INTRODUCTORY ECONOMETRICS

This highly accessible and innovative text and accompanying CD-ROM use Excel workbooks powered by Visual Basic macros to teach the core concepts of econometrics without advanced mathematics. These materials enable Monte Carlo simulations to be run by students with a click of a button. The fundamental teaching strategy is to use clear language and take advantage of recent developments in computer technology to create concrete, visual explanations of difficult, abstract ideas. Intelligent repetition of concrete examples effectively conveys the properties of the ordinary least squares (OLS) estimator and the nature of heteroskedasticity and autocorrelation. Coverage includes omitted variables, binary response models, basic time series methods, and an introduction to simultaneous equations. The authors teach students how to construct their own real-world data sets drawn from the Internet, which they can analyze with Excel or with other econometric software. The Excel add-ins included with this book allow students to draw histograms, find *P*-values of various test statistics (including Durbin–Watson), obtain robust standard errors, and construct their own Monte Carlo and bootstrap simulations. For more, visit www.wabash.edu/econometrics.

Humberto Barreto is the DeVore Professor of Economics at Wabash College in Crawfordsville, Indiana. He received his Ph.D. from the University of North Carolina at Chapel Hill. Professor Barreto has lectured often on teaching economics with computer-based methods, including short courses with the National Science Foundation's Chautauqua program. The author of *The Entrepreneur in Microeconomic Theory*, Professor Barreto has served as a Fulbright Scholar in the Dominican Republic and is the Manager of Electronic Information for the History of Economics Society.

Frank M. Howland is Associate Professor of Economics at Wabash College. He earned his Ph.D. in Economics from Stanford University. He was a visiting researcher at the Fundación de Estudios de Economía Aplicada in Madrid, Spain. Howland has extensive experience designing spreadsheets and Visual Basic macros for intermediate microeconomics, statistics, and corporate finance courses. His academic research covers a variety of topics, including government crop support programs and the economics of higher education.

INTRODUCTORY ECONOMETRICS

Using Monte Carlo Simulation with Microsoft Excel®

HUMBERTO BARRETO
Wabash College

FRANK M. HOWLAND
Wabash College

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521843195

First published 2006

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Barreto, Humberto, 1960–
Introductory econometrics : using Monte Carlo simulation with Microsoft Excel /
Humberto Barreto, Frank M. Howland.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-521-84319-5 (hbk.)
ISBN-10: 0-521-84319-7 (hbk.)
1. Econometrics. 2. Monte Carlo method – Data processing. 3. Microsoft Excel (Computer file)
I. Howland, Frank M., 1958– II. Title.
HB139.B376 2006
330′.1′518282 – dc22 2005014585

ISBN-13 978-0-521-84319-5 hardback
ISBN-10 0-521-84319-7 hardback

Humberto Barreto

Para mi familia, Tami, Tyler, Nicolas, y Jonah

Frank M. Howland

To my parents, Bette Howland and Howard Howland

Contents

Preface

"I hear and I forget. I see and I remember. I do and I understand."

Confucius[1]

The Purpose of This Book

We wrote this book to help you understand econometrics. This book is quite different from the textbooks you are used to. Our fundamental strategy is to use clear language and take advantage of recent developments in computers to create concrete, visual explanations of difficult, abstract ideas.

Instead of passively reading, you will be using the accompanying Microsoft Excel workbooks to create a variety of graphs and other output while you interact with this book. Active learning is, of course, the goal of the Excel files. You will work through a series of questions, discovering patterns in the data or illustrating a particular property. Often, we will ask you to create your own version of what is on the printed page. This is made easy by the many buttons and other enhancements we have incorporated in the Excel workbooks.

You may be worried that learning econometrics will be a long, hard journey through a series of boring and extremely puzzling mathematical formulas. We will not deny that acquiring econometrics skills and knowledge takes real effort – you must carefully work through every Excel workbook and pay attention to detail – but introductory econometrics has little to do with complicated mathematics, nor need it be boring. In fact, the core of econometrics relies on logic and common sense. The methods presented in this book can be used to help answer questions about the value of education, the presence of discrimination, the effects of speed limits, and much more.

[1] This quote is frequently attributed to Confucius, but it is not in the *Analects* (a collection of excerpts and sayings), which were compiled by his followers after his death.

