

Making Everything Easier!™

Econometrics

FOR
DUMMIES®
A Wiley Brand

Learn to:

- Grasp the techniques and applications of econometrics
- Make sense of complex concepts and material
- Use specialized software to apply your acquired econometrics skills

Roberto Pedace, PhD

*Associate professor in the Department
of Economics, Scripps College*



Get More and Do More at Dummies.com®



Start with **FREE** Cheat Sheets

Cheat Sheets include

- Checklists
- Charts
- Common Instructions
- And Other Good Stuff!

To access the Cheat Sheet created specifically for this book, go to
www.dummies.com/cheatsheet/econometrics

Get Smart at Dummies.com

Dummies.com makes your life easier with 1,000s of answers on everything from removing wallpaper to using the latest version of Windows.

Check out our

- Videos
- Illustrated Articles
- Step-by-Step Instructions

Plus, each month you can win valuable prizes by entering our Dummies.com sweepstakes. *

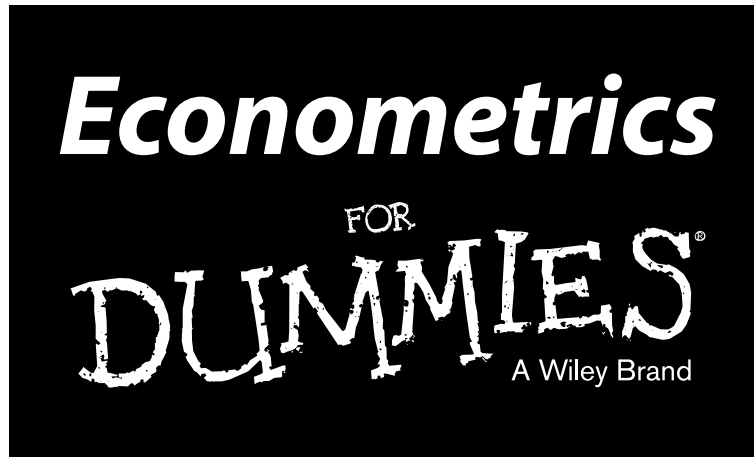
Want a weekly dose of Dummies? Sign up for Newsletters on

- Digital Photography
- Microsoft Windows & Office
- Personal Finance & Investing
- Health & Wellness
- Computing, iPods & Cell Phones
- eBay
- Internet
- Food, Home & Garden

Find out “HOW” at Dummies.com

*Sweepstakes not currently available in all countries; visit Dummies.com for official rules.





by Roberto Pedace



Econometrics For Dummies®

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2013 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, the Wiley logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc., and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002.

For technical support, please visit www.wiley.com/techsupport.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2013934761

ISBN 978-1-118-53384-0 (pbk); ISBN 978-1-118-53387-1 (ebk); ISBN 978-1-118-53388-8 (ebk); ISBN 978-1-118-53391-8 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

About the Author

Roberto Pedace is an associate professor of economics at Scripps College in Claremont, California. Prior to joining the faculty at Scripps College, he held positions at Claremont Graduate University, the University of Redlands, Claremont McKenna College, and the U.S. Census Bureau. He holds a PhD in economics from the University of California, Riverside.

Roberto regularly teaches courses in the areas of statistics, microeconomics, labor economics, and econometrics. While at the University of Redlands, he was nominated for both the Innovative Teaching Award and the Outstanding Teaching Award. At Scripps College, he was recognized for his scholarly achievements by winning the Mary W. Johnson Faculty Achievement Award in Scholarship.

Roberto's academic research interests are in the area of labor and personnel economics. His work addresses a variety of important public policy issues, including the effects of immigration on domestic labor markets and the impact of minimum wages on job training and unemployment. He also examines salary determination and personnel decisions in markets for professional athletes. His published work appears in the *Southern Economic Journal*, the *Journal of Sports Economics*, *Contemporary Economic Policy*, *Industrial Relations*, and other outlets.

Roberto is also a soccer fanatic. He's been playing soccer since the age of 5, paid for most of his undergraduate education with a soccer scholarship, and had a short semi-professional stint in the USISL (now known as the United Soccer League). He continues to participate in leagues and tournaments but now mostly enjoys sitting on the sidelines watching his children play soccer.

Dedication

To my wife, Cynthia, for supporting me emotionally and being a wonderful mother to our children. To my children, Vincent and Emily, for brightening up my days.

Author's Acknowledgments

None of this would have been possible if my professors hadn't motivated me and given me a solid foundation in economics. My undergraduate adviser at California State University, San Bernardino, Thomas Pierce, opened my eyes to the world of economics and gave me wonderful advice in preparation for graduate school. I was fortunate to have taken several courses from Nancy Rose and Mayo Toruño, who helped me see economics in a different light when standard theory just wasn't helping me understand certain aspects of the world. Kazim Konyar was the first to introduce me to the realm of econometrics and helped me understand how it could be a powerful complement to economic theory. At the University of California, Riverside, Aman Ullah's uncanny ability to make advanced econometric theory comprehensible to a first-year graduate student solidified my interest in the topic. Finally, in his labor economics course and as my dissertation adviser, David Fairris taught me the art of using econometrics to address important economic policy issues.

Many of my econometrics students deserve special gratitude. Several of them stand out: Lora Brill, Megan Cornell, Guadalupe De La Cruz, Matthew Lang, Chandler Lutz, India Mullady, and Stephanie Rohn. Some became friends, a few colleagues, and a couple coauthors, but all inspired me to think of effective approaches to making econometrics accessible, useful, and interesting.

I thank Sean Flynn, my friend and colleague, for believing that I'd be the best person to write this book and Linda Roghaar, my literary agent, for listening to Sean and having faith in my ability to complete the project.

The folks at Wiley have also been incredibly supportive. In particular, I'd like to thank Jennifer Tebbe, my project editor, for working with me every step of the way and keeping me motivated to stay on track with my deadlines. No matter how long the tunnel, she always helped me see the light at the end. Erin Calligan Mooney, my acquisitions editor at Wiley, also helped me get through my sample chapter and ensured that it would meet the standards of others on the editorial team. My copy editor, Caitie Copple, and technical reviewers, Ariel Belasen and Nicole Bissessar, were ideal for this project. Their "eagle eyes" were instrumental in finding my mistakes and improving the finished product.

