æ

æ

æ

The Political Methodologist

Newsletter of the Political Methodology Section
American Political Science Association
Volume 7, Number 1, Fall 1995

Contents

Notes From the Co-Editors

R. Michael Alvarez
California Institute of Technology

Nathaniel Beck
University of California, San Diego

Over two years ago, we took on the task of editing *TPM* with a promise that we would try to deliver the type of high-quality section newsletter we have all grown accustomed to under the leadership of past editors — Gary King, Larry Bartels, and Charles Franklin. We hope that the four issues of *TPM* which you've received have met those expectations!

After this issue, the job of editing *TPM* passes to Jonathan Nagler, University of California, Riverside. A committee of section members selected Jonathan, and we believe they have made a perfect choice. You'll hear from Jonathan briefly below and in coming months about the direction he wants to take this newsletter.

So, in our final issue, we provide a series of interesting articles and reviews. We begin with an issue which we have covered before, that of replication in political science. Brian Roberts and Jay Dow recount their attempts in graduate courses on methods, to replicate the work of others. Their experiences should underscore the importance of replication in political science research.

We then turn to a new area for political science, that of computational or numerical analysis. From the increasing number of papers which rely on computer simulations for analysis of complicated political models, it is clear that this is a technique which will be increasingly popular as the price of computing power continues to fall. We have two articles, one by Ken Kollman and the other by Richard Timpone and Charles Tabor. Both these articles should stimulate discussion about computational methods in political research. We also present a short article by Christopher Mooney on simulating random variables in GAUSS.

Continuing with our desire to provide reviews of the software packages which political methodologists use, we have two reviews of STATA, a package which many of our colleagues use. We also review two of the most popular data transfer packages, DMBS/COPY and STAT/TRANSFER. Last, we present the results of our informal internet survey of what software packages methodologists actually use in their everyday work.

Last, we have a series of section announcements and book reviews. We hope you have found the reviews in the past *TPM* issues as useful as we have.

We have had a lot of fun editing *TPM* . It would not have been possible without support from many individuals. In particular, we would like to thank Larry Bartels for giving us this great job, Charles Franklin for his assistance with the *TPM* latex macros, and Abby Delman, without whom we would never have gotten one issue of *TPM* done and in the mail. We also thank John Ledyard and the California Institute of Technology for their support of this newsletter.

Notes From the New Editor

Jonathan Nagler
University of California, Riverside

Since *TPM* is clearly not broken, there will be no attempt to fix it. Rather there will be a continuation of the effort to present information that members of the section will find useful and/or interesting. This will include textbook reviews and software reviews. In particular, I would like to do a comparison of the standard 'first regression course' texts available. And there are several software packages that have never been covered in *TPM* yet that might be of interest to many readers: RATS, Matlab, Mathematica, Maple, and SAS/ETS. I am also looking for articles on how to do things that methodologists need to do to produce or present research that might not be covered in a basic graduate econometrics text. This would include topics such as: the use of graphs (both traditional graphs and some of the newer types of graphs being produced) to present data, meaningful ways to present estimates of non-linear models, and programming suggestions on how to do a particular piece of data analysis.

Submissions in LaTeX are strongly encouraged. This is especially true if you are submitting something that has many equations in it. Other electronic formats can be accepted; but if you are not sending LaTeX you should produce a left-adjusted ascii-version of your document to go with whatever your native word-processor format is. Submissions should be sent via email to: nagler@wizard.ucr.edu. Suggestions for what anyone would like to see in future issues of *TPM* are also welcome!

A Replication Experience

Brian Roberts
University of Texas at Austin

In the spring of 1992 I taught a graduate seminar on Public Opinion and Voting Behavior in which students were charged with replicating a series of empirical studies in leading political science journals. This was a new teaching strategy for me but it was dictated by my need to put together a course on a topic about which I had only superficial knowledge (Returning Texas after a year at Caltech I was informed by the graduate secretary that one of the seven new hires made in my absence had signed up to teach the graduate political economy seminar that I normally taught in the spring semester and, besides, she pleaded, there was desperate need for a Voting Behavior course).

The key to this seminar was the omnipresence of National Election Survey (NES) data in empirical voting behavior studies; with ICPSR membership it is a trivial exercise to retrieve these data. Without the need to reproduce proprietary data or approach authors for copies of their data it was possible to attempt several replications over the course of the semester.

There were eight students in the seminar and over the course of the semester they were assigned to ever changing two or three person teams tasked with the replication of different articles. Topics included turnout, primary & general election voting, congressional & presidential voting, economic voting, negative voting, and partisanship. As a final class project each student was asked to replicate and extend (update or respecify) an article of their own choice. In total the class tried to replicate eight or ten different articles (a number of students returned to earlier team replications for their individual projects).

In no case did a team or individual successfully replicate the empirical results of a published article. Some efforts were more successful than others. Asked to specify a model to explain relative success, RHS variables would include most prominently various measures of documentation. Few articles offered complete or accurate roadmaps for would-be replicators.

A survey of my students suggests that the most common lacunae included directions for recoding variables, treatment of missing values, reference to specific NES variables, and presentation of summary statistics. In only one or two cases were the students able to come within 10% of the reported number of observations in an article. This was particularly frustrating because matching the number of reported observations is a necessary but by no means a sufficient condition for replication.

I should note that there certainly was variation in the empirical skills of the students but in putting together the teams I did my best to ensure a mix of experience. Casual observation suggests that empirical skill had a small but positive effect on the success of the final individual replication / extensions.

Finally I should note that although profoundly frustrating, this seminar is considered by many of the students to have been one of the most professionally valuable experiences they had in graduate school. It's safe to say that they walked away from the seminar with less skepticism at the reported results than frustration at the absence of enough information to replicate them, something they won't let happen with their own work. Most important, they learned from their team projects that they never want to co-author a paper for the rest of their lives.

The Use of Replication in a Graduate Methodology Course

Jay K. Dow
University of Missouri – Columbia

Last semester I taught statistical inference, aptly described by Renee Marlin-Bennett as the "course no one wants to take."[1] This is one of two quantitative methodology courses required of University of Missouri political science graduate students. The class brings students up to speed in probability and statistics before they proceed the linear models course, and provides the background to learn methodological skills not taught in our program.[2] The prerequisite is an upper-division data analysis course. Consequently, students enter the class familiar with elementary statistics and regression, and, to varying degrees of proficiency, able to use computers. Several students had also completed research design, with one concurrently enrolled in first semester econometrics.

[1]Renee Marlin-Bennett "Teaching the Courses No One Wants to Take" *The Political Methodologist* 3:2 (Fall, 1990): 14–15.

[2]The texts used were Robert V. Hogg and Elliot Tanis *Probability and Statistical Inference, 4th ed.* (New York: MacMillan Publishing Company, 1993), and Gary King *Unifying Political Methodology* (New York: Cambridge University Press, 1989). Students with little or no background in calculus were also encouraged to purchase Daniel Klepper and Norman Ramsey *Quick Calculus* (New York: John Wiley and Sons, 1985).

The overwhelming enthusiasm of my captive audience notwithstanding, the primary obstacle to effective teaching in this course is student skepticism that learning about probability and statistical distributions is important for reading literatures and conducting research. Not anxious to perpetuate the stereotype by basing course evaluations solely on homework and examination problems, I incorporated a term project that required students to replicate an empirical study. The replication promised to motivate the subject, develop research skills, and, allowing sufficient time to heal all wounds, proved to be an enjoyable exercise that conveyed a better understanding of the practice of political science.

The project requirements were motivated by the first rule of replication: one can't — at least not easily. This is true for seasoned practioners, as witnessed by the generally unsuccessful efforts of Dewald et. al. to replicate a series of studies published in the *Journal of Money, Credit and Banking*, and it is certainly true for students constrained by the semester calendar.[3] Consequently, I imposed considerable structure on the assignment. The replicated study, for example, had to be selected from a set of journals that regularly publish empirical research, and cited in a subsequent study indexed in the Social Science Citation Index.[4] This focused the project on outlets accustomed to evaluating statistical analysis, and opened the possibility that other studies might provide clues on data work not obvious in the original article. In addition, students had to undertake all necessary data collection, and could not contact the study's author(s) to obtain his or her data. This was intended to encourage students to become familiar with standard data sources in the discipline, and to replicate studies based on archived data readily available on campus. Finally, students were required to develop a journal detailing the study, data collection and the process of replication. The journal was updated and returned twice during the semester, and provided information useful for assessing the progress of the projects.

I also offered suggestions to make the project easier. Foremost, I recommended the replication of studies offering detailed discussions of the data analysis, including presentations of summary statistics, intermediate results, tables and graphs. In addition, the replicated studies should conduct non-trivial, yet not overly sophisticated analysis. The benefit of the assignment derives from the exercise of replication, not in learning statistical techniques that extend beyond the scope of the class. I was also willing to negotiate the portions of a study to be replicated on a case by case basis. This gave students the opportunity to replicate parts of studies otherwise too ambitious given the course requirements. Finally, I encouraged students to replicate studies they found substantively interesting.

The completed projects were presented in a poster session during the last class meeting. The posters reported (1) an abstract of the article, (2) a description of the steps used to replicate the article (an abbreviated final version of the journal) (3) findings, including side by side comparisons of the tables, graphs and other presentations in the original study, (4) a discussion of steps needed to replicate the study beyond that offered in the original article, and, as was invariably necessary, (5) an assessment of why the study could not be replicated including a discussion of the substantive implications of the student's findings.

Of twelve completed projects, none produced an exact replication. Four or five students replicated portions, or obtained results close to those in the original study. These more successful projects typically obtained, or nearly recreated, summary and test statistics, and coefficient estimates of comparable magnitude and direction to those in the original study. An additional three or four projects were conducted in the same spirit, but clearly specified models or otherwise conducted analysis that differed from the original study. The remaining projects did not obtain meaningful results.

Beyond the general difficulty of the assignment, there were several commonly encountered barriers to replication. The most frequent was difficulty in deciphering or recreating authors' variable coding rules. The sometimes cursory discussion of coding required students to spend hours determining how missing values were assigned, variables rescaled and similar considerations. Similar, although less common, problems arose due to ambiguity regarding exact data sources used in studies (e.g. page numbers from statistical abstracts, questions from both the pre and post ANES, etc.). These problems often led to discrepancies, sometimes minor, sometimes significant, in the number of observations between the original study and replication, and, hence, differences in all subsequent analysis. Some discrepancies appeared easy to resolve, but proved otherwise. One student replicating a recent study from the committee outlier literature calculated mean committee interest group rating scores that differed, albeit slightly, from those in the published study. Several attempts to resolve these differences by accounting for possible membership changes, and other plausible explanations, were unsuccessful.

Those who came closest to replication selected articles with care and devoted considerable time and energy to the assignment. The most successful projects were based on studies conducted at moderate levels of analysis using widely available data. Among the five replications in the top

[3] W.G. Dewald, Jerry Thursby and Richard G. Anderson "Replication in Empirical Economics: The *Journal of Money, Credit and Banking* Project" *American Economic Review* 76 (1986): 587–603. Dewald et. al. produce two exact replications, and two approximate replications, out of nine studies published in the *JMCB* in the early 1980's.

[4] These journals include the usual suspects including *APSR, AJPS, JOP*, and *PRQ/WPQ*. Specialized journals such as *LSQ, Public Choice*, and *International Studies Quarterly* were also represented in the mix. The citation requirement eliminated the vast majority of studies from consideration.

tier, all were in the American sub-field, with three based on ANES data, while the other two obtained data from statistical abstracts and *Congressional Quarterly.* These studies were published in *APSR, JOP, WPQ, Public Choice* and *Public Opinion Quarterly.* Several of the less successful replications were in the fields of comparative politics and public policy, but I think this is more a consequence of selection bias in student article choice, rather than a reflection of the literatures. In particular, those experiencing the greatest problems placed substantive interest before other selection criteria and, I believe, underestimated the difficulty of the assignment while overestimating the opportunity costs of additional time spent in the library.

The pedagogic value of the project exceeded my rather high expectations. Even students unable to produce results similar to those in the original study learned much about research. The greatest benefits of the project centered on the development of research skills. Students honed practical skills such as reading and evaluating empirical studies, using archived data and conducting literature searches that will reduce start-up time and the likelihood of initiating an intractable project at thesis time. The project also fostered numerous discussions of the importance of replication in scientific research. These ranged from principles of inquiry, to the professional rewards and etiquette of replication. Finally, while the inability of early career graduate students to replicate often sophisticated studies does not necessarily reflect badly on the status of empirical research in the discipline, many analyses were well within the abilities of these students. The existence of often inexplicable differences between original studies and replications indicates the prevalence of inadvertent errors found in the *JMCB* Project holds in political science literatures as well.[5]

PS 441 – Winter 1995
Term Project: Study Replication

Political science is an empirical discipline, and major developments in the field are predicated on qualitative and quantitative analyses published in the leading journals of the profession. Unfortunately, these findings are rarely examined for sensitivity to changes in sample, interviewing techniques, data measurement and coding, model specification, estimation procedures, reporting accuracy or, in general, face scrutiny consistent with a more idealized notion of academic inquiry. This is unfortunate because this restricts the ability of scholars to build on the work of others. As part of the requirements for PS 441 you will have an opportunity to redress this imbalance by replicating the statistical analysis of an article from a leading political science journal.

This assignment fulfills three primary objectives. First, replicating an outstanding study in one's field is a useful and interesting way to learn the process of research. You are retracing the steps of a scholar whose work you respect. Second, replication forces one to become familiar with many of the standard data sources (see below) used in the discipline. This may well be important in subsequent research (e.g. dissertation) by reducing start up time or preventing initiation of an intractable project. Finally, the assignment provides the necessary "hands on" experience in data analysis expected in an advanced methodology course.

Guidelines for this project are motivated by a single concern: *Replicating a study is very difficult.* While I have not seen comparable figures for political science, an analysis of articles published in a respected economics journal indicated that the majority of the statistical analyses could not be replicated.[6] I suspect the same is true in political science. Many studies are non-replicable, or can only be replicated with great difficulty, because the author has failed to communicate sufficient information for subsequent analysis. This poses a fundamental problem for development of the discipline, and should indicate to you that the most important decision you will make in this project is your choice of study to replicate. With this in mind, I offer the following guidelines and suggestions.

