The Political Methodologist

Newsletter of the Political Methodology Section
American Political Science Association
Volume 7, Number 2, Spring, 1996

Editor: Jonathan Nagler, University of California, Riverside
Editorial Assistant: Carole Chaney, University of California, Riverside

Contents

Notes From the Editor

Jonathan Nagler
University of California, Riverside
nagler@wizard.ucr.edu

This is my first issue as editor of *TPM*. I hope it continues somewhere near the high level of quality established by my predecessors. This issue features two articles on aspects of estimation. Robert Franzese describes how to apply the Beck and Katz pooled-cross sectional estimation procedure on data with missing observations. And Neal Beck offers advice on estimating standard errors which are robust in the face of heteroscedasticity. Andrew Martin and Kevin Quinn then describe three optimization algorithms: genetic algorithims, tabu search, and simulated annealing. Mike Alvarez reports on the current and future status of the Political Methodology Web Site and Electronic Paper Server which now contains almost 100 manuscripts and served over 300 requests for papers in June alone. Finally, returning to the nuts and bolts of methodology: Annette Steinacker, Kevin Smith, George Krause, Mohan Penubarti, Andy Skalaban, and Janet Box-Steffensmeier review econometrics texts suitable for a first year graduate course in econometrics. Chris Zorn reviews a slightly more advanced text by Dougherty. And Larry Bartels offers a graduate syllabus for Research Design. This issue also includes information about: *Political Analysis*, the 1996 Political Methodology Summer Meeting, the Political Methodology Panels at the 1996 APSA Meeting, and instructions on using the Electronic Paper Server.

A Gauss Procedure to Estimate Panel-Corrected Standard-Errors with Non-rectangular and/or Missing Data

Robert J. Franzese, Jr.
Harvard University
franzese@husc.harvard.edu

[Acknowledgements: Gratitude is due to Neal Beck for providing the Beck-Katz procedures, for other suggestions, and for general helpfulness. This piece has also benefitted from the helpful criticism of James Alt, Gary King, Kenneth Scheve, and Curtis Signorino. Special thanks to Curt for his assistance in optimizing the procedure for minimal memory and time use and to Kenneth Scheve for finding an error in an earlier version of the procedure.]

Econometric analysis of time-series-cross-section (TSCS) data often present the researcher with residuals which exhibit contemporaneous correlation (i.e., residuals from different cross-sections (CS's) in the same year are correlated) in addition to the more usual time-series property of serial (or auto-) correlation and the typical cross-section property of heteroskedasticity. This non-sphericity in the error-covariance matrix implies that estimated coefficients are inefficient and that standard errors are inconsistent. Prior to Beck and Katz (1995), the usual solution was to estimate Feasible Generalized Least Squares in up to three steps, estimating a serial-correlation parameter for each cross-section, a error-variance parameter for each cross-section, and a number of contemporary correlation parameters. In principle, this solution is unbiased, consistent, and efficient in both its coefficient and standard-error estimates.

Beck and Katz show, however, that in more realistically sized data sets the procedure tends to underestimate standard errors seriously. In essence, the problem is that there are a large number of parameters in the error-covariance matrix to be estimated and relatively few observations are being leveraged to estimate them. While this is perfectly fine in that mythical land of assymptopia (actually only one province in that land–the one where the number of time-periods (TP's) goes to infinity), in practical samples it has a tendency to overstate one's certainty about coefficient estimates relative to what Monte Carlo experimentation suggests is appropriate.

They suggest, therefore, that one deal with the serial-correlation property separately, transforming the data and/or including lag(s) prior to worrying about contemporaneous correlation. [1] Panel heteroskedasticity, perhaps should be ignored; the implications of their findings regarding such non-constant variance depend on the actual sample size and the researcher's distaste for inefficiency relative to standard-error inaccuracy. In any event, it may be treated in the usual WLS way or ignored, and, while they suggest the latter is usually safer for smaller data sets, the issue is separate from what should be done about contemporaneous correlation. These findings are treated in more detail by Beck and Katz in their forthcoming *Political Analysis* article. Finally and most importantly for the present purposes, they show that the efficiency gains from the contemporaneous correlation correction are typically small to non-existent while the standard-error bias is relatively large. Accordingly, they suggest estimating standard errors according to a robust procedure which incorporates potential contemporaneous-correlation information into the coefficient variance-covariance matrix without adjusting the coefficient estimate. They show that this procedure – controlling for serial correlation in some (simple) way prior to final estimation and relying on robust standard errors, i.e. Panel-Corrected Standard-Errors or PCSE's – has admirable small sample properties relative to alternatives.

Beck and Katz have also made available computer programs in RATS or GAUSS for calculating these PCSE's. However, both of their procedures constrain the data to be analyzed to be rectangular (i.e., all CS's begin and end in the same year) and to have no missing data. This is a restriction in their programs, not in their econometric analysis. In other words, in principle, there is no reason PCSE's cannot be estimated in non-rectangular samples and/or samples with missing data. This is fortunate since almost every pooled data set has at least one or the other complication. For example, in "The Political Economy of Public Debt," I had a data set with 21 CS's and up to 35 time-periods

[1] Actually, they make a related, stronger argument that estimating a single serial-correlation parameter rather than N different parameters is probably optimal in small (i.e., real) samples.

(TP's) for a total of 550+/- usable observations, but if I had to rectangularize it somehow and remove breaks in the sample within a CS, it would have had at most 300 observations! An extreme example to be sure, but it is not rare to have to sacrifice 15% or more of one's data to rectangularize it and remove missing-data breaks. Of course, we would like to avoid this if possible. Accordingly, I have written a GAUSS procedure, "pcse.g", which will accept such data and produce PCSE estimates along with the usual regression output. The procedure also adds a number of options which I think researchers commonly dealing with TSCS data will find useful.

Beck and Katz's PCSE estimate of coefficient standard errors is defined by:

$$(X'X)^{-1}X'(\frac{E'E}{T} \otimes I_t)X(X'X)^{-1}$$

where E is the TxN (number of TP's x number of CS's) matrix of OLS residuals, $\otimes$ is the Kroenecker product operator, and I_t is an identity matrix of size T. Notice that E'E/T is just an estimate of the matrix of error covariances between cross- sections. The difficulty with missing values and/or non-rectangular samples is that not each element of E'E is in fact based on T pairs of observations. My procedure solves this problem by substituting zeros for missing values in X and E, and dividing each element of E'E by the number of pairs of observations on which it is actually based (i.e., not counting the pairs where one or the other (or both) observation(s) is (are) missing). This is done by creating another TxN matrix, M, where each element is equal to one if the corresponding observation in E is not missing and zero if it is. Then, E'E divided element by element by M'M can replace E'E/T in the above formula and all is correct.