My research assistant, Anne Miles, gathered data for some of the examples I use in the book and assisted with the imaging of figures and graphs. Her turnaround time was amazing, and I'll be forever grateful for all the hard work she provided on this project. I also want to thank my friend and colleague, Latika Chaudhary, for responding immediately to an urgent request for a sample of panel data.

Last, but not least, I'd like to thank my family and friends for being patient with me while I wrote this book. I know that sometimes I wasn't myself and that I'll need to make up for lost time.

Publisher's Acknowledgments

We're proud of this book; please send us your comments at <http://dummies.custhelp.com>. For other comments, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002.

Some of the people who helped bring this book to market include the following:

Acquisitions, Editorial, and Vertical Websites

Project Editor: Jennifer Tebbe

Acquisitions Editor: Erin Calligan Mooney

Copy Editor: Caitlin Copple

Assistant Editor: David Lutton

Editorial Program Coordinator: Joe Niesen

Technical Editors: Ariel Belasen,
Nicole Bissessar

Editorial Manager: Christine Meloy Beck

Editorial Assistants: Rachelle S. Amick,
Alexa Koschier

Cover Photo: © iStockphoto.com/studiocasper

Composition Services

Project Coordinator: Sheree Montgomery

Layout and Graphics: Carrie A. Cesavice,
Christin Swinford

Proofreader: Melissa Cossell

Indexer: Riverside Indexes, Inc.

Publishing and Editorial for Consumer Dummies

Kathleen Nebenhaus, Vice President and Executive Publisher

David Palmer, Associate Publisher

Kristin Ferguson-Wagstaffe, Product Development Director

Publishing for Technology Dummies

Andy Cummings, Vice President and Publisher

Composition Services

Debbie Stailey, Director of Composition Services

Contents at a Glance

<i>Introduction</i>	<i>1</i>
<i>Part I: Getting Started with Econometrics.....</i>	<i>5</i>
Chapter 1: Econometrics: The Economist's Approach to Statistical Analysis	7
Chapter 2: Getting the Hang of Probability	21
Chapter 3: Making Inferences and Testing Hypotheses.....	39
<i>Part II: Building the Classical Linear Regression Model.....</i>	<i>59</i>
Chapter 4: Understanding the Objectives of Regression Analysis	61
Chapter 5: Going Beyond Ordinary with the Ordinary Least Squares Technique	75
Chapter 6: Assumptions of OLS Estimation and the Gauss-Markov Theorem.....	93
Chapter 7: The Normality Assumption and Inference with OLS.....	111
<i>Part III: Working with the Classical Regression Model... </i>	<i>135</i>
Chapter 8: Functional Form, Specification, and Structural Stability	137
Chapter 9: Regression with Dummy Explanatory Variables	153
<i>Part IV: Violations of Classical Regression Model Assumptions.....</i>	<i>173</i>
Chapter 10: Multicollinearity	175
Chapter 11: Heteroskedasticity.....	191
Chapter 12: Autocorrelation.....	209
<i>Part V: Discrete and Restricted Dependent Variables in Econometrics.....</i>	<i>229</i>
Chapter 13: Qualitative Dependent Variables.....	231
Chapter 14: Limited Dependent Variable Models.....	253
<i>Part VI: Extending the Basic Econometric Model.....</i>	<i>265</i>
Chapter 15: Static and Dynamic Models	267
Chapter 16: Diving into Pooled Cross-Section Analysis.....	281
Chapter 17: Panel Econometrics.....	291

<i>Part VII: The Part of Tens</i>	305
Chapter 18: Ten Components of a Good Econometrics Research Project	307
Chapter 19: Ten Common Mistakes in Applied Econometrics	315
<i>Appendix: Statistical Tables</i>	321
<i>Index</i>	327

Table of Contents

<i>Introduction</i>	1
About This Book	1
Foolish Assumptions	2
Icons Used in This Book	3
Beyond the Book	3
Where to Go from Here	4
 <i>Part 1: Getting Started with Econometrics</i>	5
 Chapter 1: Econometrics: The Economist's Approach to Statistical Analysis.	7
Evaluating Economic Relationships	8
Using economic theory to describe outcomes and make predictions	8
Relying on sensible assumptions	8
Applying Statistical Methods to Economic Problems	9
Recognizing the importance of data type, frequency, and aggregation	10
Avoiding the data-mining trap	12
Incorporating quantitative and qualitative information	12
Using Econometric Software: An Introduction to STATA	13
Getting acquainted with STATA	14
Creating new variables	18
Estimating, testing, and predicting	19
 Chapter 2: Getting the Hang of Probability	21
Reviewing Random Variables and Probability Distributions	21
Looking at all possibilities: Probability density function (PDF)	22
Summing up the probabilities: Cumulative density function (CDF)	25
Putting variable information together: Bivariate or joint probability density	28
Predicting the future using what you know: Conditional probability density	29



Understanding Summary Characteristics of Random Variables	31
Making generalizations with expected value or mean	31
Measuring variance and standard deviation	33
Looking at relationships with covariance and correlation	35

Chapter 3: Making Inferences and Testing Hypotheses 39

Getting to Know Your Data with Descriptive Statistics	40
Calculating parameters and estimators	40
Determining whether an estimator is good	43
Laying the Groundwork of Prediction with the	
Normal and Standard Normal Distributions	44
Recognizing usual variables: Normal distribution	44
Putting variables on the same scale:	
Standard normal distribution (Z)	46
Working with Parts of the Population: Sampling Distributions	47
Simulating and using the central limit theorem	48
Defining the chi-squared (χ^2), t, and F distributions	49
Making Inferences and Testing Hypotheses	
with Probability Distributions	53
Performing a hypothesis test	54
The confidence interval approach	56
The test of significance approach	57

Part 11: Building the Classical Linear Regression Model 59

Chapter 4: Understanding the Objectives of Regression Analysis . . 61

Making a Case for Causality	62
Getting Acquainted with the Population Regression Function (PRF)	63
Setting up the PRF model	64
Walking through an example	66
Collecting and Organizing Data for Regression Analysis	68
Taking a snapshot: Cross-sectional data	68
Looking at the past to explain the present: Time-series data	70
Combining the dimensions of space and time:	
Panel or longitudinal data	71
Joining multiple snapshots: Pooled cross-sectional data	72