1. Your study must be selected from the attached list of journals and have been published after 1974. This requirement does not suggest that quality statistical analysis does not appear in other leading journals. Rather, I have selected these journals because they regularly publish statistical analysis and promote editorial policies that will make replication easier. The 1975 forward restriction is largely arbitrary, but will limit your literature search and facilitate one of the other guidelines (2) listed below.

2. The study you select must have been cited at least once by subsequent research published in outlets indexed by the Social Science Citation Index (SSCI).[7] The SSCI is perhaps the most fundamental research tool in the social sciences, and if you are not familiar with it see me or a reference librarian to get up to speed. The citation should indicate the empirical findings of the original study, e.g. Jones and Smith (1992) find that .... This obviously limits your ability to replicate recently published articles. What may not be so obvious

[5]Indeed, my students did not fare poorly compared to the investigators in the *Journal of Money, Banking and Credit* project given the latter enjoyed the advantage of submitted data sets and consultation with studies' authors.

[6]Dewald, William G. et. al. "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project" American Economic Review. 76 (1986): 587–603.

[7]Self-Citations do not count. (E.g., Jones and Smith 1990 cite their 1985 paper.)

is that this requirement eliminates the vast majority of studies from consideration. One advantage of selecting a study that has been cited is that methodological issues glossed over in the original study may be reviewed in subsequent research.

3. I strongly suggest that you select a study conducted using data held by standard archiving organizations (e.g. ICPSR). Perhaps the single most significant constraint on replication is the inability to obtain or recreate the original data set used in the study under consideration. By using archived data regularly used in the profession, the initial stage of data collection is at least partially resolved. Further, you should select a study for which the original data is on campus. Examples of this include most American National Elections Studies (ANES), Eurobarometer, World Bank data and others. Data held by archiving agencies but not available on campus may be used, or you may choose to replicate a study that requires that you recreate the original data, but you assume the risk of delays in completing the assignment associated with each of these courses of action. *Under no circumstances may you replicate a study that is based on proprietary or personally held data sets (i.e. you may not contact the study's author to obtain his or her data).*

Given these provisos, consider the following suggestions.

- Select a study you find interesting. You will spend a considerable amount of time on this project, so pick a paper that is relevant to your scholarly interests.
- Select a study that conducts non-trivial, yet not overly sophisticated analysis. Avoid studies that use time series techniques, multi-equation models or dimenensional analysis. The importance of this assignment is in the exercise of replication and seeing the impact of methodology on a literature, not in learning statistical techniques that extend beyond the scope of this class.
- Select a study that clearly presents an adequate discussion of data analysis including intermediate results (e.g. tables with summary statistics, graphs, ect.). This will enable you to check your progress as you replicate the study.
- Finally, select a study that is reasonably well known and has had some impact on a literature. A good starting point may be articles you have read in classes and seminars.

Evaluation:
What you will turn in at various points in the semester:

<u>February 15</u>

1. A complete citation for the study you are going to replicate.
2. A complete citation for the study that cites the article you are going to replicate.
3. A statement detailing the data sources used in the study, if this data is available on campus or, if this information is not available at MU, what steps have been taken to obtain the data.
4. A photocopy of the study you are going to replicate.

<u>March 22:</u>

1. A <u>four</u> page report that clearly summarizes the article, data sources used, **all** data manipulation and preliminary analysis, methods of estimation, presentations (tables, figures, etc.) and findings. This summary, which will serve as your "play book" for replicating the study, should contain an ordered, detailed, presentation of all the analysis conducted in the study.
2. A one page summary of your progress in recreating the data set used in the study being replicated. The summary should indicate if you have obtained the original data, if it is in machine readable form, and overview any preliminary analysis you have conducted on this data.

<u>May 3:</u>

A final report containing all of the following information.

1. A brief summary of the article.
2. A description of your attempt to replicate the article. This is essentially a revised version of your "play book" where the revisions reflect the actual process of retracing the steps of the original researchers.
3. your findings, including "side by side" comparisons of **all** tables, graphs and other presentations in the original study.
4. If you are unable to replicate the study, a complete discussion of possible explanations why the study could not be replicated. This should include an assessment of the substantive implications of your inability to replicate the study. For example, do your findings still substantiate the major conclusions of the original study? Do they undermine the study's findings?
5. your data set and computer programs used in replication. The data set should be in ASCII format. All data files and programs should be sufficiently documented so that I can <u>easily</u> understand how you obtained your empirical results. Please turn these in on a labeled IBM 3.5" disk.

How your report will be evaluated:

This project will be graded on the basis of your ability to fulfill the objectives, including deadlines, outlined above. This does not mean that to do well on this assignment you have to be able to replicate the original study. As the description of this project indicates, replication is very difficult. However, the likelihood that you successfully replicate a study is both a function of the time and energy you devote to this assignment and the choice you make regarding the study to replicate. Consequently, I will pay particular attention to the following in evaluating this assignment.

1. The quality of your choice of study to replicate. Choosing a study to replicate requires that you balance several objectives. Choosing a paper with that presents several intermediate results, tables, graphs and the like will be judged more favorably than selecting a study that only presents the main results. Similarly, replicating a paper that presents overly simple or trivial analysis will not judged as favorably as replication of an interesting study pitched at a level consistent with a graduate level methodology course. Still, adhere to the caveat regarding the avoidance of overly sophisticated analyses.

2. Your ability to obtain the data set used in the study. This is obviously necessary for replication, and may prove to be a substantial problem for those who attempt to recreate an original data set.

3. How close you come to replicating the study under consideration. For example, do you obtain the author's intermediate results? Are your estimates close to those in the original study (perhaps indicating a coding problem in a single variable, a different optimization procedure in ML analysis). Your documented data set and computer programs will likely provide insight on why you were, or were not, able to replicate the study under consideration.

Good Luck!

Journals

General:

National/Regional Journals

American Political Science Review (American Political Science Association)

American Journal of Political Science (Midwest Political Science Association)

Journal of Politics (Southern Political Science Association)

Political Research Quarterly[8] (Western Political Science Association)

British Journal of Political Science

Canadian Journal of Political Science

Field/Subject Specific:

Comparative Politics (general):
World Politics
Comparative Politics *
Comparative Political Studies

International Relations:
World Politics
*International Organization**
International Studies Quarterly
Journal of Conflict Resolution

Miscellaneous:
Legislative Studies Quarterly
*Public Administration Review**
*Policy Sciences**
Political Behavior
Public Opinion Quarterly
American Politics Quarterly
Electoral Studies

Interdisciplinary:
Social Science Quarterly
Public Choice

* Publishes fewer statistical analyses than other journals in the category.

To: PS 441 Students
From: Jay Dow
Re: Term Projects

The week before spring break we proceeded on a no objection basis to change the final requirements for the 441 term project. In particular, we agreed to change the final requirement of the project from a report to a poster. I plan to make arrangements to display these for department viewing the week of May 3 in the form on an informal poster session. I will not, however, require you to present your work at this poster session if doing so makes you feel uncomfortable. Participation, of lack thereof, in this poster session will not affect your grade nor change my estimation of the quality of your work in any way. Still, I hope all of you elect to present your work as this is a valuable learning experience and very informal and non-threatening

[8] Formerly *Western Political Quarterly*

introduction to essentially what we as political scientists do as profession.

So there is no confusion, my expectations for the revised final project are that you will follow the format originally discussed, especially regarding presentation of the original findings, your replication results, graphics and the like, but the write-up and text will be reduced. In other words you should complete objectives 1-5 in the original assignment, but scale your presentation in a manner appropriate for a poster session.

On The Use of Computers to Develop Political Theory

Ken Kollman
University of Michigan

Introduction

We are taught in the beginning of graduate school of the importance of considering alternative hypotheses when testing social theories.[1] In principle, if a theory states that A causes B, and we demonstrate that our measurements of A and B covary in the right way, that is not enough. We ought to test for the possibility of spuriousness and other deleterious problems with the model. A and B may covary, but C, which may covary with both A and B, may be the more reasonable choice as the cause of B. When both theories (A causes B, or C causes B) are consistent with available data, then we need to derive further testable implications of the models or we need a critical test that will distinguish the two theories.

Sounds simple, right? Unfortunately, when the theories we wish to test empirically are formal or mathematical, testing alternative hypotheses can be extremely trying. The alternatives we care about may be difficult to formalize, complex models may be difficult to solve analytically, or the number of possible alternative hypotheses may be quite large, depending on the complexity of our theories. Recently, some rational choice scholars have been criticized for failing to undergo the exhaustive empirical testing many empiricists would like to see (Green and Shapiro 1994). These criticisms highlight, among other things, the difficulty of developing formal theories that are flexible enough to incorporate a wide range of assumptions yet strict enough to impose discipline on theorists.

Many alternative hypotheses we might want to test are actually variations of our models under different assumptions or parameter values. For example, Green and Shapiro (p. 183) criticize Lupia's (1994) paper on voters' information in California referenda for not testing the predictions of reference group theory of information against Lupia's own rational choice theory. Presumably, Lupia could have formalized notions from reference group theory and incorporated them into his model. This may or may not have made Lupia's model intractable, but for many alternative assumptions, such additions can create modeling "monsters," where closed form solutions are all but impossible.

In this essay, I suggest that computers can help advance theory by allowing theorists to "test" theoretical models under a variety of conditions. Ideally, one would have analytical proofs of assertions, but that is simply not possible for many of the ideas that political scientists care about. Incorporating alternative assumptions into a model often becomes difficult or impossible if the only standard is the analytical proof. As an acceptable alternative, numerical experiments, or computer simulations, can provide means to generate theoretical findings for complex models.

Analyzing Complex Models

Let us trace the history of a hypothetical theoretical model to show how computers can help advance theory. Say that the assumptions of the given formal model are represented as a 10–tuple A, where each element of A can be 1 or 0. For example, the assumptions of spatial voting models (the elements of A) could be deterministic (0) or probabilistic (1) voting, single dimensional (0) or multidimensional (1) issue spaces, two (0) or multiple (1) parties, fixed numbers of parties (0) or the possible entrance of new parties (1), and so on. Imagine that the first cut at the problem, in a famous article by So and So, shows a unique equilibrium (or attractor) for $A = \{1, 0, 0, \ldots, 0\}$. In subsequent articles, various theorists show that the equilibrium does not exist for $A = \{0, 1, 0, 0, \ldots, 0\}$ and $A = \{0, 0, 1, 0, 0, \ldots, 0\}$, but does exist for $A = \{0, 0, \ldots, 0, 1\}$. Each iteration of the research literature attacks the problem with A as an identity tuple. This means theorists are changing assumptions one at a time, but without combining several of the changed assumptions at once. An incremental approach is valuable but limited.

These various assumptions, represented by elements in the tuple, can interact. A model with $A = \{1, 1, 0, 0, \ldots, 0\}$ might lead to very different results than either $A = \{1, 0, 0, \ldots, 0\}$ or $A = \{0, 1, 0, 0, \ldots, 0\}$. Interactions among assumptions, in many cases, are not easily explored. Advances could be made in theory if researchers consider models where, for example, $A = \{1, 1, 1, 1, 0, 0, \ldots, 0\}$.

It is possible to consider more complex models using the computer, exploring the way assumptions and parameter values lead to different conclusions. Theorists may not need to be constrained by such modeling conventions as uniform

[1] Thanks to Chris Achen, John Huber, and Scott Page for helpful comments. They are not responsible for any errors in judgment or fact.

types in an incomplete information model, separable preferences in a spatial model, and strict convexity of preferences in a rational choice model. Just as advances in computers enabled everyday empirical research to break into a maximum likelihood world, so can the computer enable theorists to break into a world where draws do not have to be from a uniform distribution for every probabilistic event or action. Clearly, analytical proofs are the standard we should strive for in our models. But when our current state of mathematical sophistication is insufficient to derive analytical results, numerical experiments can be a valuable tool to advance theory. One potential of computer models is that they can permit researchers to develop intuition for how to proceed with analytical models.

To continue our example, suppose that one element of A is the choice of quadratic preferences (0), or more complex, "kinked" preferences (1) known (hypothetically) to exist from empirical research. Say results have been derived from a model assuming quadratic preferences. Do similar results occur if preferences are more complex? Can we incorporate what we know from empirical research about preferences? If numerical experiments demonstrate that results are similar when preferences are more "realistic," then the theory is strengthened. If results change dramatically, then we know that previous results rely (perhaps too) heavily on strict assumptions about preferences.

In modeling complex situations via computer, of course it is important to be incremental and rigorous, regardless of the assumptions relied upon. The simulations literature from the 1960s and 1970s foundered partially on the failure of researchers to focus on the one or two assumptions that really mattered in their models.[2]

In a number of papers I have written with John Miller and Scott Page (1992, 1993, 1994, 1995), we combine several assumptions into a model that would be, save for numerical experiments, impossible to solve analytically. Our model of two–party competition assumes that voters have varying saliences across issues, that parties are adaptive rather than fully rational, and that parties sample voters' preferences to help find winning electoral platforms. We solve a single dimensional version of our model analytically, but we then use numerical experiments to show that, for a large range of parameter values, the results from the analytical model hold up in multidimensional spaces. We even demonstrate with the computer that both discrete and continuous issue positions lead to the same results. In another paper, we analyze different political institutions in a Tiebout model, starting with a computer model and working our way towards analytical results. The complexity of the model—with voters migrating from city to city, voters voting in city elections, and parties trying to win elections anticipating voters' migrations—makes analytical results difficult. Yet the computer model offers some important clues about which assumptions matter and where to begin mathematical modeling.

In all of our research, we test results under a huge range of parameter values. Robustness of theoretical results should be considered as important as the robustness of empirical results.