Our Goals

This book embodies a new approach to teaching introductory econometrics. Our approach is dictated by our beliefs regarding the purposes of a first course in undergraduate econometrics and the most important concepts that belong in that course; our frustrations with the traditional equation-laden, proof-oriented presentation of econometrics; and our experience with computer simulation as a tool that can overcome many of the limitations of traditional textbooks.

In terms of a student's educational development, there are short- and long-term reasons for including an introductory econometrics course in the undergraduate economics curriculum. Three major short-term goals for such a course stand out. A fundamental purpose of a first course in econometrics is to enable students to become intelligent readers of others' econometric analyses. To do so, they need to be able to interpret coefficient estimates and functional forms; understand simple inferential statistics, including the sampling distribution; and go beyond accepting all results at face value. A more ambitious introductory course should teach students to conduct creditable elementary econometric research. Students should be able to gather and document data, choose appropriate functional forms, run and interpret multiple regressions, conduct hypothesis tests and construct confidence intervals, and describe the major limitations of their analyses. Finally, an introductory course should prepare some students to take a second course in econometrics. Students who will take a second course ought to come to appreciate the method of least squares, the logic of the Gauss–Markov theorem, and the distinction between finite sample and asymptotic results.

In the long term, the learning that students carry with them throughout their lives should include two basic lessons: First, economic data should be interpreted as the outcome of a data-generation process in which chance plays a crucial role; second, because they typically deal with observational studies instead of controlled experiments, economists must always worry about whether the models they estimate and, more generally, the explanations they give for economic phenomena, are subject to confounding by omitted variables. Only a few students will go on to conduct econometric research in their future careers, but they all will benefit from these two fundamental ideas as workers and citizens who must evaluate the claims of business leaders, social scientists, and politicians.

An econometrics course that aims to teach these short- and long-term lessons must impart a very sophisticated message, which in its essence is something as follows: The data we observe must be interpreted as being produced by some data-generation process. Modeling that process requires both economic and statistical theory. To recover the parameters of the model,

we use estimators. Every estimator has important properties relating to its accuracy (bias or consistency) and precision (standard error). Depending on what we know or conjecture about the data-generation process, we will want to use different estimators. Given the nonexperimental character of our data, we must always be on guard for the effects of omitted variables.

How do we convey this complicated message? Many textbooks employ mathematical formalism to teach the numerous abstract concepts in the basic story. The resulting mass of equations and theorems intimidates students, and, worse yet, hides the truly essential and genuinely difficult core logic of econometrics. In the traditional classroom format of a lecture accompanied by chalkboard or overhead projector, students lapse into desperate attempts at passive memorization. Furthermore, there is pressure to cram the course and textbook with as many of the results of modern econometric research as possible.

Our approach differs from that of the traditional textbook in three aspects: we emphasize concrete examples rather than equations to exemplify abstract concepts, active learning by using computers rather than passively reading a book, and a focus on a few key ideas rather than an attempt to cover the whole waterfront. We wholeheartedly agree with Peter Kennedy and Michael Murray, critics of the traditional approach, who have argued, first, that the crucial concept in introductory econometrics (and statistics) is that of the sampling distribution and, second, that students can only learn that concept by actively grappling with it.[2] On the basis of our teaching experience, we give a second crucial abstraction almost as much prominence as the sampling distribution: the way in which a multiple regression summarizes the relationship between a dependent variable and several independent variables, including especially the notion of ceteris paribus.

All of these pedagogical considerations led to our choice of Microsoft Excel as the central vehicle to teach econometrics. We use Excel's underlying programming language, Visual Basic, to create buttons and other tools to tailor the environment for the student. The key advantages of computer-based instruction are dynamic visualization and interesting repetition. A printed textbook may contain outstanding graphics, but on the page all charts and tables are of necessity static. In contrast, using Excel, a student can instantly redraw charts and tables after changing a parameter or taking another sample. Students can toggle through different charts depicting the same data set or go back and forth through a complicated exposition. The ability of spreadsheets to convey interesting repetition greatly increases the effectiveness of specific, concrete examples designed to illustrate general, abstract ideas. Students are able to associate the specific numbers on the screen with the abstract symbols

[2] Kennedy (1998) and Murray (1999a).

in equations and can see the workings of the general claim when the asserted result is shown to hold over and over again in specific examples.