Chapter 5: Going Beyond Ordinary with the Ordinary Least Squares Technique. 75

Defining and Justifying the Least Squares Principle	76
Estimating the Regression Function and the Residuals	77

Obtaining Estimates of the Regression Parameters.....	78
Finding the formulas necessary to produce optimal coefficient values	79
Calculating the estimated regression coefficients.....	80
Interpreting Regression Coefficients.....	84
Seeing what regression coefficients have to say	84
Standardizing regression coefficients.....	86
Measuring Goodness of Fit.....	88
Decomposing variance.....	88
Measuring proportion of variance with R^2	89
Adjusting the goodness of fit in multiple regression	90
Evaluating fit versus quality	92

Chapter 6: Assumptions of OLS Estimation and the Gauss-Markov Theorem 93

Characterizing the OLS Assumptions	93
Linearity in parameters and additive error.....	94
Random sampling and variability.....	95
Imperfect linear relationships among the independent variables	96
Error term has a zero conditional mean; correct specification.....	97
Error term has a constant variance.....	98
Correlation of error observations is zero.....	99
Relying on the CLRM Assumptions: The Gauss-Markov Theorem.....	101
Proving the Gauss-Markov theorem.....	101
Summarizing the Gauss-Markov theorem.....	109

Chapter 7: The Normality Assumption and Inference with OLS . . . 111

Describing the Role of the Normality Assumption.....	111
The error term and the sampling distribution of OLS coefficients.....	113
Revisiting the standard normal distribution.....	114
Deriving a chi-squared distribution from the random error.....	115
OLS standard errors and the t-distribution.....	119
Testing the Significance of Individual Regression Coefficients	119
Picking an approach.....	121
Choosing the level of significance and p-values	123
Analyzing Variance to Determine Overall or Joint Significance	123
Normality, variance, and the F distribution	124
The reported F-statistic from OLS	124
Slope coefficients and the relationship between t and F.....	127
Joint significance for subsets of variables.....	128
Applying Forecast Error to OLS Predictions.....	130
Mean prediction and forecast error	130
Variance of mean prediction	131
All predictions are not the same: The prediction confidence interval	133

Part III: Working with the Classical Regression Model ... 135**Chapter 8: Functional Form, Specification, and Structural Stability.137**

Employing Alternative Functions	138
Quadratic function: Best for finding minimums and maximums	138
Cubic functions: Good for inflexion	139
Inverse function: Limiting the value of the dependent variable	140
Giving Linearity to Nonlinear Models	141
Working both sides to keep elasticity constant: The log-log model	141
Making investments and calculating rates of return: The log-linear model.....	143
Decreasing the change of the dependent variable: The linear-log model	145
Checking for Misspecification.....	147
Too many or too few: Selecting independent variables.....	147
Sensitivity isn't a virtue: Examining misspecification with results stability.....	150

Chapter 9: Regression with Dummy Explanatory Variables.153

Numbers Please! Quantifying Qualitative Information	153
Defining a dummy variable when you have only two possible characteristics.....	154
Juggling multiple characteristics with dummy variables	155
Finding Average Differences by Using a Dummy Variable	156
Specification	156
Interpretation	157
Combining Quantitative and Qualitative Data in the Regression Model.....	159
Specification	159
Interpretation	160
Interacting Quantitative and Qualitative Variables.....	162
Specification	162
Interpretation	162
Interacting Two (or More) Qualitative Characteristics.....	165
Specification	165
Interpretation	166
Segregate and Integrate: Testing for Significance	168
Revisiting the F-test for joint significance.....	168
Revisiting the Chow test	169

***Part IV: Violations of Classical Regression Model Assumptions* 173**

Chapter 10: Multicollinearity 175

Distinguishing between the Types of Multicollinearity.....	175
Pinpointing perfect multicollinearity	176
Zeroing in on high multicollinearity	178
Rules of Thumb for Identifying Multicollinearity	180
Pairwise correlation coefficients	181
Auxiliary regression and the variance inflation factor (VIF)	183
Knowing When and How to Resolve Multicollinearity Issues.....	185
Get more data.....	186
Use a new model	186
Expel the problem variable(s)	188

Chapter 11: Heteroskedasticity 191

Distinguishing between Homoskedastic and Heteroskedastic Disturbances.....	191
Homoskedastic error versus heteroskedastic error	192
The consequences of heteroskedasticity	193
Detecting Heteroskedasticity with Residual Analysis	194
Examining the residuals in graph form.....	194
Brushing up on the Breusch-Pagan test.....	196
Getting acquainted with the White test	197
Trying out the Goldfeld-Quandt test	199
Conducting the Park test	201
Correcting Your Regression Model for the Presence of Heteroskedasticity.....	203
Weighted least squares (WLS)	203
Robust standard errors (also known as White-corrected standard errors).....	205

Chapter 12: Autocorrelation 209

Examining Patterns of Autocorrelation	209
Positive versus negative autocorrelation	210
Misspecification and autocorrelation	212
Illustrating the Effect of Autoregressive Errors.....	212
Analyzing Residuals to Test for Autocorrelation	214
Taking the visual route: Graphical inspection of residuals.....	215
Using the normal distribution to identify residual sequences: The run test.....	215

Detecting autocorrelation of an AR(1) process:	
The Durbin-Watson test	219
Detecting autocorrelation of an AR(q) process:	
The Breusch-Godfrey test	221
Remedying Harmful Autocorrelation	223
Feasible generalized least squares (FGLS)	223
Serial correlation robust standard errors	225

***Part V: Discrete and Restricted Dependent Variables in Econometrics* 229**

Chapter 13: Qualitative Dependent Variables 231

Modeling Discrete Outcomes with the Linear Probability Model (LPM).....	231
Estimating LPM with OLS.....	232
Interpreting your results.....	234
Presenting the Three Main LPM Problems.....	237
Non-normality of the error term	237
Heteroskedasticity.....	238
Unbounded predicted probabilities.....	239
Specifying Appropriate Nonlinear Functions:	
The Probit and Logit Models	240
Working from the standard normal CDF: The probit model	241
Basing off of the logistic CDF: The logit model	242
Using Maximum Likelihood (ML) Estimation.....	244
Constructing the likelihood function	244
The log transformation and ML estimates	245
Interpreting Probit and Logit Estimates	246
Probit coefficients.....	249
Logit coefficients.....	251

