The Political Methodologist

Newsletter of the Political Methodology Section
American Political Science Association
Volume 11, Number 2, Spring, 2003

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

This issue of *The Political Methodologist* contains an assortment of pieces particularly relevant to how we teach and use methods in the classroom. Contributions include reviews of software packages, including a detailed discussion of the R package for users of all types; reviews of books; tips on testing and on conducting experiments; and thoughts on organizing the graduate methods curriculum. In articles, Chris Adolph suggests ways to present simulation results, Norm Swanson presents an economists perspective on the state of time series analysis in political science, and Tamar London presents *games* in LaTeX in the LaTeX corner. Finally, there is a lot of section news. Most notably, Robert Erikson has agreed to serve as the next editor for *Political Analysis*, but the section is looking for a new webmaster and a new *TPM* editor.

Suzanna De Boef

Software and Books: Reviews and Previews

Stata 8 A First Look From a Personal Perspective

Nathaniel Beck
University of California, San Diego
nbeck@ucsd.edu

As most readers of *TPM* are aware, Stata Corp. began shipping Version 8 in early January. It is too soon for a full review, but readers might be interested in what the new version can do, and at least the first impressions of how well it works. This note is from the perspective of how I use Stata, both in teaching and research, and my guesses about how many readers of *TPM* use Stata. There are lots of new features that I do not touch on because they are not of interest to me; these may be of interest to others, so those contemplating Stata 8 should check out the Stata web site (www.stata.com).

Stata 8 is very clearly still Stata. Those who liked Stata 7 will like Stata 8 more (this need not have been the case, as we shall see), and those who preferred GAUSS or R to Stata 7 will not likely be moved to switch to Stata 8. Time-series analysts might find the new developments in Stata 8 worth considering, though I think at this time many would still stay with such specialized packages as Eviews or RATS; advanced cross-sectional analysts who need the specialized facilities of LIMDEP (mostly in choice models) will still find they need LIMDEP. However, users of S-Plus who have chosen to not migrate to R might find the new features of Stata 8 to cause them to consider that package as a serious contender. Why do I make these sweeping generalizations?

Stata 8 claims two huge changes, and a series of big but not huge changes. The huge changes are the ability to run from a graphics user interface (GUI) instead of via commands and high quality user controlled graphics. The GUI could have really gotten in the way (as it did when S-Plus moved from a command based system to a GUI, making it more difficult to use commands for little gain), but in Stata 8 it doesn't. If you do not like the GUI, ignore it, it takes up one button on a tool bar (and does not seem to interfere with anything, including documentation). Undergraduates might like the GUI (I used to use StataQuest for my undergraduates precisely because of the GUI), and beginning Stata users (or those of us whose memory has faded with age) might find the GUI helpful as a reminder of Stata syntax (since the GUI produces an actual command). Viewed as an add-on to the help system, the GUI might actually be useful (especially for those who do not want to truck around the 20 or so pounds of manuals that Stata now comes with). Thus, one can read the help and then while in help bring up the GUI to generate the command. But the key thing is that if you like command lines, you can totally ignore the GUI. So not such a big win for Stata, but at least not a loss.

The new graphics now move Stata up amongst the contenders. In the past I have always used S-Plus for graphics (even if I did the analysis elsewhere). Stata 8 gives the user the same full control over publication graphics that S-Plus and other high end graphics packages do. Users who want graphs in the most demanding journals (say *Political Analysis*) can now happily use Stata (and when editors come back and say make your graph half an inch less wide, the happy author can now say no problem, with Stata making it easy to resize axis labels and such). As of now those of us who want to portray three dimensions on a piece of paper cannot use Stata; it lacks the lovely perspective and contour plot commands of R/S-Plus. I hope this is soon remedied. And for those who want to interactively rotate point clouds in three space, Stata provides no solution. But for the majority of folks who are happy with plots of two dimensional objects, Stata 8 rocks.

My one lament on simple graphics is that Stata 8 does not provide a simple plot of a single time-series, nor is it completely trivial to produce one using the graphics commands (it can be done, but not in such a simple way that one would want to do so for actual time-series analysis). While Stata has handled time-series for a while (ever since it allowed for lags and first differences in a data-aware manner), Stata 8 has lots more features for analyzing temporal data (in addition to my workhorse

ARMAX model, which Stata misleading calls ARIMA.) Thus there is now a full suite of VAR routines, and also a much more state of the art series of tests for those who believe that time-series might actually have unit roots. With these new routines, and the ability to fairly easily notate lags and differences, Stata 8 could almost be thought of as a package that can happily be used by both time-series and cross-section (TSCS) analysts. But, alas, the cross-sectional orientation of Stata has not caused them to introduce the first command used in every time-series program, "sample begin end," with begin and end notated in nice conventional date forms. And at present Stata 8 does not produce nice time-series plots. The new improvements in Stata 8 may make it suitable for teaching time series in the context of a graduate methods class or for the occasional analysis of time series. But at present I do not think that Stata 8 will cause me to jettison EVIEWS for my professional work. (I note that Stata is quite quick to do upgrades, and has an active user community that is good at writing programs that might provide nice time series plots, so these issues might be remedied relatively soon, but one cannot count on that.)

Stata has always excelled at both time-series–cross-section and panel models; Stata 8 adds some new facilities for the analysis of panel models which only makes Stata better. I have always liked the ease of analyzing panel and TSCS data in Stata (the various XT commands); this advantage of Stata persists in version 8.

Stata has also been historically strong in the analysis of event history data. It, like R/S-Plus, has built its analysis on the counting process notation which makes it feasible to easily input and analyze event history data with time varying covariates. Stata has dramatically improved its event history routines in version 8, incorporating the popular (and important) frailty models into its ST suite. One consequence of this is that Stata now allows for the estimation of the popular Weibull model with gamma heterogeneity. Stata has also dramatically improved the range of parametric survival functions that can be modeled, and has also improved its analysis of the Cox proportional hazards model (particularly in the graphics area). I used to think that R/S-Plus was the package of choice for the semi-parametric analysis (Cox) of event history data analysis; I now think that Stata 8 is as good as these. Stata 8 has also caught up to LIMDEP on the parametric side, and may provide the strongest combination of parametric and semi-parametric methods of any of the leading packages.

Stata 8 also comes with an improved programming and matrix facility. While I still think that serious Monte Carlo analysis requires a dedicated matrix language such as GAUSS (at many costs), Stata 8 has removed a variety of impediments that had caused me to avoid using Stata for pedagogical simulations in various graduate classes. While there is some learning curve for writing these simulations, once one gets the hang of doing it, it becomes quite easy to use simulations to show students that the mathematical properties of estimators that we derived actually hold (yes, I know this is an odd view of logic, but one that seems correct in practice).

These programming changes also make it easier to write maximum likelihood routines in Stata 8. While I have not written any complicated maximum likelihood code in Stata 8, there is no doubt that Stata 8 is fine for having students code some simple likelihood estimations "by hand;" the code written now corresponds quite nicely to the way many of us teach maximum likelihood. Thus, for courses that are now primarily Stata based, there is no need to switch to GAUSS or R to have students hand code a few models. It is also now reasonably easy (with some learning curve) to have students write the various standard econometric routines directly in a matrix language (though Stata 8 clearly does not have all the matrix facilities that make GAUSS or R or MATLAB so powerful). There is no question that Stata 8 is fine for convincing students that $(X'X)^{-1}X'y$ makes sense. Stata 8 also seems to have the various numeric options and such that should make it a suitable package for doing serious maximum likelihood research, and many scholars have used Stata like this for a few versions. But I have not had occasion to try out Stata 8 on a serious maximum likelihood program.

As noted, Stata 8 is still Stata. Other than some kernel density estimators, it does not do much non-parametrically. It is still likelihood based, has nothing either fully Bayesian or MCMC (nor does it seem to do Monte Carlo evaluation of integrals) and is pretty linear (no GAMs, let alone neural nets). It clearly does not do some specialized things such as hierarchical modeling or LISREL that some folks do. On the good side, Stata remains a very friendly way to do almost all the analysis done by 95% of political scientists (and perhaps 95% of the analyses of the other 5%), and it has a nice community of Stata users providing additional features (though it may lose out to the R/S-Plus community here). Like R, Stata makes it easy to incorporate new updates and community written code via the web. The graphics in Stata 8 are a huge improvement, the new event history methods are a serious improvement, programming is much more logical (once one figures it out!) and the GUI does no harm. There are lots of other new routines that individual researchers will care about, though most of them are fairly specialized. I had hoped to be able to teach time-series using Stata 8, but it falls just a bit short of being ideal there.

Finally, one can really avoid visits to the gym by simply carrying around the Stata 8 manuals (though it would violate the new California textbook weight code).

As before, many if not most users will find the good Users Manual plus the excellent on-line help (and maybe the GUI) all they need. Stata 8 has now broken out the time-series commands, event history commands, survey commands, programming commands, cluster commands and time-series–cross-section commands into separate manuals. This is probably sensible, since most researchers either use one of these manuals all the time or almost never (and the on-line help is adequate for the latter type). But this still leaves four volumes of other commands (including all the regression type commands). Perhaps this is all good, but the trend here is ominous. At some point Stata will realize that the P in PDF does stand for portable. Other than weight, the documentation remains up to Stata's previous high standard. And for those who cannot locate one of the 10 or so volumes of documentation, the help system and the GUI may well be adequate. It should also be noted that when all else fails, Stata provides excellent (and timely) human help by email (probably better than for any other package I know of).

To sum it all up, go back and read the first paragraph.

Stat/Transfer

Stata 8 allows much more flexibility in terms of labels and variable names than did previous versions. So Stat/Transfer has a new version which handles data files written for Stata 8. (Stata is very careful here to allow Stata 8 to read older Stata files.) I find Stat/Transfer an invaluable tool for moving data between a variety of formats (either because I get data in lots of different formats, or because even if I am using R, I prefer to do my initial construction of the data set in Stata).

Review of Andrew Gelman and Deborah Nolan's *Teaching Statistics: A Bag of Tricks*

Brandford S. Jones
University of Arizona
bsjones@email.arizona.edu

Teaching Statistics: A Bag of Tricks. by Andrew Gelman and Deborah Nolan. Oxford University Press, Oxford, 2002; 320 pp.; $40.00; ISBN 0-19-857224-7 (paper).

On an exam in an undergraduate-level statistical methods class, I once asked students to construct an 80 percent confidence interval for the mean. The purpose of the question was to evaluate whether or not students could use a t table in conjunction with a data set (as well as figure out the 1 and 2-tail probability area for the appropriate interval). In the days preceding the exam, I had covered in some detail how to construct confidence intervals, what the implications were for selecting relatively large or small α, and importantly, how to ascribe some meaning to the information given by the interval (i.e. the "in repeated samples" idea) as it pertained to hypothesis testing. In teaching this material, I noted that a significance level of 1, 5, or 10 percent is typically chosen and then dutifully discussed the trade-off between Type I and Type II errors. The students seemed to follow the material and their problem sets seemed to indicate they could competently construct, say, 95 percent confidence intervals and then give a reasonably accurate interpretation of their intervals. I felt good about myself.

And then came the exam and the 80 percent confidence interval. I recall a few students shifting uncomfortably in their seats, occasionally shooting daggers in my direction. One student came up to me and said "it was unfair they had to compute an 80 percent interval." After all, the student implied, "we only learned 95 percent." The student was unhappy with my response: "it's the same as a 95 percent interval, just 15 percent smaller." In grading the exams, I noticed a high correlation between those few students shifting uncomfortably in their seats and their subsequent (in)ability to construct an 80 percent confidence interval. I felt less good about myself ... and then came student X's answer. To quote directly: "there's no such thing as an 80 percent confidence interval. There are only 90, 95, or 99 percent confidence intervals." I felt bad about myself.

Now to be fair about my students, *most* had little or no trouble constructing the interval (though more had trouble interpreting it); nevertheless, it only took a few students to make me realize that my lectures and textbook-style coverage of confidence intervals were not getting through to some students. The question arose: how could I do better next time? Enter *Teaching Statistics: A Bag of Tricks* by Andrew Gelman and Deborah Nolan.

This book, recently published by Oxford University Press, addresses head-on the problems and pitfalls commonly faced by teachers of introductory statistics classes. Moreover, the book is filled with literally hundreds of teaching tips, classroom demonstrations, student projects, and exercises designed precisely to help alleviate student confusion, anxiety, and misunderstanding of the application and interpretation of statistical methods. Importantly, the "bag of tricks" they provide is meticulously documented such that teachers could incorporate these activities into the classroom relatively easily.

As the title suggests, the intended audience of the book is teachers of undergraduate-level statistics and research design courses. The book is divided into three sections. The first section of the book (consisting of nine of the book's 16 chapters) deals specifically with topics typically covered in most introductory statistics classes. Material in these chapters include descriptive statistics, regression and correlation, probability and probability models, and statistical inference. Each chapter in this section of the book provides several (sometimes dozens) of illustrations of classroom activities and demonstrations that can be used to convey otherwise arcane or nontransparent information in an accessible and informative way. All of the examples—or tricks—Gelman and Nolan discuss have been applied in actual classroom settings. This is useful because the authors in several places note how the various teaching tricks have evolved based on their own classroom experience. One gets the feeling that the classroom activities they present will actually *work*.

When I first started reading the book, I was initially dubious. After all, statistics classes typically require a lot of ground to be covered and so where would I find the time to implement these activities? Further, in political science statistics classes, my experience has been that students enter the class with extremely weak (or nonexistent) mathematical backgrounds and almost certainly no background in statistics. So despite Gelman and Nolan's best intentions, who could ever implement all (or even some) of these classroom activities when so much basic material had to be covered in a relatively short period of time? Yet after just a few pages of the book, it became apparent what the authors were doing. The "bag of tricks" they propose are really a series of small-scale classroom activities, usually involving student collaboration, that can quickly and efficiently illustrate issues pertaining to such topics as sampling, computing probabilities, and yes, constructing and interpreting confidence intervals. Nearly all of the activities the authors suggest can be completed within a single classroom period and most of the activities can be completed in a matter of minutes. Also, since the activities they propose usually involves collaboration, their tricks allow for interactive learning.

As there are so many tips and suggestions the authors make, it would do the book injustice for me to attempt to describe them in the context of this review. For example, in their chapter on multiple regression, Gelman and Nolan give about a dozen suggestions, examples, and discussion topics covering such issues as interpreting regression coefficients, statistical interactions, transformations of the variables, and uncertainty in the model. Again, note that nearly all of the suggestions they make can conceivably be completed in a relatively short period of time. One exception to this rule comes in their chapter on "Statistical Literacy" (Chapter 6). I found this chapter to be particularly interesting if for no other reason than the tricks they suggest directly address a common question I hear among political science undergraduates: "what is the relevance of statistics?"

Gelman and Nolan describe several activities and outside class assignments that require the student to systematically evaluate statistical research that has been published in newspapers. Most of their chapter on statistical literacy provides *detailed* documentation of possible projects and class assignments that can be used in conjunction with media coverage of statistical information. Moreover, the authors provide stylized examples of student reports and delineate specific kinds of questions students should address when evaluating statistical information, as reported in the media.

The second section of the book consists of two chapters and revolves roughly around the question of how to put together a course that incorporates the "bag of tricks" espoused in the previous nine chapters. This section was useful because the authors not only make suggestions for possible syllabi using many of the specific activities they previously document, but also, the section is helpful because the authors discussed the *problems* they have encountered when applying these activities in the classroom. That is, they provide some practical advice on the actual implementation of such activities. This advice, I think, will prove to be quite useful, particularly if one is converting from standard "textbook" coverage of statistical material to a more interactive approach to teaching, as advocated by these authors.

Finally, the third section of the book picks up where the first section leaves off. The tips and tricks suggested in the first section of the book typically center on topics that are perennial to introductory statistics courses. The tips and tricks given in the third section of the book deal with departures from standard introductory material. In the four chapters comprising this section, Gelman and Nolan deal with such topics as Bayesian statistics, sampling theory, advanced probability, and projects for courses on mathematical statistics. The suggestions given in these chapters are as useful and provocative as those made in first section of the book; however, given the nature of most introductory statistics courses in political science—courses that are often an amalgam of research design and statistics—the first two sections of the book would seem to be most relevant.

Ultimately, I found *Teaching Statistics* to be an invaluable reference. The suggestions made by the authors have a singular goal: to make a difficult subject relevant and accessible. The bag of tricks fosters collaborative research among the students and the examples they give enliven concepts that for many students, often prove to be lifeless—like 80 percent confidence intervals.

Review of Andrew Gelman, John Carlin, Hal Stern and Donald Rubin's *Bayesian Data Analysis*

Joshua D. Clinton
Princeton University
clinton@princeton.edu

Bayesian Data Analysis by Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin. Chapman & Hall: New York. 1995; 526pp; $59.95; ISBN: 0412039915

Bayesian Data Analysis deserves a place on the bookcase of any political scientist interesting in bringing empirical models to data. It is well written, intelligibly focused and sensibly motivated using examples of relevance to political scientists. In so doing, it is an excellent look at the Bayesian statistical paradigm and the modelling benefits that such offers.

The authors' stated goal is to articulate and demonstrate using motivating examples the three steps of Bayesian statistics: "(1) setting up a full probability model using substantive knowledge, (2) conditioning on observed data to form a posterior inference, and (3) evaluating the fit of the model to substantive knowledge and observed data" (p. xvii). The authors describe the Bayesian approach in terms of the pragmatic advantages that it offers researchers interested in complex problems.

The book itself contains four parts. Part I introduces the Bayesian paradigm and distinguishes the approach from a frequentist approach. Bayesian estimation in single and multi-parameter models are presented and the authors discuss the assessment of Bayesian estimators using large-sample frequentist notions. Part II introduces hierarchical models, model checking and sensitivity analysis, how to analyze various study designs using Bayesian methods, and regression models. Part III deals with computation of Bayesian methods using posterior mode approximations (e.g., EM), posterior simulation methods, and MCMC methods. Part IV deals with a selection of specific models for robust inference and sensitivity analysis, hierarchical linear models, generalized linear models, multivariate models, mixture models, and models for handling missing data.

