

THE POLITICAL METHODOLOGIST

NEWSLETTER OF THE POLITICAL METHODOLOGY SECTION
AMERICAN POLITICAL SCIENCE ASSOCIATION
VOLUME 11, NUMBER 1, FALL, 2002

Editor: SUZANNA DE BOEF, PENNSYLVANIA STATE UNIVERSITY
sdeboef@psu.edu
Editorial Assistant: HEATHER L. ONDERCIN, PENNSYLVANIA STATE UNIVERSITY
hlo114@psu.edu

Contents

Notes from the Editor	1
Graduate Teaching Options	2
Frederick J. Boehmke: Teaching Advance Graduate Methods	2
Kristian Skrede Gleditsch: Review of <i>Applied Regression Analysis</i>	4
Ryan Baker: Review of Fox and the <i>R and S-Plus Companion Guide</i> from the Graduate Student Perspective	5
Web Based Statistics Books	6
EITM	7
Aldrich: EITM	7
Kunicova: Reflections	10
Oleg Smirnov: The EITM Experience	12
Karama: EITM and Dissertation Work	13
Robert F. Trager: Not Just for Techies	14
Articles	15
Jeff Gill: Wielding L ^A T _E X To Greater Effect	15
Colin Elman, David Collier and Henry E. Brady: A New Qualitative Methods Section	20
Jonathan Wand: Evaluating the Consequences of Assumptions Using Simulations	21

Section Activities	26
Section Awards	26
Webmaster Needed	26
2003 Summer Methods Meetings	26
ANES Update	27

Notes From the Editor

This issue of *TPM* includes the usual assortment of contributions on teaching, research, and section news and also melds perspectives from formal and qualitative methods with the typical focus on quantitative methods. In particular, David Collier and Henry Brady present a portion of the petition to form a new APSA organized section on Qualitative Methodology. The section's plans complement and coordinate with the Political Methodology Section. With a more formal bent, the articles continue last spring's attention to EITM, with perspectives on the first summer program on the empirical implications of theoretical models from John Aldrich and from four student attendees from different graduate programs and at varying stages of their graduate career. Students offer their wisdom on combining formal and empirical work for those considering applying to attend future summer programs and discuss the strengths of the program for one's graduate career and beyond. On the more familiar empirical side of things, Jonathan Wand offers his suggestions for presenting simulation results. With increasing computing power, more sophisticated estimators, and advances in agent-based modeling comes increasing recourse to computer simulations. Often the simulations leave the analyst with pages and pages of computer output, unclear how to

Review of John Fox's *Applied Regression Analysis, Linear Models, and Related Methods*

Kristian Skrede Gleditsch
University of California San Diego
kgleditsch@ucsd.edu

Applied Regression Analysis, Linear Models, and Related Methods. by John Fox (Sage: Thousands Oaks: Ca, 1997; 597pp; \$79.95. ISBN: 080394540x.)

The textbook typically chosen for a political science graduate course tends to be an introductory econometrics text, with an emphasis on applications to economics. This book is written by a sociologist who also holds a position in a statistics department. The examples used in this book illustrate the relevance of statistics to social sciences beyond economics and are quite different from those one would normally find in an introduction to econometrics text. The structure of the book is more in line with statistical textbooks than books on introductory econometrics. It emphasizes learning statistical theory in combination with data analysis, and makes extensive use of modern visualization techniques. The book has many strong sides and features that make it an attractive choice to social scientists and political scientists.

Rather than starting with the usual one chapter review of statistical concepts before delving into regression analysis, Part I starts with non-parametric data analysis. Fox first shows how a conditional distribution of Y given X can be fitted to observed data. This is followed by an overview of data analysis by graphical methods, and a review of methods for data transformation and their rationale.

Linear regression is then introduced in Part II. With the background from Part I, the actual meaning of what a linear model, $E(Y|x) = \mu$, entails is better put into context. Many texts simply introduce regression by assuming a linear relationship with no introduction. Here, the material in Part I makes the linear regression model come across as less arbitrary, and the ability to approximate linearity through transformations and non-parametric varieties of linear regression is clear to the reader before taking on the effort to work with the theory of linear models.

Part III of the book is devoted to regression diagnostics and a discussion of various problems that may occur in applied regression. This part of the book is particularly strong and raises many issues that are not often covered in much detail in econometrics texts.

Part IV of the book introduces extensions of the linear regression model to limited dependent variables, time series, GLS, and non-parametric regression. It also contains a very useful chapter on bootstrapping and model cross validation.

