

Review of An Introduction to Modern Econometrics Using Stata by Baum

Austin Nichols
Urban Institute
Washington, DC
anichols@ui.urban.org

Abstract. This article reviews *An Introduction to Modern Econometrics Using Stata* by Christopher F. Baum.

Keywords: gn0036, econometrics, textbook, ivreg

1 Introduction

An Introduction to Modern Econometrics Using Stata, hereafter IMEUS, by Christopher F. Baum, is useful and informative, somewhat difficult in parts, and unusually correct and up to date. Econometrics is often defined as the set of statistical methods of interest to economists, which does little to restrict its scope. IMEUS begins from first principles of data management, or nearly so, and hence to a more experienced Stata user may often read more like an introduction to Stata by using modern econometrics. Both topics—Stata and econometrics—are of somewhat narrow interest, perhaps, and dense and difficult to navigate for the uninitiated, but the book manages to make its topics accessible. Full of good advice and worked examples, IMEUS assumes no prior Stata knowledge, and it focuses on the practice of econometrics rather than the theory. For most purposes, it has neither the breadth nor the depth to stand alone as an econometrics textbook. It would, however, make a useful companion to a more densely written text in a graduate course, and I fully expect to be assigning it as a required text soon.

2 Summary

The preface aims the book at applied researchers in economics and finance who wish to learn basic econometrics and use Stata, and it assumes prior exposure to an advanced undergraduate econometrics course on the level of Wooldridge (2006). The stated target audience probably already knows much of the material in the book, although having it collected is handy. A more likely audience is those who would like to become applied researchers—many are in the first year of graduate studies in economics or a related discipline. These folks have taken a class in probability and statistics, and a second on regression methods, and perhaps a class that used the phrase “maximum likelihood”, but these students have not really ever had to design an empirical research project, nor have they hammered data into the shape that would be useful in fitting an appropriate model.

Fittingly, the first 67 pages cover data management and the use of common commands and functions. The next 150 pages cover linear cross-sectional regression, diagnostics, robust standard errors, and instrumental-variables models. Chapter 9 offers a brief treatment of panel-data models, namely, `xtreg`, `xttest0`, `hausman`, `xthtaylor`, `xtivreg`, `xtabond`, `xtabond2`, `sureg`, and `rollreg`. Chapter 10 covers several models of discrete and limited dependent variables. The end of each chapter offers a handful of exercises, and two appendices cover infiling and programming in Stata.

3 Strengths

The emphasis throughout is on practicality, with worked examples using downloadable data, no theorems or proofs, and only a little matrix algebra. This approach makes the book readable and allows room for discussion about the interpretation of models and the vagaries of real data—a welcome change from the typical introduction to econometrics.

This approach is at its best in the central chapters, the 150 pages encompassing chapter 4 (“Linear regression”), chapter 5 (“Specifying the functional form”), chapter 6 (“Regression with non-i.i.d. errors”), chapter 7 (“Regression with indicator variables”), and chapter 8 (“Instrumental-variables estimators”). Nearly half of that space is devoted to Stata output and direct interpretation thereof.

The treatment reflects the common concern in applied work by economists and other social scientists about endogeneity or correlation of the error term with explanatory variables (pervasive because of the use of observational data instead of experimental data) leading to biased and inconsistent coefficient estimates. Of particular value is the treatment of instrumental-variables estimators in chapter 8, in some sense continuing a thread begun in Baum, Schaffer, and Stillman (2003). The unified treatment and progression from linear regression, through corrections to standard errors, and into instrumental variables, makes this an exceptionally useful textbook treatment.

Chapters 5 and 7 cover specification and issues in interpretation in clear prose and with helpful examples, so they should be particularly helpful to readers who have not had much experience running regressions in Stata or interpreting the output.

The first three chapters have a wide overlap with the *Stata User’s Guide* but offer a quick and easy entry into thinking like a Stata user. These chapters make the book useful as an introduction aimed at beginners, although the technical style of writing marks it as a graduate text.

4 Weaknesses

There are a variety of minor typographical errors, most of which will not lead anyone astray but do give a reader pause. Equation 6.6, for example, has a footnote averring “there is no error in (6.6)” but confuses the matrix and vector notation. Presumably, the ardent pedants among us can use the errata page on the Stata Press web site as a bookmark.

Some errors may be open to interpretation. In chapter 2, the section on “Handling missing data” claims that any function of missing data produces missing data (but generate a running sum by using the `sum()` function or use the `max()` function to see counterexamples) and does not mention methods of imputation or related estimation techniques. Although a book’s space limitations make some hard choices necessary, some mention of imputation, and perhaps reference to Allison (2002) or Royston (2005), is in order in the discussion about missing data.

There are several places where the text could be expanded. There is little information in IMEUS on models for qualitative dependent variables, and I found questionable advice on computing marginal effects in those models. These problems could benefit from adding references to documents on the Stata web site and to Long and Freese (2006).

5 Possible future strengths

Any book reflects the interest of the author, and IMEUS is no exception. What some readers see as omission, others will regard as desirable parsimony. Nevertheless, some marginal improvements on the existing text could make it stronger and might be feasible. It is easy to imagine a web site containing not only an up-to-date list of errata but also additions or revisions to the text under consideration for later editions, available as preprints. Some of these could be written by Baum, and some could be submitted by interested readers, and the collection could assume the shape of a moderated Wikipedia site. Many of the additions or revisions might be on the order of one paragraph, or the size of a sticky note, that could be stuck into a copy of the first edition long before the advent of a second.