As the book was written in 1995, some sections are less relevant than others given the present computing environment. In particular, although the computational details in Part III may prove of interest to those interested in writing their own procedures to estimate specific Bayesian models, most is of little use to scholars interested in using publicly available software (e.g., winBUGS) to estimate Bayesian models. However, aspects of Part III may prove of interest to scholars desiring to know what the computer is doing (and why). In addition, although consciously written to contain only a select number of applications rather than a comprehensive "cookbook" of Bayesian models, classes of models that are increasingly applied to political science data (e.g., survival analysis and correlated data models) are entirely or largely absent from *Bayesian Data Analysis*.

A critical question all authors face is the right balance between discussion and mathematical results. Whereas many econometric texts can be overwhelming to students and practitioners alike due to a seemingly endless sequence of formal claims and proofs of the claims, *Bayesian Data Analysis* takes a much more discursive approach. However,"more discursive" does not imply either a lack of rigor or difficulty. In fact, the book explicitly assumes that readers posses strong skills in probability (particulary with respect to the use of probability distributions), statistics and linear algebra. As the authors themselves note "Although introductory in its early sections, the book is definitely not elementary in the sense of a first text in statistics" (xv). While readers possessing these skills will find the extended discussions useful, readers lacking these skills may find themselves hopelessly lost, as readers cannot turn to the mathematics to gain understanding or insight.

The author's stylistic choice is both a strength and a weakness. On the one hand, the approach makes it very accessible to readers interested in understanding the Bayesian approach without getting lost in notation. On the other hand, because the book does not adopt a claim-proof technique, readers interested in specific details are sometimes required to look elsewhere. Also, although the discursive approach means that readers weak in probability theory are able to get a sense of the Bayesian approach, the lack of explicit derivations permits the possibility that they may not fully understand the reasons for the results. Chapter 4, dealing with "Large-Sample

Inference and Connections to Standard Statistical Methods" nicely illustrates these points. The chapter contains an excellent discussion of the evaluation of Bayesian estimators using frequentist assessments (e.g., consistency), including descriptions of the conditions under which the described results fail. However, the results are almost entirely described rather than proved (although Appendix B does sketch some of the proofs) and readers desiring detail will be left wanting. This is not to say that the text is inadequate in any sense; however the book does seem to favor readers who prefer "more text, less math" over readers who prefer "more math, less text."

Another useful pedagogical aspect of the text is its inclusion of a "Bibliographic Note" following each chapter. More than simply a listing of cited works, these notes both reference additional material (necessarily current only as of 1994) and briefly discuss how the additional material relates to the discussion in the text.

An especially attractive strength of the text is its reliance on a variety of examples. The 38 models that the book examines are illustrated using 33 examples. Concepts are frequently introduced and motivated by way of an example. For example, the desirability of hierarchical models is motivated explicitly using experimental data on rat tumors. Hierarchical estimation is motivated by describing how such a model permits researchers to more fully exploit information in the available data. By introducing readers first to the modelling problem by way of an example, the authors nicely illustrate the reasons for turning to a Bayesian methodology.

Furthermore, unlike most texts that are largely concerned with examples of questionable relevance to the questions we are likely to ask of the data we confront in political science, *Bayesian Data Analysis* uses several political science examples. For example, the text illustrates methods using: pre-election polling data, forecasting U.S. presidential elections, estimating incumbency advantage, and multiply imputing for missing data in a Slovenian opinion poll.

In addition to motivating new models by referencing examples, several examples are revisited throughout the book. This provides a nice sense of continuity; readers are able to follow the development of more sophisticated methodologies to handle the questions being asked of the data. For example, the rat tumor data used to introduce hierarchical modelling in chapter 5 is then extensively used in chapter 5 to demonstrate how a full probability model is derived, estimated and summarized. In chapter 9 the book returns again to the example to illustrate how the same problem can be analyzed using hierarchical logistic regression. Chapter 10 returns to the estimation of the rat tumor data to demonstrate the use of importance sampling in estimating posterior distributions.

In several cases the authors rely on an extended treatment of an example to motivate a model and demonstrate how to apply the steps of Bayesian inference. For example, in chapter 16 the authors use data on the reaction times of schizophrenics and non-schizophrenics to demonstrate how to model and estimate data using a two-component mixture model. Extended discussions describing the actual application of Bayesian methods to actual data and questions being asked of the data permits the authors to move the discussion beyond the presentation of abstract statistical models and estimation procedures. So doing provides readers a sense of what is entailed in applying Bayesian methods.

Although the authors rely heavily on examples, the book is not simply a collection of examples analyzed using various Bayesian models. The book's primary focus is on illustrating concepts of Bayesian methodology, not providing a series of analyzed examples that readers can use to inform their own modelling decisions. Consequently, readers interested in the question of "what model should I use to answer my question," should look elsewhere. *Bayesian Data Analysis* is instead pitched at readers interested in answering "what does adopting a Bayesian methodology entail and what are some reasons for analyzing data using a Bayesian methodology."

Although recent offerings are both more current (and tailored) to the present computing environment and describe classes of models beyond those considered in *Bayesian Data Analysis*, *Bayesian Data Analysis* provides an excellent introduction to the nature of Bayesian inference. As opposed to being simply a reference book of Bayesian models, *Bayesian Data Analysis* is a book that describes how and why to "go Bayesian," and what such a decision entails. In sum, this book is a highly desirable addition to the shelves of all political scientists interested in the foundation and application of Bayesian methodology.

Review of Jeff Gill's *Bayesian Methods: A Social and Behavioral Sciences Approach*

Jack Buckley
State University of New York at Stony Brook
sbuckley@ic.sunysb.edu

Bayesian Methods: A Social and Behavioral Approach by Jeff Gill. Chapman & Hall/CRC, Boca Raton, 2002; 459 pp.; $69.95; ISBN 1-58488-288-3 (hardcover).

In the past few years an increasing number of political scientists (among others) have recognized the advantages of Bayesian statistics for social science data applications. Some have embraced the philosophical foundations of the Bayesian approach (e.g. the "subjective" definition of probability and recognition of the fallacy of fixed population parameters in hypothetical repeated sampling). Others have recognized the advantage of formally incorporating the results of prior research or expert knowledge in their analyses. Still others, indifferent to the philosophical underpinnings of the approach, nevertheless adopt Bayesian methods for their flexibility and ability to estimate otherwise intractable models through computer simulation. One thing is certain: political scientists and other researchers in the social sciences need to familiarize themselves with Bayesian data analysis, if only to stay abreast of their own literature.

In *Bayesian Methods: A Social and Behavioral Sciences Approach*, Jeff Gill's goal is create a textbook that presents applied Bayesian data analysis in an accessible way to a target audience of practicing social scientists and advanced students, without expecting or requiring particular mathematical sophistication. If this sounds like a difficult task, it is. Nevertheless, Gill's book succeeds in many dimensions and fills an important gap between the more technical treatments of the subject and older, outdated introductory texts. After using the book for a semester-long course in applied Bayesian data analysis for advanced political science graduate students, I can quite comfortably recommend it to instructors of a similar course, or for those seeking to augment their methodological toolbox through self-study.

There are several obvious strengths of the book. First, and perhaps most importantly, the book for the most part manages to walk the narrow line between overly sophisticated statistical language and notation, and the overly simplified, "Bayes for dummies" approach. Gill has a strong background in applied statistics and it shows in his ability to explain complex concepts with clarity while providing footnotes and appendices pointing the curious or more advanced reader to additional bibliographic material. Examples of this abound, including his treatment of the somewhat technical literature on Bayesian robustness analysis, and of the more esoteric diagnostics for the detection of nonconvergence in Markov Chain Monte Carlo (MCMC) estimation. It is hard to overemphasize the value of Gill's contribution here—most of the literature on Bayesian statistics beyond the most introductory levels is all-but-impenetrable to the average social scientist, and Gill's ability to summarize and present sections of this literature without overwhelming the reader is invaluable.

Additionally, where Gill suspects that the average reader's background may be insufficient for an upcoming technical topic, he presents chapter-length introductions to get them up to speed. Examples include sections on the general linear model and on basic (non-Bayesian) Monte Carlo methods for numerical maximization and integration. I found these sections to be helpful for my students, but they can be easily skipped or assigned as supplemental material for a more experienced audience.

Another strength of the book is the inclusion of relevant examples illustrating every major concept with data from the social and behavioral sciences. I found that these examples were excellent in helping students to grasp unfamiliar ideas and estimation procedures. Moreover, Gill has made the computer code (all using free, publicly available software packages such as R and WinBUGS) and data for replication of the examples available either in print in the book or from his website, which allowed me to run them in real-time during class to illustrate concepts from the day's lecture. Additional datasets and code are available in the exercises at the conclusion of each chapter, enabling instructors to assign practical homework problems that allow the students to conduct real analyses with meaningful data.

Yet another strength of Gill's approach in *Bayesian Methods* is his focus only on the elements of Bayesian data analysis that are relevant to his audience. For example, Bayesian decision analysis, long a central concept in the development and application of Bayesian methods, is nevertheless (and correctly, in my opinion) ignored as largely irrelevant to quantitative, empirical social scientists with a background primarily in applied maximum-likelihood estimation. On the other hand, practical MCMC noncovergence diagnostics, a topic that is not commonly addressed in great depth in similar texts (yet a topic vital to the daily application of Bayesian methods to real data) has its own 46-page chapter.

Perhaps my favorite chapter in the book is Gill's discussion of Bayesian priors and whence they come. For most researchers, the concept of the formal incorporation of prior information is the Great Mystery of the Bayesian approach. Gill does a wonderful job of surveying a broad and often contentious literature about the many ways in which informative priors can and should be used, and when the need to specify priors is better regarded as a nuisance. Particularly interesting is the book's brief discussion of the elicitation of priors from substantive experts for use in analysis. This is an idea that has seen little application in political science but that has tremendous potential for the quantification and empirical study of "small n" questions in a variety of subfields, including comparative politics, international relations, and political history/American political development. Moreover, the techniques used for prior elicitation are interesting in their own right to political psychologists and students of elite decision making.

I have several minor quibbles or frustrations with the book. First, although I laud Gill's aforementioned inclusion of many examples interspersed throughout the text, some of the examples are a bit too complex for the concept that they are intended to illustrate. For instance, in a section introducing the mechanics of the Gibbs sampler, Gill presents an adapted version of a well-known article that estimates a changepoint between two Poisson processes. While this an interesting model (and perhaps useful to some applied researchers in the audience), its use as an example somewhat obfuscates the already sufficiently complex subject of the section.

Another minor issue I have with the book is its treatment of multiple linear regression. I suspect that the methodological background of most potential readers (at least in political science) is primarily a course on probability theory and introductory statistics, a course on least-squares linear regression and regression diagnostics, and (perhaps) additional coursework on maximum-likelihood estimation set in the framework of what to do when the least-squares assumptions are violated. For these readers, such as my students, much of their understanding of applied data analysis is thus centered on the least-squares estimator and the classic linear model. I found, for my students at least, that I needed to recast sections of Gill's book with this in mind. For example, I spent several classes using the same simple multiple regression model to illustrate many different points, instead of adopting Gill's interesting but confusing diversity of examples.

A final quibble is with the homework exercises at the conclusion of each chapter. Gill adopts Don Knuth's logarithmic scale for homework difficulty assessment, a somewhat tongue-in-cheek approach that nevertheless provides a reasonable measure of exercise complexity. My problem is that too many of the exercises are beyond the ability of the average reader, at least without an inordinate amount of work (in the language of the Knuth scale, there seem to be too many problems rated 30-40, "difficult, significant effort required"). Furthermore, at the time of this writing, no instructor's solution guide to the homework is available (although I understand that Gill plans to remedy this in the near future).

In conclusion, let me point out that these critiques are very minor relative to the overall strengths of *Bayesian Methods*. On the whole, I have been quite pleased with the book, and my students, I believe, have been equally satisfied. I plan to use the book when teaching similar courses on applied Bayesian methods in the future, and I recommend anyone considering creating a similar course to review this book for adoption.

Review of Jeff Gill's *Bayesian Methods: A Social and Behavioral Sciences Approach*

Kosuke Imai[1]
Harvard University
kimai@fas.harvard.edu

Bayesian Methods: A Social and Behavioral Approach by Jeff Gill. Chapman & Hall/CRC, Boca Raton, 2002; 459 pp.; $69.95; ISBN 1-58488-288-3 (hardcover).

Jeff Gill's *Bayesian Methods: A Social and Behavioral Sciences Approach* represents a rare contribution by a political scientist to the collection of statistics textbooks. The book contains a number of examples, exercises, `R` and `BUGS` codes (more are available on Gill's website), and references. Another unique feature is that the equation-by-equation derivation of conditional posterior distributions is shown for several standard Bayesian models. Readers may feel that some materials of the book overlap with what is covered in existing standard textbooks such as Gelman *et al.* (1995) and Carlin and Louis (2000). Nevertheless, those who have little prior exposure to statistics and computing should find the book extremely useful. Below, I comment on each chapter while paying attention to how the materials of the book differ from those of the two texts mentioned above.

In the first chapter, Gill presents his original view about the connections between social science research and Bayesian statistics. Specifically, he argues that Bayesian inference is appropriate for the social sciences because of its ability to formally incorporate subjective knowledge as prior distributions. This contrasts with a common view that such requirement is a drawback to Bayesian inference because researchers may not possess such knowledge or wish to ignore it. (Even if such knowledge exists, it may be difficult to express it in the form of probability distributions.) I would also add that a primary advantage of Bayesian approach lies in its computational flexibility that allows for the reliable estimation of sophisticated models.

The second chapter reviews likelihood inference via the Generalized Linear Models (GLMs) of McCullagh and Nelder (1989). Chapter three presents univariate Bayesian models such as the Beta-Binomial model. The fourth chapter is devoted to the Gaussian linear and Student t models with covariates. Although these chapters somewhat resemble those of Gelman *et al.* (1995),

[1] `http://www.people.fas.harvard.edu/~kimai`

the main difference is that Gill shows the equation-by-equation derivation of conditional posterior distributions for these standard Bayesian models.

Chapter five discusses the role of prior distributions in Bayesian analysis. Gill starts with conjugate priors and acknowledges their limitations. Then, he presents various non-informative priors including uniform prior and Jeffreys' invariant prior. Informative priors such as elicited prior are also presented. The explanation of these prior distributions and extensive references are useful (Carlin and Louis (2000) also present many of such priors). Given the importance of prior specification in Bayesian analysis, however, a few real data examples would have provided readers with more direct guidance. In this regard, the examples of sensitivity analysis in Chapter six are helpful.

Chapter seven presents Bayesian hypothesis testing with emphasis on Bayes factors. This chapter differs in its approach from that of Gelman *et al.* (1995) who focus on the use of posterior predictive distributions for hypothesis testing. Gill correctly argues that the main advantage of this approach is the ability to conduct the comparison of non-nested models. In this regard, an example with non-nested models would have given readers concrete advice on the computational issues as well as the choice of prior distributions, both of which are important considerations when using Bayes factors.

Chapters eight and nine present various Monte Carlo techniques as well as *EM* algorithms. When compared with Gelman *et al.* (1995), Gill spends more time talking about general properties of Markov chain Monte Carlo (MCMC). Such presentation will be particularly useful for readers who would like to get a sense of these properties but are not interested in mathematical details. Chapter eleven provides further practical tips for assessing the convergence of MCMC. The illustration of various diagnostic plots and statistics is very helpful. Finally, Chapter ten discusses hierarchical models. The models presented in this chapter include the Poisson-Gamma, random effects logit, and two-level Gaussian linear models (These models are presented in Gelman *et al.* (1995) as well).

In conclusion, Gill's *Bayesian Methods* is a great addition to the array of teaching tools that can be used for an introductory Bayesian statistics course in social science departments. The book can be used as a supplementary text along with existing textbooks such as Gelman *et al.* (1995). The exercises, detailed derivation, and computer codes provided on the author's website as well as in the book will be of great help for beginning students.

References

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd ed. Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall.

Carlin, B. and Louis, T. A. *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. Chapman & Hall.

Review of *Methods of Social Movement Research*

Jillian Schwedler
University of Maryland
jschwedler@gvpt.umd.edu

Methods of Social Movement Research. by Bert Klandermans and Suzanne Staggenborg, eds. Minneapolis: University of Minnesota Press, 2002; 382 pp; $29.95. ISBN: 0-8166-3595-1.

In recent years, calls for methodological pluralism have tended to focus on two appeals: first, for tolerance and mutual respect of methods other than one's own favorite(s); and second (and to a lesser extent), for studies to employ two or more methods in a single study. In our puzzle-driven discipline, it seems only logical for scholars to adopt the methods that can best answer the empirical questions that animate their research interests. In reality, however, methods choice entails more than a small element of pragmatism of two types. First, there is no methodology suitable for exploring every empirical puzzle. Longitudinal survey research on the attitudes and attributes of al-Qa'eda members, for example, is likely to prove impossible to complete or at the very least hazardous to one's health. Seconds, methods choice is complicated by the considerable cost of keeping up with methodological innovations, which often prohibits scholars from gaining more than basic competence in a wider range of methodologies. Although triangulation of multiple methods is perhaps most desirable, it makes sense at a minimum for us to engage with those working on similar questions from different methodological perspectives. With this objective in mind, Klandermans and Staggenborg have put together a terrific volume that takes seriously the value of promoting methodological pluralism.

Methods of Social Movement Theory is aimed at two audiences: students and scholars of social movements seeking an overview of the strengths and weaknesses of a range of methodologies, and scholars in other areas interested in learning how theory can be developed through empirical research (vii) and the use of mixed methods. The chapters are organized broadly to move from micro to macro questions, though the editors acknowledge that most methods are useful at a variety of levels of analysis.

The introduction by the editors provides a quick overview of social movement studies, a field that some consider unified while others complain of a lack of synthesis. The individual chapters cover survey research, formal models, discourse and frame analyses, semi-structured interviewing techniques, theory-driven participant observation, use of single case studies, network analysis, historical research, protest event data analysis, macro-organizational analysis, and comparative analyses. The conclusion, co-written by the editors and Sidney Tarrow, makes a strong case for blending methodologies in empirical analysis as a means of building middle-range theory.