The book can be read at many different levels. Fox's writing is very clear, and the emphasis on social science examples in the discussion makes it possible to read much of the book without any background in statistics and probability theory. Appendices on statistics and probability theory make the book largely self-contained. At the same time, the book is also theoretically thorough, and takes the reader through the most important proofs. Fox places particular emphasis on understanding the linear regression model geometrically. The geometrical diagrams and interpretations of the linear model are very good, and there are large payoffs for someone who takes time to work through this material. Finally, Fox does a good job of stressing the essential similarities between different models as classes of the general linear model. Looking at estimation equations in matrix notation makes it is easy to see why the estimation equation for limited dependent variables models are essentially the same as for linear regression, but have to be estimated via different procedures since the model is non-linear in the predicted probabilities.

The book is also strong on practical issues in data analysis and model diagnostics. I personally see the issues covered in Part III of this text as more useful for a first regression course in political science than topics in time series which often are covered in extensive detail in many introductory econometrics texts.

Many of the good things in this book rely on matrix algebra, however, and this might ultimately might make it difficult to use the book for many courses. Although it is possible to read a great deal of this book without matrix algebra, this would force the reader to skip parts of the book that contain its strongest sides. Many of the alternatives to this text for a course text, such as Gujarati, do not require any knowledge of matrix algebra.

I like this book a great deal, but whether I would choose it as a text depends on the course. The Fox book would be particularly suitable for a second course where students already have had some prior course in statistics and it is appropriate to go over the linear model in more detail. However, I would find it difficult to use this book in a first course on regression without matrix algebra, and would probably opt for a book that does not presume any knowledge of matrix algebra. At UCSD, for example, we currently have a 10 week first introductory regression course. Since we do not have an introductory course in

probability and statistical inference, we cannot presume any background and have to start from scratch. I have ultimately decided not to use Fox's book in this course, at least for the time being.

One potential selling point of the book is that the examples are generally not drawn from economics, and they can be understood without any background in economic theory. However, the examples that Fox use – such as Duncan's social mobility data – may not be much more familiar to the average political scientist. Much of this is compensated by the availability of the data online, and the author's companion book on R/S-Plus. Using these two books in combination would allow students to get a better understanding of the issues through the use of simulation and graphs, which R and S-Plus are particularly well suited for.

My reservations aside – and they are not really reservations against the text but rather its suitability for teaching – this is a great book which deserves to be widely read. All political scientists should give it serious consideration as a possible text for political science methods courses.

Review of John Fox's *Applied Regression Analysis, Linear Models, and Related Methods* and *R and S-Plus Companion to Applied Regression*

Ryan Baker
University of Florida
fannyker98@yahoo.com

Applied Regression Analysis, Linear Models, and Related Methods. by John Fox (Sage: Thousands Oaks: Ca, 1997; 597pp; \$79.95. ISBN: 080394540x.)

An R and S-Plus Companion to Applied Regression. by John Fox (Sage: Thousands Oaks: Ca, 2002; 328pp; \$39.95. ISBN: 0761922806.)

As graduate students in political science, most of us do not have strong backgrounds in calculus, linear algebra, or math/stat theory. If we do well in our introductory data analysis courses and decide to pursue more advanced methods training, we are all too often discouraged by our first experience with an econometrics text.

We find ourselves in a 'Goldilocks-esque' dilemma where "this stats book is too easy but that one is too hard". Unfortunately, texts that are 'just right' have been few and far between.

John Fox's *Applied Regression Analysis, Linear Models, and Related Methods* helps to bridge the divide between introductory and intermediate to advanced methods courses. The book is written in a clear, concise manner and organized in such a way as to help facilitate comprehension of the material. It is important to understand that this is not an introductory text. The author clearly states that a knowledge of applied statistics including basic probability and statistical inference theory is necessary to appreciate the material in this text.

Fox has provided an excellent text book for a second and/or third semester course in a political science methods sequence. *Applied Regression Analysis, Linear Models, and Related Methods* can be used as the primary text for a course dedicated to linear regression and some basic extensions or in combination with a basic econometrics text (Greene, Gujarati, Kennedy) for a more advanced course in regression and maximum likelihood. Over the past two years, I have had the opportunity to use this text in both settings as a student and a teaching assistant. *Applied Regression Analysis* was very well received by nearly every student, regardless of their previous methods training. The fact that this text can be used at several different levels is certainly one of its major strengths.