The advantages of Excel and Visual Basic are perhaps best displayed in the numerous Monte Carlo simulations we use to approximate sampling distributions throughout the book. Visual Basic obtains the samples, computes the estimates, and draws the histograms. Students are invited to make comparisons by altering the parameters of the data-generation process or directly racing two estimators against each other.[3] We are under no illusions that students will immediately understand the sampling distribution. Thus, we employ Monte Carlo simulation whenever it is relevant and ask students to view the outcomes of Monte Carlo experiments from a variety of angles. Because it is easier to understand, we have used the fixed-X-in-repeated-sampling assumption in almost all of our Monte Carlo simulations. We also, however, demonstrate other, more realistic sampling schemes.

We also use Excel to compute regression estimates in numerous examples. Although Excel has been rightly criticized for its sloppy statistical algorithms, it is adequate for the relatively straightforward computations required in an introductory course. Where Excel is lacking, we have written add-ins – for example to draw histograms, compute Durbin–Watson statistics and robust standard errors, and obtain nonlinear least squares and maximum likelihood estimates in Probit and Logit models. We do not recommend Excel as a statistical analysis package; we use it instead as a teaching tool.

To sum up, our primary goal in writing this book was to bridge the gap between the traditional, formal presentation of econometrics and the abilities of the typical undergraduate student. Today's student is, on the one hand, uncomfortable with mathematical formalism, and, on the other, adept at visually oriented use of computers.

We take advantage of modern computing technology to teach introductory econometrics more effectively. The basic Microsoft Excel spreadsheet, as familiar to the student as pencil and paper, is augmented and enhanced by powerful macros, buttons, and links that easily facilitate complicated computations and simulations. Monte Carlo simulation is the perfect tool for conveying the fundamental concept in econometrics – the sampling distribution – to the modern audience. It produces concrete, visual output and permits exploration of the properties of estimators without sophisticated mathematics.

It is ironic that simulation- and computer-intensive numerical techniques figure prominently in frontier research in econometric theory, whereas the teaching of econometrics languishes in old-style, chalk-and-talk memorization and proof methods. Our goal is to bring the benefits of the computer revolution to the undergraduate econometrics textbook and classroom.

[3] See Murray (1999b).

Content and Level of Presentation

This book is divided into two parts: descriptive data analysis and inferential econometrics. The first part is devoted to methods of summarizing bivariate and multivariate data. We use the correlation coefficient and PivotTables to set up regression as an analogue to the average. Additional chapters on functional form, dummy variables, and multiple regression round out the first part.

In the second part of the book, we focus on the effects of chance in estimation and the interpretation of regression output. We begin with a chapter dedicated solely to Monte Carlo simulation. Throughout the second part of the book, we emphasize modeling the data-generation process. We explain the sampling distribution of the OLS estimator with Monte Carlo simulation to support the proof of the Gauss–Markov theorem. We also employ Monte Carlo simulation to demonstrate the effects of elementary violations of the classical linear model, including omitted variable bias, heteroskedasticity, and autocorrelation. Every chapter contains both contrived and real-world examples and data. Finally, we provide introductions to binary response (dummy dependent variable) models, forecasting, simultaneous equations models, and bootstrapping.

Although the entire book is essentially a study of regression analysis, our two-part organization enables us to emphasize that regression can be used to describe and summarize data without using any of its inferential machinery. When we turn to regression for inference, we are able to highlight the importance of chance in the process that generated the data.

Because we expect students to obtain and analyze data, we include detailed instructions on how to access a variety of data sets online. For example, the CPS.doc file (in the Basic Tools\InternetData folder) explains exactly how a student can extract data from the Current Population Survey (CPS) available at <ferret.bls.census.gov>. In addition, we include practical explanations of how to recode variables and construct an hourly wage variable. The CPS is an outstanding source and has provided the raw data for many excellent student papers.

Another advantage of online data resources is the ability to access the latest figures. Web links are included in our Excel files, and thus updating data is easy. Many of the workbooks have links to a variety of data sources. Online data resources have transformed our teaching and enabled students to access high-quality, timely data.