Chapter 14: Limited Dependent Variable Models 253

The Nitty-Gritty of Limited Dependent Variables.....	253
Censored dependent variables	254
Truncated dependent variables.....	255
Modifying Regression Analysis for Limited Dependent Variables	256
Tobin's Tobit.....	256
Truncated regression	258
Oh, what the heck if I self select? Heckman's selection bias correction.....	260

Part VI: Extending the Basic Econometric Model 265

Chapter 15: Static and Dynamic Models267

Using Contemporaneous and Lagged Variables in Regression Analysis.....	267
Examining problems with dynamic models	268
Testing and correcting for autocorrelation in dynamic models...	270
Projecting Time Trends with OLS	271
Spurious correlation and time series	273
Detrending time-series data	274
Using OLS for Seasonal Adjustments	276
Estimating seasonality effects.....	276
Deseasonalizing time-series data.....	278

Chapter 16: Diving into Pooled Cross-Section Analysis 281

Adding a Dynamic Time Element to the Mix.....	281
Examining intercepts and/or slopes that change over time	282
Incorporating time dummy variables.....	283
Using Experiments to Estimate Policy Effects	
with Pooled Cross Sections	285
Benefitting from random assignment: A true experiment	286
Working with predetermined subject groups:	
A natural (or quasi) experiment.....	287

Chapter 17: Panel Econometrics291

Estimating the Uniqueness of Each Individual Unit	291
First difference (FD) transformation	294
Dummy variable (DV) regression	294
Fixed effects (FE) estimator.....	295
Increasing the Efficiency of Estimation with Random Effects.....	298
The composite error term and assumptions	
of random effects model	298
The random effects (RE) estimator	299
Testing Efficiency against Consistency with the Hausman Test	302

Part VII: The Part of Tens..... 305

Chapter 18: Ten Components of a Good Econometrics Research Project307

Introducing Your Topic and Posing	
the Primary Question of Interest	307
Discussing the Relevance and Importance of Your Topic	308

Reviewing the Existing Literature.....	308
Describing the Conceptual or Theoretical Framework	309
Explaining Your Econometric Model	310
Discussing the Estimation Method(s)	310
Providing a Detailed Description of Your Data.....	311
Constructing Tables and Graphs to Display Your Results.....	312
Interpreting the Reported Results.....	312
Summarizing What You Learned	313

Chapter 19: Ten Common Mistakes in Applied Econometrics 315

Failing to Use Your Common Sense and Knowledge of Economic Theory.....	315
Asking the Wrong Questions First.....	316
Ignoring the Work and Contributions of Others.....	316
Failing to Familiarize Yourself with the Data	316
Making It Too Complicated	317
Being Inflexible to Real-World Complications.....	318
Looking the Other Way When You See Bizarre Results	319
Obsessing over Measures of Fit and Statistical Significance	319
Forgetting about Economic Significance	320
Assuming Your Results Are Robust	320

Appendix: Statistical Tables 321

The Standard Normal Distribution.....	321
t-Distribution	322
Chi-Squared Distribution.....	323
F-Distribution	325
Durbin-Watson d-Statistic.....	325

Index 327

Introduction

My appreciation for econometrics grew out of my interest in trying to figure out how the world works. I discovered that empirical techniques tailored to specific circumstances could help explain all sorts of economic outcomes. As I came to understand how the theoretical structure of economics combines with information contained in real-world data, I began to see observed phenomena in a different light. I'd often ask myself questions about my observations. Could I determine whether the outcomes were random and simply appeared to be related? If I believed that two or more things I observed had a logical connection, could I use data to test my assertions? Increasingly, I found myself relying on the tools of econometrics to answer these types of questions.

I've written *Econometrics For Dummies* to help you get the most out of your economics education. By now, your classes have taught you some economic theory, but you're craving more precision in the predicted outcomes of those theories. Perhaps you're even questioning whether the theories are consistent with what you observe in the real world. I find that one of the most attractive characteristics of properly applied econometrics is that it's "school of thought neutral." In other words, you can adapt an econometric approach to a variety of initial assumptions and check the results for consistency. By using econometrics carefully and conscientiously, you can get the data to speak. But you better learn the language if you hope to understand what it's saying!

About This Book

Econometrics For Dummies provides you with a short and simple version of a first-semester course in econometrics. I don't cite the seminal work or anything from the large collection of econometric theory papers published in scholarly journals. The organization of topics may have some resemblance to traditional econometrics textbooks, but my goal is to present the material in a more straightforward manner. Even if you're taking a second-semester (advanced) econometrics course or a graduate course, you may find this book to be a useful, one-stop, nuts-and-bolts resource.

Of course, some technical sophistication is essential in econometrics. Besides, you've taken introductory economics, statistics, and maybe even intermediate economic theory, so now you're ready to show off your technical prowess. But wait a minute! Sometimes, with all the technical skills being mastered in learning econometrics, students fail to appreciate the insights

from the simplicity. In fact, you may even forget why you're approaching a problem with a particular technique. That's where this book can help.

Please note that I have tried to remain consistent with my terminology throughout the book, but econometricians sometimes have several different words for the same thing. Also, note that I use the statistical software STATA 12.1 throughout, but sometimes I refer to it simply as *econometrics software* or just *STATA*.

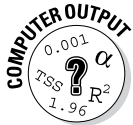
Foolish Assumptions

If you're following the normal course of action, you take an econometrics course after you complete courses on principles of microeconomics, principles of macroeconomics, and statistics. In some cases, depending on the school, you may also be required to complete intermediate economic theory courses before taking econometrics. I cover the topics in a way that accommodates some variation in preexisting knowledge, but I've had to make the following assumptions about you:

- ✔ You're a college student taking your first econometrics class taught in a traditional manner — emphasizing a combination of theoretical proofs and practical applications.
- ✔ Or you're a graduate student (or are taking an advanced undergraduate econometrics class) and would like to refresh your memory of basic econometric concepts so you can feel more comfortable with the transition into advanced material.
- ✔ You remember basic algebra, principles of economics, and statistics. I review the concepts from your statistics course that are most important for econometrics, but I also assume that a quick overview is all you need to get up to speed (and you can skip it if you're ready to dig right in).
- ✔ Numbers, equations, and Greek letters don't intimidate you. I know that on the surface using the so-called dismal science with quantitative methods isn't exactly the most attractive combination of topics. By this point in your studies, however, I'm sure you're over the fear people often have at the mere mention of these subjects.
- ✔ You'll be using some econometrics software in your class and are willing to adapt my examples in STATA to the software you're using (although chances are high you're using STATA in your class anyway).