Alternative Hypotheses

The computer is underutilized among formal political theorists, especially in deriving testable alternative hypotheses. It may be possible, for example, to model a version of reference group theory for voters in presidential primaries. An extremely simple, rational choice model of primary voting might analyze the assumption that voters choose among five or six candidates based on ideological proximity. A less simple, but more realistic, model might incorporate momentum effects in presidential primaries by including in utility functions voters' desires to vote for possible winners. A still more complex model might also include the possibility of strategic voting within a two-stage game, where the first stage is the state primary and the second stage is the general election. The model may be highly stylized at this point, but it may generate valuable counterintuitive predictions. To take one possibility, say the model predicts that a voter will choose the most proximate candidate within some threshold of ideological distance. If the voter is very close ideologically to the candidate, the voter will choose the candidate regardless of momentum or probability of winning in the general election. For voters beyond the threshold of ideological distance from a candidate but far enough from other candidates, they will tend to vote for the party candidate with the highest probability of winning the general election. In other words, each candidate's closest ideological adherents will vote for that candidate, but all other voters will vote for the expected winner. Let us say further that empirical results tend to support these predictions.

A reference group theory of voting would offer an alternative explanation for the same empirical evidence, but might provide a further set of predictions. Voters ideologically close to a particular candidate will vote for the candidate because they identify with other people who are ideologically close to the candidate. However, voters not close enough to any candidate may place different probabilities on how the candidates will fare in the general election, and these probabilities will depend on each voter's reference group. So the degree to which voters will jump on the front-runner's bandwagon will vary with sources of information and reference group.

[2]Thanks to Chris Achen for this point.

Incorporating different probability calculations among voters and the possibility of having voters exchange information with like-minded citizens may very well render unworkable any analytical solutions to a formal model of primary elections. But a computational model could analyze such a scenario. Here is a brief sketch. Assume citizens make voting decisions much like Downsian spatial voters, but with additional elements to their utility functions. Allow voters to exist on an ideological grid, with each point a policy position. Characterize voters by where they are on the grid, by how much they want to vote for winning candidates in the primary, and by how much they want their party's candidate to win the general election. Then allow them to interact with ideologically proximate citizens, updating their calculations of candidate viability according to these interactions. By interact, I mean that with some probability each citizen copies a neighbor's information. After a finite time of interaction with other citizens, have a primary election. One could even create new voters each primary election and run the series over 40 or so elections, simulating the primary season. Recording and analyzing aggregate patterns of this simulated primary voting ought to offer insights into the advantages of reference group theories versus rational choice theories of elections. One could probe which set of assumptions led to improved empirical prediction in primary voting.

Conclusion

Certainly part of the appeal of formal theory is to develop rigorous, testable models of how real political institutions and real people and organizations behave. However, analytical methods sometimes constrain the assumptions we can incorporate into models. Often we care about a complex mix of assumptions when analyzing a model and when testing a model with data. Computational modeling, if done carefully, can help in analyzing a model and in deriving testable propositions.

Admittedly, we are not far along in our techniques for modeling complex social systems. But given the potential to learn more about political and economic life, I encourage the continued development of such techniques.

References

Green, Donald and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory.* New Haven: Yale University Press.

Kollman, Ken, John H. Miller, and Scott E. Page. 1992. "Adaptive Parties in Spatial Elections." *American Political Science Review.* 86:929–37.

Kollman, Ken, John H. Miller, and Scott E. Page. 1993. "Political Parties and Electoral Landscapes." Working Paper: Northwestern University.

Kollman, Ken, John H. Miller, and Scott E. Page. 1994. "An Investigation of Adaptive Parties' Platform Convergence." Working Paper: California Institute of Technology.

Kollman, Ken, John H. Miller, and Scott E. Page. 1995. "Policy Position–Taking in Two–Party Elections." Manuscript.

Lupia, Arthur. 1994. "Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review.* 88:63–76.

Note from the Co-Editors: Interested readers of Ken Kollman's article may wish to see computational analysis of political theory in action. The Pascal program used by Kollman and colleagues in their 1992 *American Political Science Review* article is available on the Political Methodology Home Page (see Jonathan Nagler's article on "Polmeth" later in this issue of *TPM*.)

Out of a Rut and Onto The Superhighway

Richard J. Timpone and Charles S. Taber
State University of New York at Stony Brook

Is political methodology in a rut? While the query itself may offend some political scientists, posing such questions can be useful for evaluating the state of a field. How one defines a rut, though, is important in answering the question. If being in a rut implies the absence of innovation and progress or rigid adherence to inappropriate tools, then political methodology is clearly not in one. If, however, we consider a stricter test that of the adequacy of these methodological tools for examining many of the substantive questions that political scientists pose the answer is less apparent. Of course, the appropriateness of methodological applications depends on the nature of the research question.

In this paper, we will briefly describe a class of modeling tools that, while far from unknown, has received insufficient attention in political science. While significant advances have been made in the development of a variety of methods, including formal modeling, time series and structural equation analyses, many complex social or political theories remain difficult to examine with traditional techniques due to the need for closed form solutions or convergence. Computational modeling techniques may allow complex theories to be formally represented and explored. Note that we do not suggest that all research questions are this complex. Even basic empirical techniques like regression and logit/probit or simple formal models continue to provide useful substantive insight into a variety of political questions (see Achen 1992 for a counter viewpoint). We will discuss the types of research questions that may benefit more from computational modeling though, after it is briefly characterized.

The title of this article is somewhat misleading since computational modeling uses computers as more than platforms for word processing and statistical software or on-ramps to the information superhighway. Computational models are theories (broadly defined) rendered as computer programs. While this includes a diverse array of techniques, a number of their advantages over other methods are held in common (for more detail see Hastie 1988; Taber and Timpone 1994, forthcoming): (1) computational modeling enforces theoretical and operational precision while allowing for uncertainty when necessary; (2) it is more expressive than other formal languages, allowing a wider range of theoretical statements beyond standard mathematical form; (3) it facilitates the simulation of counterfactuals; (4) it offers more deductive power by allowing the theorist to explore the combined implications of several theoretical assumptions; and (5) it allows us to examine more complex theories (compare with Fiorina's 1975 advocacy of formal models in political science).

Computational modeling procedures can be summarized into three broad categories: dynamic models, knowledge based systems, and machine learning models (Schrodt forthcoming; Taber and Timpone forthcoming). Dynamic simulation models are the most familiar to political scientists and are most closely related to statistical and formal methods expressed in standard mathematical form. In this approach, theoretical processes are generally cast in the form of equations, and model behavior is examined through systematic numerical analysis on the computer.

Knowledge based and machine learning models include much of what is commonly called artificial intelligence (AI). Knowledge based approaches, including semantic networks, frame systems, and expert systems, build on the information processing framework from cognitive and organizational science. Most fundamentally, they share the knowledge representation hypothesis: all intelligent information processing is based on the application of stored and organized knowledge. Since advances in cognitive and organizational science have been influential in theories of political behavior and institutions, these methods can be particularly helpful for these research questions. The integration of social and cognitive psychological theory into political behavior can be advanced by models that formally structure political knowledge. The parallel between the cognitive revolution and computer science has led to substantial use of the latter for modeling the former. Machine learning methods, including neural network analysis, rule induction engines, and genetic algorithms, do not subscribe to the knowledge representation hypothesis. ID3, for example, can be used to induce rules from categorical data, and genetic algorithms can provide an alternative to more standard statistical optimization routines, with some advantages in avoiding local maxima.

In addition to being able to model certain theories more flexibly than through mathematical equations, computational models also allow the development of more complex theories of political processes. One of the thorniest issues for computational modelers, though, is validation. While there is no direct way of testing the validity of a theorized process, we propose three methods to examine computational process models. These include face validity (as is often done with panels of domain experts in evaluating expert systems), testing of model assumptions (including both accuracy and sensitivity analysis), and multi-level outcome testing. All models attempting to represent processes pose special problems for validation and can gain support using these techniques. An advantage of computational models, is their ability to explicitly incorporate complex processes in their representation in an attempt to substantially expand our understanding.

Having suggested what methods we include in computational modeling and some of their merits, we would like to return to our original concern with the application of methods to political research. Creating a parsimonious summary of political research may help to illustrate the potential of computational modeling for particular types of theoretic questions. Elsewhere, we have developed a taxonomy of social science research models that can be applied to the questions posed by political scientists (Taber and Timpone forthcoming). We focus on three dimensions of research: the degree of simplification, level of analysis, and time scale of model processes. The most important of these are illustrated in Figure 1.

Occam's dimension, represented by the horizontal axis in the figure, deals with the degree of simplification in a theoretical model and is named in honor of William of Occam's call to reduce questions to their core components. Although parsimony is an important tenet of research, discarding important factors or processes can lead to misunderstanding. The degree of simplification is therefore clearly an important research choice. The vertical axis highlights the level of analysis, ranging from micro to macro. We do not argue that some basic level of explanation is appropriate for all political theories. The appropriate level of analysis for a given theoretical model depends entirely on the question being examined. The temporal dimension of research questions, not included in Figure 1, ranges from nearly instantaneous processes, such as cognitive reaction times, to epochal phenomena such as societal and cultural evolution. Although there is a substantial correlation between this dimension and that for unit of analysis, they are logically distinct and include substantial variation as we move toward the center of each.

The question posed earlier in this article dealt with the adequacy of current methods for examining questions of interest to political scientists. Present approaches are best suited to answering questions that are towards the parsimonious end of Occam's dimension and towards the middle of the level of analysis and time dimensions. Formal analysis will not be able to achieve closed form solutions for more complex problems, while empirical analysis has difficulty at the extremes of the levels of analysis or time dimensions where measurement is difficult. Computational modeling will be (and already is proving) most useful in these areas. In terms of the taxonomy in Figure 1, computational modeling can extend our analyses toward the center of Occam's dimension, and toward both extremes of the levels of analysis and time dimension where processes become less directly observable.

Computational modeling can also aid in the synthesis of current theoretical understanding. We can combine formal and empirical findings into more complex structures that may be simulated using computational methods. To be truly useful, of course, these findings must be empirically tested. This synthesis of existing theories may be even more useful when theories are combined that span different levels of analysis. The combination of individual, institutional and contextual factors is often more than mere summation, and computational modeling can be very helpful in examining more complex interactions in these areas than is commonly done at present.

So, is political methodology in a rut? The techniques that are commonly employed by political scientists continue to extend our understanding of politics and provide us with greater insight into the underlying processes involved. While tuning' and incrementally modifying standard methods allows them to more adequately test a broader array of theoretical questions, we should also be looking to new classes of methods. Developing new tools, like computational modeling, may enable us to examine questions that are difficult or indirectly done with current techniques. Applying appropriate methods provides us with the soundest tests of our theories, and will result in the greatest progress in political science.

References

Achen, Christopher H. (1992) "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." Political Behavior 14 : 195-211.

Fiorina, Morris P. (1975) "Formal Models in Political Science." American Journal of Political Science 19: 133-159.

Hastie, R. (1988) "A Computer Simulation Model of Person Memory." Journal of Experimental Social Psychology 24: 423-447.

Schrodt, P.A. (forthcoming) Patterns, Rules and Learning: Computational Models of International Behavior. Ann Arbor, MI: University of Michigan.

Taber, Charles S., and Richard J. Timpone (forthcoming) Computational Modeling. Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage.

Taber, C.S. and Timpone, R. (1994) "Pictures in the PC: Computational Modeling and Political Psychology," in M.X. Delli-Carpini, L. Huddy, & R.Y. Shapiro (eds.) New Directions in Political Psychology. Greenwich, CT: JAI.

GAUSS Code to Generate Random Variables

Christopher J. Mooney
West Virginia University

In conducting Monte Carlo simulation, one must be able to generate variables from specific distribution functions. It can make an enormous difference to the outcome of a simulation experiment, for example, if a given variable has been generated as a normal variate versus one distributed as chi-square. It is also important to be able to set the parameters of these functions at specified values. These parameters may set the dispersion and location of the variable's distribution (as with the mean and variance of a normal distribution), and/or they may affect higher order characteristics (as with the degrees of freedom of a chi-square distribution). Substantive theory will usually dictate at least some characteristics of the distribution to be generated (such as whether it is to be discrete or continuous), and the design of the simulation experiment may dictate others (such as how to vary the skew of the distribution across runs).

The GAUSS code available on the Political Methodology server will generate random variables from a variety of distributions that can then be used in simulation experiments. There are three general categories of distributions here: univariate continuous, univariate discrete, and multivariate continuous. The latter category is of special interest in that the generation of multivariate distributions allows the experimenter to set a pseudo-population correlation matrix among any number of variables with ease and accuracy. Also included in this code is a procedure for conducting a Jarque-Bera test for normality, which delivers skew and kurtosis estimates for the tested variable as intermediate output. This procedure, along with the MEANC(x) and STDC(x) commands, allows for the estimation of the first four moments of a generated variable, and thereby allows for an assessment of its approximation to the theoretical moments of the specified distribution.

For further information on these and other useful distribution functions, see:

Johnson, Mark E. 1987. Multivariate Statistical Simulation. New York: Wiley.

Johnson, Norman L., and Samuel Kotz. 1970. Continuous Univariate Distributions, vols. I and II. New York: Wiley.

Johnson, Norman L., Samuel Kotz, and Adrienne Kemp. 1992. Univariate Discrete Distributions. 2nd ed. New York: Wiley.

For further information on the generation and use of these distributions in Monte Carlo simulation, see:

Mooney, Christopher Z. 1996. Monte Carlo Simulation. Sage University Paper Series on Quantitative Applications in the Social Sciences, forthcoming. Newbury Park, CA: Sage.

Rubinstein, Reuven Y. 1981. Simulation and the Monte Carlo Method. New York: Wiley.

Note from the Co-Editors: Interested readers of Christopher Mooney's article may wish to use Mooney's GAUSS code for generating many of these random variates. It is available on the Political Methodology Home Page (see Jonathan Nagler's article on "Polmeth" later in this issue of *TPM*.)

American National Election Studies on CD-ROM

Erik W. Austin
Inter-university Consortium for Political and Social Research

Steven J. Rosenstone
Center for Political Studies, University of Michigan

The study of voting, public opinion, and electoral participation in the United States has come to the desktop with the release of the ANES CD-ROM. With this new resource, researchers, teachers, and their students will have ready microcomputer access to a wealth of data from the longest-running survey in the social sciences, the American National Election Studies.

What's Included on the CD-ROM?

