Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Book Review: *Microeconometrics: Methods and Applications* and *Microeconometrics Using Stata*

Patrick Bajari & Thomas Youle

To cite this article: Patrick Bajari & Thomas Youle (2012) Book Review: *Microeconometrics: Methods and Applications* and *Microeconometrics Using Stata* , Econometric Reviews, 31:1, 107-117, DOI: 10.1080/07474938.2011.607090

To link to this article: https://doi.org/10.1080/07474938.2011.607090

Published online: 13 Oct 2011.

Econometric Reviews, 31(1):107–117, 2012

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2011.607090

Taylor & Francis
Taylor & Francis Group

BOOK REVIEW: *MICROECONOMETRICS: METHODS AND APPLICATIONS* AND *MICROECONOMETRICS USING STATA*

Patrick Bajari and Thomas Youle

Department of Economics, University of Minnesota, Minneapolis, Minnesota, USA

Cameron, Colin A., Trivedi, Pravin K. (2005) *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.
Cameron, Colin A., Trivedi, Pravin K. (2009) *Microeconometrics using Stata*. Texas: Stata Press.

1. OVERVIEW

Microeconometrics: Methods and Applications (MMA for short) by A. Colin Cameron and Pravin K. Trivedi is an in-depth, textbook-style treatment of techniques that are commonly used in applied microeconomics. The companion text *Microeconometrics Using Stata* (MUS for short) by the same authors shows how to apply these techniques in the powerful statistics software package Stata. Both texts are appropriate for Ph.D. students already familiar with the first few chapters of an introductory text such as Greene (2008), Hayashi (2000), or Ruud (2000).

Both books are organized around econometric topics. MMA has a broader coverage and depth of material than MUS, which instead focuses on having readers get their hands dirty with real data sets on the computer. Both provide useful discussions for applied economists. Students will also learn about different data types and how to load and manipulate them in Stata. This sort of practical knowledge is very useful for Ph.D. students making the transition to applied research.

Two features make MMA and MUS useful additions to the applied microeconomist's bookshelf. First, they have a broader coverage of topics and are more current than many other available texts. After a first-year econometrics course, Ph.D. students are often frustrated when

Address correspondence to Patrick Bajari, Department of Economics, University of Minnesota, 7031 A 1925 4th St. S, Minneapolis, MN 55455, USA; E-mail: bajari@umn.edu

they attempt to read journal articles or follow seminar presentations in applied microeconomics. Many of the methods commonly used by leading practitioners are omitted in their first-year texts. The broad and up-to-date coverage goes a long way towards filling in these gaps. These books are also useful for practitioners who wish to quickly get up to speed on particular methods in order to read recent research.

Second, a standard first-year text frequently omits material that cannot be formally and completely developed given page constraints. Many widely used methods are left out of standard texts as a result. By comparison, Cameron and Trivedi succinctly summarize widely used methods even if they are too advanced to formally develop in the text. As a result, Ph.D. students at least have a frame of reference for topics that they are likely to encounter in their lives after graduate school. Furthermore, the text provides detailed discussions of sticky implementation issues that are sometimes hard to formalize, but that nevertheless are likely to be encountered in practice.

A good example of this difference in style is their coverage of instrumental-variables (IV) estimators. While Cameron and Trivedi discuss the standard theory of IV included in many first-year texts, they also discuss a number of additional topics. The authors have a detailed discussion of the choice of instruments in estimating the returns to education. The choice of instruments is seldom without a bit of controversy in applied work. Many first-year texts do not have detailed discussions of the difficulties that arise when trying to find a good instrument. As a result, students may be unprepared for the reaction they will face when they first begin to use IV in their own applications. Cameron and Trivedi also discuss the advantages and disadvantages of different IV estimators, such as the Jackknife IV and Limited Information ML. While the full comparison of these estimators requires advanced theory, the authors show they are easy to implement and compare in Stata. The relevant literature is cited for those who wish to investigate the formal econometric theory.

Cameron and Trivedi also discuss certain theoretical pitfalls in applying IV, such as the weak instruments problem. Most first-year texts omit this topic since the relevant econometric theory is too advanced. As a result, students may be puzzled when they are asked to report their first-stage F-statistic when presenting their own research. Cameron and Trivedi, by comparison, provide an intuitive explanation of the weak instruments problem, discuss several alternative diagnostic tests for weak instruments, and then show how they can be implemented in Stata.

