He is a fellow of the Econometric Society and of the *Journal of Econometrics*. He is past editor of the *Journal of Business and Economic Statistics*, and past econometrics coeditor of *Economics Letters*. He has served on the editorial boards of *Econometric Theory*, the *Journal of Economic Literature*, the *Journal of Econometrics*, the *Review of Economics and Statistics*, and the *Stata Journal*. He has also acted as an occasional econometrics consultant for Arthur Andersen, Charles River Associates, the Washington State Institute for Public Policy, Stratus Consulting, and Industrial Economics, Incorporated.



# 简 明 目 录

第 1 章 计量经济学的性质与经济数据 .....	1
第 1 部分 横截面数据的回归分析 .....	19
第 2 章 简单回归模型 .....	20
第 3 章 多元回归分析: 估计 .....	60
第 4 章 多元回归分析: 推断 .....	105
第 5 章 多元回归分析: OLS 的渐近性 .....	149
第 6 章 多元回归分析: 其他问题 .....	166
第 7 章 含有定性信息的多元回归分析: 二值 (或虚拟) 变量 .....	205
第 8 章 异方差性 .....	243
第 9 章 模型设定和数据问题的深入探讨 .....	274
第 2 部分 时间序列数据的回归分析 .....	311
第 10 章 时间序列数据的基本回归分析 .....	312
第 11 章 用时间序列数据计算 OLS 的其他问题 .....	344
第 12 章 时间序列回归中的序列相关和异方差 .....	372
第 3 部分 高级专题讨论 .....	401
第 13 章 跨时横截面的混合: 简单综列数据方法 .....	402
第 15 章 工具变量估计与两阶段最小二乘法 .....	461
第 16 章 联立方程模型 .....	499
第 19 章 一个经验项目的实施 .....	605
附录 .....	
附录 E 矩阵形式的线性回归模型 .....	720
附录 F 各章习题解答 .....	734
附录 G 统计学用表 .....	743
参考文献 .....	750
术语表 .....	756

# Brief Contents

<b>Chapter 1</b>	The Nature of Econometrics and Economic Data	<b>1</b>
<b>PART 1: Regression Analysis with Cross-Sectional Data</b>		<b>19</b>
<b>Chapter 2</b>	The Simple Regression Model	20
<b>Chapter 3</b>	Multiple Regression Analysis: Estimation	60
<b>Chapter 4</b>	Multiple Regression Analysis: Inference	105
<b>Chapter 5</b>	Multiple Regression Analysis: OLS Asymptotics	149
<b>Chapter 6</b>	Multiple Regression Analysis: Further Issues	166
<b>Chapter 7</b>	Multiple Regression Analysis with Qualitative Information: Binary (or Dummy) Variables	205
<b>Chapter 8</b>	Heteroskedasticity	243
<b>Chapter 9</b>	More on Specification and Data Issues	274
<b>PART 2: Regression Analysis with Time Series Data</b>		<b>311</b>
<b>Chapter 10</b>	Basic Regression Analysis with Time Series Data	312
<b>Chapter 11</b>	Further Issues in Using OLS with Time Series Data	344
<b>Chapter 12</b>	Serial Correlation and Heteroskedasticity in Time Series Regressions	372
<b>PART 3: Advanced Topics</b>		<b>401</b>
<b>Chapter 13</b>	Pooling Cross Sections Across Time: Simple Panel Data Methods	402
<b>Chapter 15</b>	Instrumental Variables Estimation and Two Stage Least Squares	461
<b>Chapter 16</b>	Simultaneous Equations Models	499
<b>Chapter 19</b>	Carrying Out an Empirical Project	605
<b>APPENDICES</b>		
<b>Appendix E</b>	The Linear Regression Model in Matrix Form	720
<b>Appendix F</b>	Answers to Chapter Questions	734
<b>Appendix G</b>	Statistical Tables	743
<b>References</b>		<b>750</b>
<b>Glossary</b>		<b>756</b>

# Contents

Preface xii

About the Author xxi

## CHAPTER 1 The Nature of Econometrics and Economic Data 1

- 1-1 What Is Econometrics? 1
- 1-2 Steps in Empirical Economic Analysis 2
- 1-3 The Structure of Economic Data 5
  - 1-3a Cross-Sectional Data 5
  - 1-3b Time Series Data 7
  - 1-3c Pooled Cross Sections 8
  - 1-3d Panel or Longitudinal Data 9
  - 1-3e A Comment on Data Structures 10
- 1-4 Causality and the Notion of Ceteris Paribus in Econometric Analysis 10
- Summary 14
- Key Terms 14
- Problems 15
- Computer Exercises 15

## PART 1

## Regression Analysis with Cross-Sectional Data 19

### CHAPTER 2 The Simple Regression Model 20

- 2-1 Definition of the Simple Regression Model 20
- 2-2 Deriving the Ordinary Least Squares Estimates 24
  - 2-2a A Note on Terminology 31
- 2-3 Properties of OLS on Any Sample of Data 32
  - 2-3a Fitted Values and Residuals 32
  - 2-3b Algebraic Properties of OLS Statistics 32
  - 2-3c Goodness-of-Fit 35