The procedure has been tested (in Gauss 386-i v. 3.1 and 3.2.13 running under DOS and in a DOS Window under Windows 3.1) as follows. First, I created some artificial data, rectangular and without missing values. X, the independent variable is random normal. e, the true error is also random normal and independent of X. Y, the dependent variable, was set equal to X+e. Neal Beck's (thank you, Neal) GAUSS PCSE procedure was then run on these data. In this sample, my procedure produces identical results. This proves only that my procedure works in perfect rectangular data sets if Beck's does (I consider that about as safe a bet as can be found). Testing how my procedure compares to Beck's in non-rectangular data sets and/or those with missing values cannot be done because the latter cannot deal with such data to provide the baseline. I did perform a very crude test in this setting by, again generating random normal and independent data as above, only non-rectangular this time. The results are generally close to equal to the OLS results, and appear to converge toward it in larger samples, as we would expect if it were working correctly. I fear that is the best that can be done by way of a comparison to the ideal-type.

More generally, the procedure seems to operate bug-free for me (I have been using it extensively in a data set with 850+/- rows, but only 550+/- usable observations; the data are non-rectangular and have a few internal missing values). It was also used/tested this fall by Gary King's second-semester graduate statistics class without mishap following one important correction (thank you to Ken Scheve).

What the procedure can do:

- (1) Produce PCSE's in any data set where they are theoretically defined.
- (2) Automatically generate and use CS and TP fixed-effect dummies if requested. (See Smith (1995) for a cogent discussion of the interpretation of fixed-effect models in TSCS data.)
- (3) Estimate and incorporate Panel WLS information prior to applying PCSE's if requested. N.b., this is done in two stages. First, OLS is estimated. The residuals from that are squared and regressed on the CS indicators (less one of them) and a constant. The F-test from this regression can be seen as a rough-and-ready test of homoskedasticity as the null against CS-specific variance (and the DW from this auxiliary regression can be seen as an equally rough-and-ready test of ARCH(1) against the null of panel heteroskedasticity). The inverse of the square root of the fitted values from this auxiliary regression are then used as weights to transform the original data. OLS is run on the transformed data, which produces the Panel WLS results. The PCSE estimation is then applied to these results. The line in the procedure where the squared residuals are regressed on the CS indicators and related commands can be changed to a different model of error- variance if the researcher desires. They are flagged for easy recognition.
- (4) GAUSS's OLS.SRC procedure produces incorrect Durbin-Watson statistics when there are missing values in the middle of the data (Aptech has been notified). The problem is that OLS uses the packr() command to remove missing values from the data eventually passed to the part of the procedure where estimates are made. This means that later, when any time-series properties are analyzed some of the physically adjacent observations are no longer truly temporally adjacent. Thus the reported Durbin-Watson statistic is not based on the correct ordered set of residuals. This problem is fixed prior to the reporting of DW-stats in output from my procedure, but it remains in OLS.SRC.

The procedure (PCSE.G) and a brief note with specific instructions for how to use it (PCSENOTE) can now be obtained from the American University Gauss Library. Please send any comments to me at franzese@fas.harvard.edu.

References

Beck, Nathaniel and Jonathan Katz. "What to Do (and Not to Do) with Time-Series-Cross-Section Data in Comparative Politics", *American Political Science Review*, v. 89(3): 634-47.

———-. N.d. "Nuisance or Substance: Specifying and Estimating Time-Series-Cross-Section Models", *Political Analysis*, forthcoming.

Franzese, Robert. 1995. "The Political Economy of Public Debt: An Empirical Examination of the OECD Postwar Experience," unpublished manuscript.

Smith, Mark. 1995. "Time Serial Designs and Cross-Sectional Designs: Uncovering the Structural Logic of Pooled Analyses", Paper Presented at the 1995 APSA Conference.

Robert J. Franzese, Jr. is finishing his Ph.D. in the Department of Government at Harvard University. This fall he will be an Assistant Professor in the Department of Political Science at the University of Michigan. His fields are comparative and international political economy, methods, and comparative politics–developed democracies.

Reporting Heteroskedasticity Consistent Standard Errors

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

Leamer (1978) noted that when he was a gradate student at the University of Michigan the high priests who taught econometrics reverted to being ordinary sinners when they descended to the basement to work on the Michigan model of the US economy, committing exactly the same sins that they preached against while upstairs. But there are cases where the high priests have given us methods for guarding against sin, tools which could easily be used, but often are not used, by common sinners. One such tool is the use of "robust" (heteroskedasticity-consistent) standard errors. Modern computer packages make it easy to avoid the sin of computing standard errors which are inconsistent in the presence of heteroskedasticity; with just a bit more work it appears possible to compute robust standard errors which outperform OLS standard errors.

Working with a simple linear setup, let

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{1}$$

where $\mathbf{y}$ is the vector of observations on the dependent variable and $\mathbf{X}$ is the matrix of observations on the independent variables. Denote the vector of OLS residuals by $\mathbf{e}$; let ϵ be the vector of error terms having covariance matrix $\mathbf{\Omega}$.

Following any standard textbook (e.g., Greene 1993, Ch. 14), the usual OLS standard errors are given by the square root of the diagonal terms of

$$Var(\hat{\beta}) = s^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2}$$

where s^2 is the usual OLS estimator of σ^2. This is derived from the more complicated formula

$$Var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \tag{3}$$

given the assumption that the error terms are homoskedastic so that the variance of the errors is

$$\mathbf{\Omega} = \sigma^2\mathbf{I}. \tag{4}$$

Unfortunately, the OLS standard errors will not be consistently estimated in the presence of heteroskedasticity, since without Equation 4, Equation 3 does not simplify to Equation 2.[1]

White (1980) showed that it is possible to estimate consistent standard errors even in the presence of heteroskedasticity. This result is rather remarkable, since it would seem necessary to first estimate $\mathbf{\Omega}$, which has the same number of unknown parameters (the diagonal terms) as we have observations. White's breakthrough was to show that even though it is impossible to estimate $\mathbf{\Omega}$, it is possible to consistently estimate $(\mathbf{X}'\mathbf{\Omega}\mathbf{X})$ by $(\sum_i e_i^2\mathbf{x_i}\mathbf{x_i}')$, which is all that is needed. Consistent standard errors are then given by the square root of the diagonals of

$$Var(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}(\sum_i e_i^2\mathbf{x_i}\mathbf{x_i}')(\mathbf{X}'\mathbf{X})^{\mathbf{-1}} \tag{5}$$

where $\mathbf{x_i}$ refers to the vector of independent variables for the i^{th} observation.