- All 22 time-series studies conducted in 1948, 1952, 1956, and biennially from 1958-1994
- Other major NES data collections including the 1956-58-60 Panel, the 1972-74-76 Panel, the 1980 Major Panel, the 1984 Continuous Monitoring Study, the 1988 Presidential Nomination Process Study, the 1988-90-92 Senate Election Study, and the 1952-1992 Cumulative File which pools into a single 37,000 case data file variables included in three or more time-series studies
- Electronic versions of the codebooks, most with full frequencies
- SAS and SPSS data definition statements for all of the data files
- NES *Bibliography of Data Use*; the NES *Continuity Guide* to questions asked over the years; and files describing the history of NES, the research organization, and the study planning process, as well as a list of technical and pilot study reports

A Wide Range of Research Topics

The data collections contained on this CD-ROM cover a broad array of topics. Best known is the coverage of elections and electoral campaigns: vote choice; electoral participation; engagement in politics and the campaign; expectations regarding election outcomes; partisanship; evaluations of the political parties and their candidates. The National Election Studies also contain a wealth of information on Americans' opinions toward social and political issues over the nearly five decades and the administrations of ten U.S. presidents: arms control and disarmament; national defense and the likelihood of war; opinions about the Korean, Vietnam, and Persian Gulf wars; crime, the death penalty, and gun control; civil rights, integration, government aid to minorities, and affirmative action; prayer in schools, protest and urban unrest; medical care and health insurance; inflation and unemployment; the federal budget deficit; and abortion. There are also measures of political predispositions such as authoritarianism; moral traditionalism; patriotism; trust in government; political efficacy; individualism; racial prejudice; and religious attitudes and practices. Social and economic attributes such as occupation and employment, income, age, race, gender, ethnic origin, family composition and living arrangements, religious affiliation, and educational attainment, to name a few, can be found in the NES collection as well.

Special Features

The CD-ROM contains specialized front-end software (developed by the Center for Human Resource Research [CHRR] at Ohio State University) for use with the 1952-92 Cumulative Data File, the 1992 Election Study, and the 1980 Major Panel File. With this software, users can browse the codebook, select variables, and extract subsets of variables and/or cases for subsequent statistical analysis. The software permits the generation of raw data files for selected variables and cases; SPSS or SAS data definition statements corresponding to the subset selected; codebook documentation tailored to the extracted variables; and dBase III-compatible files.

A CD-ROM drive capable of reading ISO 9660-formatted CD-ROMs is required. All ASCII files on the NES CD-ROM (the raw data, the codebooks, and the SAS and SPSS data definition statements) can be accessed and read by IBM-compatible personal computers equipped with Microsoft [CD-ROM] extensions; by Macintosh computers running PC Exchange, Apple File Exchanger, or Version 7.5 of the Apple Finder; and by UNIX-based machines. The CHRR data extraction software runs under DOS (3.2 or higher).

Obtaining the ANES CD-ROM

Faculty, staff, and students at ICPSR member institutions can access the ANES CD-ROM by contacting their campus's ICPSR Official Representative, or by purchasing an individual copy of the disc directly from ICPSR. ICPSR has made available several copies of the ANES CD-ROM to each Official Representatives at its over 300 member institutions. Individuals at these member institutions may purchase personal copies of the CD-ROM for the discounted price of $30 apiece (includes shipping and handling; $35 for overseas orders).

Purchase of the ANES CD-ROM by individuals not located at ICPSR member institutions is also possible, at a price of $65 per disc (includes shipping and handling; $70 for overseas orders).

Individuals interested in ordering the CD-ROM can download an electronic copy of the order form from the ICPSR Gopher (gopher.icpsr.umich.edu) or from the NES World Wide Web homepage (http://www.umich.edu/ nes/). Individuals can also fax the ICPSR at 313-764-8041, or contact it by mail at ICPSR, P.O. Box 1248, Ann Arbor, Michigan, 48106-1248.

Announcement of Gosnell Award Recipients

Larry Bartels
Princeton University

The Political Methodology Section's first annual Harold F. Gosnell Award for Excellence in Political Methodology was awarded at the Section's business meeting at the annual meeting of the APSA in Chicago on 1 September 1995. The award is for the best methodological work presented at a political science meeting in the preceding calendar year.

The first Gosnell Award was shared by Janet M. Box-Steffensmeier of Ohio State University and Renee M. Smith of the University of Rochester for their paper on "The Microfoundations of Aggregate Partisanship," presented in July 1994 at the Eleventh Annual Political Methodology Summer Conference in Madison, and by Bradley Palmquist of Harvard University for his paper on "Respecification Approaches to Ecological Inference," presented in September 1994 at the annual meeting of the APSA in New York. The award winners were selected by a committee consisting of William Berry of Florida State University (chair), John Freeman of the University of Minnesota, and Douglas Rivers of Stanford University.

The Gosnell Award honors the distinguished career of Harold F. Gosnell, a pioneer in the application of quantitative methods in political science and the leading exponent of behavioral political science in the first half of the 20th century. Gosnell's book on *Non-Voting*, published with Charles Merriam in 1924, was based on the most ambitious sample survey of its era. His *Getting Out The Vote*, published in 1927, reported the first – and perhaps still the most elegant – experimental study of political behavior. His 1935 article in the *American Political Science Review* on "An Analysis of the 1932 Presidential Vote in Chicago" was the first study in the political science literature to employ multiple regression analysis, and also the first to employ factor analysis. Gosnell was a mainstay of the University of Chicago political science department during its heyday in the 1920s and 1930s; his distinguished students included Gabriel Almond, V. O. Key, Jr., Herbert Simon, and David Truman. He worked for the United States government in the 1940s and 1950s and taught at Howard University in the 1960s. Professor Gosnell is currently living in retirement in Maryland, and will celebrate his 100th birthday on 24 December 1995. His recollections of "The Marriage of Math and Young Poli Sci" were published in *TPM* in 1990.

Professor Gosnell is currently living in retirement in Maryland, and will celebrate his 99th birthday on December 24, 1995. The Political Methodology Section was honored to have him attend the first annual Gosnell Award presentation in Chicago along with several members of his family.

A Review of Stata v. 4.0

Charles H. Franklin
University of Wisconsin, Madison

Version 4.0 of Stata is the latest in a long line of releases of this excellent statistics package which has been around since 1985. Stata has a proven track record of providing many statistical procedures years before SAS or SPSS have them and being blazingly fast to boot. In version 4.0, Stata has produced its first Windows version (and has just announced a Windows 95 release). This version adds several statistical procedures (cross-section time-series models, generalized linear models, a new kernel density estimator, improved bootstrap and simulation programs, and a better version of the Heckman model of sample selection) as well as exploiting the Windows interface to (generally) good advantage. Best of all, the new version is at least as fast as previous ones, which were very speedy indeed.

Stata's speed comes in part from keeping data in RAM. This makes even complex statistical procedures such as multinomial logit amazingly fast. The price is that you need the memory to store the data, and for large datasets, it is possible to run out of memory. While Stata will run in less, I think 8 megs of RAM are the minimum, and you'd probably want 16MB if possible. Of course, you probably want that much for most recent software, so Stata is only one more reason to add memory.

Because Stata stores the dataset in RAM, it is particularly good for building aggregated datasets. Its match-merge capabilities are also excellent, so you can read in a dataset, aggregate it across some strata, and merge it with another dataset very easily. If you work with timeseries in which you often wish to aggregate from monthly to quarterly data, or some such, this feature is particularly inviting.

In general, Stata's data handling capabilities are very good. Its recode facility is good, though not as extensive as SAS. Once you are familiar with Stata's recode style it is easy to do even fairly complex recodes in only a few lines of code. One odd feature is that Stata has separate commands for creating a new variable versus changing values to an existing variable. The `generate` command creates a new variable, while the `replace` command changes existing values. This is easy enough once you are used to it, but can lead to apparently odd sequences such as the following:

```
generate newvar=old1 if state==1
replace newvar=old2 if state==2
```

While it takes a little practice to learn that `generate` is only used the first time you reference a new variable, and `replace` is used thereafter, once you do this is not a major problem.

A more substantial annoyance is that Stata does not always partition memory in "reasonable" ways. When it does not, manual intervention is required, which is surprising in such an otherwise excellent program. When first invoked, Stata makes a default choice of memory partitions which maximizes cases rather than variables. I have many times read in a small dataset (a hundred or so cases) and built a number of new variables. At about 50 variables, Stata balks, saying it is out of memory— an absurdity given 16 MB of RAM. The cause of the problem is that Stata has set aside space for something like 8000 observations, when I have only 100. This is not hard to fix, by issuing a command to set the maximum number of observations and/or variables. However, it is always a surprise that Stata does not automatically reconfigure memory to suit the dataset. The Stata technical support people have said on more than one occasion that they know this is an annoyance but the fix has yet to appear. You can largely avoid this problem by issuing the `memsize` command before reading a dataset. This will calculate 3 alternative memory partitions and allow you to select the one you like best. I just don't understand why this isn't automatic.

Because Stata reads all the data into RAM, it is not ideal when building very large datasets, such as an entire National Election Study. I have found that SAS is much better in these cases, because it reads only one record at a time and therefore never runs out of memory. I find that by using SAS to build subsets of larger studies, and then converting the SAS file to a Stata file, I can have the best of both worlds: SAS's superb data manipulation procedures, and Stata's lightning fast analysis routines. Stata sells a conversion program, called Stat Transfer which works reasonably well (though my last upgrade did not seem to solve some SAS-Stata problems with large datasets.) I now use DBMS/COPY for such transfers, and have had no problems even with large datasets. One advantage of DBMS/COPY is that it can read SAS for Windows files directly, so there is no need to create an intermediate transfer file. Either of these transfer utilities makes it easy to build files in SAS or SPSS and then convert the files to Stata for analysis.

The statistical and graphical routines in Stata are the best reason to buy it. Stata has, in every version, been months and years ahead of SAS and SPSS in incorporating modern statistical methods. Stata is almost as good as Limdep in its support of limited dependent variable models, and is much easier to use than Limdep. Stata covers all the standard models for limited dependent variables: binary logit and probit, ordered logit and probit, unordered logit, several alternative event count models, censored (tobit) models and survival models of various sorts. This is an impressive collection, especially given the comprehensive range of Stata's other routines, which run from standard regression to simultaneous equations to factor analysis. Stata also has a growing number of non-parametric modern regression methods, such as lowess regression. Where a specialty package such as Limdep beats Stata is in the non-standard extensions to models, such as heteroskedastic probit, for example. These are seldom needed, but would be nice to see in Stata. In the mean time, Stata runs circles around SAS, SPSS and SYSTAT for the range of advanced models it supports as well as the ease with which it estimates them.

The graphics in Stata are excellent, if not quite up to those of S+. Stata does very good scatter plots, and allows considerable customization of the plots, such as labeling points, including regression lines and labeling the axes. Stata also supports box plots, stem and leaf plots (somewhat less successfully), and one-way scatter plots. Time-series plots are also easily done. What Stata does not support are dot-plots and trellis displays which have recently been incorporated into S+. It would be nice to see these features appear in future versions of Stata. In the mean time, the graphics are more than good enough for 95% of the work you will do.

The new Windows version is a very nice interface, though some puzzling omissions detract from it a bit. Basically, the Windows version provides five windows, each of which can be resized. The output window shows the text printout for all commands. The graphics window shows all graphs. A command window is essentially a single line for typing commands. The nicest innovations are a history window which shows recent commands and a window listing all variables in the current dataset. The previous commands can be double-clicked to move one to the command window for either editing or execution. The variable list is a real improvement since you can scroll through it to find variables whose names you just can't quite recall.

Because these windows can all be resized, it is easy to arrange them to suit your tastes. Stata will save the configuration as the default so you don't have to arrange the windows each time you run the program. I have found that by using a high resolution VGA setting, I can put the output on the left half of the screen and the graphics on the right half. This leaves enough room for the other three windows below the two main ones. In this setup, Stata slightly resembles S+ with its command and output window on the left and graphics on the right. If you have a 17-inch or larger monitor, this is quite readable, and gives everything you want on the screen all at once. One trick to making this arrangement readable is to reset the fonts from Stata's defaults to ones which are a little bigger. The menu in each window allows this, though the output window is the most critical.

The major flaw in Stata's Windows version is that some obvious Windows features are missing from the five Stata windows. Most obvious is that minimize/maximize buttons are missing from the upper right corner of each window. This is particularly annoying when the windows are set up as I like them, because there are times when you want to expand a graph to full screen or do the same for an output window. It would make enormous sense to be able to hit a maximize button and see the window full screen and then be able to shrink it back to normal size. Alas, you can't. You must drag the borders of the window to full screen. You can restore everything by clicking on the windows menu, but this seems like a standard Windows feature so I am at some loss to understand why it is absent. Likewise, there is no scroll bar on the output window or the previous command window. This lack is most annoying for the output window. If you are logging output to a file, you can bring up the log file and scroll through it, but this seems a needless extra step. Since this is the first Windows version of Stata, we can hope these defects are eliminated in the next release.

Despite these annoyances, I find that I am using the Windows version virtually all the time. I am not a Windows fan, so I was dubious of the Windows version of Stata, but the ease of use has won me over. I still use the DOS version when running batch files (which are easily created from interactive sessions by logging commands only). The primary Windows advantages are the multiple windows which makes interactive exploratory work easy, and the point-and-click ease of printing graphs. In keeping with good programming practice (see Nagler's article in the previous issue of TPM), you should keep a log file of your commands, and Stata makes this trivially easy. Even if you forget to log commands, you can retrieve them during a Stata session and save them to a file. This allows interactive exploration with excellent record keeping of what was done.

In sum, I find I use Stata for probably 80% of my statistical work. Whenever I have a standard model, Stata almost certainly has the command to estimate it. Even things which are not routine, such as Heckman selection models, censored regressions and event count models are well represented in Stata. There is also a general maximum likelihood routine which will allow you to write your own likelihood function. The syntax takes a little getting used to, and it is not as fast as Gauss, but with practice it may take less time to get results than if you have to move your data over to Gauss (DBMS/COPY or STAT/TRANSFER can help) and write the Gauss program from scratch.

Finally, I have found that Stata consistently has excellent technical support. I've very seldom asked a question that they could not answer, and I almost never have to wait to get someone on the phone. Support is toll-free as well. Pricing is also inviting, especially for group purchases (I require Stata in my graduate intro stats course) and for labs. Individual faculty licenses are also reasonably priced, and there are no add-ons required before you have a complete system.

