

Preface

My teacher, Wassily Leontief, a Nobel laureate, had a great influence on me. In fact, my Harvard dissertation was on “Non-Linearization of Leontief Input–Output Analysis.” Leontief’s (1982) piece in *Science* criticizes typical models used by academic economists for assuming stationary equilibria and lamented “the splendid isolation” of academic economists from the real world. Morgan (1988) studies Leontief’s points by studying the content of economics journals such as the *American Economic Review* (AER) and argues that there is “market failure” in academic economics with excess emphasis on “fine points of logic to win approval within the guild.” Fortunately, things are changing in the new century. This book includes realistic econometric tools which model nonstationary equilibria. The book is inspired by Ernst Berndt’s (1991) hands-on type book suitable for an earlier era before the Internet.

Similar to Berndt, the book should help the reader in learning to *practice* econometrics with numerical estimation work involving computer software. Econometrics has been around as a subject of scientific inquiry since about 1933 when the journal *Econometrica* was founded by Ragner Frisch and others. It has grown a great deal in the 20th century and matured. The book uses econometric examples involving several policy issues such as divestiture of the Bell System, global warming, etc. Hence it can be of interest to all social and biological scientists, engineers, legal professionals in addition to those who use the R software. I assume that the reader has only a limited (not zero) exposure to the basics in Economics, Statistics, Finance, Computer Science, and Mathematics. Every attempt is made to derive results from first principles, to use self-explanatory notation and minimal cross-referencing, and to fully explain the mathematics used. Some R snippets allow the readers to ‘see’ what various matrix operations (e.g., Kronecker products) do and numerically verify the algebraic identities (e.g, singular value decomposition) with simple examples.

Thus, the book is structured for five types of potential uses.

1. as a textbook for graduate or advanced undergraduate econometrics courses for students with mixed backgrounds focused on applications rather than proofs.
2. as a supplemental textbook for usual econometrics /applied statistics courses in social sciences, engineering, law and Finance.
3. as a tool for any student or researcher wanting hands-on learning of some of the basic regression and applied statistics methods
4. as a supplemental material for computer science courses teaching object oriented languages including R.
5. as a reference for data sources and research ideas.

In the new century, fast linking of all scientists through the Internet is revolutionizing the exchange of scientific ideas including data and software tools. This book is intended to highlight and welcome these exciting changes. Since open source free software is particularly powerful for exchange of software tools, the book embraces R (distributed under Free Software Foundation's GNU GENERAL PUBLIC LICENSE Version 2, June 1991).

Why R?

All empirical scientists should learn more than one programming language, not some point-and-click software restricted by the imagination of the programmer. R is based on an earlier language called S developed in Bell Labs around 1979 (by John Chambers and others with whom I often used to have lunch). It is an *object oriented* Unix type language, where all inputs, data and outputs from R are 'objects' inside the computer. Also, R is an "*interpreted programming language*" not "compiled" language similar to FORTRAN or GAUSS (reviewed in Vinod, J2000c) or many older languages. Hence, all R commands (or command sets enclosed in curly braces) are implemented as they are typed (or 'sourced').

An advantage of Unix type languages is that numerical results are identical (platform-independent), and yet R is available for a wide variety of UNIX platforms, Windows and Mac systems. This provides one kind of flexibility in using R. SPlus is the commercial version of S, and R is the free version. Since Splus programs work in R, this allows second kind of flexibility for R. Third kind of flexibility of R arises from the fact that it is open source, meaning that every line of the code is available for any researcher anywhere to see, modify, criticize, etc. A black box of hidden code is unavoidable for any proprietary commercial software. Not so for R.

Anyone with enough patience to learn R can modify available software to suit a particular application. The fourth kind of flexibility of R is that it offers a package called ‘Rcmdr’ for the convenience of cursory users who want the point-and-click mode and do not wish to go beyond standard techniques. A fifth kind of flexibility is the ability to choose from a wide choice of packages without having to figure out if they are money’s worth. New packages and modules of proprietary packages are often expensive and expect users to keep on paying for ever newer versions. The buyer has to figure out if the latest version is really new, bug-free and worth buying. By contrast, free latest versions of all R packages are readily available, making life easier for users.