This looks formidable, but many packages make it very easy to compute these standard errors. Thus,

[1] The OLS standard errors will be consistent so long as the squared errors are unrelated to the squares and cross-productions of the independent variables. This is the basis for White's (1980) test for heteroskedasticity, which regresses the squared OLS residuals on the squares and cross-products of all independent variables (as well as the original variables and then assesses the significance of the resulting $N \times R^2$ from that auxiliary regression. Thus White's test is not a test for heteroskedasticity but rather for "harmful" heteroskedasticity that leads to incorrect standard errors.

for example, STATA produces robust standard errors using the command hreg (h standing for Huber, who pioneered the maximum likelihood version of robust standard errors); LIMDEP produces such errors with the Hetero subcommand to REGRESS; and RATS computes robust errors with the robusterrors option to linreg. Thus there is no reason for applied researchers not to compute robust standard errors routinely.

Should they do so? Robust standard errors have been shown to be slightly inferior to OLS standard errors when the model is homoskedastic, but markedly better than OLS standard errors in the presence of heteroskedasticity. Robust standard errors become more accurate with increases in sample size (they are, after all, consistent!). Monte Carlo studies (MacKinnon and White 1985) show that, under homoskedastic conditions, robust standard errors are about as good as OLS standard errors, so long as the sample size is large (say $N > 200$). The OLS standard errors do not become more accurate as sample size increases if there is "harmful" heteroskedasticity. Thus it seems clear that those doing large sample work should always report robust standard errors in place of the traditional OLS standard errors. Failure to do so may lead to incorrect t-statistics and hence incorrect inferences about model parameters.

The situation is less clear for smaller samples. One obvious solution is to first test for "harmful" heteroskedasticity, using the test described in the footnote. One would then use robust standard errors in the presence of "harmful" heteroskedasticity. (More general tests for heteroskedasticity may pick up "benign" heteroskedasticity and hence should not be used to decide on whether robust standard errors are required.) While it is easy to implement the White test, no package currently automates such a test.

The Monte Carlo studies reported in MacKinnon and White (1985) seem to show that there is a better small sample robust standard error, the jackknife. This estimator, denoted HC3, is obtained by calculating the square roots of the diagonal of

$$\frac{N-1}{N}(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}\mathbf{U^2}\mathbf{X} - \frac{\mathbf{1}}{\mathbf{N}}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}](\mathbf{X}'\mathbf{X})^{\mathbf{-1}} \qquad (6)$$

where $\mathbf{u}$ is a vector with typical element $\frac{e_i}{1-h_{ii}}$, the h_{ii} are diagonal elements of the "hat" matrix $(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ and $\mathbf{U^2}$ is a diagonal matrix made up of the squares of the elements of $\mathbf{u}$ along the diagonal.

The LIMDEP manual (Greene 1995, p. 262) provides a simple program to compute HC3. Users of other packages will have to provide their own matrix routines for this computation. Davidson and MacKinnon (1993, p. 554) show that a computationally simpler approximation to HC3 may be obtained by using $\frac{e_i^2}{(1-h_{ii})^2}$ for the e_i^2 in Equation 5. Those writing their own matrix routine will probably appreciate this simplification. (It may be possible to trick some programs that compute the usual White robust errors into computing HC3 by redefining the residuals as $\frac{e_i}{1-h_i i}$, though obviously this should only be done with extreme caution.)

How much differences does the computation of robust standard errors make in practice? The answer clearly depends on the amount of "harmful" heteroskedasticity in the data. As one example, I computed a regression of voting turnout by proportion black in 661 Marion County, Indiana precincts. (This data was provided by Gary King and is the basis for Figure 4.1 in his forthcoming *A Solution to the Ecological Inference Problem*. My interest here is in the regression relating racial characteristics to precinct turnout, rather than any ecological inference about individual voters.) The estimated regression slope is -0.107.

The OLS standard error for this slope is .023. The White robust standard error, as computed by LIMDEP is .017. King's Figure 4.1 and the White test both show that the errors are heteroskedastic. This heteroskedasticity leads the OLS standard errors to overstate variability by about one third. Note that in this case the OLS standard errors are bigger than the robust standard errors; in other cases the opposite relationship may hold. Thus the use of OLS standard errors in the presence of heteroskedasticity may lead researchers to be either overly confident or insufficiently confident about their findings.

Given the large sample size for this example, the jack-knifed HC3 standard error should be very close to the White robust standard error. In this example HC3 is about half of one percent greater than the White robust errors, agreeing with it to two significant digits. We would expect a great difference in smaller samples. To examine this, I randomly sampled one ninth of the Marion County precincts. This should increase estimated standard errors, as compared with the full 661 precincts, by a factor of three.

The OLS standard error for the subset of precincts is .072, 3.1 times the standard error for the full set of precinct. The White robust standard error for the subset is .044, only 2.6 times the standard error for the full data set. This is consistent with the MacKinnon and White simulations which found a tendency for the White robust standard errors to understate variability in small samples. The jack-knifed HC3 standard error is .047, about 2.8 times the full sample standard error. Thus while it still understates variability, the jack-knifed standard errors as expected, performs somewhat better than does the White robust standard error. But, more importantly, either heteroskedasticity corrected standard error is considerably more accurate than the OLS standard error. Thus users who find themselves unable to compute the jack-knifed HC3 standard errors should prefer to use packages which at least compute White's original robust standard errors. But those with access to the appropriate matrix computations should probably compute the jack-knifed HC3. While the ease of doing this varies from package to package, it is surely easy enough to do so in `LIMDEP`,

Finally, it should be noted that White's approach to standard errors which are robust to heteroskedasticity succeeds because it does not assume that the analyst knows the nature of the heteroskedasticity. Such ignorance is clearly the most common situation. But there are times, such as with time-series–cross-section data, that the analyst may have some better insight about the nature of $\mathbf{\Omega}$. Such structure can then be incorporated into Equation 3. This is the basis for the "panel correct standard errors" I developed with Jonathan Katz (Beck and Katz 1995). There are also circumstances where knowledge about the form of the heteroskedasticity can be used to improve estimation through weighted least squares. Such an approach has proven extremely useful in the analysis of time-series, where it is often the case that heteroskedasticity follows a simple autoregressive form, leading to Engle's (1982) autoregressive conditional heteroskedasticity (ARCH) model and its generalizations. But in most cross-sectional studies it is hard to parameterize heteroskedasticity. In such cases the computation of robust standard errors at least lessens the likelihood of incorrect inference.

References

Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Times-Series Cross-Section Data." *American Political Science Review* 89:634-47.

Davidson, Russell and James MacKinnon. 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press.

Engle, Robert. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987-1008.

Greene, William H. 1993. *Econometric Analysis.* Second edition. New York: Macmillan.

Greene, William H. 1995. *LIMDEP: Version 7.0 Users Manual.* Bellport, N.Y.: Econometric Software, Inc.