Cameron and Trivedi provide a detailed discussion of research by Kling (2001) on estimating the returns to schooling. They compare the IV to the Ordinary Least Squares (OLS) estimates and discuss weak instruments in the context of this detailed application. This additional material teaches

students how to critically read papers and helps them in preparing their own papers for submission to peer-reviewed journals.

Unfortunately, such references to the applied literature are casual and scarce. Examples tend to be chosen in order to clearly communicate the econometric properties of a method, but they seldom help familiarize students with a major applied literature. This is not a specific criticism of Cameron and Trivedi or of any particular first-year textbook. Instead, this suggests the need for a supplementary text which focuses on applications, which we will describe in the next section. The bottom line is that Cameron and Trivedi have provided an extremely valuable service to the profession by producing such detailed and comprehensive books.

2. SOME LIMITATIONS OF AVAILABLE ECONOMETRIC TEXTS

As an instructor, we have found that there are two important gaps in the available graduate econometrics texts. First, the texts tend to underexpose students to substantive applications that occur in the major empirical literatures. Second, they draw few links between what students learn in their econometrics courses and what they learn in their economic courses. While a broad awareness of methods common in microeconometrics is a necessary component in the training of an applied microeconomist, it is not sufficient. Students must be able to think critically about both the economic and econometric issues which occur in applied work. Without this training, students find it difficult to make the transition to writing substantive empirical applications.

An effective way to teach students how to do research is to expose them to many different empirical literatures. Students are then shown how econometrics can be used to attack a diverse array of problems in different subfields of economics. The classic Berndt (1996) follows such an approach. Each chapter is organized around a large applied literature such as those studying the capital asset-pricing model (CAPM), costs and learning curves, and the demand for electricity. In each chapter the relevant economic theory is discussed along with empirical facts, econometrics, and references to important papers.

In a chapter on wage regressions, Berndt (1996) cites 164 papers (108 within 15 years of its publication) in the course of discussing human capital theory, signaling theory, econometrics, and recent research. In developing the wage regressions motivated by theory, Berdnt discusses the econometric issues of specifying a functional form, adding dummy variables for gender, and trying to control for the omitted variable bias resulting from unobserved ability. Students see that a broad understanding of economics is indeed useful in conducting applied research.

Both professors and Ph.D. students would benefit from a 21st century equivalent of Berndt (1996) to supplement the primary text in an

econometrics course. Ideally, each chapter would focus on a major empirical literature such as hedonic home-price regressions applied to estimating the value of environmental amenities, the estimation of production functions and productivity, or differentiated product demand applied to measuring market power. A given chapter would also contain one or two data sets used in prominent papers written by a leading researcher within the past 15 years.

Finally, the text should have detailed problem sets with applications that force students to apply their econometric theory in Stata. Such a supplementary text would be invaluable in teaching students how economic theory and econometrics can be used together to explore applied problems.

These comments are not criticisms of Cameron and Trivedi's excellent work. There is no reason to expand the scope of their texts since each organized around a self-contained theme. However, a text that had substantive empirical applications that cover major empirical literatures would be invaluable as a supplement to standard econometrics texts.

3. HIGHLIGHTS OF MMA

Here we discuss some highlights of MMA. When detailed chapter outlines are available on the internet, there is no need to simply repeat the table of contents.

3.1. Part 1: Preliminaries

Although some of these topics are also covered in Wooldridge (2002), this section of MMA is unique to microeconometric textbooks because it provides an overview of how microeconometrics fits into applied microeconomics and how microeconometrics differs from other areas of statistics. The discussion in Chapter 2 on causal and noncausal models provides a brief history of microeconometrics, and a unique introduction to the issues of causality, structural relationships, and identification. The discussion in Chapter 3 about what can be learned from the types of data sets available to microeconometricians is especially informative and differs most substantially from what is available in other textbooks.

3.2. Part 2: Core Methods

This section covers the core econometric theory that is the centerpiece of first-year texts such as Greene (2008), Hayashi (2000), or Ruud (2000). However, as discussed above, MMA has a broader coverage of topics which includes more advanced methods and a more complete discussion of

sticky implementation issues. The authors cover all the standard topics and provide introductions to all the standard formal results.