If I didn't own Stata, I'd rush out and buy it. For a package as remarkably complete and as easy to use as Stata, there is no real competition. If you are only going to own one stat package, this would be my choice. And if, as I do, you own a half-dozen you will find that you use Stata more than any other for data analysis.

A Review of Stata v. 4.0

Jim Wenzel
University of California, Riverside

Despite developments in the realm of high end statistical software, many political scientists still rely on one or more general purpose econometrics packages to meet their everyday statistical computing needs. This is not terribly surprising. Programs like GAUSS, and to a lesser extent S-Plus, while offering unmatched flexibility and power, require a commitment to the mastery of programming languages the many cannot make. SAS, although incorporating what must be every statistical tool known requires its own particular set of incantations to bring it to heel and can consume prodigious amounts of system resources in the bargain. Moreover, if SAS' cost and size fail to overwhelm one's checkbook or one's hard drive, loading up on the requisite manuals is sure to inflict damage both to finances and bookshelves.

There are a number of very good general econometrics packages that may satisfy the everyday needs of users who for one reason or anther cannot commit to Gauss, S-Plus, or SAS. For example, Shazam and SST can handle most tasks and are programmable. They don't exact the kind of programming commitment Gauss does, or the disk (and bookcase) space that SAS does. Those looking for a competent and economical everyday statistical package would do well to add Stata to the list of packages under consideration.

Written and distributed by the Stata Corporation in College Station, Texas, Stata offers substantial analytic power in a compact, reasonably priced package. Actually, I should have said that Stata comes in compact, reasonably priced package. Stata has been ported to DOS, Windows, Windows 95, Windows NT, Macintosh, and Unix.[1] For owners of Intel x86 PCs the program comes in two "flavors", "small" and "intercooled". "Small" Stata is intended for users of PCs based on Intel 80286 processors or lower, as well as users of 80386 based computers without math coprocessors. It places significant limits on the size of the data sets that can be loaded, the numbers of variables that can be analyzed, macro sizes, and command lengths. By comparison, "intercooled" Stata is intended for use on 80486 or better processors and is capable of handling much larger data set (up to 2,047 variables with the number of observations limited only by available memory), larger macros (2,296 characters), and longer command lines (3,072 characters as opposed to 1,000 in "small" Stata).[2] In addition, differences in the base code between small and intercooled Stata make the intercooled version substantially faster. Unlike SAS, which can consume up to 200 megabytes of drive space, the "intercooled" version of Stata for Windows installs in less than three megabytes.

Stata has long relied on a command driven interface. Users have the option of entering commands interactively and examining the results as the appeared or in command file form, generally using a text editor to edit the command file and view the results. With the winter 1995 appearance of Stata for Windows (and, more recently, the Windows 95 and Windows NT versions) a Graphical User Interface (GUI) has been added to the program. Those who have used SPSS for Windows, or SPSS as an X client on a Unix platform, will find the look of Stata for windows faintly familiar. A large window is dedicated to the display of results while two smaller windows to the left of the display contain respectively, a list of the variables in the active data set and a log of previously issued commands and command line interface lies at the bottom of the Window. On call is a spreadsheet type data editor that allows the analyst to make minor alterations to the data on the fly.[3] However, users hoping to point and click their way through analyses are destined to be disappointed. Unlike SPSS' "statistics" menu which provides a menu of available tools, Stata requires the analyst to enter the name of the procedure to be run, after which the variables can be selected through the use of the mouse. Once a procedure has been run, however, it can be re-run with minor changes with a couple of mouse clicks. These features are helpful primarily to those doing exploratory analysis. Stata remains, at heart, a command oriented program. I, for instance, run Stata for Windows in OS/2 and only occasionally take advantage of its windowing capabilities. In all but a few instances, I simply use the interface to run command files, keeping the command file open in one EMACS window and use another to view the results.

The program has a number of strengths that make it attractive to the general user. First and foremost, Stata includes an array of analytic procedures ranging from A (one-way, two-way, and N-way analysis of variance) to W (Weibull regression). In between are included the core procedures on which most researchers depend. In addition to standard (OLS) regression techniques (including 2-stage least squares, 3-stage least squares, and weighted least squares), factor analysis, reliability (Chronbach's alpha and Kappa) the program includes routines to estimate a variety of general linear models (GLM), Poisson regression, and non-linear least squares. Also included are an array of graphical diagnostic tools.

Stata also has broad array of techniques for dealing with limited dependent variables. In addition to the industry standard logit and probit routines, Stata is able to estimate ordered probit and logit models, logistic regression, conditional logistic regression, and multinomial logistic regression.[4] For dealing with count and event data, Stata offers, in addition to Weibull and Poisson regression, implementations of tobit regression, Cox proportional hazards models, and survival analysis.

Stata incorporates its own programming language, enabling users to augment the canned routines provided in the program As a result (as with S), a cottage industry has sprung up in the development of third party additions to Stata. A number of these additions, along with general operating tips, are made available to users via a semi-monthly newsletter (subscription price approximately $35 per year, more if disk versions of software updates are included). Enhancing the program's flexibility is the fact that many of the "canned" routines Stata offers exist in the form of separate subroutines that are accessible to the user and which can be modified if necessary. Also, included is a matrix programming language and a separate maximum likelihood routine for the estimation of user specified maximum likelihood models.

Second, the program is quite fast. To illustrate I used a command file that loads a 205 variable, 6037 observation data set (approximately 2.5 megabytes), creates 22 new variables, performs 22 conditional recodes, five frequencies, four probits (averaging eight independent variables and 6000 observations), and finally, runs a set of predicted values of the dependent variables base on postulated values

[1] The Unix variant is available for DEC Alpha and RISC, HP 9000, IBM RS/6000, and Sun (SPARC) computers.

[2] "Intercooled" Stata for PCs is the functional equivalent of Stata for Unix.

[3] The emphasis is on minor. Users doing wholesale data transformations will most likely find that the editor is no substitute for a command file.

[4] To compliment these routines, Stata's "predict" command allows the user to clear the existing data from memory while retaining the parameters estimated by whatever routine was just run. The user can then enter a new data matrix containing a set of postulated values for the independent variables and Stata will return the predicted values of the dependent variable.

of the independent variables. With Stata for Windows running in a Windows session under OS/2 on an otherwise idle (more or less) 90 Mhz pentium, the total elapsed time for all procedures was 13.6 seconds. On average each probit took about 1.1 seconds. The same file when run in the Unix version of Stata on a (likewise idle) Hewlett Packard 735/125, ran in about the same time. In many ways, however, this is picking at invisible nits–in practice, the results of individual procedures available instantly.

Aside from the obvious benefits the program's speed provides, the Unix / PC comparison points to another of Stata's strong points–Stata command files and, more importantly, Stata system files are identical across platforms. There is no need to create transportable files to transfer data from Stata on the PC to Stata on the local Unix machine (or NT or Mac, for that matter). You simply ftp the system file in binary format (or transfer by disk) from one platform to the other and the data are ready to run.

This leads to the discussion of a couple of the programs weak points. Speed exacts a cost in terms of memory. Stata is fast, in part, because it loads the entire data set being analyzed into RAM.[5] This speeds up processing considerably, but places considerable demands on system resources. If the user does not have lots of RAM or an operating system that handles virtual memory well, the size of the data sets that can be analyzed is limited.

A second problem relates to Stata's inability to deal with data not in native system file form or in ASCII. It cannot read SPSS export files or SAS Transport files. This is aggravated by the programmers' tendency to change system file formats between major releases. For example, Stata 3.1 cannot read Stata 4.0 system files. Version 4.0 can, however, read and write earlier file formats, so this particular problem need not be fatal. Translating among different file formats can be accomplished with relative ease through the use of a third party data conversion program. Stata markets one such program, Stat/Transfer ($95 academic price), that does a relatively good job of translating between a number of different data formats.[6]

[5]The default memory space allocated to data by Intercooled Stata 4.0 is 1024k. Users have the option of specifying more, but the ability to do so will depend on the memory available. I generally load Stata in an eleven megabyte memory block. Configured this way I can load and use a variable.

[6]Stat/Transfer is available in either a DOS or Windows version (both of which run well under OS/2) and can translate among the following file formats; Alpha IV, Clipper, Crunch, dBase III-IV, Excel, FoxBase, Gauss, Lotus 1-2-3 (.WK1 and .WK3), Paradox, Quattro Pro, SAS Transport, SPSS Export, Stata, Symphony, and SYSTAT. In my experience Stat/Transfer does an acceptable job on conversions between SAS, SPSS, and Stata. The most recent version is NOT capable of writing to versions of Stata preceding 4.0.

In short, Stata does a number of things quite well. It brings a wealth of statistical tools to the desktop in a compact, reasonably priced package. The documentation is accurate, concise, and clearly written. If you are searching for a general statistical analysis package, Stata is well worth examining.

Stata, the Stata Graphics Editor, and Stat/Transfer are available from: The Stata Corporation, 702 University Drive East, College Station, TX 77840
Prices (academic):

- Small Stata for DOS Windows, and Macintosh - (single user) - $189.00
- Intercooled Stata for DOS Windows, and Macintosh - (single user) - $338.00
- Intercooled Stata for Unix and Windows NT - (two user base license) - $395.00
- Stata Graphics Editor (for all versions) - $ 95.00
- Stat/Transfer - $ 95.00

Moving Data Around: A Review of DBMS/COPY and STAT/TRANSFER

Nathaniel Beck
University of California at San Diego

I often receive datasets that are sent as system files for programs that I do not own (say `SPSS` or `SAS` portable files). I also frequently want to move data between formats so my students can use them. And I often would like to manipulate data using one program but analyze it with a second one. Thus, for example, I hate to manipulate, but love to analyze, data in `S-PLUS`. For the latter purposes I could write out an ASCII file and then read the ASCII file back into the second program, but this involves too much retyping, with too many possibilities for error.

There are two programs which allow you to move data easily between formats. Both work well at what they are supposed to do, and each has its advantages. `STAT/TRANSFER` is small, works with the more common data formats and allows for selection of cases and/or variables; `DBMS/COPY` is large, works with many more formats and allows for extensive manipulation as well as subsetting of the data. `STAT/TRANSFER` is also dramatically cheaper ($95 vs. $295 for academics), a not inconsequential point for readers of *TPM*. At present both programs have `DOS` and `Microsoft Windows` versions. Both have plans for `UNIX` versions.

I compare the `Microsoft Windows` (16 bit) versions here; the `DOS` versions are quite similar. `STAT/TRANSFER` requires about 200Kb of storage; DBMS/COPY requires almost 6Mb! What does DBMS/COPY do that STAT/TRANSFER doesn't? STAT/TRANSFER handles a few popular spreadsheet and database formats as well as SAS transport, SPSS export, GAUSS, S-PLUS (both UNIX and DOS), STATA and SYSTAT files. DBMS/COPY handles all of these, and a variety of less common formats (including several variants of SPSS and SAS formats). Thus, if you have need to read or produce files of BMDP or GLIM or MINITAB or SCA you have no choice, but most of us will probably find the formats converted by STAT/TRANSFER to be adequate.

Both programs transfer well and pleasantly quickly (both fast enough on my Pentium that I didn't even bother to time or compare). Both programs allow the selection of which variables and/or cases to transfer. Both programs handle missing data very well.

One particularly nice feature that is available only in DBMS/COPY is the ability to change data from double precision to single precision or integer formats. Since our data are often small integers, much space is wasted storing these in eight bytes. This can be critical for large files.

Another advantage of DBMS/COPY is that it allows for a wide range of data transformations as well as the selection or deletion of records that meet some given criterion. This might be useful if one had data with odd string variables, or dates, or other nasty things that many standard programs do not handle easily. I had hoped that this feature might allow me to keep my data in STATA system files but use DBMS/COPY to prepare subsets of data for S-PLUS or GAUSS, avoiding the clunky data transformations and subsetting routines of those programs. Unfortunately, DBMS/COPY is no better at handling social science data than is S-PLUS. Thus, while you can take the log of the Gamma function there is no simple way to recode a variable or create a dummy variable. Of course these can be done, but it is no easier to do this is DBMS/COPY than it would be in S-PLUS.

A third advantage of DBMS/COPY is that you can view the database before deciding what to transfer. This feature could prove handy in transforming a poorly documented data set (presumably not created by any reader of *TPM*).

Both vendors promise UNIX versions real soon now (with some versions currently in beta test). A program that can transform data formats across platforms would be a clear winner. Both vendors are also adding continually adding new data formats. It would be very nice if one or the other program handled LIMDEP save sets. There are some readers of *TPM* who might even care about the ability to transform SST save files.

I would think that most readers of *TPM* would want to have access to one of these programs. Many will use them sufficiently infrequently that a copy in the data lab would suffice. I would think that a data lab should pay for the additional flexibility of formats provided by DBMS/COPY. Most individual users will find STAT/TRANSFER more than adequate. In short, STAT/TRANSFER is a good bargain at $95. Those needing more features, and in particular data labs, may find it worthwhile to pay the additional $200 for DBMS/COPY.

STAT/TRANSFER is available from Circle Systems, (206) 682-3783, fax (206) 328 4788 for $95 (academic price). DBMS/COPY is available from Conceptual Software, (713) 721-4200, fax (713) 721-4298 for $295 (no academic pricing). Both are available for either DOS or Microsoft Windows 3.1 with other platforms possibly available soon.

What Software do Political Scientists Use?

Nathaniel Beck
University of California at San Diego

I wanted to know what software was actually being used by political scientists (both in teaching and in research). Being a good social scientist, I used the best social science data gathering methods I know, that is, I sent email to the people whose email addresses I could easily recall. The nonresponse rate was not considerably worse than we commonly see in overnight telephone surveys conducted by candidates from their living rooms. Having said this, the comments I got are of some interest. They are provided in (more or less) unedited form.

- Jonathan Katz, *California Institute of Technology*: I use Gauss, Stata, Splus, and SAS. I use Gauss for MLE, Stata for run-of-the-mill analysis, Splus for graphics, and SAS for data-management. Too many I think, but such is life.

- David Epstein, *Columbia University*: S-Plus and Limdep most recently; we've also used SST and TSP in the past. S-Plus does the non-linear, non-parametric stuff beautifully. Data entry is hard, the graphics are great but take a lot of work, and the program itself is wonderful for its object-oriented approach, but takes a long time to master.