Leamer, Edward. 1978. *Specification Searches: Ad Hoc Inference with Non-experimental Data.* New York: Wiley.

MacKinnon, James and Halbert White. 1984. "Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29:305=325.

White, Halbert. 1980. "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica* 48:817-838.

Nathaniel Beck is Professor of Political Science at UCSD. He is currently working on time-series–cross-section models, the application of generalized additive models and the analysis of discrete duration data. These techniques are used to study comparative political economy and international relations.

A Review of Discrete Optimization Algorithms

Andrew D. Martin and Kevin M. Quinn
Washington University
admartin@wugate.wustl.edu and
kquinn@wugate.wustl.edu

Introduction

Political scientists are often faced with optimization problems involving discrete parameter spaces, multi-modal functions, functions which are not well defined, noisy functions, and even non-differentiable functions. These problems can arise in maximum likelihood estimation (MLE), forecasting, dynamic modeling, and some types of game theoretic models. While these problems were daunting 10 years ago, recent advances in computational optimization combined with falling computer prices have brought us much closer to reasonable solutions. In these applications, we encounter optimization problems of the following general form: minimize $c(x_i)$ s.t. $x_i \in \mathbf{X}$ where $c(\cdot)$ is the objective function and $\mathbf{X}$ is the solution space[1]. The objective function $c(\cdot)$ maps the members of the solution space onto the real number line. This article serves as a review of three widely used discrete optimization algorithms that are well suited to dealing with the problems above and provides suggestions of the types of problems each method is well suited. The three techniques we review are: genetic algorithms, tabu search, and simulated annealing. These algorithms are able to provide solutions to optimization problems in which calculus based optimization is infeasible or impossible. While the focus of this review is on discrete optimization, it should be noted that two of the algorithms

[1] Note that minimizing $c(\cdot)$ is equivalent to maximizing $-c(\cdot)$.

(genetic algorithms and simulated annealing) discussed below have real-coded counter-parts.

Genetic Algorithms

Genetic algorithms (GAs) were created by Holland (1975, 1992) to study the mathematical underpinnings of adaptive behavior. While GAs were designed to simulate natural adaptive behavior, they have also proven to be very powerful optimization procedures. This twofold ability of GAs to both simulate adaptive behavior and to solve extremely difficult optimization problems has proven quite useful to several social scientists. Examples of social scientific work employing GAs include searches for optimal strategies in complicated formal models (Axelrod 1987; Miller 1989; Andreoni and Miller 1990; and Kollman, Miller, and Page 1992), and the estimation of LISREL models (Mebane et al. 1995). GAs have also been used to find solutions to systems of nonlinear equations (Shaefer, 1985).

GAs are implemented in roughly the following manner. First, choose a coding scheme which maps each element of the search space onto a unique bit vector [2]. Then randomly select a small set, $X_t \subset \mathbf{X}$, of m potential solutions to the optimization problem. Evaluate the objective function $c(\cdot)$ at each potential solution $x_i \in X_t$. The difference between $c(x_i)$ and some measure of the expected value of $c(x_i)$ (for example, $\frac{1}{m}\sum_{j=3D1}^{m} c(x_j)$) provides a measure of fitness for x_i. Then form a new set of potential solutions, X_{t+1}, by applying a series of genetic operators to X_t. Reproduction generally occurs first. This operator selects potential solutions from X_t (with replacement) on the basis of fitness (relatively more fit solutions are more likely to be selected). The crossover operator is then applied. The crossover operator randomly takes vectors of potential solutions, breaks them apart, and recombines them. For example, the crossover operator could transform {11111} and {00000} into {11000} and {00111}. Finally, the mutation operator flips elements of the solution vectors to the opposite value with some low probability. Denote the new set of solutions formed from X_t as $X_{t+1} \subset \mathbf{X}$.

Iterating this procedure a number of times yields a very powerful optimization algorithm. As Holland and Miller (1991) note, this three-part process of reproduction, crossover, and mutation may seem to be nothing more than a random search algorithm which retains the best potential solutions. As they further argue, this is in fact not the case. In order to understand why, think of each bit of a solution vector as an arm of an n-armed bandit. The problem then is to allocate trials to each of the n arms in a manner which will yield the highest cumulative payoff. Holland (1975, 1992) has shown that GAs allocate trials to building blocks in a manner which very closely corresponds to the optimal solution of an n-armed bandit problem.

Goldberg (1989) provides a non-technical introduction to GAs. For advanced discussions of additional genetic operators as well as refinements and variations of the three basic genetic operators we urge the reader to see the edited volumes by Grefenstette (1987), Rawlins (1991), and Whitley (1993).

Tabu Search

Tabu search, first proposed by Glover (1977), is a meta-heuristic used to solve both combinatorial and discrete optimization problems. For the purpose of discrete optimization problems, the heuristic used in the tabu search algorithm is a local improvement scheme, beginning with a good feasible solution. Local search starts from an initial solution $x_i \in \mathbf{X}$ and searches to find an improving solution $x_{i+1} \in \mathbf{X}$. In other words, the search attempts to find an x_{i+1} such that $c(x_{i+1}) < c(x_i)$.

Consider the case when optimizing over a discrete space $\mathbf{X}$ with respect to an *a priori* objective function $c(\cdot)$. Define $\mathbf{X}$ as the solution space which contains all of the possible solutions $x \in \mathbf{X}$. For each x_i the practitioner defines a set $N(x_i) \subset \mathbf{X}$ which denotes the neighborhood of x_i. On each iteration of the search, the objective function value is evaluated for all $x \in N(x_i)$. The entire neighborhood is searched and an improving move is chosen by selecting the move x_i which most improves the value of $c(x_i)$. That move x_i is chosen and is labeled x_{i+1}. The search moves to the next iteration, by looking at the neighborhood of the accepted move, $N(x_{i+1})$. The problem with simple local search is that traps of local optimality cannot be escaped. In a discrete space, local optimality is defined in terms of an *a priori* neighborhood structure as opposed to an $\varepsilon-$neighborhood in the continuous case. Although unimodal functions are easily optimized, multi-modal functions make local search an impractical optimization technique. To remedy this problem, simple local search is modified to accept some non-improving moves in an attempt to escape traps of local optimality.

The first of these meta-heuristics based on local search is called tabu search. Tabu search uses a memory structure (called the tabu list) that restricts the possible members of the neighborhood to which a search can progress. Thus, once a local optima is encountered, the search will not be able to revisit that area of the solution space. The tabu list must be small enough to allow the search to carefully scrutinize certain parts of the solution

[2] Real number encoded GAs are also possible. For discussions of the implementation and performance of real-coded GAs, we refer the reader to Eshelman and Schaffer (1993), Wright (1991), and Antonisse (1989). While we only discuss binary-coded GAs here for reasons of simplicity and space, it should be noted that the use of real-coded GAs has grown rapidly in the pas t few years.

space, yet large enough to prevent a return to a previously generated solution. The tabu search meta-heuristic also uses an aspiration criterion which defines a condition under which the tabu status of a certain move can be overridden. Short term memory functions are employed to intensify and diversify the search. Tabu search is allowed to run for a maximum number of iterations that is computationally practical. A comprehensive description of tabu search can be found in Glover and Laguana (1993).