Below, we mention some discussions that make this text special. It should be understood that all the standard topics and results are covered. In general, MMA is special because of its broad scope and intuitive discussions of more advanced topics.

As discussed above, the authors discuss various problems encountered in practice with IV, such as the difficulty of finding good instruments and the possibility of weak instruments. The discussions of diagnostics for detecting weak instruments and the bias in the presence of weak instruments are concrete and helpful.

The chapter on maximum likelihood and nonlinear least squares stands out by including more about estimating-equations estimators, the analogy principle, and estimators for models of the linear exponential family of densities. The treatment here provides more concrete guidance for applied work than the more theoretical treatments.

Chapter 6, on generalized method of moments (GMM) estimators, covers optimal instruments and optimal moment conditions. The treatments of empirical likelihood and estimators based on moment conditions with nonadditive errors provide the student with an intuitive discussion of frontier methods.

Chapter 7, on hypothesis testing, includes more detailed discussions of size and power than standard treatments. The authors perform Monte Carlo exercises to show in practice the distinction between asymptotic and actual size and power and show how to implement the Wald test using the bootstrap. Chapter 8, on specification tests and model selection, discusses the power of the Hausman test, pretest estimation, and data mining. The chapter closes with an insightful discussion about the role of specification testing in practice.

In Chapter 9, the authors employ their signature level of scope and detail to semiparametric methods and they produce a unique discussion among the mainstream textbooks. While a detailed discussion of the theory of these topics is not presented, students are at least introduced to important terminology and key concepts from the theory.

Chapter 10, on numerical optimization, covers some advanced methods, such as the Expectation Maximization (EM) algorithm and simulated annealing, and it provides some useful suggestions for checking code reliability.

3.3. Part 3: Simulation-Based Methods

The application of computationally intensive techniques that exploit improvements in computer hardware and software is one of the most important advances in applied microeconomics in the past two decades.

Microeconometrics has three chapters devoted to these methods. This concise summary of recent developments is very valuable for Ph.D. students and practitioners.

Chapter 11 discusses bootstrap methods. This chapter has a self-contained description of the bootstrap and a sketch of the relevant econometric theory including the consistency of the bootstrap, Edgeworth expansions, and asymptotic refinements. Uses of the bootstrap in bias reduction, computing standard errors, hypothesis testing, confidence intervals, and other topics are covered. Simulation examples are provided. Extensions to the bootstrap, such as subsampling, the block bootstrap, the nested bootstrap, recentering, and the jackknife are discussed. Applications of the bootstrap to heteroskedastic errors, panel and clustered data, and overidentified GMM models, nonsmooth estimators, and time series are covered. The chapter closes with a discussion of some barriers that can preclude the use of the bootstrap in practice.

Chapter 12 covers simulation-based estimation. After motivating these techniques, the authors discuss methods for computing integrals, including quadrature and Monte Carlo methods. The chapter then describes the mechanics of setting up maximum-simulated-likelihood (MSL) and method-of-simulated-moments (MSM) estimators. After discussing key theorems regarding consistency and asymptotic normality, the chapter provides a helpful comparison between MSL and MSM. The chapter also touches on indirect inference, importance sampling, variance reduction, and quasi-random numbers. The chapter closes with a detailed discussion of different methods for drawing random variables.

Chapter 13 covers Bayesian methods. The chapter includes an overview of some key elements of Bayesian statistics. Bayesian methods have become increasingly common in both statistics and econometrics because of their computational advantages in certain problems. The chapter also covers Gibbs sampling, data augmentation, and the Metropolis–Hastings algorithm. It is obviously difficult to adequately summarize the recent advances in Bayesian methods that have occurred in the past two decades. However, the chapter at least introduces Ph.D. students to many important concepts and illustrates how to construct the simulators for a nontrivial simultaneous equations model.

3.4. Part 4: Models for Cross-Section Data

As the introduction to the text emphasizes, the dependent variable in many applied studies is discrete, integer valued, or censored. The data may also come from a selected sample. This part covers methods used to analyze nonlinear, limited-dependent variable models. In addition to the standard topics, many semiparametric estimators are nicely treated. The authors provide a helpful discussion of the identification of selection

models using exclusion restrictions. The coverage of duration analysis is quite extensive compared to standard textbooks.