  Limdep, as usual, does qualitative dependent variables like a dream and is rather easy to use. The output is annoying, and the graphics are nonexistent.

  SST and TSP are fine workhorses, but need updating.

  (For teaching) SPSS/PC+, although we'd love to dump it

- Charles Stewart, *Massachusetts Institute of Technology*: For research, I use whatever is available on the machine I happen to be using. The workstations at MIT have SAS and STATA, with S being a new arrival. Because I do a lot of work where I have to merge and otherwise manage large data sets, I do a lot of basic work on SAS. When the data will fit on my PC, I still find SST to be a great resource, since it's simple to use and doesn't take up a lot of system overhead.

I teach the undergraduate methods class here, and for that I continue to be amazed that SAS ends up getting high marks. Particularly once students start working on their own projects, which frequently involve combining disparate data sources, the data handling capacities of SAS are much appreciated by the students. In addition, the fact that one "writes code" in SAS by default (rather than points and clicks) makes students think about what they're doing before they do it, *and* makes it easier to debug problems that they encounter. Finally, it is easier to motivate undergraduates to learn the idiosyncracies of a program which is widely used in academics and the private sector, like SAS, because they feel they are acquiring an employable skill.

- Don Green, *Yale University*: In the interest of maximizing the amount of syntax I need to remember, I use a hodgepodge of of different software packages. GAUSS for maximum likelihood applications such as Poisson regression; SAS for for most analyses of large datasets; SPSS for applications involving LISREL; Mathematica for certain Monte Carlo simulations; RATS for the limited amount of time-series analysis I do. For teaching, I tend to use MINITAB for introductory statistics courses and GAUSS for more advanced courses, as both are supported by microcomputer consultants at Yale. Ideally, each of the various research and teaching applications would be available as a front-end for Mathematica, which seems to be the most versatile program (handling, as it does, both numerical and analytic math), but such user-friendly applications packages have yet to materialize.

- John Williams, *Indiana University*: I use RATS, Stata, and Gauss, both in the classroom and for my research. For classroom use, RATS and Gauss both have problems that can be overcome through communicating with students. Most of the problems with these programs occurs due to mistakes made by our data lab. I use Gauss as a teaching tool, sort of the same way as making students actually work out problems by hand. I also use Gauss in an advanced class where students are asked to program certain procedures. RATS I use in a time series class.

- Janet Box-Steffensmeir, *Ohio State University*: I use RATS for time series, mainly LIMDEP for cross-sectional, duration, etc., and sometimes STATA for these latter analyses – all research. I am very satisfied with these.

  For teaching, I am less satisfied (probably an eternal problem) but used STATA last year and SPSS for windows this year. Who knows what I'll use next year – maybe it will be determined by the results of this poll.

- Simon Jackman, *University of Chicago*: Splus for day-to-day stuff. Is a memory pig but does soooo much cool stuff and I have a 64M HP workstation to myself. SAS for data management and rare analysis tasks. Sometimes Maple for mathematical things. Students use Splus. Used to be a huge fan of gauss... you can read my own personal struggle with this in a rag called TPM...

- Harold Clarke, *University of North Texas*: TPM isn't BYTE, and I'm not Jerry Pournelle, but like Chaos Manor, 131 Wooten Hall has lots of software lying around. So what do I use for research and teaching, you (politely) ask? For research, my decision is driven, not surprisingly, by what I want to do and what kind of data I'm torturing. For time series, Microfit 3.0 is the staple. It's easy to use (menu driven), does what I mostly need, i.e., OLS, IV, NLLS, and recursive estimation, and provides a broad array of diagnostics, including ADF's with MacKinnon's critical values. The graphics are OK, not sensational. As I said in an earlier TPM, for graphics, EVIEWS wins hands down. Given this advantage and its equivalent functionality, I'll probably be doing more and more with EVIEWS in the months ahead. RATS 4.2 gets a workout as well, especially since people like Neal occasionally write useful RATS PROCS. The new 32-bit version of RATS for Windows is VERY FAST (at least by my humble standards), and the new Johansen cointegration module (CATS) enhances its utility. Other programs, including GAUSSX, PCGIVE 8.0, SHAZAM, and TSP, come off the bench to pinch hit occasionally. PCGIVE deserves mention for its graphics, and if it comes out in a Windows version, I'll definitely give it more playing time. Since I do a lot of work with individual-level survey data, programs like LIMDEP and, yes, SPSS also are important. SPSS 6.1 for Windows is a direct port of the mainframe version, so it's much more useful than its clunky DOS cousin. When I do confirmatory factor and covariance structure analyses, LISREL 8.0 is still the word, but EQS and AMOS (both for Windows) have their strong points.

  As for teaching, I'm constrained (alas) by what the university will buy for our labs. We have the usual suspects – BMDP (students find the 2T module very intuitive for doing transfer functions), SAS and SPSS, as well as Micro-Crunch, and RATS 4.2 (DOS). Right now, my new colleague, David Leblang, and I are trying to convince the powers that be to site licence STATA for Windows. No joy yet, but we will try again later this summer. I also use Bill Greene's ET a lot. This is basically a mini version of LIMDEP, and students can buy the progam and a very useful manual for less than $40. At Essex, I get to use Microfit, and the students love it. Programs I'm currently considering are Gary King's COUNT, and Steve Hall's REG-X. Although I'm sure all TPM readers know about the former, the latter may be less familiar. It is a time series-oriented package, is menu driven, and includes modules to do simple and more elaborate Kalman filter stuff, as well as recursive estimation and Johansen analyses. It also provides all the standard Hendry-style diagnostics and has decent (again not outstanding) graphics. It's not terribly expensive, so maybe I'll be able to get a site licence for it. Definitely worth a look-see.

- Stephen Ansolabehere, *Massachusetts Institute of Technology*: I use STATA for instruction and some research. STATA offers respectable graphics, excellent implementations of standard econometric routines, good on-line help, and is available for DOS, Windows, and Unix. The main drawback is memory, which is very limited in the student version and in the professional version has to be managed more than in other programs. For more advanced problems, GAUSS or MATLAB.

  When I get the time, I will learn S-Plus and LISP-STAT.

- Phillip Schrodt, *University of Kansas*: For instructional purposes, we use SAS in one of our required methods courses and SPSS in the other, so students become acquainted with both.=20 Quantitative international studies students also learn elementary programming using Pascal (Think Pascal on the Macintosh or Turbo Pascal in Windows), and some students are using Mathematica, primarily for numerical simulation.=20

  My most common research software is Think Pascal, though I'm migrating to MetroWorks C++. For basic statistics I use a wonderful–albeit no longer marketed–Macintosh program called Exstatix; for more complicated routines I'll use Systat or occasionally SPSS. For intensive analyses–for example running thousands of cross-correlations on behaviors in an international subsystem– I usually just write my own routines. A compiled specialized program can be one or two orders of magnitude faster than using commands in a stats package, and by using previously-written code for basic operations such as eigenvalues and matrix inversion, requires little effort to write.

- Guy Whitten, *Texas A&M*: I use SAS on both unix and pc platforms for teaching and research. I also use Limdep on the pc for my own research. I am interested in starting to use Gauss sometime when I have a month to learn it (yeah right).

- Charles Franklin, *University of Wisconsin, Madison*: For research, I use several different packages. I do all data manipulation in SAS for windows (run in batch mode!). SAS has never let me down when I needed an exotic merge or had to deal with large datasets that other packages have had trouble with. I do very little analysis in SAS, however. I use STAT/Transfer to convert from SAS to the file format of my other packages (though DBMS/Copy has a new version that looks awfully good– extremely comprehensive set of file formats it can convert among, including spreadsheets and database programs as well as stat packages. I think I'll switch.)

  For most routine analysis I use Stata which is fast, has good graphics and has a wide array of limited dependent variable models. When I need a more exotic model I switch to Limdep. And if that doesn't do it, I use Gauss to write a program to estimate the model I want. I've started using S+ for it's graphics which go far beyond Stata's capabilities and for S+'s ability to estimate generalized linear and additive models plus a variety of modern regression techniques. It is a pain to switch among all these packages, but a good file conversion program makes it relatively easy to do. For teaching I use Stata, which is easy to use and is available at very attractive pricing for labs and for group purchase by students.

- James Alt, *Harvard University*: I have always used SST for teaching but will now go over to STATA to take advantage of Windows availability, better documentation, etc. I may reconsider this if thoroughly documented upgrades of SST become available. For research I have always used whatever was needed, SST as the workhorse, occasionally LIMDEP, GAUSS as necessary. I used to do far more in SAS but found it impossible to adapt smoothly to a PC version and so eventually dropped it. If what's now available had been available some years ago I would probably never have switched. SAS was a classic high-fixed-cost package, but once you figured out how to do something (and I eventually figured how to everything I wanted, even reading multiply-punched card files using column binary formats) it was a terrific operating environment.

- Gary King, *Harvard University*: For research, I use Gauss (Aptech Systems, [sales@aptech.com]) for all data analysis and programming, and for most graphics. I also use SPLUS (StatSci, [support@statsci.com]) for more sophisticated graphics. Both programs have high startup costs but relatively low recurring costs. For advanced classes, I teach students to use the same software tools I use for my research, although they are always welcome to use other software if they prefer.

- Michael Alvarez, *California Institute of Technology*: Okay, I'll admit it. I still use SST. And it's not because I'm not yet tenured, and one of my senior colleagues is one of the co-producers of SST. It's because I find that it is about the best general purpose statistics program around — and to me it has always been free!

  When I was in graduate school at Duke, we heard one word when it came to statistical programming — SAS. SAS is indeed a good statistics package, since it can do just about everything under the sun. However, SAS was not written for social scientists, and it has a number of oddities which I find puzzling, irritating, and often counter-productive. These include the need to separate data manipulation and analysis into separate steps ("the data step" and the "proc step"), the often bizarre and strange syntax, and odd aspects of procedures (ever try doing a simple binary probit — they call it normit, and unless you specify a certain subop you get coefficients with the wrong sign). SAS is a HUGE program, since the last time I installed SAS for Windows (6.10) on my PC it took up at least 25 megs of disk space. Last, for those of us who are not at institutions with site licenses for SAS, it verges on being prohibitively expensive.

  I learned about SST while in grad school (actually at my first methodology meeting), and I tried it out when I returned home. And I was hooked. Here was a fast program with simple syntax. I could run it on all sorts of platforms, and it could even work on these old IBM-XT machines we had sitting around with tiny 30 meg hard drives. I could do all of the types of analyses I needed — simple regression, two-stage least squares, binary logit and probit, and multinomial logit. The syntax for data manipulation was simple, and to me, it made a lot of sense. So I've been using it every sense.

  That said, what do I do in the course of everyday research? I do virtually all of my data access and manipulation in SST, some I do in SAS. This depends upon what I'm doing. Recently the NES has provided SAS data descriptions with their datasets, and I've found those quite useful when I need to read an entire NES study. But if I need to do many recodes, or some simple analyses, I find that I switch to SST immediately.

  For complicated analysis — maximum likelihood and graphics — I turn to GAUSS and S-PLUS. Obviously GAUSS, combined with MAXLIK and the new CML package, is the leader in maximum likelihood analysis. I have tried to do simple user-defined MLE in SST, SHAZAM, STATA and SAS, and I can say without much effort that GAUSS beats them all hands down. But GAUSS has one of the most obtuse programming languages I use, and it cannot be considered a good package for data manipulation or simple analysis.

  SPLUS is probably the best graphical analysis package I've seen. While the object-oriented approach takes a lot of getting used to, SPLUS can produce some of the best graphics of any of the programs I've used. They can also be easily output to the screen, to a printer, or in postscript format, which I like. But like GAUSS, SPLUS is a difficult language to master, and I have never really been able to do a good job with data manipulation in SPLUS.

  The last consideration is documentation. These packages all vary considerably in how adequately they are documented, and how easy it is to find answers to your questions in the documentation. I think the GAUSS manuals are almost a waste of time; the reference volumes are useful, but I personally like documentation with lots of examples (which the new MAXLIK and CML documentation has). SPLUS is much better, with excellent examples and almost step-by-step cookbook directions on many procedures. However, the SPLUS reference is difficult to use. SAS takes up about three feet of my shelf space, and I think that speaks for itself. SST, on the other hand, has a minimal set of documentation, but it is very useful. And Jeff Dubin has taken a great step in a new direction by making the SST documentation available on the World Wide Web (you can get there from the Political Methodology Home Page).

- Nathaniel Beck, *University of California, San Diego*: For time series and panel data research I use RATS. If I were just starting out I would think seriously about EVIEWS, but I have a lot invested in RATS. For Monte Carlo work only GAUSS seems to do. For graphical work and semi-parametric analysis I use S-PLUS, though I hate it.

  For teaching we are switching over to STATA. I use LIMDEP for the workshops I give on limited dependent variables and duration models. While STATA does almost all the usual things, nothing beats LIMDEP for what LIMDEP is good at. For my freshmen I use SYSTAT; they like to point and click.

- Jonathan Nagler, *University of California, Riverside*: I use SST and Gauss for my research. I have done research in Shazam before. I teach with SAS and Shazam; though I will probably replace Shazam with SST for teaching.

  SST is a great program for both research and teaching. The syntax makes sense, the output makes sense, there is alot of power both in the programming language and the statistical routines available, and it runs on everything from ancient X86 laptops to Crays (I have done both).

It is very easy to teach graduate students to use SST. I think people don't appreciate the virtue of SST essentially being a 'terminal-mode' program. But this is why it is the same whether you run on DOS or Unix. And SST is as 'windowing' as I would like since I just have a few different emacs windows open when I run SST; one with my command-file and one with my output-file. Also, SST is capable of producing perfectly decent postscript and dvi graphics output. It is certainly not the strength of SST; but the feature is there.

Finally, yes - I use SST because I *know* SST. This is not a minor point. I started using SST on a PC. I have since used it on a VAX, an AT running Xenix, an Amdahl running Unix, a Cray running Unix, a NeXT running NeXTStep, a PC running OS/2, and an HP workstation running HP-UX - and I can't wait to try it on a PC running Linux. And in switching across all those machines I have never had to alter a single file to get it to run on another platform! The syntax of the command files does not change; and sst binary datesets ('save sets') are identical across platforms. This point may be lost on DOS or Windows users; or just anyone who has not lived long enough to have switched computers a few times. But working in the Unix world one learns to appreciate continuity and cross-platform compatability.