When implementing tabu search, the practitioner must define the neighborhood structure with respect to the solution space, select the type of tabu list to be employed, and determine the aspiration criterion to be used. Practitioners also traditionally choose to employ multi-start techniques, where tabu search is re-started numerous times from different members of th e solution space. Throughout the operations research literature, there are many examples of successful implementations of tabu search as well as discussions of effective tabu structures and aspiration criteria (Cvijovic and Klinowski 1995; Glover and Laguana 1993; and Glover 1990).

Simulated Annealing

Another meta-heuristic which relies on local search is called simulated annealing. Simulated annealing was first introduced by Kirkpatrick et. al. (1983) and Cerny (1985) and has roots in the work of Metropolis et al. (1953). Simulated annealing is analogous to the annealing process in physical chemistry, when liquid metals are heated and then left to cool into a steady, organized state. Numerous successful applications of simulated annealing can be found in Collins, et. al. (1988). The simulated annealing algorithm can be described in terms of a Markov chain. The solution space $\mathbf{X}$ consists of the feasible solutions that satisfy all the constraints $x \in \mathbf{X}$. An objective function $c(\cdot)$ is defined on $\mathbf{X}$. From each state x_i a transition is a search action that combines the selection of a state $x_j \in N(x_i)$ with the decision of whether to move to x_j state or not. The neighborhood $N(x_i) \subset \mathbf{X}$ of state x_i is defined as the set of states that can be reached from state x_i in exactly one step. Thus, if the transition probability $p_{x_i x_j} > 0$, then $x_i \in N(x_j)$. Furthermore, the selection is reversible; if $x_j \in N(x_i)$ then $x_i \in N(x_j)$. In simulated annealing, each member of the neighborhood is randomly selected, and the algorithm then determines whether to move on to the next state. The decision to move to the next state depends on the values of $c(x_i)$ and $c(x_{i+1})$. The decision allows the acceptance of some non-improving moves, thus escaping traps of local optimality. The algorithm provides a chance for the search to escape from a local optimum based on an acceptance probability, which is defined as:

$$\Pr(accept\ x_{i+1}) = min\left\{1, \frac{\exp[c(x_{i+1}) - c(x_i)]}{t}\right\} \quad (1)$$

where t is the temperature control parameter. This temperature control parameter is decreased as the search progresses, thus allowing the search to settle down into a local optimum.

It can be shown that convergence to the global optimum is guaranteed if the temperature control parameter approaches 0 and an infinite number of transitions is made. However, since this convergence is quite impractical, a finite-time implementation of the simulated annealing algorithm is often used to approach the optimal solution within a reasonable amount of computation time. When implementing simulated annealing, a practitioner must define the neighborhood structure with respect to the solution space and develop a cooling schedule with which to decrement the temperature parameter t. A survey of successful cooling schedules can be found in Hajek (1988) and Collins et. al. (1988).

Simulated annealing is guaranteed to converge to the global optimum of functions defined over both discrete and continuous spaces as the cooling parameter t goes to zero. Thus, it is particularly appropriate for estimation of econometric models.[3]

Conclusion

In sum, the operations research literature provides numerous computational techniques that political scientists can implement to conquer previously intractable problems. These techniques can be applied to optimization problems encountered in the estimation of econometric models, forecasting, dynamic modeling, and some types of game theoretic models. For a comprehensive description, evaluation, and comparison of many discrete optimization techniques, we refer the reader to Ackley (1987) who empirically assesses the success of each algorithm.

References

Ackley, David H. 1987. "An Empirical Study of Bit Vector Optimization." in Lawrence D. Davis (ed.). *Genetic Algorithms and Simulated Annealing.* London: Pittman.

Andreoni, James, and John H. Miller. 1990. "Auctions Wit h Artificial Adaptive Agents." University of Wisconsin Department of Economics, mimeo.

[3]It is interesting to note that the underlying logic behind the method of simulated annealing can also be used to simulate multivariate distributions. See Chib and Greenberg (1995) for a more detailed discussion of these issues.

Antonisse, H. J. 1989. "A New Interpretation of Schema Notation that Overturns the Binary Encoding Constraint." *Proceedings of the Fourth International Conference on Genetic Algorithms.* San Mateo, CA: Morgan Kaufmann.

Axelrod, Robert. 1987. "The Evolution of Strategies in the Iterated Prisoner's Dilemma." in Lawrence D. Davis (ed.). *Genetic Algorithms and Simulated Annealing.* London: Pittman.

Cerny, V. 1985. "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Program." *Journal of Optimization Theory and Applications.* 45: 41-51.

Chib, Siddhartha and Edward Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician.* 49: 327-335.

Collins, N.E., R.W. Eglese and B. L. Golden. 1988. "Simulated Annealing - An Annotated Bibliography." *American Journal of Mathematical and Management Sciences.* 8: 209-307.

Cvijovic, Djurdje, and Jacek Klinowski. 1995. "Taboo Search: An Approach to the Multiple Minima Problem." *Science.* 267: 664-666.

Eshelman, L. J., and J. D. Schaffer. 1993. "Real-Coded Genetic Algorithms and Interval-Schemata." in L. Darrell Whitley (ed.). *Foundations of Genetic Algorithms 2.* San Mateo, CA: Morgan Kaufmann .

Glover, Fred. 1977. "Heuristics for Integer Programming Using Surrogate Constraints." *Decision Sciences.* 8: 156-166.

Glover, Fred. 1990. "Tabu Search: A Tutorial." *Interfaces.* 20: 74-94.

Glover, Fred and Manual Laguana. 1993. "Tabu Search." in Colin R. Reeves, (ed.) *Modern Heuristic Techniques for Combinatorial Problems.* Oxford: Blackwell Scientific Publications. pp. 70-150.

Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading, MA: Addison Wesley.

Grefenstette, John J. (ed.) 1987. *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms.* Hillsdale NJ,: Lawrence Erlbaum Associates.

Hajek, B. 1988. "Cooling Schedules for Optimal Annealing." *Mathematics of Operations Research.* 13: 311-329.

Holland, John H. [1975] 1992. *Adaptation in Natural and Artificial Systems.* Cambridge, MA: MIT Press.

Holland, John H., and John H. Miller. 1991. "Artificial Adaptive Agents in Economic Theory." *American Economic Review.* 81: 365-370..