Most of the chapters do not offer much new in terms of any particular methodology. A few do suggest possible advances for their method in the study of social movements, such as the chapter on survey research. For the most part, however, the authors provide an overview of their approach illustrated with examples from full-length studies. Most chapters also systematically assess what sorts of questions the method is appropriate to explore. This attention to strengths and weaknesses has tremendous practical value, and the authors take seriously the appeal for methodological pluralism by suggesting where other methods might pick up where the method under review leaves off. The chapters also address the value of single case studies and paired comparisons as well as large N analyses, multi-country studies, and formal models. Each chapter includes an individual bibliography that lists key texts for that methodology as well as good examples of scholarship that employ it.

This volume has several important strengths. In terms of organization, it is a welcome relief that the editors did not divide the various methodologies between quantitative or mathematical approaches, on the one hand, and qualitative or interpretive approaches, on the other. This sort of binary distinction, which is unfortunately common, obscures the wide diversity of approaches. In this regard, three important themes run through the various chapters: unit of analysis, methodological diversity, and use of empirical evidence.

Social movement studies might appear to newcomers as a cohesive field in which investigators agree on "social movements" as the unit of analysis, but the chapters in this volume illustrate the diverse foci of social movement research. For example, some scholars take individual movements as the unit of analysis and undertake intensive studies of the "lives" of particular movements, from their inception through their rise and later decline. This approach tells us a great deal about movement dynamics, but it necessarily underplays attention to other dimensions of mobilization. Completely different pictures emerge when the focus is shifted to networks, events (including protest events), narratives, macro-organizational structures, or other units of analysis. And studies that focus on a single unit of analysis can be explored through a range of methods: networks, for example, can be evaluated using archival data, in-dept interviews, sampling, surveys, and matrixes. Happily, this volume captures and emphasizes this diversity.

The chapters also explore the practical issue of building theory through use of empirical data. In this regard, the contributions provide lots of practical ideas that will be useful for students as well as established scholars. Print media, for example, have proven extremely useful for the study of various dimensions of social movement activities, but what can they provide evidence of, and how can that evidence be analyzed? The contributions to this volume illustrate that newspapers have been analyzed through computer event-data coding with extremely interesting results, but they are also important resources for discourse analysis, case studies, and participant-observation field work. Likewise, the role of culture and identity has be explored through discourse and frame analyses focusing on the interpretations of the text, but elements of culture can also be examined by coding texts for content analysis and modeling narratives as a form of opportunity structure.

In this regard, the central plea of the volume is for methodological pluralism—not as a peace offering, but because methodological triangulation provides an extremely important mechanism for testing hypotheses, revealing political processes, and addressing the limits of every methodological approach. The focus on middle-range theory may disappoint scholars interested in generating the broad covering laws of nomothetic theory, and ethnographers who reject comparative methods altogether will likewise be frustrated. But the contributors to this volume make a strong case for promoting methodological pluralism: not only through tolerance, but ideally through cross-method collaboration.

The Encyclopedia of Social science Research Methods: A Preview

Michael S. Lewis-Beck
University of Iowa
michael-lewis-beck@uiowa.edu

In the ancient Greek, the word "encyclopedia" means an "all-encompassing education." The forthcoming *Encyclopedia of Social Science Research Methods* aspires to provide something of an all-encompassing education within its sphere, including qualitative as well as quantitative approaches to social science research. To be published by Sage, fall 2003, it will consist of about 2400 pages, in four volumes. Editors are myself, and two leading sociologists, Tim Futing Liao (University of Illinois and University of Essex, England) and Alan Bryman (University of Loughborough, England). The Editorial Board is interdisciplinary, composed of internationally known researchers in economics, education, history, political science, psychology, public health, and sociology. On the board from political science are Nathaniel Beck (University of California, San Diego) and Michael Alvarez (California Institute of Technology). While contributors have come from all the social sciences, many are from political science. Suzanna De Boef, Editor of this newsletter, as well as Editorial Assistant Heather Ondercin, are among those in our discipline who have made significant contributions.

There are over 1000 contributions, from Abduction, the first entry, to Z-test, the last entry. No matter the data source — archives, biographies, content analysis, conversations, documents, diaries, experiments, field observations, informal interviews, informants, simulations, surveys, yearbooks — it receives coverage. Further, all methodological approaches, quantitative or qualitative, big or small, mainstream or not, are entertained. Thus, for example, there are entries on subjects as diverse as Action Research, Behavior Coding, Case Study, Delphi Technique, Ethnography, Feminist Research, Game Theory, Heuristic Inquiry, Impact Assessment, Jackknife Method, Key Informant, Logical Positivism, Macro, Narrative Analysis, the Other, Participant Observation, Q Sort, Random Number Generator, Stochastic, Total Survey Design, Unit of Analysis, Venn Diagram, Weighting, X variable, Yule's Q, Zero-Order. For the quantitative-minded, as most readers of this newsletter are, statistical topics are abundantly treated. Here is a smattering: ARIMA, Bayes Factor, Causal Modeling, Degrees of Freedom, Error Correction Models, Factor Analysis, General Linear Modeling, Heteroscedasticity, Identification Problem, Joint Correspondence Analysis, Kurtosis, Least Squares, Markov Chain, Nonparametric Statistics, Outlier, Parameter Estimation, Quadratic Equation, R-squared, Seemingly Unrelated Regression, t-distribution, Unit Root, Varimax Rotation, Weighted Least Squares, Yule's Q, Z-score. From these examples, one sees that complicated, state-of-the-art topics are treated, as are the simpler, but important, bread-and-butter topics.

While each of us may be very knowledgeable in some, or even many, of the myriad of entries, there probably remain areas where we need to learn more, either for our own work, or to better guide our students. The very first encyclopedia, *l'Encyclopèdie*, edited by Diderot and d'Alembert, and published in France beginning in 1751, had several goals. The first was "to examine everything" (Il faut tout examiner). While we certainly cannot claim to have examined everything under the sun in the world of social science research methods, we have nevertheless spaded a good deal of ground. A second goal of the original encyclopedia was to provide readers with more scientific tools, so they could think for themselves. We hope we have done that.

Teaching and Research

A Solution for Testing Students in Math-oriented Classes

Kristin Kanthak
University of Arizona
kanthak@u.arizona.edu

Our students never believe us, but it is true: We hate giving examinations as much as they hate taking them. And the rule is doubly true in political science classes with a mathematics orientation. Nervous students crowd faculty offices the day before the examination, bleary-eyed with tales of all-nighters. We try not to make quick movements, for fear we might set off imminent tears. Students worry they will make a mathematical error in line 3 of their work, making their answer in line 27 nonsense. We worry that we will have to dig through 24 lines of nonsense to figure out where the mistake was. They spend long periods of post-exam waiting time wondering what their grade is, while we muscle our way through grading all n of the exams. Something has to give.

This semester, it gave. I currently have 55 undergraduates enrolled in a game theory class. I don't have anywhere near the muscle to grade 55 mathematics-oriented exams. Happily, there is another way: the "multiple-choice plus" exam. It is a traditional multiple-choice exam, but the "plus" is that students must provide a two-to-three sentence justification for each of their answers. If they know the answer is right, the justification is a recitation of how they found the answer, often simply the equivalent of "showing their work" on a traditional exam. If they find the answer by process of elimination, they explain how they eliminated "bad" answers. If they are truly guessing, this is readily apparent in their justification (or lack thereof).

A multiple-choice plus exam looks exactly like a traditional multiple-choice exam. In my classes, the midterm exam comprises 10 questions, and I provide three or four possible answers for each question. The final is the same format, with double the questions (my students have double the time for final exams). Students receive five points for each correct selection, and five points for the justification, for a total of ten points per question. Therefore, a student who selects the wrong choice, but provides a perfect justification (I'll give an example of how this is possible below) gets the same number of points as a student who makes the correct selection, but cannot tell me why it is right.

There are six key reasons why I like the multiple-choice plus format, and why it is a saving grace this semester, both for me, as well as my 55 budding game theorists.

It decreases math anxiety. Most political science students are fearful of doing mathematics, particularly in the already anxiety-inducing context of an exam. Indeed, many of them have not taken a math class since high school and specifically selected their major in the hopes of avoiding math. They often have no way of confirming their answer on an exam question is correct, and thus spend endless exam time redoing simple calculations that they probably had correct the first time. I would rather they spend that time thinking about the topic of the class, not arithmetic.

And multiple-choice plus, like any multiple-choice exam, does that. The student calculates the answer, and then consults the list of potential responses. If the student's answer is among the choices, she simply notes it, writes the justification, then moves on. If the answer is not there, she knows she has made a mistake. If she has time, she can go back and figure it out. If not, she can describe what she did in the short response section of her answer, thus allowing at least partial credit. Beyond this, simply knowing that verification of an answer is forthcoming likely decreases the anxiety students feel while they are finding the answer. Possibly, this will decrease the chances that the arithmetic error will arise in the first place.

It avoids the "search for the dropped sign." I think I am like most professors in a mathematics-oriented political science class in that I believe understanding the material is more important than being able to do simple arithmetic without error. Though I think the latter is important as well, I am quite comfortable leaving the job to second grade teachers, who are better at teaching arithmetic than I am, anyway. To that end, I believe a well-graded exam in a math-oriented political science class requires that we score math errors as less grievous sins than errors that arise directly from a lack of understanding of the course material. This goal seems clear, but implementing it for a traditional mathematics-oriented exam is tricky, and, more frustrating, extremely time-consuming. I end up poring through hastily-written lines of math, searching for the moment that a "-1" became a "+1," or, often, finding that what looks at first like a simple math error was actually more severe.

Because students know immediately during a multiple-choice plus exam when they have made errors, they do not hand in examinations with incorrect answers, thus leaving the instructor to find the error. They find the error themselves. If they cannot do so, they describe what they did. I can read the description, not the math. If it is right, that translates to partial credit.

It gives credit for educated guesses. One of my pet peeves associated with traditional multiple-choice exams is how to handle guessing. Often, students are not sure of the answer, but are able to eliminate some of the responses. Or, students know exactly how to solve the problem, but have made some pesky math error, so their response does not match any of those given, thereby requiring them to guess. Educated guesses are indiscernible from incorrect guesses made with no priors. And I am not confident that the central limit theorem will kick in at only ten questions per exam.

The multiple-choice plus format allows students who are not sure of the right answer, but have some information that is relevant to the question, to get credit for the information they *do* have. For example, following is a verbatim response from an undergraduate in a previous game theory class, when asked to find a mixed strategy Nash equilibrium: "I know probabilities go from 0 to 1, and my probability is negative, so it is not right!! I messed up somewhere, but I can't find it. The big payoff for Down, Left for Column has me thinking that Row has to play Up A LOT to make Column indifferent, so I don't think A and B are right." The student guessed wrong, but got a full five points for the justification. In my estimation, she showed a greater understanding of how mixed strategy equilibria work than a student who had correctly applied the formula. A traditional multiple-choice exam would have resulted in a score exactly the same as that of a student who knew nothing about the problem.

It decreases credit for lucky guesses. The other side of the guessing coin is that some students get lucky and pick the right answer, even when they have no idea how to solve the problem. And again, the central limit theorem is cold comfort with only ten questions per exam. Furthermore, it seems that students often greatly exaggerate the probability of their selecting the correct answer on a traditional multiple-choice exam. The fact that these students know that half their points come from correct justifications may make them likely to study harder.

Further, when grading multiple-choice plus, I can differentiate between random guessers and good students with bad luck. Another verbatim response from a game theory undergraduate on a multiple-choice plus exam: "A comes first, so I pick A." That is zero points for the justification (I give no credit for knowing the alphabet), even though A happened to be the right answer. So this student received the same amount of credit for the question as the mixed-strategy student above. Certainly, one can argue that the student with the good justification deserves *more* points, but remember that a traditional multiple-choice test would have given her zero points, while "A comes first" would have gotten full credit. And more important, at least in my estimation, I did not have to take the time to work through her incorrect math to determine that she understood the material upon which she was being tested.

It is a snap to grade. One of the most popular features of traditional multiple-choice exams is the ease of grading. Simply compare the students' answers to the correct ones, and a ten-question exam is graded in a few seconds. Multiple-choice plus is not quite that simple, but grading is nowhere near as tedious as that associated with exams with open-ended questions.

When I grade the exams, I first grade each one like a traditional multiple-choice exam. Then when I reread the exams to check the justifications, I already know which answers are correct, which is a majority of the responses in a majority of the cases. I know that reading these answers can consume much less time. In most cases, if they got the answer correct, they have a correct justification, meaning that I can lightly skim these justifications to assure that there are no problems. I'll often take off points if a statement in the justification is blatantly wrong, or if it is clear the student did not know how to do the problem. Then I spend more time on the incorrect justifications, so that I can point out to students exactly where their understanding of the material falls short. This operation allows me to grade exams quickly, while assuring that information gaps are clear when the students get their graded exams returned to them.

Students like it. Obviously, popularity with students is not the top reason to implement a pedagogical tool. But in this case, it is a win-win scenario. Students often comment on teaching evaluations that they like this form of testing. I am therefore convinced that using multiple-choice plus increases my teaching evaluation scores. At the same time, multiple-choice plus saves me time, both before the exam by decreasing the number of office visits from overwrought students, and after the exam by cutting down on grading time. Further, I think students learn the material better from taking multiple-choice plus exams than from other types of testing. And, it is a freebie in terms of increasing my teaching evaluations. My evaluations go up, my students learn more, and I free time up that I can dedicate to research. Particularly as a junior faculty member, that's gold.

Tips for the First Time Experimentalist

Rebecca B. Morton
New York University
rebecca.morton@nyu.edu

Increasingly I (and other experimentalists) get an email, phone call, or office visit from a graduate student or a colleague that begins with the following statement: "I have an idea for an experiment, can you give me advice on how to do it?" Getting this question is great, as I love to see the use of experiments in political science research and have argued in a previous issue of this publication and elsewhere the benefits for the discipline from increasing experimentation (Morton 1999). Nevertheless, answering this question is never easy and when the editor of *The Political Methodologist* suggested a short article with tips for graduate students and other first time experimentalists, I both jumped at the chance and worried about whether what I would say would be adequate. Given that caveat, below is how I would answer the question:

- Tip 1. Carefully identify the research question you want to ask and in what way you plan to use experiments to answer that question.

One of the main differences between experimental research and research using naturally occurring data is that in most experiments the design stage of the experiment in a sense replaces the need for a complex statistical analysis of the data - the statistical theory goes in the front end of the research rather than later after the data is generated. Ideally, the researcher uses control and randomization such that the data generated by the experiment can be almost "clean" and easy to analyze. This means that the design stage of the experiment is crucial and not to be treated haphazardly. It is beyond the space of this short essay to discuss the statistical theory of experimental design and many good texts on the basics of design exist, see for example Campbell and Russo (1999), Davis and Holt (1993), and Friedman and Sunder (1994). A first time experimenter should consult these texts.

Designing the experiment will also require you to think carefully about who the subjects will be in the experiment (will you use undergraduates or draw from a pool of subjects outside of the university?), how you will recruit the subjects, how you will pay the subjects, and whether you plan to use deception or not in the experiment. The answers to these questions depend on the research question and the reasons behind using an experiment to answer that question. It is also important that the researcher consider at this time the ethics of his or her proposed research design — that is, experimentation involves using real human subjects and applying a manipulation to them and thus a moral responsibility to treat the subjects with respect and not to endanger them without their consent. While experiments conducted by political scientists have not achieved the notoriety of some medical and psychological experiments of the 20th century (for example the Tuskegee syphilis experiments), many worry about the use of deception in some laboratory experiments as well as the potential of manipulation of real world political outcomes in more recent field experiments. If you plan to use deception or manipulation in such a way as to have measurable real effects on your subjects, you need to think carefully about the ethics of your design.

- Tip 2. Find out the requirements for doing human subjects research at your university and begin the process of getting approval for your research.

In order to run an experiment you will need approval from your university's human subjects review board even if you plan to use your own students in a class you teach and the experiment is not funded by any outside resources (unless the goal is purely educational and not related to a research question). Because of the problems that medical experiments have generated and the complex ethical issues involved in human subject research, universities' review boards can be fairly strict and rightly so. It is a good idea to begin to investigate what is required at your university for human subject research. For example, at NYU all activities involving human subjects, whether funded or non-funded, including dissertations, master's theses, pilot studies, class projects, and non-funded faculty-directed research must be reviewed and approved by the university's Institutional Review Board prior to the commencement of the research if the proposed research meets any of the following conditions:

- The research is sponsored by the University, or
- The research is conducted by or under the direction of any University employee, or agent (for example, faculty member, researcher, or student) in connection with his other institutional responsibilities or
- The research is conducted by or under the direction of any University employee or agent (for example, faculty member, researcher, or student) using any University property or facility or
- The research involves the use of the University's non-public information to identify or contact human research subjects or prospective subjects or
- The research involves the use of the University's students, employees or facilities.

Researchers at NYU are required to take an online tutorial on human experimentation and to pass an online exam to qualify for approval. Many other universities have similar setups and as this can take some time (it is not uncommon for review boards to ask for revisions or clarifications on experimental designs before granting approval), the sooner you begin the process of securing human subject approval the sooner you will be able to conduct your research.

- Tip 3. Examine the existing experimental literature on similar research questions.

In many cases, experiments related to your idea may have been already run. This is particularly true if the basis for your experimental idea is a public goods game, a prisoner's dilemma game, a bargaining game, or a voting game. Unfortunately, this may take more than the usual literature search since experimental research crosses disciplines and important experiments on many of these games can be found in sociology, psychology, and economics journals as well as in political science publications. This is true even if the theoretical perspective underlying the model is generally thought of as discipline specific — e.g. experiments in public goods games have been run by researchers from all four of these disciplines. Good summaries (although quickly becoming out of date) of the literature on economics experiments are found in Kagel and Roth's *Handbook of Experimental Economics.* Davis and Holt, mentioned above, also review the economics literature. I don't know of such an up to date handbook for the other disciplines, although one may certainly exist.