In presenting the material, we focus on getting a few key ideas exactly right and explaining these concepts as simply as possible. Instead of concisely writing a result in terse mathematical notation, we walk you through the commonsense logic behind the formula. We also repeat the same idea. Often,

we will use hypothetical data to demonstrate a point and then use an example from the real world to show how it has been applied.

To read this book and use our materials, you need a working knowledge of elementary statistics. You should understand the standard deviation (SD) and be able to read a histogram. You should be familiar with the standard error (SE) of the sample average. To help those in need of a little brushing up, we have included a chapter that reviews inferential statistics and the explicit modeling of a data-generation process. Finally, this book will make more sense if you have had at least an introductory economics course.

Of course, you also have to know how to use Excel and work with files on a computer. You do not need to be an expert, but we expect you to be able to create formulas and make a chart. Our materials will introduce you to much more advanced uses of Excel, and it is fair to say that, by working through this book, you will become a more sophisticated user of Excel.

Conclusion

The installed base of Microsoft Excel (and Office) is staggering. At any given time, many different versions of Excel are in use. Exactly counting all of the versions in use, legally purchased, and pirated is impossible. Table 1 shows the Gartner Group's estimate of the breakdown of Windows Office versions in use in 2001.

The materials in this book require Excel 97 (or greater). Each version will have differences in functionality and, especially, display, but the files packaged in this book should work with versions of Excel from 97 to 2003.

- To determine the version of Excel you are using, execute Help: About Microsoft Excel.
- To find the previous versions of Excel, visit <www.microsoft.com/office/previous/excel/default.asp>.
- To find the latest version of Excel, go to <office.microsoft.com/home/>.
- To compare versions of Office products, see <www.microsoft.com/office/editions/prodinfo/compare.mspx>.
- To search Microsoft's extensive Knowledge Base for questions about Excel, visit <support.microsoft.com/> and click on the Search the Knowledge Base link.

Office 95	10%
Office 97	55%
Office 2000	35%
Yearly Upgrade	15%

Table 1. Estimated Microsoft Office installed base in 2001.
Source: www.infotechtrends.com/

Acknowledgments

This book is a collaborative effort of several professors in the Wabash College Economics Department. Humberto Barreto and Frank Howland are the principal authors of this book and share equal responsibility for its content. In addition, Kealoha Widdows made extensive contributions to the book, and Joyce Burnette revised many of the illustrations, examples, and computer files.

Michael Einterz provided excellent error checking and test drove many examples. He added clear instructions for data sources and improved our presentation greatly. Matthew Schulz helped us polish the exercises and answers. The book also reflects comments and reactions of several generations of Wabash College students. We thank them all.

Scott Parris of Cambridge University Press deserves our gratitude. He took a chance on an idea far outside the mainstream, and we hope it pays off. Thank you, Scott, for your support and helpful feedback. We are grateful to Katie Greczylo, our production manager at TechBooks, and John Joswick, our copy editor.

We also thank our Web site designer Jeannine Smith. Her ability to write sophisticated computer code combined with artistic flair resulted in an excellent Web site (www.wabash.edu/econometrics) that is functional and fashionable.

A major inspiration for our work is *Statistics* by David Freedman, Robert Pisani, and Roger Purves (W. W. Norton & Company, 3rd edition, 1998). We have tried to adapt the same approach to the study of econometrics that Freedman et al. use in teaching statistics. That is, we emphasize the importance of basic concepts and reject the memorization of boring formulas. We have taken the central metaphor of their book, the box model, and extended it to handle the classical linear model. Finally, we follow their visual and verbal approach to the material.

We would appreciate your criticisms and suggestions.

Humberto Barreto and Frank Howland
barretoh@wabash.edu and
howlandf@wabash.edu

Crawfordsville, Indiana
2005

References

Kennedy, P. (1998). "Teaching Undergraduate Econometrics: A Suggestion for Fundamental Change," *American Economic Review* **88**(2): 487–491.

Murray, M. (1999a). "Taking Linear Estimators Seriously in Introductory Economics." Conference on Teaching Undergraduate Econometrics: Does Monte Carlo Work?, Middlebury College.

Murray, M. (1999b). "Econometrics Lectures in a Computer Classroom." *Journal of Economic Education* **30**(3): 308–321.