- 2-4 Units of Measurement and Functional Form 36
  - 2-4a The Effects of Changing Units of Measurement on OLS Statistics 36
  - 2-4b Incorporating Nonlinearities in Simple Regression 37
  - 2-4c The Meaning of "Linear" Regression 40
- 2-5 Expected Values and Variances of the OLS Estimators 40
  - 2-5a Unbiasedness of OLS 40
  - 2-5b Variances of the OLS Estimators 45
  - 2-5c Estimating the Error Variance 48
- 2-6 Regression through the Origin and Regression on a Constant 50
- Summary 51
- Key Terms 52
- Problems 53
- Computer Exercises 56
- Appendix 2A 59

### CHAPTER 3 Multiple Regression Analysis: Estimation 60

- 3-1 Motivation for Multiple Regression 61
  - 3-1a The Model with Two Independent Variables 61
  - 3-1b The Model with  $k$  Independent Variables 63
- 3-2 Mechanics and Interpretation of Ordinary Least Squares 64
  - 3-2a Obtaining the OLS Estimates 64
  - 3-2b Interpreting the OLS Regression Equation 65
  - 3-2c On the Meaning of "Holding Other Factors Fixed" in Multiple Regression 67
  - 3-2d Changing More Than One Independent Variable Simultaneously 68
  - 3-2e OLS Fitted Values and Residuals 68
  - 3-2f A "Partialling Out" Interpretation of Multiple Regression 69

3-2g	Comparison of Simple and Multiple Regression Estimates	69
3-2h	Goodness-of-Fit	70
3-2i	Regression through the Origin	73
3-3	The Expected Value of the OLS Estimators	73
3-3a	Including Irrelevant Variables in a Regression Model	77
3-3b	Omitted Variable Bias: The Simple Case	78
3-3c	Omitted Variable Bias: More General Cases	81
3-4	The Variance of the OLS Estimators	81
3-4a	The Components of the OLS Variances. Multicollinearity	83
3-4b	Variances in Misspecified Models	86
3-4c	Estimating $\sigma^2$ Standard Errors of the OLS Estimators	87
3-5	Efficiency of OLS: The Gauss-Markov Theorem	89
3-6	Some Comments on the Language of Multiple Regression Analysis	90
	Summary	91
	Key Terms	93
	Problems	93
	Computer Exercises	97
	Appendix 3A	101

## CHAPTER 4 Multiple Regression Analysis: Inference 105

4-1	Sampling Distributions of the OLS Estimators	105
4-2	Testing Hypotheses about a Single Population Parameter: The $t$ Test	108
4-2a	Testing against One-Sided Alternatives	110
4-2b	Two-Sided Alternatives	114
4-2c	Testing Other Hypotheses about $\beta_j$	116
4-2d	Computing $p$ -Values for $t$ Tests	118
4-2e	A Reminder on the Language of Classical Hypothesis Testing	120
4-2f	Economic, or Practical, versus Statistical Significance	120
4-3	Confidence Intervals	122
4-4	Testing Hypotheses about a Single Linear Combination of the Parameters	124
4-5	Testing Multiple Linear Restrictions: The $F$ Test	127
4-5a	Testing Exclusion Restrictions	127
4-5b	Relationship between $F$ and $t$ Statistics	132
4-5c	The $R$ -Squared Form of the $F$ Statistic	133
4-5d	Computing $p$ -Values for $F$ Tests	134
4-5e	The $F$ Statistic for Overall Significance of a Regression	135
4-5f	Testing General Linear Restrictions	136

4-6	Reporting Regression Results	137
	Summary	139
	Key Terms	140
	Problems	141
	Computer Exercises	146

## CHAPTER 5 Multiple Regression Analysis: OLS Asymptotics 149

5-1	Consistency	150
5-1a	Deriving the Inconsistency in OLS	153
5-2	Asymptotic Normality and Large Sample Inference	154
5-2a	Other Large Sample Tests: The Lagrange Multiplier Statistic	158
5-3	Asymptotic Efficiency of OLS	161
	Summary	162
	Key Terms	162
	Problems	162
	Computer Exercises	163
	Appendix 5A	165

## CHAPTER 6 Multiple Regression Analysis: Further Issues 166

6-1	Effects of Data Scaling on OLS Statistics	166
6-1a	Beta Coefficients	169
6-2	More on Functional Form	171
6-2a	More on Using Logarithmic Functional Forms	171
6-2b	Models with Quadratics	173
6-2c	Models with Interaction Terms	177
6-2d	Computing Average Partial Effects	179
6-3	More on Goodness-of-Fit and Selection of Regressors	180
6-3a	Adjusted $R$ -Squared	181
6-3b	Using Adjusted $R$ -Squared to Choose between Nonnested Models	182
6-3c	Controlling for Too Many Factors in Regression Analysis	184
6-3d	Adding Regressors to Reduce the Error Variance	185
6-4	Prediction and Residual Analysis	186
6-4a	Confidence Intervals for Predictions	186
6-4b	Residual Analysis	190
6-4c	Predicting $y$ When $\log(y)$ Is the Dependent Variable	190
6-4d	Predicting $y$ When the Dependent Variable Is $\log(y)$	192