Icons Used in This Book

Throughout the book, you may notice several different icons along the left margin. I use them to grab your attention and make the book easier to read. Each icon has an important function.



If you see this icon, it means I'm applying the techniques of a particular chapter or section with STATA. I briefly summarize the data I'm using to produce the output, show you how to format the data or create the variables required for the analysis, and point you to the most important components of the output.



I use this icon to signal that the information that follows is essential for your success in applying econometric analysis. To the extent possible, I explain these important, big-picture ideas in a nontechnical manner. However, keep in mind that this book is about econometrics, and therefore some technical sophistication may be required for even the most basic principles.



This icon appears next to information that's interesting but not essential for your understanding of the main ideas. You're welcome to skip these paragraphs, but if your econometrics class is more theory based (something that usually depends on the professor's preferences), you may need to spend more time with this material.



I use this icon to indicate shortcuts that can save you time or provide alternative ways of thinking about a concept.



This icon flags information that helps you steer away from misconceptions, common pitfalls, and inappropriate applications of a particular econometric technique.

Beyond the Book

You may not always have your e-reader or a copy of this book handy, but I'm guessing you have almost constant access to the Internet courtesy of a smartphone or tablet. That's why I include a wealth of accessible-from-anywhere additional information at www.dummies.com.

In need of some of the most useful formulas in econometrics? Looking for a breakdown of how you can give your econometric model some flexibility? Head to www.dummies.com/cheatsheet/econometrics to access this book's helpful e-Cheat Sheet, which covers these topics and more.

But that's not all. Because *econometrics* is synonymous with *forecasting* in some fields, I've put a bonus chapter online at www.dummies.com/extras/econometrics. It's all about helping you hone your forecasting skills so you can select the right method to predict an outcome based on the information you have and later vet the accuracy of your forecast.

Where to Go from Here

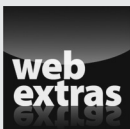
Unlike most books, you don't need to start at the beginning and read through to the end in order to gain an understanding of fundamental econometric concepts. Simply turn to the topic that most interests you. Are you struggling with the intuition or justification for a particular type of econometric model? Do you think that a specific econometric tool will help you reveal more insights from your data? You can find that topic in the table of contents or the index and then jump right to it.

Maybe you're not puzzled and are simply curious about the various tools econometrics has to offer for data analysis. Feel free to browse through the chapters. Maybe an interesting paragraph or a fascinating equation will catch your eye and give you ideas about approaching a problem — hey, it's possible!

If your statistics knowledge is rusty, I recommend you begin with the first couple chapters. On the other hand, if your experience with statistics wasn't a good one, you'd like to avoid disturbing flashbacks, and you're confident in your ability to catch on quickly, then by all means start at any other point. No matter where you start, you'll never look at data the same way after learning econometrics (for better or for worse!).

Part I

Getting Started with Econometrics



For Dummies can help you get started with lots of subjects. Visit www.dummies.com to learn more and do more with *For Dummies*.

In this part . . .

- ✓ Get familiar with the approach economists use when investigating empirical issues — not controlled experiments that never seem to contradict standard statistical assumptions.
- ✓ Find out the basic commands you need to work with data files in STATA 12.1, a popular form of econometric software, and discover the syntax structure for executing estimation commands.
- ✓ Review the probability concepts that are most relevant for your study of econometrics: topics that focus on the properties of probability distributions and their use in calculating descriptive statistics of random variables.
- ✓ Reinforce your knowledge of statistical inference so you can be better equipped to use surveys and other forms of sample data to test your hypotheses and draw conclusions.

Chapter 1

Econometrics: The Economist's Approach to Statistical Analysis

In This Chapter

- ▶ Discovering the goals of econometric analysis
 - ▶ Understanding the approach and methodology of econometrics
 - ▶ Getting familiar with econometrics software
-

Welcome to the study of econometrics! The Econometric Society, founded in 1930, defines econometrics as a field based on a “theoretical-quantitative and empirical-quantitative approach to economic problems.” This mouthful means that, at times, econometricians are mathematicians and use complex algorithms and analytical tools to derive various estimation and testing procedures. At other times, econometricians are applied economists using the tools developed by theoretical econometricians to examine economic phenomena.

In this chapter, you see that a distinguishing feature of econometrics is its development of techniques designed to deal with data that aren't derived from controlled experiments and, therefore, situations that violate many of the standard statistical assumptions. You also begin to understand that, under these circumstances, obtaining good quantitative results depends on using reliable and adequate data as well as sound economic theory.

And because computers and econometric software are now commonly used in introductory econometrics courses, I also devote a section of this chapter to introducing basic commands in STATA (version 12.1), a popular econometrics software program. This software allows you to immediately apply theoretical concepts and enhance your understanding of the material.

Evaluating Economic Relationships

Economics provides the theoretical tools you use to evaluate economic relationships and make qualitative predictions of economic phenomena using the *ceteris paribus* assumption. You may recall from your previous courses that the *ceteris paribus* assumption means that you're keeping everything else constant. Two examples among numerous possibilities are:

- ✓ In microeconomic theory, you'd expect economic profits in a competitive market to induce more firms to enter that market, *ceteris paribus*.
- ✓ In macroeconomic theory, you'd expect higher interest rates to reduce investment spending, *ceteris paribus*.



Econometrics ties into economic theory by providing the tools necessary to quantify the qualitative statements you (or others) make using theory. Unknown or assumed relationships from abstract theory can be quantified using real-world data and the techniques developed by econometricians.

The following section explains how econometrics helps characterize the future and describe economic phenomena quantitatively, and then I clarify why an econometrician must always make sensible assumptions.

Using economic theory to describe outcomes and make predictions

One of the characteristics that differentiate applied research in econometrics from other applications of statistical analysis is a theoretical structure supporting the empirical work.



Econometrics is typically used to explain how factors affect some outcome of interest or to predict future events. Regardless of the primary objective, your econometric study should be linked to an economic model. Your model should consist of an outcome of interest (dependent variable, Y) and causal factors (independent variables, X s) that are theoretically or logically connected to the outcome.