3.5. Part 5: Models for Panel Data

This section covers the standard theory and recent research on estimators for the parameters of linear and nonlinear panel-data models. Much of this material is now standard and covered by Greene (2008) and Wooldridge (2002). When discussing dynamic models, the authors have a useful discussion on the distinction between true state dependence and unobserved heterogeneity.

3.6. Part 6: Further Topics

Chapter 24 covers stratified and clustered samples. In practice, survey data sets are seldom based on random samples of the population. This chapter covers weighting schemes and the problem of endogenous stratification. In addition, techniques for clustering standard errors, such as cluster-robust standard errors, are presented. Different models for clustered data, diagnostics for clustering, and hierarchical linear models are also covered.

Chapter 25 covers treatment evaluation. This topic is not covered in standard introductory econometrics textbooks and is an important addition given the wide use of these methods. This chapter discusses commonly used estimators such as matching, propensity-score methods, control-function estimators, regression-discontinuity-design, and difference-in-difference estimation. The chapter contains a fairly detailed discussion of the identification assumptions required for the alternative estimators. The different estimators of treatment effects are carefully compared in an example of the effect of training on earnings.

Chapter 26 covers the important topic of measurement error. This chapter starts with a discussion of the errors-in-variables model and a derivation of the biases from measurement error in regression and linear panel-data models. The authors discuss potential strategies for correcting for measurement error including IV methods and replicated data. Finally, measurement error in a number of different nonlinear models, such as discrete choice or count regression, is discussed.

4. SUMMARY OF CONTENTS OF MUS

Cameron and Trivedi's MUS is a valuable companion to MMA and an excellent introduction to Stata in its own right. MUS has several key advantages over more traditional Stata manuals. The first and striking advantage is the comprehensive coverage of topics, with the authors

discussing a very wide array of estimators, tests, diagnostics, and visual tools. The material is naturally organized around the concerns of an applied microeconomist. For example, in the case of IV methods, the authors describe a wide number of estimators and their relative merits, give guidance on interpreting their output, and then introduce a number of recent tests for instrument weakness and validity.

A second advantage is the coverage of topics that are increasingly important in the applied microeconomics literature, but are not typically covered in Stata textbooks. The chapters on simulation methods, nonlinear optimization, and the bootstrap are especially valuable in reducing the student's barriers to entry in using these important methods. Simulation methods are used regularly throughout the book, encouraging students to run Monte Carlo exercises to investigate the many estimators and tests discussed in the text.

Finally, as the structure of MUS resembles MMA, students can refer to the other text for more econometric details. MUS is also incredibly clear, with the authors nearly always showing relevant output for the methods they consider. They also cover many recent and valuable user-written Stata commands that may be omitted in standard reference manuals, and supply datasets on their Web site and helpful problems to get students quickly up and running in Stata.

The first chapter covers a basic introduction to Stata. In addition to presenting the standard material, the authors provide practical advice for conducting an empirical investigation in Stata. They suggest a three-step process of writing do files, running them, and examining the resulting log file. They provide useful tips, such as to take user-generated commands from the *Stata Journal* as they are more thoroughly vetted.

Chapter 2 discusses data management and graphs. The authors show how to include a number of nonparametric methods, such as kernel regression, in graphs and scatterplots. As nonparametric methods are becoming increasingly common, yet are seldom included in econometrics or Stata texts, this is a helpful addition. It is unfortunate that these methods are not given a lengthier discussion.

Chapter 3 describes the standard linear regression. Data on the U.S. Medicare program is used throughout the chapter while the authors show how to describe the data, construct table of summary statistics, and perform specification analysis in addition to running regressions. The attention given to what should be done before and after linear regressions is very valuable. The authors close the chapter by showing how to use sampling weights, common in most survey datasets.

Cameron and Trivedi cover simulation methods in Chapter 4, reflecting the growing importance of simulation methods in microeconometrics. This section is a key advantage of their text over others, which rarely cover simulation methods in detail. Simulation is also

an important pedagogical tool, giving students the ability to simulate their own data and run Monte Carlo exercises on the estimators discussed in the book. The authors show just how widely these tools can be used, in simulating estimators, computing integrals, and in estimating the size and power of hypothesis tests in finite samples.

Chapter 5 gives a detailed summary of Generalized Least Squares, and a helpful discussion of heteroskedastic, clustered, and correlated errors that often arise in practice. The authors also provide a stand-alone discussion of using survey data in Stata, using a complex, geographically stratified health survey as an example.