The main point is that most simple straightforward experimental designs of political situations like public good games have been run even though the literature may not be well known in political science circles. Therefore, the literature search required for a potential experimenter is more complicated than for a project in an area that is more directly defined as a political science research area. That said, it has also been my experience that even in this case (i.e. the basic underlying game or problem for the experiment has been evaluated in a number of ways before by researchers in other fields) the approach the political science researcher takes to the problem is generally distinctive enough to merit more experimentation, so this is not meant as a discouraging step, merely a highly necessary one so that you don't waste time (and money) rediscovering known facts. Moreover, the existing literature may help you see a better way to examine the research question you have identified or lead you to a better question to ask.

- Tip 4. Write up the instructions for subjects and run a "trial" with friends and/or willing students, not for publication.

Friedman and Sunder (1994) provide examples of instructions and many published experimental papers should contain the instructions used in appendices. More than likely in your literature search (discussed above) you have found similar experiments to your own design. Using instructions that are based on these can allow you to use the earlier results in an informative way to analyze your own results and increase the applicability of your work. Since experiments are costly both in terms of money and time, trial runs are crucial to eliminate bugs. Furthermore, if you are seeking outside funding for the experiment, you are more likely to receive this funding if you have some preliminary results or trials to report as well as detailed instructions for the experiment.

- Tip 5. Get advice from an experimentalist.

While the literature search above can help tremendously as well as the references mentioned here on experimental research, once you have a design and a trial run, before going forward with the experiment, it is a good idea to "float" what you have done so far to experimentalists who are working on similar problems or may be interested. If you plan on applying for resources from say the NSF these researchers are likely to get the proposal anyway and knowing their comments in advance cannot only help you conduct better research but also increase your probability of getting funded. You may want to attend the meetings of the Economic Science Association or the International Society of Political Psychology, both groups hold their annual meetings during the summer, in order to make contacts with experimentalists. Finally, good luck!

References

Campbell, Donald T. and M. Jean Russo. 1999. *Social Experimentation.* Sage Publications: Thousand Oaks, CA.

Davis, Douglas D. and Charles A. Holt. 1993. *Experimental Economics.* Princeton University Press: Princeton, NJ.

Friedman, Daniel and Shyam Sunder. 1994. *Experimental Methods: A Primer for Economist.* Cambridge University Press: New York.

Kagel, John H. and Alvin E. Roth, eds. 1995. *The Handbook of Experimental Economics.* Princeton University Press: Princeton, NJ.

Morton, Rebecca B. 1999. *Methods and Models.* Cambridge University Press: New York.

Some Thoughts on Graduate Political Science Curricula in Quantitative Methods

Christopher Zorn[1]
Emory University
czorn@emory.edu

Introduction

A little over two years ago, the political science Ph.D. program at Emory faced a decision of a sort becoming increasingly common in our discipline. At the time, Emory students were required to take a single quantitative data analysis course, one which covered basic univariate and

[1]My thanks to Cliff Carrubba, Mike Giles, Jeff Gill, Andrew Martin, Eric Reinhardt and Dan Reiter for their discussions about quantitative methods pedagogy.

bivariate statistics and "regression appreciation". Over the years, in addition to offering more and more regular courses on topics beyond that covered in the required course, the department had frequently sent students to both the ICPSR and Essex summer programs. At the same time, there was some question about whether our students were receiving the training they needed.

In the fall of 2000, in the face of the increasing level of sophistication in quantitative methods among our students' job–market competitors, we considered revising our decades–old single–course requirement to require a second. That proposal (advanced by two junior faculty members, and ultimately defeated) raised a host of questions at the core of graduate pedagogy in quantitative methods, and our field more generally: questions relating to the proper role of methods in political science graduate education, to the relative pluses and minuses of mandated courses versus electives, and to more specific issues of course organization, structure, and content. Some of the fruits of those discussions are presented here.

What Do We Do?

Its fair to say that, as a subfield, political methodology is among the most cognizant of pedagogical concerns. Articles in *TPM* routinely address questions related to teaching, and teaching–related issues appear regularly on H-PolMeth as well. At the same time, I suspect we all feel we could be doing what we do, better; this is particularly true in the area of graduate methodology training, often the most important contribution readers of this newsletter make to their department and to the discipline. But how might we get better? One place to start the inquiry is to determine exactly what it is we do when we teach nascent political scientists quantitative methods. To get a sense of the way(s) political science departments structure their graduate course offerings in quantitative methods, I conducted a short, highly unscientific "survey" in January and early February of 2003. In brief, I sent e-mails to people involved in teaching such courses at 26 Ph.D. granting departments ranked in the "Top 25" by *U.S. News and World Report.*[2] The survey itself consisted of four short questions:

1. How many courses in quantitative methods/statistics are your Ph.D. students *required* to take? (Note: Your answer should NOT include courses in formal/game theory, or in research design /epistemology).
2. What material does your "first" course in quantitative methods cover?
3. What "advanced" courses (i.e., those beyond linear regression) does your department *regularly* offer?
4. In what other departments/programs, if any, do your students commonly get "outside" training in statistics?

Of the 26 departments surveyed, 22 responded (85 percent – a rate the survey research folks tell me isn't too shabby).[3] Their responses are enlightening, and provide both answers and additional questions.

Quantity and Content

At least two things are immediately apparent from the responses to the first two survey items. First, there is a remarkable balance to the answers to question one: of the 22 departments, six had no required statistics course, eight had one, seven required two courses and one (SUNY–Stony Brook) required four.[4] Many falling into the first category are departments which adopt a laissez–faire approach to coursework more generally, requiring few courses in any field. Second and related is the relationship between the number of required courses and the content of that course (or courses). Here, two models become apparent.[5] The first, which I'll call the "Big Tent" approach, predominates in departments producing students across all the various subfields and approaches of political science. As a result, "Big Tent" curricula are designed (and I use the term loosely) to address the needs of a highly heterogenous group of graduate students, from the committees–in–Congress gang to students of LGBT influences on the political culture of 11th–century rural Mongolia, or the ethnography of prison rodeos. As a result, such courses necessarily endeavor to be all things to all people; in particular, they are forced to strike a balance between being a comprehensive (and nontechnical) introduction to data analysis and serving as a math primer for more quantitatively–oriented students. The second approach treats the first course solely as an introduction to the math necessary for more advanced courses; call this the "Boot Camp" model. These introductions cover little

[2]The selections were based on the 2002 rankings; there was one tie.

[3]Thanks to my friends and enemies at Michigan, Washington University, Princeton, Rochester, Yale, Wisconsin, Minnesota, Columbia, Indiana, Stanford, UCLA, Ohio State, Chicago, Harvard, University of Washington, Northwestern, Duke, UCSD, MIT, Illinois, SUNY – Stony Brook and Cornell for their responses. I'm happy to share the information from this survey with anyone who might be interested.

[4]This finding is similar to that of Burns (1990), who found that 15 of the 22 departments she surveyed more than a decade ago required at least one course in methodology.

[5]Plutzer (2002) finds five models extant, though I suspect his powers of discrimination are greater than mine.

or no regression, focusing instead on probability and distribution theory, calculus and linear algebra, and statistical inference; they are almost always offered in departments that require a second course (typically, a course on some variant of regression). In addition, these courses tend to occur more frequently in "boutique" departments, particularly those which place an emphasis on American politics and/or quantitative and game–theoretic approaches to the study of politics. Herron (2002) provides a nice description of such a program. There are, of course, variants on these two types, as well as important institutional differences (quarter versus semester systems, etc.). A few schools, for example, offer pre–first–semester "math camps" covering basic maths and other topics; a few others integrate research design topics into their first statistics course. On balance, however, most "first courses" emphasize introductory data analysis, a necessity given the inconsistent backgrounds and aptitudes of most first–year political science graduate students.

Beyond OLS

Responses to the query regarding "advanced" classes were also interesting. Every department surveyed save two[6] offered at least one "advanced" course on a regular basis.[7] The modal and median number of such courses offered was two, with a maximum of four; note, however, that a number of departments offer generic "topics" courses which vary in subject matter from one term to the next, and that these were counted as a single course here.

The most common such course by far was some variation on maximum likelihood estimation, typically covering generalized linear models (logit/probit, event count models, and so forth). Other specific topics, though mentioned far less frequently, are nonetheless a testament to the wide range of approaches quantitative political scientists have adopted in their work (see Table 1). And, not surprisingly, the topics offered tend to track closely with the interests and technical abilities of a department's faculty. Finally, a number of respondents indicated their intention to expand their department's offerings in the future; such expansions tended to be geared towards contemporary "hot topics" such as measurement models, nonparametrics, and Bayesian approaches.

[6] And one of these was a member of the ITV consortium; see below.

[7] "Advanced" appears in scare quotes since the definition of that term is itself highly variable across departments. Generally, "advanced" corresponds to elective courses; for most, this is equivalent to "post–OLS regression," but in a few cases – those where the only one course is required, and where that course offers little or no coverage of regression techniques – the regression class itself is considered "advanced".

Table 1: **"Advanced" Topics Courses Regularly Offered**

Topic	Frequency
MLE / GLMs	16
"Regression"	8
"Topics"	5
Time Series Analysis	4
Bayes/MCMC	3
Panel / TSCS Models	2
Structural Equations	1
Measurement Models	1
Survival Analysis	1
Computational Models	1
Dimensional Analysis	1

Table 2: **Departments where Political Scientists Learn Statistics**

Department/Program	Frequency
Economics	17
Statistics / Applied Statistics	14
Sociology	6
Psychology	5
Business School	3
Biostatistics / Biometrics	2
Mathematics	2
Public Policy	1
ICPSR	1

Going "Off the Reservation"

Political scientists have long gotten methods training outside of their field, and that tradition continues today. At the same time, my survey found wide variation in both the extent and the location of such training. With respect to the frequency of students' enrollment in such courses outside of political science, responses ranged from "they don't go outside much" and "rarely" to "nearly all of ours students take at least one outside methods course." In general, however, outside training appears to be more the exception than the rule.

When they do go outside their department, the survey found that students are most likely to take courses in economics, with statistics running a close second. The first is not surprising; the second is somewhat more so, particularly given that significantly fewer survey respondents mentioned psychology, sociology or business as loci for their students' graduate statistics training. Also, while only one department mentioned the summer program at ICPSR (and none the program at Essex), those programs also draw a significant number of students from both the surveyed and other departments.

Implications, Innovations and Suggestions

The results of this survey are suggestive of a number of matters facing graduate methods curricula in political science; here, I'll mention just a few.

"Two–Tracking", or, What about the theorists?

I believe the key issue facing most graduate programs' methodology training is the question of balance. As noted above, often the most significant challenge for the instructor of a first–semester methods course is how to provide the mathematical rigor necessary for students who will (by force or by choice) go on to more advanced courses while still offering a broad (and nontechnical) enough overview of data analysis techniques for those who will not. One solution is simply to ignore one subpopulation or the other; that is, teach (e.g.) a "Boot Camp" course and let those with no need for or interest in additional courses shift for themselves. This is certainly an efficient approach, but perhaps not one likely to endear us to our substantive colleagues (not to mention the graduate students themselves). A different approach is "two–tracking" of students. This approach takes many forms, from requiring one or more "tools" courses which can be fulfilled by language proficiency to simple exemptions from quantitative methods requirements for (e.g.) normative theorists. The latter approach is both relatively common and one which offers benefits to all concerned; one respondent stated flatly that "(W)e don't force the squishy folks to do a serious stats course, which makes them happy and also makes life easier for instructors." In addition, at least one full–service department in the survey is currently moving to an explicitly two–track system: students will choose a first course of the "Big Tent" or "Boot Camp" varieties, depending on their plans for future coursework. Provided one's department is big enough to offer such courses regularly (typically, every autumn), and that students get good advice about the proper "track" given their substantive interests, this is an attractive alternative.

Other Issues

At least two other matters deserve a brief discussion. First, it's important to bear in mind that this survey was limited to quantitative methodology courses; it tells us nothing about qualitative methods, game theory, and the like. As I mention above, qualitative methods in particular ought to be incorporated as a "full partner" into graduate methods curricula, particularly as an increasing number of departments (including, happily, my own) begin to offer such courses on a regular basis. At Emory, for example, we recently restructured our minor–field qualifying exam in methodology: the exam now covers three fields (game theory, qualitative methods and statistics), and students choose two of the three in which to take the exam. The result is the active inclusion of a broad range of types of scholars into the fold of "methodologists."

Another major issue facing political scientists is the inability (with rare exceptions) of any single department to provide the breadth and depth of training needed today. The statistical "arms race" taking place even among primarily substantively–focused junior researchers makes it difficult if not impossible for even large departments to offer courses across the range of possible topics.

What to do? Clearly, going outside one's department provides one alternative, albeit one with tradeoffs (see e.g. Smith 1992). ICPSR and Essex are also good alternatives, for departments able to send their students there. Some universities have created cross–disciplinary institutes which, among other things, offer courses in statistics and research methods; the University of Washington's *Center for Statistics and the Social Sciences* (`http://www.csss.washington.edu/`) is a prime example. A somewhat different program is the Interactive Television Program in Advanced Political Methodology at Ohio State, Minnesota, Illinois and Wisconsin (Box–Steffensmeier et. al. 1997; see also the course website at `http://psweb.sbs.ohio-state.edu/faculty/jbox/ITV/ITVHome.html`). ITV courses have addressed such topics as duration models, measurement, cross–level inference and even qualitative methods, and provide an efficient alternative to replicating the same course across several institutions.

References

Box–Steffensmeier, Janet M., John Freeman, Kathy Powers, W. Phillips Shively and Brett Sutton. 1997. "The Multi-site Interactive Video Curriculum in Advanced Data Analysis for Political Science." *The Political Methodologist* 8(1):2–8.

Burns, Nancy Elizabeth. 1990. "Methodology in Graduate Political Science Programs." *The Political Methodologist* 3(1):9–10.

Herron, Michael. 2002. "Teaching Introductory Probability Theory." *The Political Methodologist* 10(2):2–4.

Plutzer, Eric. 2002. "First Things First (whatever "first" happens to mean): Syllabus Choices for Statistics I." *The Political Methodologist* 10(2):4–6.

Smith, Charles E. Jr. 1992. "Graduate Training Outside the Department: Taking Full Advantage of the Methodological 'Trade Deficit'." *The Political Methodologist* 5(1):7–9.

Computing With R

R For the Political Methodologist

Simon Jackman
Stanford University
jackman@stanford.edu

Nine years ago I contributed an essay to *The Political Methodologist* comparing `GAUSS` and `Splus`(Jackman 1994). `Splus` was the new kid on the block at that point; in the early 1990s, I could literally count on one hand the number of political scientists using `Splus`.

What a difference a decade makes. `Splus` made serious inroads into quantitative social science (and other fields) through the 1990s. And now its free competitor, `R`, is widely used in political science methods classes, and is now the package-of-choice in the advanced classes at the ICPSR Quantitative Methods Summer School (see the accompanying article in this issue of *The Political Methodologist*). To clarify: `Splus` is a commercial product, a superset of the `S` language for data analysis developed at Bell Labs. `R` is largely a "GNU `S`", developed by some of the same people who developed `S`, plus a large group of public-spirited statisticians and programmers (many of whom had contributed libraries to `Splus`). Basically, the goal of the `R` project was (and remains) to take the `S` language to the masses, using many features of `S` as the foundation of an open-source and freely-available statistics package.

In this short essay I give a brief overview of `R` as a tool for advanced researchers and methodologists. As I did in 1994, let me stress at the outset that comparing software for the *TPM* readership is asking for trouble: it is curious how zealous or protective some of us get about the tools of our trade. But the trade and the tools evolve, and getting emotionally invested in a piece of software is just not worth it (and frankly, reflects something more serious than a methodological problem, but I digress). I've used lots of different programs thus far, and imagine that I'll keep doing so over the years ahead. And at any given time point, its extremely rare that I find one package does it all. This is especially true for methodologists (we're more or less expected to know our way around lots of types of data and models, and to keep current with new approaches), and so while `R` does many things for me, it doesn't do quite everything: every so often I'll use `SAS` for getting a massive data set in shape for analysis (although I've been able to do most of my data management in `R` in recent years); when I want to implement a Markov chain Monte Carlo approach (aka "Bayesian simulation") I use `WinBUGS`, or, for my work on roll call data with Josh Clinton and Doug Rivers, I use my own `C` program, `ideal` (see `http://jackman.stanford.edu/ideal`).

At the same time that `Splus` and `R` have been making headway into the social sciences, so too has `STATA`. `STATA` has the strength of being fairly easy to use, with many models and options from econometrics implemented and added with successive releases. For instance, for panel data, `STATA` sports an impressive array of models, options and post-estimation commands (the `xt...` family of commands). In general, the kinds of things we encounter in intermediate to advanced econometrics texts are more likely to have been implemented in `STATA` than in `R`: e.g., a Cochrane-Orcutt or Prais-Winsten correction for auto-correlated residuals in a linear regression, Durbin's h statistic, Newey-West standard errors, Heckman's two-stage selectivity estimator. But frankly, it is surprising is how much econometrics is in `R` across various user-contributed libraries, given how little econometrics there was in `Splus`, at least in 1994: for instance, in the `strucchange` library one can find Huber-White heteroskedasticity-consistent standard errors, recursive residuals and Chow tests; in the `lmtest` library there are many tests of iid disturbances and linear functional form from the econometric literature (e.g., Breusch-Pagan and Goldfeld-Quandt tests of heteroskedasticity, and the Durbin-Watson and Breush-Godfrey tests for autoregressive disturbances); in the `tseries` library there are ARCH and GARCH models, and numerous tests of stationarity, normality, and non-linearity; tests of general linear hypotheses are supported in the aptly named `gregmisc` library; the `systemfit` library implements systems of equations methods (SUR, two and three stage least squares; see also the `sem` library); the Kalman filter and ARIMA modeling can found in the `ts` library, along with many other smoothers and filters for time series. Since many political scientists learn data analysis via econometrics, they will be pleased to know `R` supports many of the models and specification tests from that literature.