Relying on sensible assumptions

A variation of a famous joke about economists goes as follows: A physicist, a chemist, and an economist are stranded on an island with nothing to eat.

A can of soup washes ashore. The physicist says, "Let's smash the can open with a rock." The chemist says, "Let's build a fire and heat the can first." The economist says, "Let's assume that we have a can opener. . . ." Despite the joke, making assumptions about reality can help you construct logical arguments and predict outcomes when specific preexisting conditions apply. In econometrics, however, making assumptions without checking the feasibility of their reality can be dangerous.



Making too many assumptions about preexisting conditions, functional form, and statistical properties can lead to biased results and can undermine the estimation accuracy you're trying to accomplish. Although you have to make some assumptions to perform your econometric work, you should test most of them and be honest about any potential effects on your results from those you can't test.



Testing predictions from economic theory or logical reasoning is rarely a straightforward procedure. Observed data don't tend to be generated from a controlled experiment, so testing economic theory is challenging in econometric work because of the difficulty in ensuring that the *ceteris paribus* (all else constant) assumption holds. Consequently, in applying econometrics, you need to give considerable attention to the control (independent) variables you include in the analysis to simulate (as closely as possible) the *ceteris paribus* situation.

Applying Statistical Methods to Economic Problems

Most econometrics textbooks assume you've learned all the statistics necessary to begin building econometric models, estimating, and testing hypotheses. However, I've discovered that my students always appreciate a review of the statistical concepts that are most important to succeeding with econometrics. Specifically, you need to be comfortable with probability distributions, the calculation of descriptive statistics, and hypothesis tests. (If your skills are rusty in these areas, make sure you read the material in Chapters 2 and 3.)

Your ability to accurately quantify economic relationships depends not only on your econometric model-building skills but also on the quality of the data you're using for analysis and your capacity to adopt the appropriate strategies for estimating models that are likely to violate a statistical assumption. The data must be derived from a reliable collection process, but you should also be aware of any additional limitations or challenges. They may include, but aren't limited to

- ✓ **Aggregation of data:** Information that may have originated at a household, individual, or firm level is being measured at a city, county, state, or country level in your data.
- ✓ **Statistically correlated but economically irrelevant variables:** Some datasets contain an abundance of information, but many of the variables may have nothing to do with the economic question you're hoping to address.
- ✓ **Qualitative data:** Rich datasets typically include qualitative variables (geographic information, race, and so on), but this information requires special treatment before you can use it in an econometric model.
- ✓ **Classical linear regression model (CLRM) assumption failure:** The legitimacy of your econometric approach always rests on a set of statistical assumptions, but you're likely to find that at least one of these assumptions doesn't hold (meaning it isn't true for your data).



Econometricians differentiate themselves from statisticians by emphasizing violations of statistical assumptions that are often taken for granted. The most common technique for estimating an econometric model is ordinary least squares (OLS), which I cover in Chapter 5. However, as I explain in Chapters 6 and 7, a number of CLRM assumptions must hold in order for the OLS technique to provide reliable estimates. In practice, the assumptions that are most likely to fail depend on your data and specific application. (In Chapters 10, 11, and 12, you see how to identify and deal with the most common assumption violations.)

In the following sections, I describe how familiarity with certain characteristics of your data can help you build better econometric models. In particular, you should pay attention to the structure of your data, the way in which variables are measured, and how quantitative data can be complemented with qualitative information.

Recognizing the importance of data type, frequency, and aggregation

The data that you use to estimate and test your econometric model is typically classified into one of three possible types (for further details on each type, see Chapter 4):

- ✓ **Cross sectional:** This type of data consists of measurements for individual observations (persons, households, firms, counties, states, countries, or whatever) at a given point in time.

- ✓ **Time series:** This type of data consists of measurements on one or more variables (such as gross domestic product, interest rates, or unemployment rates) over time in a given space (like a specific country or state).
- ✓ **Panel or longitudinal:** This type of data consists of a time series for each cross-sectional unit in the sample. The data contains measurements for individual observations (persons, households, firms, counties, states, countries, and so on) over a period of time (days, months, quarters, or years).



The type of data you're using may influence how you estimate your econometric model. In particular, specialized techniques are usually required to deal with time-series and panel data. I cover time-series techniques in Chapter 12, and I discuss panel techniques in Chapters 16 and 17.



You can anticipate common econometric problems because certain CLRM assumption failures are more likely with particular types of data. Two typical cases of CLRM assumption failures involve heteroskedasticity (which occurs frequently in models using cross-sectional data) and autocorrelation (which tends to be present in models using time-series data). For the full scoop on heteroskedasticity and autocorrelation, turn to Chapters 11 and 12, respectively.

In addition to knowing the type of data you're working with, make sure you're always aware of the following information:

- ✓ **The level of *aggregation* used in measuring the variables:** The level of aggregation refers to the unit of analysis when information is acquired for the data. In other words, the variable measurements may originate at a lower level of aggregation (like an individual, household, or firm) or at a higher level of aggregation (like a city, county, or state).
- ✓ **The *frequency* with which the data is captured:** The frequency refers to the rate at which measurements are obtained. Time-series data may be captured at a higher frequency (like hourly, daily, or weekly) or at lower frequency (like monthly, quarterly, or yearly).



All the data in the world won't allow you to produce convincing results if the level of aggregation or frequency isn't appropriate for your problem. For example, if you're interested in determining how spending per pupil affects academic achievement, state-level data probably won't be appropriate because spending and pupil characteristics have so much variation across cities within states that your results are likely to be misleading.

Avoiding the data-mining trap

As you acquire more data-analysis tools, you may be inclined to search the data for relationships between variables. You can use your knowledge of statistics to find models that fit your data quite well. However, this practice is known as *data mining*, and you don't want to be seduced by it!



Although data mining can be useful in fields where the underlying mechanism generating the outcomes isn't important, most economists don't view this approach favorably. In econometrics, building a model that makes sense and is reproducible by others is far more important than searching for a model that has a perfect fit. I reinforce the importance of building sensible models in Chapter 4 and provide some specific examples of common economic models in Chapter 8.

Incorporating quantitative and qualitative information

Economic outcomes can be affected by both quantitative (numeric) and qualitative (non-numeric) factors. Generally, quantitative information has a straightforward application and interpretation in econometric models.