The documentation used to be pretty poor; but now that it is available as hyper-text I would say SST has the best documentation of any package available. Also, some of the nicest and most helpful people belong to the community of SST users.

I use Gauss because of the Maxlik package. But I hate Gauss. The documentation is criminal. When I started using Gauss it took forever to find out 'where my data was' within the code. I still don't know half the time.

I used SAS to teach for the first time this past quarter. I did this because we felt some desire to make sure graduate students learned a 'standard' package, and one that would handle just about anything they wanted to do. It was fine to teach the basic regression course with. But I can't imagine using it beyond that. It is big, cumbersome, has a very bad feel when trying to write code, and produces ugly output.

The next package I learn (someday) will be SPlus. Everyone says its so cool . . .

- Douglas Rivers, *Stanford University*: SPLUS gives the best feel for what a graph should look like. I use SST, C and Loessplot (the last bought at a bookstore for $29.99, my idea for the right price for software).

Review of William Greene's *Econometric Analysis* (Second Edition. New York: Macmillan, 791pp.)

Timothy Amato
University of Iowa

William Greene's "Econometric Analysis" (EA) has quickly become a standard textbook in first year graduate econometrics sequences in economics programs. Rather than summarize and critique the contents of the book (these types of reviews are available in economics journals like "Econometric Theory" and elsewhere) this brief review will discuss the role that EA can play in graduate methods sequences in political science. Most of my observations stem from having used the book as a required text in the third and final course in the methods sequence here at the University of Iowa.

The strengths of EA are the breadth of coverage, the clarity of the exposition and the use of numerical examples to supplement and illustrate the statistical and econometric theory it presents. Chapters 5-9 and 13-17 cover, at a reasonably advanced level, the topics that would be covered in a typical intermediate regression class-bivariate and multiple regression, simple nonlinearity and model misspecification, and generalized least squares (GLS). The remaining chapters cover what one would typically put into an advanced course and includes general asymptotic theory and maximum likelihood, intrinsically nonlinear models, discrete and limited dependent variables, and simultaneous equations models. A student who has been through EA in its entirety, understands the presentation, and can solve many of the end of chapter problems will be well prepared to teach methodology at the graduate level and to do careful empirical work in their own substantive research. The relevant question for political methodology is whether this book can be used as an effective teaching tool given the length of our typical methods sequence and the mathematical preparation of an average graduate class.

As I alluded to above, a good deal of the material in EA is not really new to students. This is because multiple regression, GLS, and model misspecification, which forms the core of chapters 5-9 and 13-17, is covered in intermediate regression courses taught from texts like Kmenta, Maddala or Pindyck and Rubinfeld. What will be new to students is the level of presentation, which makes liberal use of matrices and vectors, and the breadth of coverage, which takes the reader through many recently developed techniques such as generalized method of moments (GMM) estimation, robust estimation of standard errors in the presence of heteroskedasticity or autocorrelation of unknown form, and general diagnostic tests based upon the Lagrange

Multiplier (LM) principle. Since most of these methods are now available on "canned" statistical packages it is quite important that these procedures be given a textbook and classroom presentation so that students can gain some insight into when these techniques will be appropriate in their own research and how to interpret their results.

My own experience with using EA in an advanced methods class is somewhat mixed. Many good students don't like the level of presentation and/or the sheer volume of material that is covered. A common reaction here was a clear preference for an alternative text like Johnston which, while presented at roughly the same technical level, covers much less material and is not up to date on many topics. As an instructor I am not sure why one would prefer Johnston to Greene as I think that in all regards EA is the superior book for learning about econometric theory. However, as a student, I might have a somewhat different opinion. EA covers a great deal of material and in this sense is more like a "handbook" rather than a textbook of econometrics. It may be that because of the breadth of coverage one is forced to try and keep track of too much material, not all of which is relevant to a particular topic or lecture. In this sense, one of EA's strengths can also be a liability if it is not used properly.

My response to the criticism that EA is too technical or mathematical is that methodology is a technical subject and there's not much one can do here short of switching fields. My response on the breadth of coverage issue has been to give extremely detailed syllabi with specific page readings for specific topics (and also to give references from other texts as well). This means that students don't need to try and digest entire chapters on, say, heteroskedasticity when I am lecturing on some specific problem dealt with in one part of that chapter.

As a final observation on whether this text should be adopted for courses in our graduate methods sequences it is important that students have prior exposure to a matrix presentation of the general linear model in their intermediate classes. If they haven't, one will spend an excessive amount of time teaching the rudiments of matrix algebra at the expense of econometric theory. Without some prior exposure to matrices students will probably have a difficult time appreciating much of what Greene has to say and a different text such as Kmenta might be a more appropriate choice.

Review of Achen and Shively's *Cross-Level Inference* (University of Chicago Press, 1995)

Bradley Palmquist
Harvard University

Christopher Achen and Phillips Shively's newly published "Cross-Level Inference" is not just the most recent contribution by political scientists to the discussion of a long-standing problem, it is the best comprehensive work on the subject in the social sciences. The book establishes new standards for any researcher who does aggregate data analyses to understand individual-level processes. No practitioner of ecological regression can justifiably neglect what Achen and Shively teach us about the importance of microlevel foundations, the sources of aggregation bias, and the range of estimation strategies beyond simple Goodman regression.

This is not to say that "Cross-Level Inference" is satisfyingly complete. On this I believe the authors would concur. Some of the *best* discussion, in fact, is an honest admission of what is still not understood. One example is the whole tangled web of how to weight ecological regressions. Is it better to use the population size of the aggregate unit, some estimate of internal variability, or no weights at all? After convincingly arguing that the efficiency issues that might lead one to GLS are minor in many practical applications, the answer to the question of whether one should estimate the desired parameters without population weighting remains unclear.

Much of the book is based on a series of articles that the two authors have produced over the years. But do not assume, therefore, that there is nothing new. The book goes well beyond these earlier pieces. One way to provide a road map to the variety of topics addressed is to divide the chapters into diagnosis and remedy. Chapters 1 through 4 and 9 are mostly diagnosis. Compared to my recollection of the articles that preceded them, much has been added, rethought, and rewritten. Chapters 2 through 4, especially, provide the most complete and penetrating understanding of ecological regression as a method to estimate voter transition rates yet presented. This may be the best place anywhere in the literature to find out what the heck so often goes wrong with ecological regression. The assembly of new insights and sound judgement in these sections is impressive. This is hard-won territory that has been gained inch by inch.

I also think Achen and Shively have opened up some of the sampling and probability space issues more clearly than anyone previously. They strongly recommend that the problems be framed by a superpopulation model. Whether or not this becomes the consensus approach, the discussion provokes much more careful attention to the statistical status of in-the-sample values and expectations of both individual and constituency parameters and disturbances.

After showing what can go wrong, chapters 5 through 8 take on the challenge of doing better. Chapters 5 and 6 elaborate and illustrate flexible extensions to basic Goodman regression. Are you doing ecological regressions and want to move beyond the 1950s? This might be the first section to check out after the diagnosis chapters. Chapter 7 is a clear presentation of the Thomsen-Achen factor analytic approach to cross-level inference, and as such is somewhat separate from the ecological regression material. But for problems like estimating voter transition rates it may be preferable. Similarly, Chapter 8 provides alternatives to ecological regression. It is a compilation of various tabular approaches such as the Davis-Duncan method of absolute bounds which have been previously extended by Shively in a number of papers. I am impressed by how much can be done with feet firmly planted on the ground from this perspective. How lucky you can get depends on the particular problem, but these methods should not be ignored. Achen and Shively do make linkages between these alternative techniques and ecological regression which are informative as well.

I find one aspect of Achen's and Shively's approach a bit limiting. They have chosen to circumscribe their characterization of Ecological Regression as a technique suitable to those situations where the microlevel relations can be represented in a contingency table. It has always seemed to me that the cross-level inference issues were best addressed from the more general perspective of regression models of any form, whether there was a contingency table representation or not. Indeed, over time I have come to regard Goodman's contribution in the 1950s as somewhat less signficant than that for which he is usually given credit. Perhaps a bit uncharitably one could say that he merely pointed out that it is possible to estimate the proportions of a contingency table by running a regression with dummy variables. In the context of a wholly micro-level analysis this is not news. It just happens that he was discussing a situation where the model at one level (micro) could also be represented as a contingency table while at the other (macro) only the regression form is available. To be sure, he began a fruitful discussion of what conditions might be necessary or sufficient to allow valid cross-level inference. But filling in the cells of a crosstab does not capture the problems of aggregate regression more generally.

Several points, however, can be made in defense of Achen's and Shively's focus. One of the consistent themes of both the diagnosis and remedy sections is that there is no single solution to the difficulties of doing ecological regression. Rather, close acquaintance with the peculiarities of the data at hand, and external substantive and theoretical information and assumptions are required to make reasonable cross-level inferences. Hence they find it useful to emphasize one typical application, the estimation of voter transition rates, about which there is both accumulated empirical experience and a greater measure of social scientific theory than is often available. All of the "how to" chapters (5 to 8) are nice illustrations of the interplay of technique and understanding that Achen and Shively recommend. It is also true that in Chapter 4 some linkage is made between the special features of this specific case and more general characterizations of cross-level inference.

I am not sure who will be most likely to pick up this book. For serious political methodologists it is clearly a must read. Practitioners of aggregate analysis without a strong probability and statistics background may well find it hard going, however. Either that, or they will have to have the confidence to skip from one practical pointer to another. This latter approach would at a minimum provide one with much more dependable guidance than relying on the old, sometimes unfounded, advice that was handed out in *Journal of Interdisciplinary History* or *Historical Methods*. We can only hope (vainly, no doubt) that naive postulations of macro-level regression models which simply mirror their micro-level analogues will no longer be taken seriously. There is no point in explaining the behavior of counties and other "accidental geographic aggregations" with no theoretical status as actors. Another service to the community is provided in the discussion that exposes naive aggregate analyses of contextual effects as nearly worthless. In fact the discussion on contextual analysis in the final chapter could be fruitfully read by someone with no interest in the other issues. (Contextualists take heart. There are alternatives.)

Not all the answers are in. There are complementary (competing?) approaches to aggregate data analyses. But any discussion and even disagreements can now begin from a much more informed jumping off point. In any case, "Cross-Level Inference" is for now the most comprehensive and innovative single reference on the subject. Don't leave the more familiar home of individual-level analysis without it.

Review of New Time-Series Texts: David Hendry, *Dynamic Econometrics* (Oxford University Press, 1995. $40 (paper). 869pp.; James Hamilton, *Time Series Analysis* (Princeton University Press, 1994. $55 (cloth). 799pp.

Nathaniel Beck
University of California, San Diego

Time series would have felt blessed by the publication of either of these books; both showing up in the space of a year is as almost as good as celebrating Hannukah and Christmas. And like those two celebrations, each of these books brings a different set of gifts. Since the two books are complementary, I discuss them separately.

Dynamic Econometrics (DE) is THE long awaited text detailing the London School of Economics approach to time series analysis. This approach had its roots in the work of Denis Sargan; David Hendry is its leading current exponent. LSE econometrics is particularly concerned with modelling (as opposed to estimation) issues.

In Hendry's view there are four levels of knowledge or investigation. The first is concerned with probability theory or the properties of data generation processes (DGP). Once the DGP is known, the investigator is then concerned with estimation issues. Most econometrics texts are devoted to issues in the theory of estimation. Unfortunately theory does not usually (ever?) provide clear guidance about the DGP so we are forced to deal with issues of modelling. It is this concern that marks LSE econometrics. Finally we are often concerned with issues of forecasting. *Dynamic Econometrics* deals with the first of these three issues, with forecasting considered in a promised companion volume (Clement and Hendry, *Economic Forecasting*, Cambridge University Press).

The joy of DE is that it presents a unified treatment of time series econometrics from the LSE perspective. Until now I used Harvey's *The Econometric Analysis of Time Series* (MIT Press, 1990) as the standard reference for modern econometric time series analysis. While Harvey's book is admirable, it suffers from being the second edition of a book originally written in 1980. Time series analysis in 1995 is very different from time series analysis in 1980. Thus Hendry's book is extremely welcome.

Like LSE econometrics, DE takes probability and statistics seriously. The text contains an excellent discussion of the mathematical prerequisites. While this discussion is not easy, it does not assume that the reader already has either a Ph.D. in mathematics or a complete knowledge of the subject coming in. DE contains a clear set of definitions and discussions of most of the mathematical arcana so necessary for modern time series analysis. But while the mathematics is clearly and correctly laid out, Hendry happily does not make a fetish of being so mathematically precise that only the chosen few have any hope of reading the book. The mathematical setup of dynamic models occupies about half the text.

The second main section is a one hundred page treatment of estimation issues. While there is some review of general estimation issues, most of the section is devoted to dynamics, and, fittingly, error correction approaches. This section contains the least amount of new new material in the book. This is appropriate, since the estimation strategy of the LSE school is straightforward maximum likelihood.

The third main section is concerned with modelling issues. This section contains an excellent discussion of encompassing (model choice) and an interesting, but less compelling, discussion on "real world" modelling.

The book closes with a long set of appendices (about a quarter of the book) on the various mathematical and statistical prerequisites. While these appendices are more useful than the typical twenty page appendix, those unfamiliar with the material covered will find them tough going. The appendices do serve the admirable purpose of providing a complete set of definitions of important statistical concepts. As one example, the appendices provide a complete and accurate statement of the various types of convergence of sequences of random variables. But these appendices should not be confused with a good text on the mathematical and statistical foundations of econometrics.

DE would make a fine text for a time series course following a standard introductory econometrics sequence. Instructors providing such a course would have to give up on some cherished, but old fashioned topics (e.g. Box-Jenkins as a special approach). My guess is that very few Ph.D. programs in political science will have the room (or the appropriate number of students) for such a course. (At UCSD we surely do not.)