Since `R` is modeled on `S`, it shares many of the strong points of `Splus` that I wrote about nine years ago: some object-orientated design features, a strong emphasis on graphics and visualizing data, and a steady flow of innovation (both computational and statistical) from the applied statistics community. For example, consider the following R code fragment, with comments:

```
plot(y ~ x)      ## scatterplot
identify(x,y)    ## interact with scatterplot,
                 ## click to identify points
reg1 <- lm(y ~ x)  ## reg1 is a linear
                   ## model object
abline(reg1)     ## overlay regression line
                 ## on scatterplot
plot(reg1)       ## plot knows what to do with
                 ##  reg1(diagnostic plots)
```

The `plot()` command is generic, and sees that it has been passed (1) a formula, in the first call to `plot()`, and so produces a scatterplot; (2) a fitted regression in the second call, or more precisely, an object of class `lm`, and so hands off to `plot.lm()` which produces a series of diagnostic plots (partial residual plots, residuals against fitted, and influence statistics). Many other classes of R objects have a `plot` method, including data frames and matrices (all pairwise scatterplots), discrete or character variables (barplots). My current favorite is from the `MCMCpack` library contributed by (political science's own) Andrew Martin and Kevin Quinn: when passed an object of class `mcmc`, `plot` will produce posterior density plots and diagnostic traces for parameters estimated via Markov chain Monte Carlo. Another useful generic command is `summary()`. In short, the goal of this object-oriented approach to program design is to reduce the workload for the user: one command does lots of different things, with the software "smart enough" to figure out what the command means in different contexts.

Next to its price, one of the key reasons to use R over Splus is the ease of writing functions. R has much more permissive scoping rules than Splus (*lexical* scoping versus *static* scoping in Splus, meaning that one need not worry about whether variables used by a function have been declared "local" or "global" or whatever). This amplifies one of the great strengths of packages like Splus and R: *user-extensibility* or "writing your own programs". Many quantitatively-inclined political scientists will never have to extend the functionality of their statistical software, but that is probably less so for readers of *TPM*. Indeed, if your notion of data analysis runs to more than estimating coefficients and t-statistics, or if a write-up is more than cutting-and-pasting tables of estimated parameters, noting which have more "stars" (asterisks) than others, then from time-to-time you'll find yourself programming, if only a little. Big substantive payoffs often come from computing "auxiliary quantities of interest"[1], but knowing just what those quantities might be in any given application is hard to know in advance, meaning that pre-programmed functions are only going to get you so far. Moreover, we should be driving the software, not the other way around: the methodological frontier should not be the drop-down menu of program P: or, as I said in 1994 (if a little too earnestly):

> once methodological problems start being perceived or even defined in terms of what one's favorite software does well, then the software has stopped being a tool, and has become a crutch, and at worse a shackle.

The implication is that easy programming and flexibility is key for a serious statistical computing environment.

To demonstrate function writing in R, suppose we want to compute the area under the receiver-operator characteristic (ROC) curve, a goodness of fit measure for binary classifiers (including, as a special case, logit and probit models), and closely related to other measures of association such as Somer's D, Gamma, and τ_a. These goodness-of-fit measures are already present in R via Frank Harrell's `Design` library, but let us proceed for purposes of exposition. Say we have a model that produces $\hat{p}_i = \widehat{\Pr}(y_i = 1)$ and we classify $\hat{y}_i = 1 \iff \hat{p}_i > k$ and $\hat{y}_i = 0$ otherwise, for some threshold $k \in [0, 1]$. The ROC curve plots the TPF (true positive fraction, TPF = 1 - FNF), versus the FPF (false positive fraction) as k varies over $[0, 1]$. The resulting function is defined on the unit square, and the area under the ROC curve C is interpreted as a measure of the classification success of the model. A value of $C = .5$ indicates random predictions and a value of $C = 1$ indicates perfect prediction. This is a useful and widespread measure of classification success and one that we'd like to be able to compute any time we fit a binary response model, suggesting that coding it up as a function is worthwhile. Suppose we've fitted a binary response model in R using the usual `glm` command, for generalized linear models, e.g.,

```
logit1 <- glm(y ~ x, family=binomial)
```

producing an object of class `glm`. The following function will compute C, the area under the ROC curve:

```
rocarea <- function(glmobj){
                     ## pass GLM object
  p <- predict(glmobj,type="response")
                     ## predicted probs
  y <- glmobj$y      ## actual zeros and ones
  n <- length(y)     ## self-explanatory
  ones <- y==1       ## indicator for the 1s
  n1 <- sum(ones)    ## how many actual 1s
```

[1] Note the enthusiasm for this type of add-on functionality King, Tomz and Wittenberg (2000) provided to STATA users.

```
  n0 <- n - n1       ## how many actual 0s
  C <- (mean(rank(p)[ones]) - (n1+1)/2)/n0
                     ## the Formula
  C                  ## return C
}
```

exploiting the fact that

$$C = \frac{n_1^{-1} \sum_{\{i:y_i=1\}} r_i - \frac{n_1+1}{2}}{n - n_1}$$

where n_1 is the number of observations with $y_i = 1$, $r_i \in \{1, \ldots, n\}$ are the ranks of the predicted probabilities, and n is the sample size. To use the function we'd type

```
rocarea(logit1)
```

and the answer would appear on screen.

`R` is also much better about memory usage than `Splus`; iteration and looping (essential to simulation-based estimation and inference) was and remains infeasible in `Splus`, but it is much more plausible to use `R` for simulation. The general purpose optimizer in `R` also improves over that in `Splus`; the user need only supply a function to compute the objective function (e.g., a log-likelihood) and the `optim()` function will not only find optimal parameter values, but will also compute a Hessian matrix (minus the inverse of the Hessian is the usual estimate of the variance-covariance matrix of MLEs) without the user supplying functions for first and second derivatives; the optimizers in `Splus` would not provide a Hessian without functions to evaluate at least the first derivative of the objective function. `R` also has the very nice property that its binary data files (`.RData` files) can be automatically read across all hardware and operating systems: e.g., `R` for Windows reads the files created by `R` for whatever else (various Unixes, linux, and my own favorite, Mac OS/X).

On the other hand, `Splus` has a nice GUI, at least for its Windows product; indeed, the GUI is one of the ways `Splus` has been providing "value-added" over its `S` engine, fighting back against its free competitor, `R`. There are a number of projects underway for GUIs for `R`, `Splus` also has a nice set of hooks into `Excel` (permitting easy data import/export), whereas `R`-to/from-`Excel` is more circuitous. The `foreign` package in `R` imports/exports data sets from `Stata`, `SPSS` and `SAS`, among others. But in short, much of the "Plus" part of `Splus` that made `Splus` famous came from the public domain, and has been ported to `R` (e.g., generalized linear models, survival analysis, time-series, multivariate analysis, classification, bootstrapping, various smoothers and density estimators, multiple imputation for missing data and so on); almost everything I used to do in `Splus` I can do in `R`.

In summary, if you are attracted to `Splus` because of its fundamental strengths (easy user-extensibility, object-oriented design philosophy, interacting with data via visualization, and its widespread use among statisticians) and you're not afraid of the command-line (a CLI, not a GUI), then you'll like `R` and may already be using it. Otherwise, consider downloading the base distribution of `R` from `http://lib.stat.cmu.edu/R/CRAN` and see for yourself; the price is hard to beat.

References

Jackman, Simon. 1994. "GAUSS and SPlus: a comparison." *The Political Methodologist* 6:8-13.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analysis: Improving Interpretation and Presentation." *American Journal of Political Science* 44: 341-355.

The Times They *R* Achanging: The State of Advanced Methodology Curriculum at ICPSR

Bob Andersen
University of Western Ontario
randerse@uwo.ca

John Fox
McMaster University
jfox@mcmaster.ca

Charles Franklin
University of Wisconsin
franklin@polisci.wisc.edu

Jeff Gill
University of Florida
jgill@polisci.ufl.edu

This last summer at the ICPSR Summer Program we implemented a new program to integrate the `R` statistical package into a newly designated suite of high-end courses. The purpose was to thoroughly immerse students into advanced material through the use of what is now the de-facto research language in applied and research statistics: `S` . The `S` statistical language was originally created by AT&T Bell Laboratories with two available implementations: `S-Plus` the (expensive) commercial offering focused towards corporate users, and `R` the free offering directed at academic users (ironically there is no real quality difference). Our broader goal was to provide a turn-key curricular experience focused around advanced computation for more advanced graduate students in social science programs who are seeking to dramatically increase their methodological tools.

We are happy to announce that the advanced program will be repeated for the summer of 2003. With the support of Hank Heitowit and Bill Jacoby, the courses will be offered with the same focus on advanced computing in `R`. The course schedule will again be configured to guarantee that students can select any two in the series as well as the special `R` lecture series.

Prerequisite knowledge was a minor issue this past summer. Our experience demonstrated that students for this program should have already taken a basic regression course and be at least familiar with elementary linear algebra and calculus (although not necessarily at an advanced level). Although the `R` training starts from a basic level, some experience with statistical computing is helpful. Other than these issues, no special knowledge is necessary.

By design, this cluster of advanced courses provides a tightly integrated training experience–students attend the seminar in `R`, and typically take two of the three advanced statistics courses. Teaching assistants, class assignments, and computing support is all coordinated across the courses. The result is that students see the connections between the courses, rather than view them as independent offerings. Because all three courses will share the `R` foundation, there is a common computational environment. Most importantly, this common language is more than just a shared computing interface–it is a common way to think about the structure of models, estimation, and display of data. Thus immersing students in this common culture allows them to substantially increase their sophistication in both statistical modeling and computing during the four weeks of the first session.

The new "Track 4" series of courses, all offered for the first four week term, are Maximum Likelihood, Bayesian Methods, and Advanced Methods. These are complemented by a two week overview of the `R` basics given by John Fox in the evenings, covering `R` basics, linear and generalized linear models in `R`, an introduction to `R` graphics, and an introduction to programming in `R`. Each of the advanced offerings combines statistical and mathematical theory with real, hands-on applied analysis that shows exactly how to implement the method. Specifically:

- **Maximum Likelihood Analysis.** This long-running standard bearer has been changed to use `R` for all applied data analysis assignments. Lectures include an emphasis on a variety of limited dependent variables models in `R`, estimation by maximizing the likelihood function, and graphical analysis and presentation, as well as core statistical theory for maximum likelihood estimation.
- **Bayesian Methods for the Social and Behavioral Sciences.** This course uses `R` extensively to demonstrate the mechanics and theory of Bayesian computation. Many standard and conventional Bayesian analyses can be done in `R`, exploiting its rich set of data manipulation and optimization tools. In addition, the most commonly used language for Bayesian stochastic simulation (MCMC) is `WinBUGS`, which has an intentionally `R`-like syntax, and two very useful `WinBUGS` diagnostics (`BOA` and `CODA`) are implemented in `R`. The end-result is that students are able to fully understand, specify, and code their own Bayesian models.
- **Regression III: Advanced Methods.** This "modern" approach to linear and related models focuses on graphical analysis, comprehensive diagnostics, non-parametric fits, and a deeper theoretical understanding than most regression offerings. Although this is a regression course, it is distinguished from other ICPSR regression offerings in that it presents a more advanced and sophisticated look at the linear model and related methods targeted towards students interested in recent developments.

Computing in the ICPSR Summer Program has traditionally been eclectic, employing a wide range of statistical software. As a way of ensuring that students in the three workshops were all proficient in the same language, John Fox taught an evening program on `R` basics. More than 100 participants attended the `R` lectures, which took place over a two week period. `R` was installed in the ICPSR Windows-based computer labs, and a CDROM with `R` for Windows and Windows binaries for all of the packages on CRAN was made available to each participant. In addition, students were encouraged to submit assignments written in LaTeX. There was an additional evening program to introduce LaTeX typesetting for those unaccustomed to the program.

Why would we pick the `R` implementation of the `S` language? `R` is not merely a powerful computational tool, but is a common language that can help researchers appreciate the linkages between these statistical topics. The format of model specifications is not twisted around pull-down menu selections and button options. Instead the format of commands resembles the way we think about statistical models *analytically.* Therefore there is a strong connection between theoretical model specifications and `R` language implementations.

`R` is now central to graduate education in many of the best statistics departments. There is a reason for this. The flexibility and power of the `S` language is expressed not only in model specification but also in the intuitive nature of *object oriented* programming. The extensive set of packages available for downloading is important: many of these are written by the statisticians (and others) that originally developed the technique. For instance: Rob Tibshirani wrote the bootstrap package, David Scott wrote the average shifted histograms package, Peter Rousseeuw wrote the clustering package, and there are others written by Sanford Weisberg, Brian Ripley, Robert Gentleman, Douglas Bates, David Firth, and Luke Tierney (just to name a few). Furthermore, the dominance of the S language in statistics journals means that the provided training is in the use of a tool that is almost certain to be the basis for statistical applications for decades to come.

Our goal is nothing less than to change the statistical lives of our students. We want to change the way they see statistical modeling (whether regression, MLE or Bayes) and we want to change the tools they use so fundamentally that their work will literally never look the same. Students leave the program with a CD containing the free R statistical environment, a free TEX environment, and a collection of tools that the students can continue to use back home. The result is a lasting impact on their research that can be directly attributed to the ICPSR Summer Program.

Quasi-Likelihood Models in R

Jeff Gill
University of Florida
jgill@polisci.ufl.edu

Model Specification R

One of the best features of the R implementation of the S statistical language is that it has a uniform expression for parametric model specification. Model formulas defining variable relationships are essentially the same whether the researcher is specifying a linear model, generalized linear model, generalized additive model, or even some nonparametric forms. This general form is given by: `Y ~ X1 + X2` within the specific model being called. There is more flexibility here than it initially appears. Placing `-1` on the right-hand-side excludes the intercept. The plus sign here is just the most basic specification: changing `+` to `*` gives an interaction term *and* the main effects for `X1` and `X2`, or using `:` in its place just gives the interaction effect, terms can be nested with `%in%` (or equivalently `/`), and `-` to include all of `X1` not in `X2`. It is also sometimes convenient to embed functions into the specification, such something like: `sqrt(Y) ~ cos(X1) + I(X2 / X3)`, where the `I()` function protects the division operation from being interpreted as a nesting specification.

There are more ways that model formulas can be specified (described nicely in Chambers and Hastie (1993) Ch.2, and Fox (2002) Ch.4). Wrapped around this construction is the particular model being run: `lm`, `glm`, `gam`, `nls`, `nlme`, etc. Within that function the user can specify: the data used, factor contrasts, the link function for glms, offsets for fixing constant certain coefficients, random effects terms, weights, starting values for the algorithm, and in some cases the form of the optimization method used. So for example, a gamma glm might look like:

```
copper.stage1 <- glm(COPPERPRICE ~ INCOMEINDEX
    + ALUMPRICE + INVENTORYINDEX + log(TIME),
    family=Gamma,data=copper.dat,
    control=glm.control(epsilon=0.0001,
    maxit=10, trace=F))
```

where the `glm.control` sub-function stipulates: the convergence threshold, the maximum number of *iteratively weighted least squares* (the workhorse glm numerical estimation method) iterations, and whether or not to print steps to the screen. The specification `family` is the means by which the glm link function is identified, and other common forms are obtained by simply replacing `Gamma` here with: `poisson`, `gaussian`, `inverse.gaussian`, `binomial`, and others. Chapter 7 of what is certainly the best *statistical* reference to the S language, Venables and Ripley (1999), contains an explanation of these and other forms.

The purpose of this article is to demonstrate the power and flexibility of the R environment with a specific example of an under-appreciated model form. Even though the theory for quasi-likelihood models can be a little involved (but given below), the means of specifying a quasi-likelihood glm in R are very nearly trivial. For instance, the gamma glm given above can be made into a quasi-likelihood model with the change:

```
copper.stage1 <- glm(COPPERPRICE ~ INCOMEINDEX
    + ALUMPRICE + INVENTORYINDEX + log(TIME),
    family=quasi(link="inverse",var="mu^2"),
    data=copper.dat, control=glm.control(
    epsilon=0.0001, maxit=10, trace=F))
```

where to preserve the gamma link parameter we stipulate that the link is "inverse," and the general `quasi` function allows three other types of variance terms: `mu(1-mu)`, `mu`, and `mu^3`.

Quasi-Likelihood

Wedderburn (1974) introduced the concept of "quasi-likelihood" estimation to extend the standard generalized linear model of Nelder and Wedderburn (1972) to the circumstance when the parametric form of the likelihood is known to be misspecified, or only the first two moments are definable. The goal is to create a more flexible form that retains desirable GLM properties (i.e. those described in Fahrmeir and Kaufmann 1985 and Wedderburn 1976).

Suppose that we know something about the parametric form of the distribution generating the data, but not in complete detail. Obviously this precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation. Wedderburn's idea was to develop an estimation procedure that only requires specification for the mean function of the data and a stipulated relationship between this mean function and the variance function. This is also useful in a Bayesian context when we have prior information readily at hand but only a vague idea of the form of the likelihood.

Instead of taking the first derivative of log likelihood with respect to the parameter vector, $\boldsymbol{\theta}$, suppose we take this derivative with respect to the mean function in a generalized linear model, μ, with the analogous properties:

- $E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = 0.$
- $Var\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = \frac{1}{\phi v(\mu_i)}.$
- $-E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_i^2}\right] = \frac{1}{\phi v(\mu_i)}.$

Therefore what we have here is a linkage between the mean function and the variance function that does not depend on the form of the likelihood function, and we have a replacement for the unknown specific form of the score function that still provides the desired properties of maximum likelihood estimation as described. Thus we imitate these three criteria of the score function with a function that contains significantly less parametric information: only the mean and variance.