Vinod (J1999b, J2000c, J2003c & d, J2004d & e) discuss numerical accuracy issues. R is believed to be numerically one of the most accurate languages available. The accuracy of its nonlinear algorithms and random number generators is reasonable. Although not perfect, R language is progressing fast. I am convinced that R is headed to become the *lingua franca* of all applied statistics, including econometrics.

The book provides numerous snippets containing R code explaining what the code does, with detailed comments hoping to simplify and speed up the learning curve. Since the idea of learning a programming language like R may seem daunting to disinterested students, this book does some spoon feeding. It suitably motivates the reader by making the learning a bit more hands-on and fun. The reader can first do any one of the dozens of computing tasks in our snippets (e.g., run projection pursuit regression), look at the results, and choose to learn only the relevant features of interesting tasks at leisure. My students find it fun to discover hidden jewels inside R by direct searches on Google’s website and by doing illustrative examples inside the contributed packages.

Advantages of R for Replication of Published Research

There is a new emphasis on replicable empirical work, for the benefit of all students and researchers, anywhere in the globalized world. Under the able editorship of Ben Bernanke (current Chair of the US Central Bank called the Federal Reserve Bank) reproducible empirical work is being emphasized at the *American Economic Review* (AER), the main journal of the American Economic Association. Bernanke (2004) cites McCullough and Vinod (J2003d) in his “editorial statement” requiring all AER authors to submit both data and software code. Indeed, it would be beneficial for the profession if any researcher anywhere is able to replicate published quantitative results from all Economics journals. *Econometrica*, *Journal of*

Political Economy and many top journals are following the lead of AER and require authors to submit both data and software. I hope that this book facilitates the movement toward replicable econometrics by making R more easy and fun. Replication with free R will be much simpler and fast for international students, eliminating possible delays in paying for software in different currencies.

Obtaining R and its documentation files

Go to <http://www.r-project.org>. On the left side, under “Download” click on the CRAN (Comprehensive R Archive Network) link. Pick a mirror closest to your location (e.g., Pittsburgh). Assuming that you have a Windows PC, under “Download and install R” click on Windows. Now, under subdirectories, click on ‘base.’ Next, click on the fourth item under “in this directory:” Do not be confused by options. Click on R-2.7.0-win32.exe (or similar latest setup program, size over 25 megabytes). Depending on your Internet connection, this process will take some time. The ‘setup’ program creates an Icon for R-Gui (graphical user interface). The point is that anyone with an Internet connection sitting almost anywhere in the world can get R on any number of computers completely free of charge. This is perhaps too good to be true. Many books, some free manual and other documentation is also available by clicking on a link in the left column under “documentation” at the R Homepage. The Wiki for R at (<http://wiki.r-project.org>) is particularly useful for newcomers.

Packages within R

R comes with the “base” package and additional contributed “contrib” packages have to be explicitly requested from the R website. The contributed packages are written by statisticians, computer scientists and econometricians from around the world and contain several functions for doing operations plus many illustrative data sets. They are freely available on demand for noncommercial purposes. All R packages are required to follow a nice readable format describing all their functions. Each function is described with details of inputs and outputs, references to the literature and generally does have examples. If one is curious to know exactly how the function was implemented, usually all one has to do is write the name of the function and the entire program can be seen. The user then has an option to modify the program, except that beginners will find that many programs are hard to understand, since their expert authors use their ingenuity in writing efficient (fast, not necessarily easy to read) code. By contrast, the code in the snippets in this book should be easy to read.

In addition to user manuals, some packages also have “vignettes,” which are fully worked out examples explaining the usage and interpretation of results. (type ‘vignette()’ to know what is available). I always download and see both the user manuals and vignettes from the r-project website for contributed packages to fully understand what each package does and its full potential. Downloading the latest version of any package itself into R takes only a couple of clicks and automatic updates are also similarly available from the R-Gui on the user’s desktop. The Journal of Statistical Software of the American Statistical Association has the URL (<http://www.jstatsoft.org/>) for free download with a number of articles dealing with R. A recent volume 27 (July 2008) edited by Achim Zeileis and Roger Koenker deals with ‘Econometrics in R.’ It has excellent articles describing the following packages: plm, forecast, vars, np, Redux, sampleS-election, pscl.