Kirpatrick, S., C.D. Gelatt and M.P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science.* 220: 671-680.

Kollman, Ken, John H. Miller, and Scott Page. 1992. "Adaptive Parties in Spatial Elections." *American Political Science Review.* 86: 929-937.

Mebane, Walter R., Jasjeet S. Sekhon, and Martin T. Wells. 1995. "The Robustness of Normal-theory LISREL Models: Tests Using a New Optimizer, the Bootstrap, and Sampling Experiments, with Applications". Paper presented at the 1995 Political Methodology Summer Conference.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. "Equations of State Calculations by Fast Computing Machines." *Journal of Chemical Physics.* 21: 1087-1092.

Miller, John H. 1989. "The Co-Evolution of Automata in the Repeated Prisoners Dilemma." Sante Fe Institute Economics Research Program Working Paper No. 89-003.

Rawlins, G. J. E. (ed.) 1991. *Foundations of Genetic Algorithms.* San Mateo, CA: Morgan Kaufmann.

Shaefer, C. G. 1985. "Comparison of Methods for Solving Nonlinear Equations." Research Memo RIS-24r. Cambridge, MA: Rowland Institute for Science.

Whitley, Darrel L. (ed.). 1993. *Foundations of Genetic Algorithms 2.* San Mateo CA: Morgan Kaufmann.

Wright, A. 1991. "Genetic Algorithms for Real Parameter Optimization." in *Foundations of Genetic Algorithms.* G. J. E. Rawlings, ed. San Mateo, CA: Morgan Kaufmann.

Andrew D. Martin received an A.B. in mathematics and government from The College of William and Mary. He is currently pursuing a Ph.D. in political science at Washington University in St. Louis. His academic interests include American politics, formal theory, and political methodology.

Kevin M. Quinn received an B.A. in political science from The Johns Hopkins University and an A.M. in political science from Washington University in St. Louis, where he is presently a doctoral candidate. His

academic interests include comparative politics, formal theory, and political methodology.

Polmeth — You've Come a Long Way Baby

R. Michael Alvarez
California Institute of Technology
rma@crunch.caltech.edu

In many ways it is difficult to remember life without a direct pipeline into the Internet and the World Wide Web. It was just within the past few years that most people in political science have migrated full-force onto the Internet and the Web, and it is becoming quite clear now that both are shaping the way we engage in research and how we interact professionally, in quite profound ways.

One of the developments of interest to political methodologists has been the rapid evolution of our Political Methodology World Wide Web server and our "polmeth" discussion group, both maintained by Jonathan Nagler at the University of California, Riverside, through generous support by the National Science Foundation and UC Riverside. The purpose of this article is to take a brief look at the progress of *Polmeth* in the past year. I want to present some statistics on the usage of *Polmeth* which clearly document the dramatic and rapid effect which *Polmeth* has had on political methodology in the past year, and then present a few ideas for future development of our professional and research connections to the Web.

A Brief History

Polmeth began without much of a bang in the spring of 1994. A number of people began an email discussion that spring focused on both the desireability and the functionality of providing a centralized place where people could deposit the papers which they were to present at upcoming political methodology summer conferences, and at other national meetings. A number of important issues were raised in these discussions:

- Where would the paper repository be located, and how could it be maintained?
- How could we encourage (or worse yet, coerce) our colleagues into using this internet service instead of making endless copies of papers, hauling them on airplanes bound for their next meeting, and passing them out at the meeting instead of distributing them beforehand?
- What formats would we use? How could people easily distribute machine-readable versions of their papers without running the risk that the contents could be easily altered?

To examine the practical issues behind the development of a true Internet paper distribution system, we began two simple experiments. We convinced the section leadership that this practice would help advance methodology intellectually — and that they should thereby encourage people who were presenting their papers at the 1994 Summer Methodology Conference to provide machine-readable versions of their papers to the participants of the meeting, before the meeting. We set up an anonymous ftp directory at Caltech where paper presenters could place machine-readable versions of their papers, and an email reflector where they could send an email which would be "bounced" to every meeting participant.

To resolve the practical issues of paper format, we asked people to provide, at bare minimum, a version of their paper which could be printed on any HP laser jet. We also asked people to provide a Postscript version of their paper, if possible. Our thinking was that virtually everyone we knew had either an HP laser jet, or an Apple-style laserwriter. Thus, these two formats should cover the bases.

The experiment was a great success. Almost all of the papers presented at the 1994 Summer Methodology Conference were uploaded to our anonymous ftp server before the meeting; after the meeting another paper or two were added (they still are available on *Polmeth*!). As each paper was uploaded, we verified the integrity of the upload by checking that we ourselves could print the paper; if the upload was successful, we notified every meeting participant of the availability of the paper. The only problems we encountered in our experiment were "persuading" our colleagues that it was in their best interest to "post" their paper before the meeting, and some difficulties associated with the improper use of non-binary transmissions and retrievals. At the 1994 Conference, we had an open discussion of this experiment, and the meeting participants were virtually unified in their recommendation that we attempt to expand this service.

The Development of *Polmeth*

The rapid evolution of the Web in late 1994 and early 1995 facilitated this task. The members of our informal discussion group agreed that the development of a Web server for the Political Methodology community was the right direction to take. With the support of the NSF, Nagler was able to set up *Polmeth* on a high-speed Hewlett-Packard workstation in April 1995; after that, *Polmeth* was in business!

The early offerings on *Polmeth* were scarce. There were a series of links to other "interesting" sites, information about the 1995 Summer Methodology Conference, and a place for researchers to deposit and retrieve machine-readable working papers. Again, with the help of the Methodology Section leadership, we used the 1995 Summer Conference as our second major experiment. 1995 Meeting Participants were strongly encouraged to use the new *Polmeth* archive for distributing their papers; almost all participants did (and the papers can still be accessed)! Additionally, a number of individuals placed methodology-related working papers on the archive for distribution purposes.

The scheme of our 1994 experiment was used as an outline for the new paper archive. Now, when an individual "uploads" their machine-readable working paper, they set into motion a chain of automatic events: the paper is placed in the appropriate directory, the abstract is converted to an html file, links to the paper and abstract are established, and an email is sent to all current members of the "polmeth" discussion group announcing availability of the paper and describing how to retrieve it. If the author submitted a postscript version of the paper, then people can read the entire paper directly over the net without actually downloading it to their disk or printing it.