Qualitative variables are associated with characteristics that have no natural numeric representation, although your raw data may code qualitative characteristics with a numeric value. For example, a U.S. region may be coded with a 1 for West, 2 for South, 3 for Midwest, and 4 for Northeast. However, the assignment of the specific values is arbitrary and carries no special significance. In order to utilize the information contained in qualitative variables, you'll usually convert them into *dummy variables* — dichotomous variables that take on a value of 1 if a particular characteristic is present and 0 otherwise. I illustrate the use of dummy variables as independent variables in an econometric model in Chapter 9.

Sometimes the economic outcome itself is qualitative or contains restricted values. For example, your dependent variable could measure whether or not a firm fails (goes bankrupt) in a given year using various firm characteristics as independent variables. Although standard techniques are sometimes acceptable with qualitative and noncontinuous dependent variables, usually they result in assumption violations and require an alternative econometric approach. Flip to Chapters 13 and 14 to discover the appropriate techniques for situations when your dependent variable isn't continuous.

Using Econometric Software: An Introduction to STATA

Specialized software makes the application of econometric techniques possible for anyone who's not a computer programming genius. Keep in mind that several good software options are available to you and that, as a good economist, you should weigh the cost and benefits of each. Of course, the type of software you ultimately end up working with in your introductory econometrics course depends on what your professor uses for his research or finds to be the easiest to integrate into the course. I rely on STATA extensively in my academic research and use it exclusively in my econometrics courses, but your professor may employ EVIEWS, SAS, or some other program.

Because I find STATA to be the best software, it's what I use exclusively in this book. It provides an excellent combination of a user-friendly interface, consistent structure in syntax, and simple commands to implement all the techniques you learn about in econometrics, and it's available for a variety of platforms or operating systems.

STATA can be used as a *point-and-click* software (like you would use Excel or most other software these days). With point-and-click, you can use the icons and menu bar at the top to execute tasks. However, over time, you're likely to prefer using STATA as a *command-driven* program because it's faster and easier. When used in this manner, you perform tasks by providing STATA with specific syntax on the command line (using lowercase letters for the commands). In this chapter, I explain both methods, but in the later chapters, I rely almost exclusively on the command-driven approach.

The following sections show you some STATA commands that allow you to get started with the software. (Note that I introduce STATA commands as needed in other chapters.)



My coverage of STATA is not exhaustive. The supporting documentation consists of a User's Guide and several Reference manuals (thousands of pages), so clearly I can't cover every facet of STATA that you may use in econometrics. However, if you run into an obstacle, the manuals are easy to use and provide good examples. With STATA running on your computer, you also have access to the Help menu and online documentation.

Getting acquainted with STATA

In this section of the chapter, I show you how to open data files, log your modifications to data, and save your data files.

Creating and saving STATA datasets

In order to begin doing any exploratory data analysis or econometric work, you need a dataset that's in STATA format (*.dta). If you're downloading data from an online source, you may be able to obtain the data in STATA format. Many econometrics textbooks also give you access to data files in STATA format. In addition, the STATA program is preloaded with examples that you can use to familiarize yourself with the basic commands.



After opening STATA, you can access the sample datasets by selecting **File** ⇨ **Example Datasets...** If you want to open any other dataset that's already in STATA format, select **File** ⇨ **Open** and then choose the file you want to work with. On the command line, you can open a STATA dataset by typing "use *filename*" and hitting return.

If you're inputting data manually or downloading it in a non-STATA format, then you can use one of two methods to read it into STATA:

- ✓ Select **File** ⇨ **Import**: This option can be used if the data is in Excel, SAS XPORT, or Text format. You select the appropriate format of your raw data, and then you're prompted to select the file you'd like to import into STATA.
- ✓ Select **Data** ⇨ **Data Editor**: This option opens an editor that resembles a spreadsheet. You can paste columns of data into the editor or input data manually.



If you import a dataset that wasn't originally in STATA format, you need to save the dataset in STATA format in order to use it again, particularly if you inputted data through the editor and want to avoid replicating all your efforts. Also, if you made any changes to an existing STATA dataset and want to retain those changes, you need to save the revised dataset. I recommend you select **File** ⇨ **Save As** (or type "save *new filename*" on the command line) and choose a new name for the modified file. That way if you accidentally delete a variable or drop observations, you can always go back to the original data file.

Viewing data

Before you begin doing econometric analysis, make sure you're familiar with your data. After all, you don't want to estimate an econometric model with data that's mostly incomplete or full of errors.

In version 12.1 of STATA, the default setting allows you to open a dataset as large as 64 megabytes (MB) and containing up to 5,000 variables. If your dataset is larger than 64MB, you need to increase the memory allocated to STATA by typing “set memory #m” on the command line, where # is the size of your dataset in MB. Similarly, if your dataset contains more than 5,000 variables, you need to type “set maxvar #” on the command line, with # being the number of variables in your dataset.

The **Data** tab in the menu bar contains most of the elements you need in order to get acquainted with your data. After opening a STATA dataset, you'll regularly use the following commands:

- ✓ Select **Data** ⇨ **Describe data** ⇨ **Describe data in memory** or type “describe” on the command line and hit return: STATA shows you how many observations and variables are contained in the dataset. In addition, it lists the names and types (numeric or string) of all the variables.
- ✓ Select **Data** ⇨ **Describe data** ⇨ **Summary statistics** or type “summarize” on the command line and hit return: With this command, STATA provides you with basic descriptive measures for all the numeric variables in your dataset. Specifically, you get the number of observations with nonmissing values, mean, standard deviation, minimum value, and maximum value for each variable. **Note:** The string variables contain letters, names, or phrases, so no mean or standard deviation can be calculated for them.

In Figure 1-1, I use the “describe” and “summarize” commands to view the fundamental characteristics of my dataset.



The **Data** tab or “describe” and “summarize” commands provide the basic information you use for your econometric analysis. Examine the tables containing the descriptive information and make sure that all the values are sensible. In other words, make sure that the minimum, maximum, and mean values are feasible for each variable in your dataset.