A function that satisfies these three conditions is:

$$q = \frac{y_i - \mu_i}{(\phi)v(\mu_i)} \qquad (1)$$

(reference: McCullagh and Nelder 1989, p. 325; Shao 1999, p. 314). The associated contribution to the log likelihood function from the i^{th} point is defined by:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi v(\mu_i)},$$

so finding the maximum likelihood estimator for this setup, $\hat{\boldsymbol{\theta}}$ is equivalent to solving:

$$\begin{aligned}\frac{\partial}{\partial \theta}\sum_{i=1}^{n} Q_i &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{(\phi)v(\mu_i)}\frac{\partial \mu_i}{\partial \theta} \\ &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{(\phi)v(\mu_i)}\frac{\mathbf{x}_i}{g(\mu_i)} = \mathbf{0},\end{aligned}$$

where $g(\mu)$ is the canonical link function for a generalized linear model specification. In other words we can use the usual maximum likelihood engine for inference with complete asymptotic properties such as consistency and normality (McCullagh 1983), by only specifying the relationship between the mean and variance functions as well as the link function (which actually comes directly from the form of the outcome variable data).

As an example suppose we assume that the mean and variance function are related by stipulating that $\phi = \sigma^2 = 1$, and $b(\theta(\mu_i)) = \frac{\theta(\mu_i)^2}{2}$, so $v(\mu) = \frac{\partial^2 b(\theta(\mu_i))}{\partial \theta(\mu_i)^2} = 1$. Then it follows that:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{(\phi)v(\mu)} = -\frac{(y_i - \mu_i)^2}{2}.$$

The quasi-likelihood solution for $\hat{\boldsymbol{\theta}}$ comes from solving the quasi-likelihood equation:

$$\frac{\partial}{\partial \theta}\sum_{i=1}^{n} Q_i = \frac{\partial}{\partial \theta}\sum_{i=1}^{n} \frac{y_i - \theta}{2} = -\sum_{i=1}^{n} y_i + n\theta = \mathbf{0}.$$

In other words, $\hat{\boldsymbol{\theta}} = \bar{y}$, because this example was setup with the same assumptions as a normal maximum likelihood problem but without specifying a normal likelihood function.

In this way quasi-likelihood models drop the requirement that the true underlying density of the outcome variable belong to a particular exponential family form. Instead, all that is required is the identification of the first and second moments and an expression for their relationship up to a proportionality constant. It is assumed that the observations are independent and that the mean function describes the mean effect of interest. Even given this generalization of the likelihood assumptions, it can be shown that quasi-likelihood estimators are consistent asymptotically equal to the true estimand (Fahrmeir and Tutz 2001, p. 55-60, Firth 1987; McCullagh 1983). However, a quasi-likelihood estimator is often less efficient than a corresponding maximum likelihood estimator and can never be more efficient: $V_{quasi}(\theta) \geq [I(\theta)]^{-1}$, where $I(\theta)$ is the Fisher information from the maximum likelihood estimation (McCullagh and Nelder 1987, p.347-8; Shao 1999, p.248-57).

Despite this drawback with regard to variance, there are often times when it is convenient or necessary to specify a quasi-likelihood model. A number of authors have extended the quasi-likelihood framework to: *extended quasi-likelihood* models to compare different variance functions for the same data (Nelder and Pregibon 1987), *pseudo-likelihood* models which build upon extended quasi-likelihood models by substituting a χ^2 component instead of a deviant component in dispersion analysis (Breslow 1990; Carroll and Ruppert 1982; Davidian and Carroll 1987), and models where the dispersion parameter is dependent on specified covariates (Smyth 1989). Nelder and Lee (1992) provide an informative overview of these variations. It is also the case that quasi-likelihood models are not more difficult to compute (Nelder 1985), and the `R` package has pre-programmed functions that make the process routine.

A Detailed Empirical Example of Quasi-Likelihood Estimation in `R`

A relatively well-known example of a glm specification that is improved by a quasi-likelihood specification is the ship damage dataset from McCullagh and Nelder (1989, p. 204). The data are provided as 19 rows corresponding to the observed combinations of type of ship and year built and 4 columns, as follows: ship type, coded 1-5 for A, B, C, D and E, year built (1=1960-64, 2=1965-69, 3=1970-74, 4=1975-79), months of service, ranging from 63 to 20,370, and damage incidents, ranging from 0 to 53. Note that there are no ships of type E built in 1960-64. The outcome variable is the number of damage incidents (`Accidents`).

Here I will replicate McCullagh and Nelder's results using a quasi-likelihood approach which they justify by noting (p. 206) that "For the random variation in the model, the Poisson distribution might be thought appropriate as a first approximation, but there is undoubtedly some inter-ship variability in accident-proneness." The primary modeling difference, as stated before, is simply the inclusion of the "quasi" statement. This has some different forms depending on the link:

```
quasi(link="identity",variance="constant")
quasibinomial(link="logit")
quasipoisson(link="log")
```

There is some flexibility with regard to the link specification. The most general and flexible form, `quasi`, requires the link and variance functions only, with the defaults given above. The `quasibinomial` form allows the link to be asserted as `logit`, `log`, `probit`, and `cloglog`. The `quasipoisson` form allows `identity`, `log`, and `sqrt`. The purpose of separating out the last two types is that these forms do not fix the dispersion parameter at one, and thus are able to capture over-dispersion.

Start by setting up the data structure:

```
ships.df <- data.frame(read.table(
    "http://web.clas.ufl.edu/~jgill/
        GLM.Data/ship.data",header=T))
ships.df$Type <- factor(ships.df$Type)
ships.df$Construction
    <- factor(ships.df$Construction)
ships.df$Start <- factor(ships.df$Start)
attach(ships.df)
```

where we specifically indicate the factors in the dataframe. The default unordered contrast in `R` is "treatment" which is what McCullagh and Nelder use. This can be verified (or changed) with `options()$contrasts`. They also stipulate that the coefficient on years of service (`Period`) is known to be 1, thus requiring the use of an offset to replicate their model. The model is run with the simple command:

```
ships.out <- glm(Accidents ~ Type+Construction
    + Start + offset(log1p(Period)),
    family=quasipoisson, control=glm.control(
    epsilon=0.0001,maxit=100))
```

where the results are given by:

```
summary(ships.out)
Deviance Residuals:
     Min       1Q   Median      3Q     Max
-1.67595 -0.65993 -0.09363 0.37454 2.79094

Coefficients:
                Estimate Std. Error t value
(Intercept)     -6.40691    0.25347 -25.277
TypeB           -0.54261    0.20702  -2.621
TypeC           -0.68788    0.38248  -1.798
TypeD           -0.07675    0.33870  -0.227
TypeE            0.32499    0.27499   1.182
Construction65   0.69721    0.17444   3.997
Construction70   0.81880    0.19787   4.138
Construction75   0.45335    0.27176   1.668
Start75          0.38463    0.13785   2.790
---
(Dispersion parameter for quasipoisson family
    taken to be 1.359193)
```

These are exactly McCullagh and Nelder's results (except that they round). For details about interpreting glm results I can (not surprisingly) recommend my favorite book on the topic (Gill 2000), or some of the standards listed in the reference section.

Redux

The point here has been to show the ease with which model specifications can be as complex as necessary, but still easy to implement in `R`. While one leaves behind the "point and click" world of more primitive statistical software, the gain is quite obviously increased power and flexibility.

References

Breslow, N. 1990. "Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models." *Journal of the American Statistical Association*85: 565-571.

Carroll, R. J. and Ruppert, D. 1982. "Robust Estimation in Heteroscedastic Linear Models." *Annals of Statistics* 10: 429-441.

Chambers, John M. and Trevor J. Hastie. 1993. *Statistical Models in S* New York: Chapman and Hall.

Davidian, M. and Carroll, R. J. 1987. "Variance Function Estimation." *Journal of the American Statistical Association* 82: 1079-1091.

Fahrmeir, L., and Kaufmann, H. 1985. "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models." *The Annals of Statistics* 13: 342-368.

Fahrmeir, L. and Tutz, G. 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models.* Second Edition. New York: Springer.

Firth, David. 1987. "On the Efficiency of Quasi-Likelihood Estimation." *Biometrika* 74: 233-245.

Fox, John. 2002. *An R and S-Plus Companion to Applied Regression.* Thousand Oaks, CA: Sage Publications.

Gill, Jeff. 2000. *Generalized Linear Models: A Unified Approach.* Thousand Oaks, CA: Sage.

McCullagh, Peter. 1983. Quasi-Likelihood Functions. *Annals of Statistics* 11: 59-67.

McCullagh, P. and Nelder, J. A. 1989. *Generalized Linear Models.* Second Edition. New York: Chapman & Hall.

Nelder, J. A. 1985. "Quasi-likelihood and GLIM." *Lecture Notes in Statistics* 32: 120-127.

Nelder, J. A. and Lee, Y. 1992. "Likelihood, Quasi-likelihood and Pseudo-likelihood: Some Comparisons." *Journal of the Royal Statistical Society, Series B* 54: 273-284.

Nelder, J. A. and Pregibon, D. 1987. "An Extended Quasi-Likelihood Function." *Biometrika* 74: 221-232.

Nelder, J. A. and Wedderburn, R. W. M. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society, Series A* 135: 370-385.

Shao, Jun. 1999. *Mathematical Statistics.* New York: Springer-Verlag.

Smyth, Gordon K. 1989. "Generalized Linear Models with Varying Dispersion." *Journal of the Royal Statistical Society, Series B* 51: 47-60.

Venables, W. N. and B. D. Ripley. 1999. *Modern Applied Statistics with S-PLUS* New York: Springer-Verlag.

Wedderburn, R. W. M. 1976. "On the Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models." *Biometrika* 63: 27-32.

Wedderburn, R. W. M. 1974. "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method." *Biometrika* 61: 439-447.

Articles

What to Do With Time Series: A Few Ideas from an Economist

Norm R. Swanson
Rutgers University
nswanson@econ.rutgers.edu

At this juncture in the historical development of the analysis of time series in political science it is clear that little needs to be said concerning the potential usefulness of applying newly developed statistical methodology to the analysis of time series data. Indeed, even an economist bereft of knowledge of the state of the art in political science (such as myself) need look no further than the google search engine on the internet to see that a vast volume of literature in political science in recent years has addressed relevant time series issues such as the presence (or not) of unit roots in the processes describing variables of interest, the usefulness (or not) of vector autoregression (VAR) and vector error correction (VEC) models for describing groups of variables, and the implications and uses of uncovering Granger causal and related predictive relationships among political science variables as well as between political science, economic, and financial variables, for example. Some recent papers that explore these and related topics in the discipline include, to name but a few: Beck (1992), Beck (1994), Beck and Katz (1995), De Boef and Granato (1997), De Boef and Granato (2000), Enders and Sandler (1993), Freeman (1983), Freeman, Williams and Lin (1989), Granato and Krause (2000), Smith (1992), and William (1990). Relevant textbooks (in the area of econometrics and time series analysis) include: Granger and Newbold (1986), Hamilton (1994), and Harvey (1991), for example.

Given the plethora of information out there, then, in this letter I instead discuss what I view as a few of the misunderstood and least well applied aspects of the methodology of time series analysis. As is well known at this juncture, I believe, univariate autoregressive moving average (ARMA) processes has been heavily favored for modeling time series dynamics since the seminal book of Box and Jenkins, published in 1970. Linear VAR models, and subsequently restricted VAR models called VEC models (which account for cointegrating relations among variables that are individually "nonstationary") have similarly received considerable attention since the 1980s and 1990s, respectively. At the current stage in the game, methodologists are mostly interested in developing new tools for analyzing data in these frameworks, including evaluation of predictions and predictive densities, for example.

Consider the VAR model. An unrestricted vector autoregression of order p (called a VAR(p)) has n endogenous variables and n equations, with p lags of each variable in each equation. Under stationarity, these sorts of models can be estimated using unrestricted equation by equation least squares (when all variables appearing in each equation are the same), or using seemingly unrelated regression techniques, otherwise. Thereafter, so called impulse response functions (IRFs) can be constructed by examining the effect over time on all variables of a unit increase in the shock of one variable (fixing all other shocks to zero), where the shock is defined as the "orthogonalized" error of the equation in question. For further discussion of IRFs, the reader is referred to Swanson and Granger (1997). It should perhaps be noted that one interpretation of a VAR model is as a reduced form version of some underlying structural model. It is in this context that one can see why contemporaneous correlation among the errors of a VAR usually exists, and why one must be careful when specifying and analyzing the "shocks" used to construct IRFs. In particular, the contemporaneous correlation arises because only lagged variables are used as regressors in the equations of the model; current variables are not. Now, there are many ways to construct appropriate shocks in a VAR (see the Swanson and Granger paper for discussion), and hence VAR models are sometimes called "atheoretical" models. Put another way, while the VAR models are clearly potentially useful for forecasting and simulation, unless restrictions either corresponding to those implied by some underlying structural model or corresponding to some sort of a priori (theoretical political science) knowledge are placed on the coefficient matrices in the VAR model, one must be very careful if and when interpreting the magnitudes, significance, and political meaning of such coefficient matrices. Of course, in the context of forecasting, the underlying structural parameters do not need to be recovered, as instead we are interested only in including enough information in our conditioning information sets so as to ensure that the "best" predictions possible can be made. (By "best", I mean that a loss function should be constructed to evaluate predictions, and the loss function should be suitable to the task at hand; simply using a mean square error loss function is in many cases inappropriate.) Thus, in such cases we need only be assured that the linkages among the variables in our model remain stable over time (for example, that there are no structural breaks in our model); thereafter, the prototypical unrestricted VAR model discussed thus far may certainly be expected to perform as well as more restrictive structural models, for example. Further, given that there is no guarantee that some given structural model truly represents the system of interest, in many instances it is perhaps most sensible to construct forecasting (and possibly also simulation) models using unrestricted VARs.

As a final note with regard to the above discussion, it should perhaps be stressed that in practical applications the best use of the VAR model often involves beginning with an unrestricted VAR. After initial estimation, lag specification (using model selection criteria such as the Schwartz Information Criterion, for example, when one is interested in forming more parsimonious models for forecasting analyses, or using the Akaike Information Criterion when one is interested in inference on the coefficients in the model for purposes of evaluation of political theories, for example), etc., certain variables might then be found insignificant based on the use of t- and F-tests, for example. These variables can then be removed from the system by setting the coefficients in the coefficient matrices associated with the variables in question to be zero (i.e. impose zero restrictions). Thereafter, other restrictions associated with cointegration can be imposed.

Finally, simple precepts from political theory can be applied by making sure that all remaining variables have coefficients with the "correct" signs and "reasonable" magnitudes, given a priori expectations concerning the signs and magnitudes of simple bivariate correlations between each explanatory variable and the dependent variable in each equation, for example. One important caveat of such "fine tuning", though, is that the VAR model is now a restricted VAR, and as such must be estimated using maximum likelihood under general coefficient restrictions (under covariance stationarity), and using a further modified version of maximum likelihood when there are cointegration restrictions.

As stated above, under the assumption of covariance stationarity, the unrestricted model can be estimated using standard maximum likelihood procedures. This result holds *even if* the errors in the model are not assumed to be normally distributed, a surprising result shown by Halbert White, although in this case, the estimator is called the quasi-maximum likelihood estimator (QMLE), and standard errors of the coefficients need to be modified (relative to their simpler form based on the usual least squares estimator) in order to carry out correct inference using the estimated coefficients. The modification involves, for example, using the heteroskedasticity and autocorrelation consistent (HAC) estimator of Newey and West or the "White" heteroskedasticity consistent standard errors (see e.g. Hamilton (1994).

In many (if not most) cases, the variables of interest are actually not covariance stationary, and processes describing the individual variables of a model may contain unit roots (see below discussion for testing for unit roots). In practice, when the variables are "nonstationary" (e.g. have a unit root), the VAR model can be fitted using differenced variables. However, in certain cases, the difference VAR should be augmented to include additional regressors. This is the case when linearly independent combinations of the "integrated" variables are in fact (covariance) stationary, and in such cases, the linearly independent combinations, in lagged form), should all be included as extra regressors in each equation in the VAR model. Subsequent VEC model estimation using ML is outlined in detail in chapter 19 of Hamilton (1994).

As a side issue, it is important to note that both in the univariate and the multivariate cases, using levels data in regressions estimated using standard techniques (such as maximum likelihood) results in classical inference (such as the use of standard t-statistics, F-statistics, and correlation measures) being invalid. This is the reason why practitioners often use differenced variables when their levels counterparts are found to be nonstationary, and has led to the use of the terminology "spurious correlation"; high correlations can be shown to exist amongst completely unrelated variables are nonstationary - try modelling the same variables in differences, though, and correlations immediately vanish! However, when prediction is the objective, then inferences of this type may not be relevant, and so "levels" models can often still be used, with little effect other than a small cost in terms of estimator efficiency. In practice, before estimating and specifying VAR models, each individual variable is usually tested for a unit root using a unit root test. The most commonly applied unit root test is the so-called augmented Dickey-Fuller unit root test (see Hamilton (1994)). Of course, if "unit roots" are found, it does not mean that the simple test regression used to construct the Dickey-Fuller test is the "correct model". Rather the unit root test simply signals the existence (or not) of a particular characteristic of our variable(s) of interest. Additionally, no one has ever claimed that the "true" model is a model with a unit root. Rather, our objective should be to find the "best" approximation to the truth possible, and if the "best" approximation happens to involve assuming that there are unit roots in the processes describing the variables, then so be it! Of final note on this issue is that any linearly independent cointegrating vectors (i.e. the weights of the linear combinations of the nonstationary variables that result in a stationary "combined" variable) found are not unique, as scalar multiples of these cointegrating vectors can be constructed which also span the cointegrating space. For this reason, economic interpretation of cointegrating vector magnitudes should be done with extreme caution (i.e. estimation routines for cointegration spaces use an arbitrary normalization).