What is quite amazing (and which should provide paper authors a great incentive to use *Polmeth*) are the statistics on downloads (retrievals) of papers in the nine month period ending May 28, 1996. In this period there were 2203 retrievals! The month-by-month statistics are:

- September 1994: 185 downloads
- October 1994: 239 downloads
- November 1994: 249 downloads
- December 1994: 132 downloads
- January 1996: 143 downloads
- February 1996: 143 downloads
- March 1996: 179 downloads
- April 1996: 668 downloads
- May 1996: 265 downloads

The bulge in April of 1996 was from the Midwest meeting - where each participant on any of the methodology panels was encouraged to submit their paper. Approximately 20 papers from that conference were submitted, and generated heavy traffic. [Note: One change to the structure of *Polmeth* occured in 1995 when the archive was switched from being 'conference-based'; now all papers are just uploaded into the annual 'working-papers' directory. A search-mechanism is currently being worked on to allow for searches by: author, keywords, and individual conferences.]

Given that there are roughly 550 dues-paying political methodologists, these are truly amazing usage statistics. These statistics indicate that each dues-paying methodologist could have downloaded almost 4 papers during this period; to slice the numbers differently, in January at least 20% of methodologists might have downloaded a paper from *Polmeth*. (Almost all of these downloads are papers; however, these figures include 104 downloads of back-issues of *The Political Methodologist*).

The papers which people are downloading off of *Polmeth* are also interesting to examine. The major categories of papers are those from the past two Summer Methodology Conferences (1994 and 1995), from the American Political Science Association Annual Meetings in 1995, and two directories containing working papers not necessarily associated with particular meetings (papers uploaded in 1995 and 1996). The access statistics for these categories of papers are:

- Summer 1994 papers: 68 downloads
- Summer 1995 papers: 84 downloads
- APSA 1995 papers: 280 downloads
- Working Papers 1995: 789 downloads
- Working Papers 1996: 731 downloads

These statistics should eliminate any idea that people might have that *Polmeth* isn't a perfect place for your papers for distribution!

Within the past year, a number of new and exciting things have been down with *Polmeth*. For example, the entire program for the 1995 APSA meetings was placed on *Polmeth* before the 1995 meetings in New York. An amazing number of people accessed the program from the Web: since the 1995 APSA program was placed on *Polmeth* there have been 39,406 "hits". Obviously, the number of people who have accessed the 1995 APSA program vastly exceeds the number of practicing political scientists in the nation! There is a good amount of information from the 1995 Summer Methods Conference on *Polmeth*, and 1164 people have "hit" the general information page about the 1995 Conference, 281 have "hit" the program from that conference, 830 have "hit" the listing of papers presented, and 120 have "hit" the list of names of meeting participants. The preliminary program for the 1996 Political Methodology meeting was recently made available on the archive.

Now *Polmeth* seems a great success, and it has been integrated into the daily research and professional activities of many people. There currently is a great deal of meeting-related information on *Polmeth*, there are links to most relevant Web-based data archives, and there are links to the Web sites of most of the major statistical software companies. *Polmeth* today is about more than just the distribution of working papers!

Associated with the *Polmeth* web site is the *Polmeth* discussion group. With roots in the mail reflector we used in 1994, today we now have a full-fledged email "list server" devoted to the discussion of political methodology.

There are approximately 525 subscribers to the "Polmeth" discussion group, and topics range from the use of statistical software packages, to specific questions people have about statistical techniques, to discussions of the teaching of methodology to our students. Whenever a paper is posted to *Polmeth*, a notice about the paper's availability is sent to this discussion group, which means that by placing your paper in the *Polmeth* paper archive, you can reach at least 525 people![1]

The Future of *Polmeth*

There is little doubt that *Polmeth*, as one of the activities of the Political Methodology Section, is having a major impact of our research and professional activities. With the continued evolution of the Web, and with increasing numbers of people connected to the Web in the office, the classroom, at home, and on the road, it is clear that we have to maintain the momentum we currently have achieved with *Polmeth*, to insure that it continues to be a valuable resource for political methodologists and the larger political science community.

What can we do to achieve this? First and foremost, **we currently have a valuable resource which can be utilized more completely.** While there are a good number of papers available from *Polmeth*, and while people are clearly downloading these papers for their use, we only have a small fraction of the papers which are written each year which are of interest to political methodologists specifically, and political scientists more generally. We need to see more of these papers posted to *Polmeth*, since by facilitating the rapid flow of information, new ideas and new findings, we will only make our profession more rigorous and more scientific.

Second, **if you know of a resource which should be added to *Polmeth*, let us know!** While there are lots of useful resources and links off of *Polmeth*, new resources come "on-line" every day on the Web. If there is a resource you use, let us know so that we can link *Polmeth* to that resource.

Third, and less immediate, **we need to think more about the types of downloadable resources which ought to be available on** *Polmeth*. Today many professors put their syllabi and other course materials on their personal Web pages — why not use *Polmeth* as a profession-wide archive of useful syllabi and course materials? Other than the common use of syllabi, it would be productive to see people provide access to the problem sets they use to teach methodology, and the trial datasets and assignments. Also, *Polmeth* could be used as a resource for the distribution of specialized computer code or software. Currently, there are limited software and computer code resources in our profession, and *Polmeth* could be expanded to cover these resources as well.

I'll conclude this historical overview with just two points. One is that *Polmeth* is available for the use of the political methodology and political science communities — please use it! The other is that the section is committed to continuing the development of *Polmeth*, but that requires your help. If you have ideas or suggestions, please send them to myself (rma@crunch.caltech.edu), directly to Jonathan Nagler (nagler@wizard.ucr.edu), or to the entire polmeth mailing list (polmeth@wizard.ucr.edu).

[1] Editor's note: At least some of whom may read and cite your work!

Review of *Basic Econometrics*, 3rd ed. Damodar Gujarati (McGraw Hill, 1995, 838 pages).

Annette Steinacker
University of Miami
asteinacke@sba01.msmail.miami.edu

Gujarati's *Basic Econometrics* is a good introductory textbook for students with little math or statistics background. I have assigned it as the text for the second semester in a required quantitative methods sequence for Master students in public affairs and in public policy (at Columbia and Georgetown). It could also be used for a required basic regression class in a political science Ph.D. program, where there are similar problems of diversity in levels of student interest and mathematical background.

One major advantage of *Basic Econometrics* is that it does not require an understanding of matrix algebra or calculus. Students just being introduced to statistics do not also have to struggle with learning new notation and mathematical concepts. Proofs and derivations are in appendices. There is an optional chapter that repeats the development of the multivariate model using matrix algebra. Gujarati also provides fairly substantial written explanations of the ideas as well as presenting the equations. Students new to reading math often find this helpful. There is no coverage of probability or univariate statistics. Another text, such as Wonnacott and Wonnacott's *Introduction to Statistics for Business and Economics*, is needed for this material.

All of the examples in Gujarati are from economics, primarily production and demand functions. This is less of a problem for public policy students than it may be for political science students since policy students also take economics courses. The problem sets following individual chapters are of limited use – computer applications with real data tend to be of more value to students. There is a companion text, *Basic Econometrics: A Computer Handbook Using Shazam*, that provides computer-based problems. Unfortunately this is only available for the Shazam program; a similar volume based on a more widely used computer program such as SAS would be helpful.