Organization of the Book

The book begins with an introductory chapter using production functions to illustrate relevance of nonlinear functions of regression coefficients. We include preliminary data analysis with recent tools (Cook’s distance) multiple regression methods, singular value decomposition, collinearity problem and ridge regression. Chapters 2, 3 and 5 deal with time series analysis with the standard topics and some sophisticated ones including: autoregressive distributed lags (ARDL), chaos theory, mean reversion, long memory, spectrum analysis, ergodicity, stationarity, business cycles with imaginary roots of AR(2), impulse response, vector autoregression (VAR) models, stochastic diffusions, cointegration, Granger causality testing and multivariate techniques including canonical correlations.

Chapter 4 discusses expected and non-expected utility theory and implications for Finance and other areas. It has new tools for measuring up to fourth order stochastic dominance shown to study ‘prudence.’ Chapter 6 has detailed derivations of simultaneous equations theory including k-class estimators, limited information maximum likelihood (LIML) and the ‘identification’ problem, with hands on examples. The novelty of Chapter 7 on ‘limited dependent variables’ is that besides economists’ favorite Tobit and Heckman estimators, it explains the less familiar general linear model (GLM) viewpoint used by Biostatisticians with their ‘link functions’ implying superiority of logit over probit. We also discuss survival models for Oil Company CEOs.

Chapter 8 has sophisticated consumer theory including Wiener-Hopf dynamic optimization and Kernel estimation. Chapter 9 explains the traditional bootstrap, its limitations and several newer tools including double and maximum entropy bootstraps (with easy to use R code). Chapter 10 has generalized least squares (GLS), generalized method of moments (GMM), vector autoregressive moving average (ARMA) models, ‘estimating function’ and related pivot functions with examples. Chapter 11 deals with nonlinear models and explains how to use ‘projection pursuit regression’ for money demand equations.

The topics covered are influenced by a need to illustrate them with examples within the size limitations of a book and my own familiarity with the topic. Admittedly somewhat vain, the long list of my own and joint papers is separated from the list of other authors. The list uses the prefix J for journal articles, P for published proceedings, B for books or chapters in books and U for unpublished but widely circulated pieces. Since it is not possible to include additional papers of mine, an Appendix lists a thematic classification of my papers. It shows that the book could not accommodate some themes and state space (Kalman filter) modeling from Vinod (B1983, 1990, J1995c) using the R package ‘sspir.’

An important selling point of the book is that it has numerous program snippets in R. My hope is that the snippets have useful practical information about implementing various theoretical results. The reader is encouraged to read them and treat the snippets as templates. I hope readers will modify the snippets to apply to different and more interesting data and models. This is the sense in which we call this book “hands-on.” The reader can readily copy and paste each snippet into R while reading the discussion and see R work out the results first hand. For brevity, very few of the outputs appear in the printed book.

I am grateful to the following former/current students for detailed comments: Erik Dellith on Chapter 1, Brian Belen on Chapter 5, Caleb Roepe on Chapter 7 and Diana Rudean on Chapter 9.

Many snippets in the book have been tried by graduate students at Fordham University. The students have found it fun to try them on some data sets of their own choice. It is important not to treat the snippets as black boxes to blindly perform some tasks. Professional programmers try to write the most efficient code in the sense that it works fast and has as few lines as possible. The snippets in this book are not at all professional. Instead, they should be viewed as tools for learning to use R in applied work. Hence instead of merely copying and pasting them, my students were

encouraged to slightly modify each snippet to suit a distinct problem. Such a hands-on approach makes it fun. A purchaser of this book need not type the snippets, since a CD in a sleeve inside the back cover contains all snippets as text files

I welcome suggestions for improvements to the content of the book and/or the snippets. Although improved code to professionals means faster running with fewer lines, I prefer a readable code with some intuition for the steps used and lots of comments. Of course, I will give proper credit to the person(s) suggesting any improvements. On that note, Professor Peter C. B. Phillips, the eminent econometrician from Yale University who has seen a preprint of this book, suggested that I should clarify that my snippets often use the assignment symbol “=” from GAUSS and FORTRAN, instead of the “<-” symbol preferred by R and S-plus professionals. My reasons for using the “=” symbol are that it works in R, requires less typing by me and takes up less space in the printed book. Thus our snippets are not portable to S-Plus until the user replaces “=” by the symbol “<-” in the body of the snippet, except for occasional ‘lists’ associated with new R functions. I hope that the reader will appreciate the reduced drudgery made possible by our R snippets and treat Econometrics as a fun subject.