Thus I think the main audience for DE will be appropriately trained graduate students and junior faculty who will read the text on their own. The joy of DE is that it is so well written, and has so many examples, that the well motivated and well trained graduate student will find Hendry's approach of the greatest value. All the examples in the text use Hendry's excellent `PC-GIVE` program. Students using the text on their own should use `PC-GIVE` to rework the examples and try the various techniques. (While `PC-GIVE` would be best for this purpose, other programs, such as `EVIEWS` could be substituted.) While more senior faculty would also benefit from DE, it is less likely that they will have the time required to profit from the unified and mathematically correct treatment of dynamic models.

Many will find the strength of DE to be its major weakness. Given its commitment to the LSE approach, it either gives short shrift or no shrift to other topics. It contains no discussion of Kalman filtering or spectral analysis and little discussion of such topics as ARCH or switching regimes. Fortunately, readers interested in a more encylopediac approach can turn to James Hamilton's *Time Series Analysis* (TSA). Unlike DE, the motivated graduate student would find it difficult to use TSA as a general text on dynamic models. All of us, however, will benefit from it as a reference tool.

TSA has 22 excellent chapters, each devoted to a single topic, starting with stationary ARMA models and concluding with switching regime models. In between there are good chapters on statistical foundations, and excellent chapters on spectral analysis, vector autoregressions, Kalman filtering, non-stationary time series (trends, unit roots and cointegration) and ARCH. Those doing research involving any of these approaches or techniques will thank Hamilton for a clear and concise treatment (and set of references) of the various topics. The conciseness of the chapters will make them tough going for the uninitiated. Those wishing to know the state of the art will find no substitute for TSA. The hypothetical time constrained senior faculty member will be especially thankful for Hamilton's book. (As is befitting for a discussion of econometrics books, I am making the fanciful assumption that our time constrained senior faculty member has the mathematical skills necessary to fully appreciate Hamilton's concise treatment.)

In brief, anyone serious about modern time series analysis will want to own both books. *Dynamic Economics* is a joy. There is no question in my mind that those newly coming to time series analysis will be well served by Hendry's text; it has no competitors. Those already doing time series analysis will find Hamilton's complete treatment of the state of the art to be indispensable. Neither book is for the mathematically challenged.

Political Analysis, Volume 5

John Freeman
University of Minnesota

Contents of *Political Analysis*, Volume 5 (University of Michigan Press, John Freeman, editor):

- "Knowledge, Strategy, and Momentum in Presidential Primaries". Henry Brady.
- "Attitudes, No Opinions, and Guesses", John Jackson.
- "Testing the Effects of Paired Issue Statements on The Seven-Point Issue Scales", William Jacoby.
- "Dynamic Analysis with Latent Constructs", Paul Kellstedt, Gregory McAvoy, and James Stimson.
- "Issues and the Dynamics of Party Identification: A Methodological Critique", Eric Shickler and Don Green.
- "The Contamination of Responses to Survey Items: Economic Perceptions and Political Judgments", Nathaniel Wilcox and Christopher Wlezien.
- "A Correction for an Underdispersed Event Count Probability Distribution", Rainer Winkelman, Curtis S. Signorino, and Gary King.

The Political Methodology Electronic Paper Archive

Jonathan Nagler
University of California at Riverside

Introduction

The previous issue of *TPM* contained an annoucement of an electronic paper archive at UC Riverside. That archive has been operating succesfully for several months now, and includes papers from both the Political Methodology Summer Conference, APSA Meeting, and working papers. These are all available to be downloaded. Submissions of papers that might hold any interest for people reading this newsletter are strongly encouraged. There is no minimal amount of 'methodological content' required: the archive is not moderated in any way.

Because there are many benefits for authors submitting their papers, the archive is designed to put the 'burden' of doing things correctly on the author rather than the reader. Printed below are the full instructions for submitting a paper. They may look long and tedious; but in fact once you read the instructions you can probably produce the necessary files to submit a paper in less than 10 minutes. And that is a one-time cost: after that your paper can be retrieved by countless readers without any effort on your part. Several things happen when you submit a paper. First, an abstract of the paper is emailed to everyone on the political methodology list-server (right now this is approximately 375 people). Second, the abstract is placed on the political methodology web-site as an html document. Third (provided a postscript version of the paper is provided by the author), a version of the paper is made available for viewing over the world-wide web. Fourth, the paper is archived and can be retrieved by anyone via the world-wide web or anonymous ftp and then printed. All of this happens automatically, with no human intervention.

There is a simple reason authors should perceive it to be in their interest to submit papers: it gives one's work the most possible exposure. Most of us have seen statistics on how infrequently much social science scholarship is ever sited. This is a good way to make your work less likely to suffer that fate. Also, this is a good way to make your work easily available to people rather than having to reproduce and mail copies.

Instructions for submitting papers are contained below. References used here are assuming you are submitting a paper for the 1995 political methodology conference. For other conferences, the term 'summer95' would be replaced by the appropriate conference (i.e., 'midwest95', 'summer94', 'apsa95', etc) name. If you are submitting a working paper, you would substitute $`working_papers95'$ for 'summer95.' [You should modify that last instruction appropriately as the calendar warrants.]

Summary of Submission Instructions

1. Produce a postscript version of your paper and a version in a format that will print on any HP laserjet compatible printer. This can be done with most any word processor.
2. Create a plain text (ascii) version of your abstract with some appropriate information (author, title, keywords) on top.
3. Produce versions of your paper in any format that you think readers might find useful (MS-Word, Wordperfect, postscript, HTML, etc).
4. Follow the naming conventions for your files that are described below.
5. Place all of these files into a single zip file (software to do this is available via ftp from the Political Methodology Site).
6. Use anonymous ftp to transmit the single zip file to the political methodology site at UC Riverside.
7. Send email to polmeth-owner@wizard.ucr.edu with notification that the zip file has been submitted.

File Formats

First, paper authors are asked to produce postscript and/or HP printing format (PCL5) versions of their files and any extra materials (figures, etc). This is the most important part of the submission process. An ever-growing number of people have postscript printers, and with the appropriate software in place *the postscript file can be viewed on-screen over the web.* HP files are not as versatile, but almost everyone has the ability to print a file which has been converted into an HP printing format. Most word-processors offer you the option to 'print' your file to disk; doing so creates a file in the format of whatever printer you choose. You would choose a postscript printer to produce a postscript file, and you would choose a standard HP printer (such as an HP 4L) to produce an HP file. (Instructions for Producing Postscript and HP files using Microsoft Word for Windows are included at the end of these instructions.)

If you cannot create a postscript file with your word-processor it is strongly suggested that you send two versions of your paper: one in the native format of your word-processor, and the other as plain text (ascii).

Note: lots of people have 300dpi printers. You should try and produce a file that can be printed on one of these printers. Any 600dpi printer will handle a 300dpi file; the converse is not true. If your default printer is set to a 600dpi printer (such as an HP 4P), you might want to add a printer option for an HP 4L or other 300dpi printer. This is an issue with HP files; it is not an issue with Postscript files.

The Abstract

Second, include a plain-text (ascii) version of the abstract. The abstract can take any form you would like. However, the first 3 lines of the abstract file must be the following (the double-quotes are required):

```
author = "Last-name-of-first-author,
  first-name-of-first-author,
all other author's names"
title = "Title-of-your-paper"
keywords = "any,keywords,you,choose,
    separated,by,commas"
```

Example:

```
author = "Alvarez, R. Michael and
  Jonathan Nagler"
title = "Correlated Disturbances in
 Discrete Choice Models:
 A Comparison of  Multinomial
 Probit Models and Logit Models"
keywords = "econometrics, logit,
    multinomial probit, gev,
discrete-choice, monte-carlo"
```

Multiple Formats

Third, the above files are the 'bare-minimum' requested from you. **YOU ARE INVITED TO SUBMIT YOUR PAPER IN AS MANY FORMATS AS YOU LIKE** (provided you send them all in a single .zip file). Other formats that some readers might find convenient: Word-Perfect, MS-Word, postscript, HTML, dvi.

Naming Conventions

Fourth, please use the following conventions for file names for papers and extra material. All names should follow the 8.3 format: being no more than 8 characters, followed by a period (.), followed by an extension of no more than 3 characters. File names should be *lower-case*, (except the README file). Notice that you will need to change the naming conventions appropriately as the calendar year changes.

First Part of File-Name:

- For the file containing the body of the paper, please use: the first 5 letters of the first author's last name, followed by '95'. If the author has another paper circulating already using this name; then use the first 5 letters of the first author's last name, followed by '95b'.
- For the file containing the ascii abstract of the paper, please use: the first 5 letters of the first author's last name, followed by '95'. If the author has another paper circulating already using this name; then use the first 5 letters of the first author's last name, followed by '95b'.
- For figures and tables that you may need to put in separate files, please use the first 5 letters of the authors last name, followed by 'tnn' or 'fnn' for table number nn and figure number nn.
- For the zip file containing all files, please use the first 5 letters of the author's last name, followed by '95'. Again, if the author has another paper on the archive already using this name, then use '95b'.

Extensions:

- For all postscript files, please use the extension .ps.
- For all HP printing files please use the extension .hp.
- For the abstract, please use the extension .abs.
- For all plain text files (ascii) – with the exception of the abstract – please use the extension .txt.
- For all WordPerfect files, please use the extension .wp.
- For all Microsoft-Word files, please use the extension .msw.
- For all dvi files, please use the extension .dvi.

Creating the Zip File

Fifth, please place the following into a single .zip file:

1. Your paper (in as many formats as you choose).
2. An ascii version of the abstract.
3. Any figures and tables (again, as many formats as you choose).
4. A README file describing the different files you include; this is a good place to include your email address.

You can do this with Info-zip, PKZip, unix zip, or a program compatible with these three. If you do not have any of the three recommended programs, you can probably get them from your local computer support person; or you can download them via anonymous ftp from /pub/utils at wizard.ucr.edu. Please do not use Gnuzip, Arj, LHA, or other archiving systems.

A good way to create your .zip file is to create a subdirectory on your machine and place all the files that will go into the zip file in that subdirectory (say /mypaper). Then the following syntax will work with pkzip:

```
C:/mypaper> pkzip alvar95.zip *.*
```

NOTE: It is very important that you include an ascii version of your abstract with the correct filename (ending in .abs) and containing the author, title, and keyword lines in the .zip file!!

Example - Here is what might be contained in a file alvar95.zip:

- alvar95.abs : abstract of Alvarez's paper.
- alvar95.ps : Text of Alvarez's paper, post-script format.
- alvarf1.ps : Figure 1 of Alvarez's paper, .ps format.
- alvarf2.ps : Figure 2 of Alvarez's paper, .ps format.
- alvar95.hp : Text of Alvarez's paper, .HP format.
- alvarf1.hp : Figure 1 of Alvarez's paper, .HP format.
- alvarf2.hp : Figure 2 of Alvarez's paper, .HP format.

- alvar95.txt : Text of Alvarez's paper, plain-text (ascii) format.
- README : File explaining what is in each of the above files.

Anonymous FTP

Sixth, use anonymous ftp to connect to wizard.ucr.edu, and place the .zip file in:

```
/pub/polmeth/summer95/incoming
```

Be sure to use binary transmission!! See ftp instructions below.

The ftp sequence looks like this on a unix machine; it could look slightly different on your machine.

```
% ftp wizard.ucr.edu
ftp> login: anonymous
ftp> password: yourusername
ftp> binary
ftp> cd pub/polmeth/summer95/incoming
ftp> put filename.zip
```

Notify Polmeth-Owner

Seventh, send an e-mail message to polmeth-owner@wizard.ucr.edu announcing that you have placed the .zip file on wizard, and include the README file in your email message. Your .zip file will be placed in the appropriate directory, your abstract will be mailed to the polmeth list-server, and your abstract will be made available via the World-Wide-Web as an HTML document. [Please do not mail your abstract to the polmeth list-server before your .zip file is moved by the polmeth-owner.]

Retrieving Papers

If you subscribe to the political methodology list-server (polmeth@wizard.ucr.edu) you should receive abstracts of papers when they are in placed on the archive. Should you wish to retrieve the paper for an abstract you have received you have two options. First, you can point your web-browser to the Political Methodology Home Page at: http://wizard.ucr.edu/polmeth/polmeth.html. This will lead you to the .zip file containing the paper you want. This method is **far** superior to using anonymous ftp. If you do not have Netscape, Mosaic, or an alternative Web-Browser you should get one ASAP. [Note: Netscape is free to academic users.]

Retrieving Papers using Anonymous ftp

The ftp sequence looks like this on a unix machine; it could look slightly different on your machine.

```
% ftp wizard.ucr.edu
ftp> login: anonymous
ftp> password: yourusername
ftp> binary
ftp> cd pub/polmeth/summer95
ftp> get filename.zip
```

Once you get the .zip file, you need to extract its contents. You can unzip it on a unix machine with unzip installed as follows:

```
% unzip filename.zip
```

On a PC using PKunzip this would look like:

```
C:>pkunzip filename.zip
```

If you do not have an unzip program, you can probably get one from your local computer support person; or you can download one via anonymous ftp from /pub/utils at wizard.ucr.edu.

Printing a Paper

The typical command for printing papers in HP format on a DOS machine is:

```
c:> COPY filename.hp /b lpt1:
```

To print postscript files directly you must have a postscript printer:

```
C:> COPY filename.ps /lpt1:
```

On a Novell network, check with your system-person, or try the following:

```
c:>NPRINT filename.hp /notabs
```

The "/b" (or "notabs") is important so that that DOS (or Novell) knows this is a binary file.

The Political Methodologist
Division of Humanities and Social Sciences
Mail Code 228-77
California Institute of Technology
Pasadena, CA 91125

The Political Methodologist is the newsletter of the Political Methodology Section of the American Political Science Association. We gratefully acknowledge the support of the California Institute of Technology in helping to defray the editorial and production costs of the newsletter and the assistance of Abby Delman and Paula Green.

Subscriptions to *TPM* are free to members of the APSA's Methodology Section. Please contact APSA to join the section. Dues are $8.00 per year.

Submissions to *TPM* are welcome. Articles should be sent to the new editor by e-mail (nagler@yoda.ucr.edu) if possible. Alternatively, submissions can be made on diskette as plain ascii files sent to Jonathan Nagler, Department of Political Science, University of California – Riverside, Riverside, CA 92521-0118. LaTeX format files are especially encouraged. The deadline for submissions for the next issue is January 31, 1996.