Chapter 6 focuses on IV methods. The authors begin by developing the classical theory of IV estimators. The coverage is wider than most other books, including the Jackknife-IV estimator, and it includes a helpful comparison of the estimators. The main advantage of this chapter over most other texts lies in the authors' extensive coverage of weak instruments, a situation of considerable importance in applied work. They discuss the practical advantages and disadvantages of four different methods of testing for weak instruments in the context of real-world datasets. They also show the sensitivity of coefficients to the choice of instruments when the model is just identified, which is another important lesson for students.

Chapter 7 examines quantile regression, both in theory and in Stata. The treatment here is more standard, similar to Wooldridge (2002), but also contains more estimators and tests, and shows how they can all be easily run in Stata. The authors also take advantage of their earlier simulation chapter to simulate examples and offer several helpful ways to plot estimates from quantile regressions.

Chapters 8 and 9 introduce linear panel-data models. This is a topic of considerable importance in practice, because many applications use panel data. Once again Cameron and Trivedi do a commendable job of discussing an unusually large number of estimators and their relative advantages and disadvantages. They also provide helpful discussions about interpreting the Stata output, interpreting available diagnostic methods, and checking the degree to which a panel dataset is unbalanced. Their discussions of helpful visual tools are another advantage over other texts that describe formal properties of tests and estimators and give less guidance for how to explore the processes that generated their data.

Chapter 10 covers the nonlinear regression methods that are available in Stata. The initial treatment is standard, although once again including a wide array of estimators, but the main value of this chapter lies in its discussion of predicted values and marginal effects. This discussion helps students decide what statistics to report.

There are many nonlinear models that are not prepackaged in Stata, however, and a general treatment requires the optimization

of a user-specified criterion function. To the Stata novice, nonlinear optimization can seem a very formidable task. Cameron and Trivedi do a superb job in lowering the barriers to entry in their Chapter 11 treatment of nonlinear optimization. They cover traditional maximization algorithms and common pitfalls, as well as more advanced debugging advice and more obscure worries. They provide examples and describe the basic maximum likelihood and more advanced optimizers that come prepackaged in Stata.

Chapter 12 concerns testing, and the authors again make good use of simulation. The authors cover the important, but again seldom discussed, bootstrap and other simulation-based methods in Chapter 13. The authors show how these methods can be used to compute a wide array of other statistics, such as estimator bias. This chapter equips the student with a very valuable set of tools.

Chapters 14–16 cover binary-outcome, multinomial-outcome, and selection models. Once again, these chapters stand out for their the breadth and for the useful discussions of marginal effects and model diagnostics. The authors also investigate regressor endogeneity in binary-outcome models, a common situation in applied microeconomics, and structural methods to deal with this problem. They miss an opportunity to discuss control-function approaches, which is an increasingly popular method of handling endogeneity in a large class of nonlinear models.

The authors also include a rich and detailed account of count-data models, including nonlinear IV methods developed using the optimization tools previously discussed. In the final chapter, the authors examine nonlinear models for panel data, with an extended treatment of count-data models.

The authors also include helpful appendices on programming and Stata's powerful matrix language Mata.

5. CONCLUSION

Cameron and Trivedi have written two excellent books which help fill the gap between what students are taught in their econometrics courses and how these ideas are implemented in practice. The broad coverage of material helps expose graduate students to methods they will surely encounter after graduate school. The emphasis on issues of practical importance to applied economists helps students make the transition to research. Clear discussions of prepackaged and general methods in Stata help get the student up and running, both in implementing these methods and in discovering the processes that generate their data.

We have outlined the need for an additional text for training graduate students, a modern version of the Berndt (1996) text, which also introduces students to large swathes of the current applied literature. Students would learn from seeing how difficulties in applying econometric

methods are overcome in practice. This additional text, along with the excellent books by Cameron and Trivedi, would be invaluable for students making the transition to conducting applied research.

REFERENCES

Berndt, E. R. (1996). *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: Addison-Wesley.

Greene, W. H. (2008). *Econometric Analysis*. Upper Saddle River, NJ: Pearson.

Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.

Kling, J. R. (2001). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business and Economic Statistics* 19:358–364.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.

Wooldridge, J. (2002). *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: MIT Press.