



Journal of Statistical Software

February 2009, Volume 29, Book Review 14.

<http://www.jstatsoft.org/>

Reviewer: Dirk Eddelbuettel
Debian Project

Applied Econometrics with R

Christian Kleiber and Achim Zeileis
Springer-Verlag, New York, NY, 2008.
ISBN 978-0-387-77316-2. 222 pp. USD 54.95.
<http://CRAN.R-project.org/package=AER>

Applied Econometrics with R by Kleiber and Zeileis is a welcome and timely addition to the fast-growing *Use R!* series of R titles published by Springer. As noted by the authors, this appears to be the first book¹ that focusses on R and econometrics. Researchers in quantitative social sciences in general, and econometrics in particular, have often favoured scripting languages such as GAUSS or Stata, or packages such as EViews. Introducing R to this particular audience could therefore be a well-appreciated title among the growing number of publications about R.

The book is comprised of seven sections. A brief introductory chapter gently guides the reader around the R environment via succinct examples that highlight several focal areas of the book. This includes the wide use of graphics for data exploration and visualization. It also covers more recent methods in applied statistics or econometrics. For example, quantile regressions are already used in the second example in the *Introductory Session* 1.1 and hence before the reader has reached Section 1.2, *Getting Started!* The chapter continues with material on how to work with R, a very brief overview of data and object types, and is followed by short sections on the help system, the development model and the history—all of which combines into a well-rounded overview.

Chapter 2 faces the delicate task of introducing R to readers who may have never used it (yet are familiar with basic statistics and econometrics). This chapter is the longest chapter in the book, though not by far. It accomplishes what it sets out to do. I found it odd that the introduction of `data.frame` objects came so much later than other basic types, and after both *R as a Programming Language* and *Formulas*, but the chapter still succeeds in covering an admirable amount of material, including an introduction to graphics as well as a section on exploratory data analysis before a section with exercises concludes the chapter.

The next four chapters provide introductions to the essential toolkit of applied econometrics. Chapter 3 covers linear regression. This starts with the standard linear model and is followed by multiple linear regression. From there, Kleiber and Zeileis take the reader to partially linear

¹However, it should be noted that [Farnsworth \(2008\)](#) has been available as a manuscript since 2004.

models (using **splines**) and then move on to the important topic of factors and interactions before addressing the special cases of regression with time series and panel data, respectively. Sections on systems of linear equations and exercises conclude this chapter.

Chapter 4 deals with model diagnostics and alternative regression models—possibly the area in which R is strongest compared to the common alternatives mentioned above. Two sections—*Regression Diagnostics* and *Diagnostic Tests*—cover testing for departures from model assumptions. This is followed by ways to address such departures—*Robust Standard Errors and Tests* and *Resistant Regression*—before *Quantile Regression* is re-introduced with slightly more coverage than in the introductory chapter. Exercises once again conclude the chapter.

The next chapter turns to microeconometrics. R follows statistical practice in using generalized linear models (via the `glm()` function) that encompass the familiar logit and probit models which are commonly used in social sciences as particular parameterizations. This topic, just like many others in the book, is exposed clearly. The coverage then extends nicely to binary dependent variable models, count data models (including zero-inflated Poisson and negative binomial models). The next section covers censored variables (and the `tobit()` wrapper for the code from the **survival** package providing this functionality). This is followed by a section on extensions such as multinomial and semi-parametric models before exercises conclude.

Chapter 6 is devoted to time series. This chapter builds on so-called *naive* models, covers *classical* (Box-Jenkins) modeling and moves on to modeling non-stationary series, unit roots and cointegration. This leads naturally to another source of nonlinearities in dynamic models: structural change, an area of particular expertise for Kleiber and Zeileis. Again, extensions (on structural models and GARCH volatility models) follow and exercises conclude.

The last chapter concentrates on one of the strengths of the R system: the ease with which one can turn from a user into a programmer in order to extend one's analyses. This is particularly well done and shows by means of examples over three sections how simulations, bootstrapping, and likelihood maximization can be set up without much effort in R. Another section covers reproducible research using **Sweave**, a system to combine code and its write-up in the same document. Exercises again conclude the chapter.

So is this a good introduction of R for econometricians? Absolutely—with a well-rounded selection of available methodologies, both classic and current, and a good focus on introducing graphical methods, as well as gently covering more novel and therefore less familiar approaches, it fulfills its task with aplomb. The writing style is conversational without being shallow. A special mention goes to the companion CRAN package **AER**. Readily installable in R, it provides datasets and example code which are used throughout the book to illustrate different methods. The datasets are often well-known examples from the literature that had been provided in numerous other textbooks, and that are now easily accessible as a collection alongside the example code. The reader can therefore analyze and study all the examples from the book, taking a small step towards reproducible research.

Lastly, can we conclude the review without any gripes? Almost. If anything, the book is too short on details given the provided space. Many topics are covered in just one to two pages which cannot provide anything but a first overview. Some other methods such as matching, generalized method of moments or estimation by simulation are not covered at all.

But many an econometrician will—and should—be introduced to R through this book by Kleiber and Zeileis book. It is about time.

References

Farnsworth GV (2008). “Econometrics in R.” Unpublished manuscript, Kellogg School of Management, Northwestern University. URL <http://CRAN.R-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>.

Reviewer:

Dirk Eddelbuettel
Debian Project
Chicago, IL, United States of America
E-mail: edd@debian.org
URL: <http://dirk.eddelbuettel.com/>