Exogenous variables can be added as additional regressors to VAR and VEC models. For example, the ML estimation procedures designed for VEC models are valid when dummy variables denoting day of week or seasonal effects are put into the models. When additional "exogenous" variables are added to AR and VAR models, the models are often re-named ARX and VARX models, respectively. Further discussion of this and related topics is given in Johnston and DiNardo (1997), and the references cited therein. An issue touched on above concerns the importance of examining the characteristics of any estimated VAR model prior to implementation. For example, simulations based on estimated models may yield useful guidance concerning the correlation implied among the variables in the model. If the magnitudes and or signs of these correlations are substantially different from those observed in historical data, then the model specification stage should be re-visited; further coefficient restrictions may be warranted; and, for example, some explanatory variables may be highly colinear leading to "nonsense" coefficient estimates. In this case the model is "overparameterized" and some extraneous explanatory variables should perhaps be dropped, etc. These sorts of considerations are important when the models are being used to simulate political scenarios, for example, as correlations among simulated variables in the model should make sense in such contexts, and nonsense coefficients leads to nonsense correlations. Of course, if the sole objective of the modelling exercise is the construction of predictions, for example, then the "cost" of including highly collinear explanatory variables is not necessarily great at all. In such a context the main objective should instead be to ensure that as much relevant conditioning information is included in the models as possible. That said, it should be stressed that in financial and economic applications it invariably turns out the more parsimonious models usually perform better - the same may be the case in political science applications. For a discussion of prediction with VAR and related models, please see Chao, Corradi, and Swanson, (2001), Corradi and Swanson (2002), and the references contained therein. These papers delineate the close connection between the notions of Granger causality and predictive ability, and in the context of our VAR model, a very important class of exclusion restrictions that may be useful when specifying VAR models are restrictions implied by the application of "Granger causality" tests. In particular, in the context of VAR models, Granger causality can be equated with the notion that one variable is useful for forecasting another. Using "in-sample" Granger causality tests (i.e. in-sample F-tests of parameter restrictions), may be very misleading, however. Granger points out that if one is truly interested in prediction, then one should carry out rolling ex-ante prediction exercises, say by re-estimating a model using increasing windows of data and producing a 1-step ahead prediction after each new model is estimated. Such a procedure allows one to compare a sequence of real-time predictions with actual data, forming a sequence of real-time forecast errors which can be used in model selection analyses, for example. The papers by Chao, Corradi and Swanson mentioned above discuss these sorts of methods, and develop tests useful in these sorts of contexts.

Overall, the analysis of time series, a field of study only made popular over the last 40 years or so, continues to see many new advances, and in the future I fully expect that further advances will come about as political scientists, economists, statisticians, and others attempt to further unravel the intricacies of dynamic interaction among variables of interest to us all!

References

Beck, Nathaniel. 1992. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3: 51-87.

Beck, Nathaniel. 1994. "The Methodology of Cointegration." *Political Analysis* 4:237-245.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (or Not to Do) With Time Series-Cross Sectional Data." *American Political Science Review* 89(3): 634-647.

Chao, John, Valentina Corradi, and Norm Swanson. 2001. "An Out of Sample Test for Granger Causality." *Macroeconomic Dynamics* 5: 589-620.

Corradi, Valentina and Norm Swanson. 2002. "A Consistent Test for Nonlinear Out of Sample Predictive Accuracy." *Journal of Econometrics* 110: 353-381.

De Boef, Suzanna and Jim Granato. 2000. "Testing for Cointegrating Relationships with Near-Integrated Data." *Political Analysis* 8: 99-117.

De Boef, Suzanna and Jim Granato. 1997. "Near-integrated Data and the Analysis of Political Relationships." *American Journal of Political Science* 42: 619-640.

Enders, Walter and Todd Sandler. 1993. "The Effectiveness of Antiterrorism Policies: A Vector Autoregression Intervention Analysis." *The American Political Science Review* 87: 829-844.

Freeman, John R. 1983. "Granger Causality and the Time Series Analysis of Political Relationships." *American Journal of Political Science* 27: 327-358.

Freeman, John R, John T. Williams, and Tse-min Lin. 1989. "Vector Autoregression and the Study of Politics." *American Journal of Political Science* 33: 842-877,

Granger, C.W.J. and Paul Newbold. 1986. *Forecasting Economic Time Series.* Academic Press.

Granato, Jim and George A. Krause. 2000. "Information Diffusion wit the Electorate: the Asymmetric Transmission of Political-Economic Information." *Electoral Studies* 19: 519-537.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton University Press.

Harvey, Andrew. 1991. *The Economic Analysis of Time Series.* Oxford University Press.

Johnston, Jack and John DiNardo. 1997. *Econometric Methods.* McGraw Hill.

Smith. 1992. "Error Correction, Attractors and Cointegration: Substantive and Methodological Issues." *Political Analysis* 4: 249-258.

William, John T. 1990. "The Political Manipulation of Macroeconomic Policy." *American Political Science Review* 84: 767-796.

Visual Interpretation and Presentation of Monte Carlo Results

Christopher Adolph[1]
Harvard University
cadolph@fas.harvard.edu

Tables vs. graphics in Monte Carlo presentation

Analytical methods are the best way to learn the properties of an estimator since they cover the entire parameter space. But the analytical route is not always available. Proofs may be intractable without unreasonable simplifying assumptions, and many properties of statistical models hold in the limit only. In either case, we turn to Monte Carlo (MC) experiments—testing models on artificial datasets with known properties—to assess the likely performance of an estimator in empirical work. While analytical methods provide complete characterizations across the (usually infinite) parameter space, Monte Carlo surveys are generally incomplete, and thus potentially misleading. To mitigate this key disadvantage, we must thoroughly explore the (relevant) parameter space and devise clever techniques for presenting as much of this space as possible on the printed page.

Unfortunately, though the increasing savvy of political methodology has brought more and better Monte Carlo work, it is not always presented clearly or thoroughly. In particular, MC results often appear in unwieldy tables rather than elegant graphics. Tables are ideal for presenting small quantities of data whose precise values are worth seeing. This is the opposite of the situation in MC work, where the "data" are potentially limitless (just add more parameter values) and precision arbitrarily high (just add more simulations) but usually uninteresting. Researchers may miss patterns, readers' eyes glaze over, and the results remain bound in a straightjacket of rows and columns. Moreover, tables discourage exploration of more than a few parameter combinations, while pictures enable researchers to present far more comprehensive findings. Perhaps the only legitimate use for tables in presenting MC work is to list many different statistics from a single model and data generating process. Usually we are interested in how one or a few statistics (such as mean squared error and bias) vary across models or scenarios, and graphics should be used instead.

I surveyed recent years of *Political Analysis* (2000–2002), the *American Political Science Review* (1998–2002), and the *American Journal of Political Science* (1998–2002), and found 24 articles reporting Monte Carlo experiments. Of these, 12 reported results in tables only, 9 used graphs only, and 3 used a combination. In all, nine articles could have substituted graphics to their advantage (on the criterion of presenting a statistic under many models or scenarios), while in other cases a large number of non-comparable outcome measures left no alternative to tables.

To help the quality of Monte Carlo presentations keep up with the sophistication of the experiments themselves, I propose five guidelines for MC graphics. I also define five graphic styles which help show the comparative performance of models over the parameter space, even when the models and parameters are many.

Five principles for visual displays of Monte Carlo results

1. **Maximize resolution.**

   The main reason to use graphs is that they allow higher data density than tables, and often gain readability in the bargain. In a table, you might only try two values on a given parameter to save space. But in a graph, running many MC scenarios just makes the pixels smaller and patterns clearer.

2. **Get the whole picture, and nothing but the picture.**

   Try to canvass the "whole" parameter space (infinities and asymptopia aside). But also take advantage of any logical or substantive limits on the parameters in applications of interest. Within these limits, a full factorial design (considering all combinations

[1] I am grateful to Robert Fannion and Gary King for helpful suggestions. Ph.D. Candidate, Department of Government, Littauer Center, North Yard, Harvard University, Cambridge MA 02138; `http://chris.adolph.name,cadolph@fas.harvard.edu`.

	Mean Squared Error			
	Pooled		Fixed	Random
N	Logit	Index	Effects	Effects
2	0.57	0.93	0.89	0.51
4	0.54	0.61	0.72	0.43
6	0.54	0.53	0.62	0.41
8	0.52	0.49	0.58	0.37
10	0.54	0.43	0.49	0.37
50	0.53	0.11	0.12	0.31

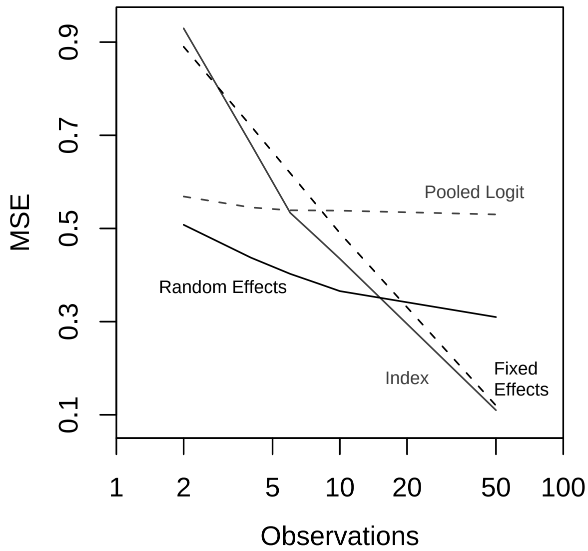

Figure 1: *Contour plot of model performance and the tabular alternative.* Monte Carlo results on ideal point estimators taken from Bailey (2001). The original display is at the left, with a new display, using the recommendations of this article, at right. The plot shows loess-smoothed contours (unsmoothed contours would also work). Note the $\log_{10}$ scaling of the horizontal axis, which avoids illegible compression of the crucial small sample results.

from some set of hypothetical values on each parameter) is especially illuminating if you have the computer power.

3. **Focus on interesting patterns.**

   Investigate repeated patterns, then use these patterns to condense the parameter space. N-dimensional space gets much smaller when everything in it is an apple, orange, or pear.

4. **Cheat the curse of dimensionality.**

   ... using small multiples (arrays of similar graphics), shading, and creativity (see Tufte, 1990, 2001).

5. **Make the results usable.**

   Ideally, MCs help researchers decide which model to use in specific situations. If your results show when certain models should be used or avoided based on knowable quantities, make these recommendations clear and easily referenced. They will be the most used part of your article.

Graphical alternatives to tabular tyranny

The remainder of this article grapples with the key challenge of Monte Carlo presentation: many parameters (p_1, p_2, ...), many models, and only two-dimensional paper to put them on.

Model performance plots

The first model performance plot we consider is an estimate-vs-truth plot (EvT). This is a special case in which the measure of performance is an estimate of p_1. The one-parameter EvT plot is a simple scatter-diagram, in which a pattern of points on or near the 45° line indicates good performance. This plot is ideal for demonstrating whether $\hat{p}_1 - p_1$ is independent of variation in other parameters. For each value of p_1 in the experiment, run trials with diverse values of all other parameters, then check whether the results cluster on or near the 45° line. Additional models can be distinguished through different symbols or colors, and patterns of dependence through arrays of EvT plots. For an example EvT plot, see Adolph et. al. (2003).

EvT is suited to the special case where the performance measure is the estimate of a parameter. Generally, we want to show model performance on some arbitrary metric Q for parameters p_1 and p_2. To reduce these three dimensions to a sheet of paper, we have at least two options: contour plots and image plots. We can plot Q against p_1 for a given value of p_2, producing a "performance contour" at that level of p_2. (The contour itself may be a loess-smoothed curve or a line "connecting the

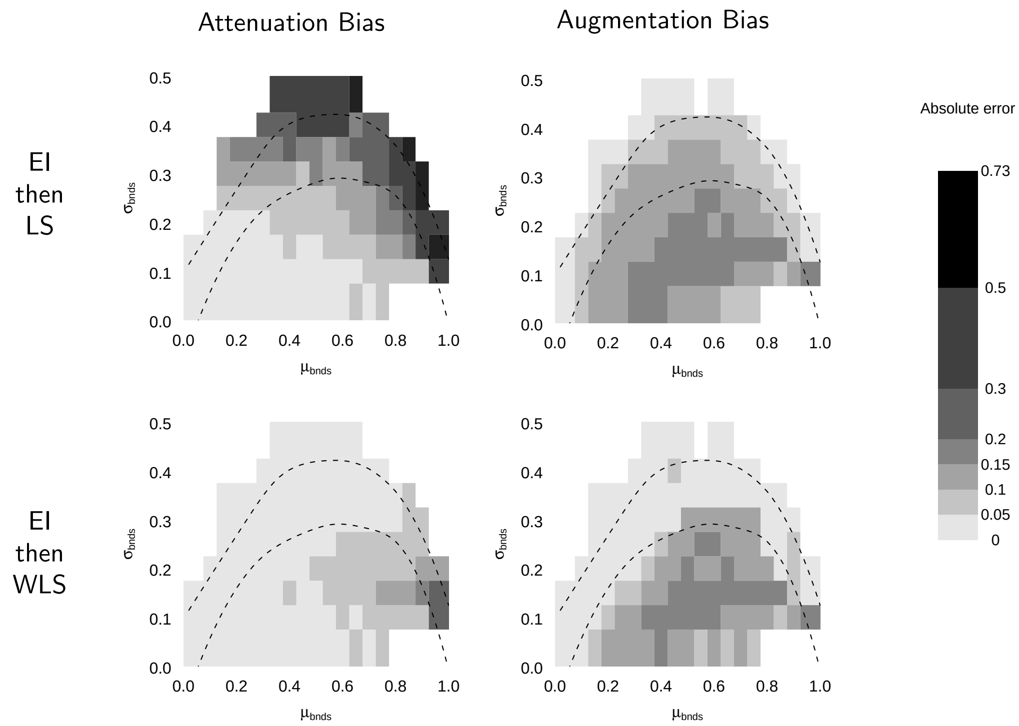

Figure 2: *Image plots of model performance.* Image plots which show absolute error in estimates derived from two different models (EI followed by least squares and EI followed by weighted least squares) over three parameters (the average bound width μ_{bnds}, the standard deviation of bound width σ_{bnds}, and whether the data were generated to produce attenuation bias or augmentation bias). Dark areas indicate poor performance, lightly shaded errors good performance, and blank areas cases not included in the experiment. The dashed lines indicate where datasets in the real world tend to fall in terms of μ_{bnds} and σ_{bnds}. For more details, and a treatment of the same data using contour plots, see Adolph et. al. (2003).

dots"). Then by selecting various levels of p_2, we can map out performance in p_2 space, though with coarser detail than is available in p_1 or Q space. As always, we can accommodate p_3, p_4, etc., through arrays of contour plots.

The contour plot can also be used to directly compare several models on a single parameter. Bailey (2001) investigated the performance of four estimators of legislators' ideal points when only a few votes are available, presenting the table in Figure 1. Bailey concludes that the random effects estimator is superior for small samples, while the fixed effects estimator is better as the number of votes grows. This can be gleaned from the table, but is immediate when we redisplay Bailey's results as a contour plot. The plot also draws attention to the point at which random effects ceases to be the best choice ($N \approx 20$). For more examples of contour plots in MC work, see King and Zeng (2001) and Adolph et al (2003).

An alternative plot focuses on the parameters and categorizes Q. This produces a "map" of performance over precise values of p_1 and p_2 where shading indicates the level of Q. Additional models or parameters can be easily accommodated by small multiples, a design for which these "image" plots are particularly suited. Figure 2 illustrates this approach using Monte Carlo results on second-stage ecological inference estimators (see Adolph et al, 2003 for further details). These plots demonstrate the five principles advocated above. First, they *show the whole picture*, while highlighting those parameter combinations likely to occur in practice (the area inside the dashed lines). Second, image plot's inherent capacity for high resolution allows presentation of a matrix of scenarios covering the whole space in two parameters (μ_{bnds} and σ_{bnds}). Third, the graphs *focus on interesting parts* of the remaining parameter space (here, each column of plots represents a different "worst case" scenario for model failure). Fourth, *arrays of plots* show 5 dimensions (error, μ_{bnds}, σ_{bnds}, type of bias, and model). Finally, the plots *provide a guide for practical research*, allowing researchers to make better informed modelling decisions.

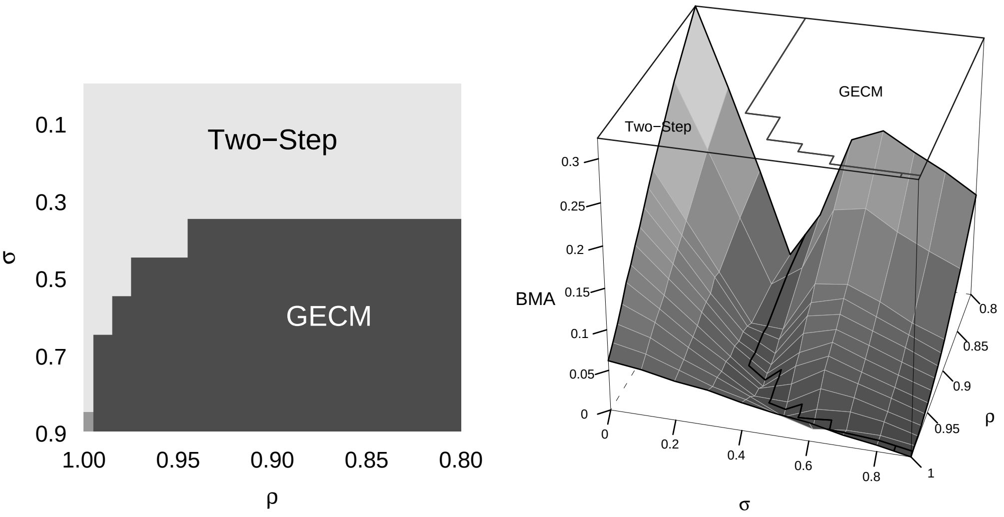

Figure 3: *Best Model plots.* These plots report MC results on two time series estimators, the Generalized Error Correction Model (GECM) and the Engle-Granger two-step method (Two-Step) from De Boef (2001). The left plot shows which model has minimum bias ("best model") at each combination of autocorrelation (ρ) and simultaneity (σ). (The models have equivalent performance in the bottom left corner). The right plot shows the advantage of the best model (BMA) over the second best model for each parameter combination. If the true parameter values are known, the left plot shows which model to use. If the parameter values are unknown, the right plot aids in deciding which model is likely to minimize bias.