A student's primary concern with a statistics test generally involves the problem sets at the end of the chapters. Here again, John Fox has done a wonderful service for students and teachers alike. The exercises range from trivial to demanding and are directly linked to specific sections in the text. The organization of the exercises is sometimes clumsy in that the exercises are spread throughout the chapters, rather than grouped at the end, but this is a minor inconvenience. Overall, the exercises thoroughly incorporate the important principals in the chapters. The data analysis questions are particularly useful in this respect. Professor Fox has posted all of the data used in the examples and the exercises on his website, allowing students to replicate his findings and to use real world examples when learning the principles of regression analysis.

Although *Applied Regression Analysis, Linear Models, and Related Methods* is a wonderful teaching/learning tool on its own, its true strength comes from its use in combination with Fox's *R and S-Plus Companion to Applied Regression*. Not only does this combination teach regression in an intuitive, hands-on fashion, but it also provides an excellent introduction to the statistical computing environment R. Helping to move students away

from point-and-click stats packages may well be one of Fox's most important contribution to methods training in the social sciences.

The *R and S-Plus Companion to Applied Regression* is arguably the best introduction to the S language for social science graduate students. As is the case with the *Applied Regression Analysis* text, all data sets used in the *Companion* are available on-line. The *Companion* is loosely organized to follow the text and includes a fabulous index of both subject headings and S language functions. Fox has also included a package of S functions written explicitly to be used in combination with the book.

The *R and S-Plus Companion to Applied Regression* introduces the S language in an accessible, lucid manner. The book begins with the basics, such as data entry and simple manipulations and transformations. The chapters move on to discuss linear regression, regression with dummy variables and interactions, and linear model diagnostics. Later chapters introduce more advanced regression techniques, including GLM's and related diagnostics. The last two chapters include a very useful discussion of graphing and function writing, two of R's most impressive features. Throughout the book, Fox has included an extensive number of clear examples, which students can replicate and modify.

Together, John Fox's *Applied Regression Analysis, Linear Models and Related Methods* and the *R and S-Plus Companion to Applied Regression* have made a fantastic contribution to the world of quantitative social science methodology. I would strongly encourage instructors of introductory and intermediate methods courses to consider *Applied Regression Analysis* as a primary text. For students and professors that have yet to discover the wonders of the S language, the *Companion* is a must read.

Web Based Statistics Books

Recently, Phillip Schrodt posted a link to a web-based statistics book. At that time he also asked if anyone else knew of good web sites offering similar resources on the PolMeth listserv. There were several responses we found useful as references that could be recommended to students or used as teaching tools. We also felt that it might be beneficial to compile them into one place so everyone does not have to hunt through archived e-mails to find

them. In general, all the sites provide a nice overview of many basic statistical concepts. However, many of the sites go beyond what a general textbook can provide by including interactive graphs and demonstrations. Thank you to all that posted a response to Phillip Schrodt's question.

Heather L. Ondercin

- **HyperStat Online Textbook**

In suggesting this web site Rick Almeida stated "there is everything an introductory statistics student might need." HyperStat provides information on a wide variety of different topics. But in addition to the basic information, it contains many links to other statistics web sites. Recommended by Rick Almeida, University of Missouri — Columbia.

www.davidmlane.com/hyperstat/index.html

- **The Rice Virtual Lab in Statistics (RVLS)**

This site provides fun, interactive examples of basic statistical concepts such as regression and goodness of fit. Recommended by Rick Almeida, University of Missouri—Columbia.

www.ruf.rice.edu/~lane/rvls.html

- **Seeing Statistics**

This web-based book covers material from univariate statistics through regression. The examples are from a range of different disciplines including political science!! The draw back to this web-book is that it requires the user to purchase a copy, compared to the other sites that allow free access. However, this may provide a more effective learning or teaching tool than traditional text books. Recommended by Caroline Tolbert, Kent State University.

www.seeingstatistics.com

- **StatSoft**

StatSoft starts with very simple statistical concepts but also contains information on more advanced statistical concepts. Recommended by Paul Manna, University of Wisconsin-Madison, who believes one of the nice features of StatSoft is that it can be completely downloaded.

<http://www.statsoftinc.com/textbook/stathome.html>

- **BMJ Statistics at Square One**

This site is clearly written and provides good information on some basic statistical concepts such as confidence intervals and t-test. It also contains a few chapters on more advanced concepts such as duration analysis. The downside to this web site is