You can also use the “list” command on occasion, but be careful with it because it displays the value for every variable and every observation. In other words, it displays the entire dataset. With a large dataset (thousands of observations and dozens of variables), this list isn't likely to help you find errors unless you refine the list to a specific observation using an “if” statement or by subscribing (I discuss this in the later “Creating new variables” section).

```
. describe
```

Contains data from `/Applications/Stata/ado/base/c/census.dta`

obs: 50 1990 Census data by state
vars: 13 6 Apr 2011 15:43
size: 2,900

variable name	storage type	display format	value label	variable label
state	str14	%-14s		State
state2	str2	%-2s		Two-letter state abbreviation
region	int	%-8.0g	cenreg	Census region
pop	long	%12.0gc		Population
poplt5	long	%12.0gc		Pop. < 5 year
pop5_17	long	%12.0gc		Pop. 5 to 17 years
pop18p	long	%12.0gc		Pop. 18 and older
pop65p	long	%12.0gc		Pop. 65 and older
popurban	long	%12.0gc		Urban population
medage	float	%9.2f		Median age
death	long	%12.0gc		Number of deaths
marriage	long	%12.0gc		Number of marriages
divorce	long	%12.0gc		Number of divorces

Sorted by:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
state	0				
state2	0				
region	50	2.66	1.061574	1	4
pop	50	4518149	4715038	401851	2.37e+07
poplt5	50	326277.8	331585.1	35998	1708400
pop5_17	50	945951.6	959372.8	91796	4680558
pop18p	50	3245920	3430531	271106	1.73e+07
pop65p	50	509502.8	538932.4	11547	2414250
popurban	50	3328253	4090178	172735	2.16e+07
medage	50	29.54	1.693445	24.2	34.7
death	50	39474.26	41742.35	1684	186428
marriage	50	47781.4	45130.42	4437	218864
divorce	50	23679.44	25094.01	2142	133541

Figure 1-1:
Examining
data two
ways in
STATA.

Keep in mind that the results section of STATA, by default, displays approximately one page of output. STATA then prompts you with the “-more-” message. Hitting the return key allows you to see an additional line of output, and hitting the spacebar shows another page of output. If you don’t want STATA to pause for “-more-” messages, type “set more off” on the command line. Subsequent output is then displayed in its entirety.

Interpreting error messages

If you make a mistake with a command, STATA responds with an error message and code. The error message contains a brief description of the mistake, and the code has the format `r(#)`, where `#` represents some number. Reading the error message and carefully examining the command that resulted in the error usually helps you arrive at a solution. If not, the codes, known as a return codes, are stored in STATA, and clicking on the code allows you to obtain a more detailed description of the error.



The outcome of a command can be identified quickly by looking at the colors of the text in the results area (the middle portion of STATA’s interface). If you see the color red, it means something has gone wrong and you should correct your mistake before moving on.

Stopping STATA

When you occasionally want to terminate a process in STATA, you can just click the **Break** button on the toolbar (right below the menu bar). Stopping STATA may be appropriate if an estimation procedure doesn't converge to a result or you change your mind about the command you'd like to execute and don't want to wait until the process is complete. After you stop STATA, your data remains in memory, and you can continue with any command.

In Figure 1-2, I use the "list" command to see each observation in the dataset. However, after I see a few of the observations, I decide that I don't need to see more observations one by one. I click the **Break** button to stop the command.

Preserving your work

Saving your commands and resulting output in a log file is one of the most essential things you can get into the habit of doing while using STATA. You can do it by selecting **File ⇨ Log ⇨ Begin...** from the menu bar and then assigning the file a name or by typing "log using filename" on the command line and hitting return. After you complete the work you want to save, select **File ⇨ Log ⇨ Close** or type "log close" on the command line and hit return. Your log files are given a .smcl file extension.



```
. list
```

1.	state	state2	region	pop	poplt5	pop5_17	pop18p	pop65p	popurban	medage
	Alabama	AL	South	3,893,888	296,412	865,836	2,731,640	440,015	2,337,713	29.30
death				marriage				divorce		
35,305				49,018				26,745		

2.	state	state2	region	pop	poplt5	pop5_17	pop18p	pop65p	popurban	medage
	Alaska	AK	West	401,851	38,949	91,796	271,106	11,547	258,567	26.10
death				marriage				divorce		
1,604				5,361				3,517		

3.	state	state2	region	pop	poplt5	pop5_17	pop18p	pop65p	popurban	medage
	Arizona	AZ	West	2,718,215	213,883	577,604	1,926,728	307,362	2,278,728	29.20
death				marriage				divorce		
21,226				30,223				19,908		

Figure 1-2:
The break
action in
STATA.

```
4. —Break—
r(1);
```

In Figure 1-3, I open a log file, execute a “summarize” command, and close the log file. I can examine the contents of the log file by selecting **File** ⇨ **View...** from the menu bar and then choosing my log file.

```
. log using "/Research/Econometrics for Dummies/ExampleData/Chapter1.smcl"
```

```
name: <unnamed>
log: /Research/Econometrics for Dummies/ExampleData/Chapter1.smcl
log type: smcl
opened on: 29 Dec 2012, 19:04:55
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
state	0				
state2	0				
region	50	2.66	1.061574	1	4
pop	50	45181.49	4715038	401851	2.37e+07
poplt5	50	326277.8	331585.1	35998	1708400
pop5_17	50	945951.6	959372.8	91796	4680558
pop18p	50	3245920	3430531	271106	1.73e+07
pop65p	50	509502.8	538932.4	11547	2414250
popurban	50	3328253	4090178	172735	2.16e+07
medage	50	29.54	1.693445	24.2	34.7
death	50	39474.26	41742.35	1604	186428
marriage	50	47701.4	45130.42	4437	210864
divorce	50	23679.44	25094.01	2142	133541

Figure 1-3:
Saving
log files in
STATA.

```
. log close
name: <unnamed>
log: /Research/Econometrics for Dummies/ExampleData/Chapter1.smcl
log type: smcl
closed on: 29 Dec 2012, 19:05:21
```

Using STATA’s viewer, you can always go back to your log file to see how you modified the data or any statistical estimates you may have previously calculated. You can also copy and paste from your log file to any other file, or you can simply print your log file.



Don’t forget to close your log file when you’re done with the work you want to retain. Otherwise, everything you do in STATA continues to be written to the log file you opened, which may create an unnecessarily huge file.

Creating new variables

After you compile your data, you’ll likely want to create new variables for the analysis. Your econometric model may specify that a variable should be measured in logs, or you may need to use a squared term for a quadratic function