Possible additions might include more talk of exploratory analysis, including the kinds of graphs that would be helpful in model building, although chapter 5 shows `graph matrix` and `avplots`. A more thorough treatment of `twoway`, or even just enough discussion to enable the uninitiated to produce figure 4.1 (a scatterplot with an overlaid linear fit), and perhaps a list of required reading in the collection of help files available to all Stata users, seem warranted. Some mention of `kdensity` and `lowess` or `locpoly` also seems called for. The discussion of exploratory graphics could be short, given the existing reference to Mitchell (2004).

The focus in IMEUS is on hypothesis testing, but many readers could benefit from more discussion of two-way tabulation or confidence intervals. One short section at least should address survey estimators—one wonders how a budding applied econometrician would test the equality of proportions by using survey data without having such obscure tricks as

```
. svy: tab y x, row se
. test _b[p12] = _b[p22]
```

explained to them.

Someone teaching beginners might prefer some explication of common pitfalls in writing Stata code, as well as methods of checking errors in Stata code, including such niceties as `set trace on`. For example, when the `scalar` command is introduced, Baum might mention that a scalar whose name can be interpreted as an abbreviated variable name cannot be referred to without the `scalar(name)` notation, as noted by Kolev (2006).

Proposed future sections could contain more discussion of panel methods, e.g., the use of the `cluster()` option in fixed-effects models to correct for arbitrary intrapanel serial correlation, as discussed by Arellano (1987) and Kézdi (2004), and to calculate heteroskedasticity-robust estimates, as discussed by Stock and Watson (2006). Clustering is mentioned in chapter 7, in the context of cross-sectional clustering, but is not mentioned in chapter 9 (on panel methods), and Baum provides only three sentences on `xtivreg`, with no mention of `xtivreg2`.

Although diagnosis of weak instruments appears in chapter 8, Baum proposes no methods for inference robust to weak instruments, such as those discussed by Nichols (2006). To be fair, this is an active area of research, and some of the methods were published after IMEUS went to press, notably the conditional likelihood-ratio test implemented for Stata by Mikusheva and Poi (2006).

Other short sections might provide information on `poisson`, `treatreg`, `ivprobit`, `cdsimeq`, `pscore`, `nnmatch`, `dfl`, `jmp`, `oaxaca`, `devcon`, and other relevant commands, both official Stata and not. Future editions should probably mention regression discontinuity, event history analysis (or long difference versus first-difference models), survival analysis, and mixed models, although each of these could easily become a long chapter in itself. The last two sections could, of course, provide brief introductions to the class of models and then refer the interested reader to Cleves, Gould, and Gutierrez (2004) and Rabe-Hesketh and Skrondal (2005), respectively.

6 Conclusion

This book is an excellent and carefully written collection of much useful material. I will refer to it often and recommend it to others, but I would prefer more detail in the last two chapters. I hope that a second edition will mention more models, even if the relevant sections were short and referred the reader elsewhere.

An introductory book provides an overview, leaving advanced material to more specialized texts. Baum's book starts with the manipulation of raw data and builds to instrumental variables and generalized method-of-moments estimators with HAC (heteroskedasticity- and autocorrelation-consistent) standard errors. Web supplements would make a fine, although nonessential, addition.

7 References

- Allison, P. D. 2002. Missing data. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.
- Arellano, M. 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49: 431–434.
- Baum, C. F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press.
- Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1–31.
- Cleves, M., W. Gould, and R. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata*. Rev. ed. College Station, TX: Stata Press.
- Kézdi, G. 2004. Robust standard error estimation in fixed-effects panel models. *Hungarian Statistical Review* Special number 9: 96–116.
- Kolev, G. I. 2006. Stata tip 31: Scalar or variable? The problem of ambiguous names. *Stata Journal* 6: 279–280.
- Long, J. S., and J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.
- Mikusheva, A., and B. P. Poi. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 6: 335–347.
- Mitchell, M. 2004. *A Visual Guide to Stata Graphics*. College Station, TX: Stata Press.
- Nichols, A. 2006. Weak instruments: An overview and new techniques. Boston, MA: 5th North American Stata Users Group meetings. <http://www.stata.com/meeting/5nasug/wiv.pdf>.
- Rabe-Hesketh, S., and A. Skrondal. 2005. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press.
- Royston, P. 2005. Multiple imputation of missing values: Update of ice. *Stata Journal* 5: 527–536.
- Stock, J. H., and M. W. Watson. 2006. Heteroskedasticity-robust standard errors for fixed effects panel data regression. NBER Technical Working Paper No. 323. <http://www.nber.org/papers/T0323/>.
- Wooldridge, J. M. 2006. *Introductory Econometrics: A Modern Approach*. 3rd ed. New York: South-Western.

About the author

Austin Nichols is an economist at the Urban Institute, a nonprofit, nonpartisan think tank. He occasionally teaches statistics and econometrics, and he has used Stata almost daily since 1995. His research interests include poverty, social insurance, tax policy, and demographic outcomes such as fertility, marital status, health, and education.