Best model plots

Arrayed image plots handle arbitrarily many models or parameters, but if many models are considered at once, the ensuing pages of graphs will try the patience of readers. We need a presentation that shows the "big picture", leaving the task of "zooming in" on interesting features or patterns to selected model performance plots. In this case, I recommend a summary graphic which shades the "best" model at each point in the parameter space. (Lest we interpret "best model" too rigidly, where models are approximately indistinguishable, the plot should list all contenders.) Once the "best" model is identified, a researcher could follow up with an appropriate selection of model performance plots. Alternatively, a 3D "best model advantage" plot shows how much better the best model performs than the second-best at every point in the parameter space. Models covering more volume in this plot are safer bets.

To create example best model plots, I drew on De Boef's (2001) work on two time series estimators—the Engle-Granger two-step method and the Generalized Error Correction Model (GECM)—applied to highly autoregressive data. De Boef notes that both methods may be inconsistent if the data generating process is not quite permanently memoried, and that simultaneity in the errors of the time series exacerbates this problem. De Boef ran MC experiments with varied persistence (ρ) in the time series and covariance (σ) in shocks to the explanatory and dependent variables. De Boef employs useful and elegant 3D plots to show how the coverage of confidence intervals for each model varies with ρ and σ, but resorts to large tables to report bias in estimates of the long-run relationship between time series. Making a "best model" plot from these results helps show which model is less biased under different circumstances (in this case, GECM if simultaneity is high, and Engle-Granger otherwise). A "best model advantage" plot shows that the gap between the models' performance grows as the persistence ρ declines. The difference in performance also displays a regular, though non-linear, relationship with simultaneity.

How to do it

Using graphics instead of tables will make your Monte Carlo results more complete, clear, and usable. These advantages can be gained without much effort, since the five plotting styles proposed in this article (MP EvT, MP contour, MP image, BM image, and BMA) can all be produced by the software package `SeeMC` (Adolph, 2003). `SeeMC` runs in R and Gauss, and is available at `http://chris.adolph.name`. A standalone version is in the works.

NB: Formal modelers will doubtless find "best model" image plots reminiscent of graphs of comparative statics from game models (see, e.g., Deiermeyer and Stevenson, 2000). `SeeMC` also provides an easy way to make these plots.

References

Adolph, Christopher. 2003. *SeeMC*. R and Gauss software for visualization of Monte Carlo results, available at `http://chris.adolph.name`.

Adolph, Christopher, and Gary King with Michael C. Herron and Kenneth W. Shotts. 2003. "A consensus on second-stage analyses in ecological inference models." *Political Analysis* 11.1: 87–95.

Bailey, Michael. 2001. "Ideal point estimation with a small number of votes: A random-effects approach." *Political Analysis* 9.3: 192–210.

De Boef, Suzanna. 2001. "Modelling equilibrium relationships: Error correction models with strongly autoregressive data." *Political Analysis* 9.1: 78–94.

Deiermeyer, Daniel, and Randolph Stevenson. 2000. "Cabinet terminations and critical events." *American Political Science Review* 94.3: 627–640.

King, Gary, and Langche Zeng. 2001. "Logistic regression in rare events data." *Political Analysis* 9.2: 137-163.

Tufte, Edward. 1990. *Envisioning Information.* Graphics Press.

Tufte, Edward. 2001. *The Visual Display of Quantitative Information.* Graphics Press. 2nd ed.

The LaTeX Corner: Games in LaTeX

Tamar London
Pennsylvania State University
trlondon@psu.edu

One of things that attracted me to LaTeX was a desire to create nice looking game trees. Extensive form games that I have created in MS-Word have left me unsatisfied. I had to eyeball the length and angle of the branches, as well as the location of labels. In LaTeX you specify the exact location, slope and length of branches. The end result, in my opinion, is more professional looking.

Let me preface this piece by saying that I am new to LaTeX. As such, I am sure there are many more qualified than I am to write this. On the other hand, I have found that to those familiar with it, LaTeX becomes second nature and memories of a painful start have long faded. To me, these memories are fresh. I will assume, therefore, that users are familiar with LaTeX, but (like me) are still new to it. For those that are unfamiliar with LaTeX, I recommend Chan H. Nam's article in a previous edition (10:1) of *The Political Methodologist*. I will also assume that users are using a windows based platform (simply because I am unfamiliar with unix and macintosh). Most of what I go over should translate easily to the other platforms.

Foreplay

To make trees using LaTeX you will need the basics: LaTeX and a text editor. I recommend the MikTex implementation for Windows. To edit the LaTeX text you need a text editor. WinEdt is the one most commonly used with LaTeX. Once you download WinEdt and MikTex they are fully integrated.

You will also need a package capable of drawing extensive games.

The Birds and The Bees

If your goal is to create basic trees then I recommend using egame. This style was created by Martin Osborne and can be found on his website: `http://www.chass.utoronto.ca/~osborne/latex/index.html`.[1]

To install egame you simply need to save the file to you local Tex tree. For example, create a folder called "egame" under `C:\texmf\tex\LaTeX` and save the egame.sty file there. Your computer might try to save it as an html document. Don't let it. Make sure it is saved as a .sty file. You then need to let MikTex know that this file has been downloaded. You can refresh MikTex by going to `Accessories/MikTex_Options` in WinEdt and clicking "refresh."

To create a game you will need to put the `usepackage{egame}` command in the preamble (between `documentclass{article}` and `begin{document})`.

[1] Osborne has a very straightforward style for normal form games, as well.

On his site, Martin Osborne outlines how to use egame, and I recommend reading the egame.pdf file. I will summarize some of the basic features of egame.

You begin the figure with following command:

```
\begin{figure}[htb]
\hspace*{\fill}
\verb+\begin{egame}(x,y)[0.1mm]
```

"x" and "y" represent the size of the figure in units of .1mm. (You can change the default unit, but it is probably a good idea to just stick with it.)

Next you need to tell LaTeX where to place branches, in what direction they will go, and how long they should be. I use the game in Figure 1 to go over the steps of creating a simple game tree.

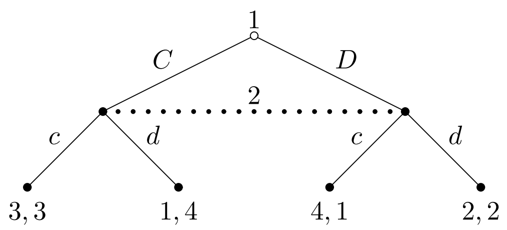

Figure 1: Prisoner's Dilemma in Extensive Form

We begin by creating a figure area that is 10cm wide and 3cm tall.

```
\begin{figure}
\hspace*{\fill}
\begin{egame}(1000,300)
```

Next we place a call for a branch (or branches). The branches will begin in the middle, and near the top, of the figure area (500,250). The branches will be sloped in a 2:1 ratio (2,1), and will be of horizontal length 200 (2cm).

```
\putbranch(500,250)(2,1){200}
```

It is worth noting a few things. First, branch length is given by *horizontal* length. This makes computing the location of future branches much simpler than if it were the actual branch length. Second, branches are placed symmetrically, so the 2:1 ratio refers to both branches. If there were three branches, one would be placed straight down the middle. Egame cannot give branches arbitrary slopes. Third, the default direction is for branches to go down. (One can add an optional direction command to change the direction of the game.)

The above command does not place the branches, just puts a call for them. The next command labels the node "1", inserts two branches and labels them "C" and "D."

```
\iib{1}{$C$}{$D$}
```

You can place as many as three branches in one command. For a tree with four or more branches, combine commands. (For example, placing 2 branches with slopes of ratio 4:1 over 2 branches of slopes 2:1 will produce a tree with four branches.)

The left branch ends at the coordinates (300,150), 200 units to the left from where we started, and half of that length down. So the next set of branches begins there. We give these branches a 1:1 slope and horizontal length 100.

```
\putbranch(300,150)(1,1){100}
```

We do not label this node. The branches are labelled "c" and "d" and payoffs of (3,3) and (1,4) are assigned to each terminal node.

```
\iib{}{$c$}{$d$}[$3,3$][$1,4$]
```

Similarly, another set of branches begins at (700,150).

```
\putbranch(700,150)(1,1){100}
\iib{}{$c$}{$d$}[$4,1$][$2,2$]
```

Finally, an information set begins at (300,150) and extends the length of 400 units. It is labelled, "2."

```
\infoset(300,150){400}{2}
```

We end the game with:

```
\end{egame}
\hspace*{\fill}
\caption[]{Prisoner's Dilemma in Extensive Form}
\label{fig:PD}
\end{figure}
```

The "label" command is optional. It allows you to refer back to the figure, and ensures consistency between the figure and the text.

Hardcore

For more complicated trees Osborne has created egameps.sty - also available for download on his site. Egameps uses PSTricks, which is a graphics package. Egameps allows for a variety of tree styles, which include the use of arrows and color. The basic commands in egameps are virtually identical to those of egame. You can also create trees directly in PSTricks, but Egameps simplifies the procedures.

The downside of egameps is that it is a bit more complicated to get started. You need to download PSTricks. This is available on the ctan site (ctan.org). (You might already have PSTricks, as it comes standard with the "large" and "full" versions of MikTex.)

If you have already downloaded the small version of MikTex you can add pstricks through the MikTex package manager
`(Start/Programs/MikTex/MikTex_project_manager)`.

PSTricks uses a postscript viewer, so you will also need to download ghostscript and ghostviewer. (Once installed, an icon of a little ghost appears on the WinEdt toolbar.)

You will need to add the following commands in the preamble of your LaTeX document:

```
\usepackage{pstcol}
\usepackage{pstricks}
\usepackage{egameps}
```

(Make sure the `\usepackage{egameps}` command comes after the others. You will get an error otherwise.) After creating your document in WinEdt you need to LaTeX it. This creates a dvi file. Next, convert the dvi file to a postscript file. Finally, you can view and print the postscript file in ghostviewer. There are icons for each of these actions in the WinEdt toolbar.

I Told You So

I conclude with a few caveats. First, there are many optional commands that I have not covered, such as altering the location of the labels relative to the branches, and changing the style of the nodes. These are all documented on Osborne's site. Second, there might be other good packages for drawing trees out there. I have asked a few political scientists what they use and all used Osborne's egame and egameps. Osborne himself was not familiar with other packages.

Finally, let me offer a bit of advice. Before diving in and creating a tree in LaTeX, invest a bit of time charting out your tree on paper. Include the coordinates of nodes and the slopes of branches (width to height ratio). This will make typing up the commands much simpler.

References

Nam, Cham H. 2001. "Using LaTeXfor the First Time." *The Political Methodologists* 10(1):32-35.

Osborne, Martin. 2003. *Manual for egames.sty* `http://www.chass.urtonto.ca/~osborne/latex/egame.pdf` (1 April, 2003).

The Society for Political Methodology
ad indicia spectate

Section Activities

New *Political Analysis* Editor Appointed

I'm very pleased to announce that Bob Erikson of Columbia University has agreed to serve as the next editor of *Political Analysis*, succeeding Neal Beck effective with Volume 12, Number 1 (appearing February, 2004). Erikson was chosen by a selection committee consisting of Political Methodology Society President Jonathan Nagler, current *Political Analysis* editor Neal Beck, former Society president Gary King, Society vice president Simon Jackman, *TPM* editor Suzanna De Boef, and former *TPM* editor John Londregan. Professor Erikson previously served as editor of the *American Journal of Political Science.* Submissions should continue to go to Neal Beck till June 30, 2003.

As everyone reading this list knows, the move of *Political Analysis* from being an annual volume to a quarterly journal has been tremendously successful under Professor Beck's editorship. The committee looks forward to continued success with the change in editors.

Searching for New *TPM* Editor

The section is searching for a new editor for *The Political Methodologist*. *TPM* is considered by many to be the gold standard of APSA section newsletters. It includes book reviews, software reviews, general news of interest to those interested in political methodology. It is published twice a year, and the typical term of an editor is two years. The editor's home institution generally supports the production and mailing costs of the newsletter. If you are interested in serving, or would like to nominate others, please contact Jonathan Nagler (*jonathan.nagler @nyu.edu*). If you have questions about what the editorship entails, you can ask the current editor (Suzanna De Boef) or previous editors.

Searching for New WebMaster

After many years of doing an incredible job, Jeff Gill is *finally* tiring of providing 24/7 service to the Political Methodology Website (`http://web.polmeth.ufl.edu/`). If you are have great ideas for the site, and are interested in serving as WebMaster, or would like to nominate others, please contact Jeff Gill (*jgill@polisci.ufl.edu*). We are assuming the WebMaster's institution would host the site.

Political Methodology Conference

The 20th Annual Summer Political Methodology Conference will be held July 17-19, 2003 at the University of Minnesota in Minneapolis. The meeting is sponsored by the Society for Political Methodology, the APSA Political Methodology Organized Section, and the National Science Foundation. The Department of Political Science and College of Liberal Arts at University of Minnesota are hosting the conference and providing substantial support. The National Science Foundation recently approved funding supporting conferences thru 2005. Janet Box-Steffensmeier and Jonathan Nagler are Principle Investigators on the grant, to be administered thru New York University.

NorthEast Methodology Program

The third annual meeting of the NorthEast Methodology Program (NEMP) took place at NYU on Friday April 25. The schedule for the meeting is below:

```
Lunch:    11:30 - 12:30

Paper 1:  12:30 - 2:00
          Jeff Gill, University of Florida.
          "Fundamentals of Bayesian Inference."

Paper 2:  2:20 - 3:50
          Bear Braumoeller, Harvard University.
          "Modeling Causal Complexity with
          Boolean Logit and Probit."

Paper 3:  4:15 - 5:45
          Kevin Clarke, University of Rochester.
          "A New Test for Nonparametric
          Model Discrimination"

Happy Hour 6:00 - 7:00
```

The meeting allowed faculty and graduate students to see related cutting-edge papers on political methodology presented in a format where there is time for serious discussion among a good sized group. It also allowed for pedagogically useful presentations (such as Jeff Gill's presentation at this meeting) for grad-students or faculty not familiar with the topic.

If you would like to be on the mailing list for future meetings, please contact Jonathan Nagler (*jonathan.nagler@nyu.edu*).

EITM Competition Announcement

The National Science Foundation recently announced a special competition that involves Empirical Implications of Theoretical Models (EITM). The EITM competition's deadline is June 12, 2003. The details of the competition can be found at: `http://www.nsf.gov/pubsys/ods/getpub.cfm?nsf03552`

ICPSR: Summer Program Preview

There are a number of recent changes in the ICPSR Summer Program in Quantitative Methods of Social Research that might be of interest to you, your colleagues, and students.

The Program will offer a 4 week workshop titled "Regression Analysis III: Advanced Methods." This class will cover regression analysis with much more emphasis on graphical displays, diagnostics, and non-parametric fits than does the more common econometric perspective (Instructor: Bob Andersen). A companion course will be offered on "Statistical Computing in S," covering R, S, and S-PLUS computing packages (Instructor: John Fox).

The Program will again be offering a 4 week course on "Bayesian Methods for the Social and Behavioral Sciences." This is one of the few venues that will combine these statistical methods with social science applications. (Instructor: Jeff Gill, University of Florida)

Another recent addition is "Complex Systems Models in the Social Sciences," sometimes this area is referred to as "adaptive systems" or "agent-based models." (The type of modeling often identified with the Santa Fe Institute). Instructors are Ken Kollman, and Scott Page, University of Michigan.

There are also one week courses on:

- Categorical Analysis: Models for Binary, Ordinal, Nominal, and Count Outcomes (Instructor: Scott Long).
- Network Analysis (Instructor: Stanley Wasserman).
- Mixed Models for Categorical Data.
- "LISREL" Models (Instructor: Ken Bollen).
- Spatial Analysis (Instructor: Luc Anselin).
- Census 2000.
- Latent Growth Curve Analysis.

Standard 4 week courses include:

- Maximum Likelihood Estimation (Instructor: Charles Franklin).
- Advanced MLE (Instructors: Adam Berinsky, Michelle Claiborn, and Christopher Zorn).
- Scaling & Dimensional Analysis (Instructor: Bill Jacoby).
- Time Series (Instructor: Genie Baker).

Finally there is a 2 course sequence in formal modeling:

- Game Theory (Instructor: Ethan Bueno de Mesquita and Cathy Hafer).
- Rational Choice Theories (Instructor: Jim Johnson).

These are just some of the highlights of the 2003 program. You can find the full Program Announcement and the on-line registration form in the Program web site: `www.icpsr.umich.edu/sumprog/`

If you have further questions, please contact:

Henry Heitowit
Director Educational Resources ICPSR
P.O. Box 1248
Ann Arbor, MI 48106-1248
voice: 734/998-9888
FAX: 734/998-9889
`http://www.icpsr.umich.edu/sumprog/`

THE POLITICAL METHODOLOGIST
Department of Political Science
Pennsylvanian State University
University Park, PA 16802

The Political Methodologist is the newsletter of the Political Methodology Section of the American Political Science Association. I gratefully acknowledge the support of the Department of Political Science of the Pennsylvania State University in helping to defray the editorial and production costs of the newsletter.

Subscriptions to *TPM* are free to members of the APSA's Methodology Section. Please contact APSA (202 483-2512, `http://www.apsanet.org/about/membership-form-1.cfm`) to join the section. Dues are $25.00 per year and include a free subscription to *Political Analysis*, the quarterly journal of the section.

Submissions to *TPM* are always welcome. Articles should be sent to the editor by e-mail (sdeboef@la.psu.edu) if possible. Alternatively, submissions can be made on diskette as plain ascii files sent to Suzanna De Boef, Department of Political Science, 108 Burrowes Building, Pennsylvania State University, University Park, PA 16802. LaTeX format files are especially encouraged. See the *TPM* web-site [http://web.polmeth.ufl.edu/tpm.html] for the latest information and for downloadable versions of previous issues of *The Political Methodologist*.

TPM was produced using LaTeX on a PC running MikTex and WinEdt.