The book is adequate in its presentation of the basic linear regression model; the strength of the text is in how it covers violations of the model assumptions, and introduces several more advanced topics. There are two major and two minor limitations in the treatment of the general linear model. First, derivation of the OLS estimators is poor. The logic of minimizing the sum of the squared residuals as a criteria for choice of the sample regression function is well developed, but then the OLS estimators themselves appear with only a sentence or two saying they were found by differentiating an equation. The actual derivation in the appendix is still very short. Most instructors will want to expand on this – either a more detailed mathematical derivation or an intuitive explanation behind the process of finding the minimum value of an equation.

Second, the motivation for presenting the assumptions underlying the model and their impacts on the OLS estimators is poorly developed. Gujarati simply states a set of ten assumptions in the introduction to the bivariate model, then asserts at the end of the chapter that if the assumptions hold the regression estimators will meet the conditions of the Gauss-Markov Theorem and be the best, linear, unbiased estimators. Students must read about fifty pages before there is any indication why they should care about these assumptions, and then the idea that efficiency and unbiasedness are desirable properties of an estimator is downplayed. There also is no attempt at this point in the text to connect individual assumptions to their impact on either feature.

The two minor criticisms concern arrangement of material and inclusion of several marginal topics. A better treatment of summary statistics would be helpful. In the first chapter (of two) on extensions of the bivariate model to the multivariate case, Gujarati discusses R^2 and adjusted R^2 as measures of the goodness of fit of the regression model. He points out their limitations and that they are not universally considered important, but he does not develop any other options. The standard error of the estimate is mentioned in several places in the book but not really as an alternative measure. The author lists several other goodness of fit measures (such as Mallow's C_p) here, and again in the chapter on testing alternative model specifications, but he does not discuss them at all. This leaves students with no other alternative but focusing on the adjusted R^2 if they want to report results on the model overall.

Finally, Gujarati covers several tangential topics while developing the basic regression model. The chapter on hypothesis testing in the bivariate model (Chapter Five) offers nothing new; all students will have seen this material in more detail than given here, even if the tests did not involve regression coefficients. The author does not need to devote this much space to basic hypothesis testing. He also introduces the use of logarithm transformation of variables in an early chapter on the bivariate model (Chapter Six). Since in most courses the discussion of functional form is likely to come much later, at least after development of the multivariate model, this early placement is inconvenient. It allows Gujarati to accurately model economic functions that use log variables in his succeeding examples, so students will see them throughout the remainder of the book – making it difficult to skip the section and return later.

Despite the limitations discussed above, Gujarati does devote more of his book to development of the basic linear model than several other comparable texts (such as Pindyck and Rubinfeld) and handles it in an accessible form for most beginning students. But many instructors will want to elaborate on the mathematical relationships underlying the model and the reasons behind our concern with the assumptions, perhaps consolidating the number of assumptions presented as well.

The real strength of the book is the treatment of violations of the assumptions. In these chapters, Gujarati finally makes the connection between the impact of each assumption on the desirable features of an estimator. Each chapter follows the same format: the implications of violating an assumption are stated, methods of diagnosing if it was violated are discussed, and possible corrections are offered for the problem. There is a fairly extensive set of diagnostics for each violation with some discussion of the advantages and disadvantages of each. The variety of diagnostic options makes the book a useful reference tool, but beginning students can be overwhelmed with the choices. Guidance from the instructor on which methods are most commonly used is helpful. Also for some of the more sophisticated diagnostics, such as the condition index, Gujarati cannot explain the logic behind their use because of the mathematics involved. So either students will ignore them or potentially misuse them in rote applications. The value of either choice seems to suggest not including those diagnostic tools.

Model specification is covered reasonably well for an introductory book. The impacts of omitted and irrelevant variables, as well as measurement error, are treated in one chapter. The coverage is generally good, but the author does leave the impression that we can determine variables are irrelevant, and therefore drop them from the model, if they fail to meet a standard 95% significance test. While he does caution against adding variables based on their high t-values through use of stepwise regression, and stresses the importance of relying on theory to build models, he does not argue against dropping variables that fail standard hypothesis tests. A greater focus on the trade-off between biased coefficients from incorrectly omitting a variable and decreased efficiency from including irrelevant variables, possibly within the context of the mean squared error criteria, might be helpful.

A new chapter in the third edition covers choice between models, rather than simply the choice of variables for a model. The chapter is really too short to deal with these specification issues, but if the instructor wants to cover the subject the text provides a brief foundation. Two main topics are presented: Leamer's approach to specification issues, and choice between non-nested models. The introduction to Leamer's approach to different types of specification searches is really too brief to be comprehensible for students. Gujarati uses on an example from Leamer to illustrate how different model specifications are used depending on the question the researcher wants to answer, even when all questions involve estimation of the same demand function. However, each hypothesis and its specification is presented in one brief paragraph. Even some of the hypotheses are not clear, let alone the relation between the goal of the specification search and the hypothesis test being conducted. Leamer's original presentation would be at least as understandable to students. The discussion of choice among non-nested models primarily illustrates the potential for conflicting or ambivalent findings on which model is "better."

Gujarati does a very good job of introducing several advanced regression topics, primarily dichotomous dependent variables, simultaneous equations, and time series. The strength of these chapters is the clear explanation of the nature of the problems and the limitations of using simple OLS estimation when they exist. Students can get an understanding of the problem, the intuition of what the estimation technique needs to do to correct it, and references to more extended treatment of the topic in advanced texts. There is not enough detail in this introductory text for students to actually apply the appropriate estimation techniques (MLE is not covered for example) but they will be warned away from simplistic approaches that would be incorrect in the face of these problems.

These limitations are particularly true for the treatment of dichotomous dependent variables (one chapter) and time series analysis (two chapters, new to the third edition). They are less true of the information on simultaneous equations. These three chapters are an excellent introduction to the material. Gujarati gives one of the clearest explanations to be found in a basic text of the identification problem, the rank and order conditions, and the estimation processes of 2SLS and instrumental variables. Of the advanced topics, this one is handled most extensively and should require the least additional information from other sources. Based only on this material, students would be able to handle some basic research using simultaneous equations. One caveat is that after an entire book where matrix algebra is not used, Gujarati does need to bring in the idea of a determinant when defining the order condition for identification.

These sections on advanced topics also make *Basic Econometrics* a useful reference book for the average political science practitioner. They assume no initial knowledge of the topic, they are easier to read than sections from other texts because of the absence of matrix notation, and references to more advanced treatments are provided in the footnotes. This is a good starting point to understand if the topic is one you need to pursue in more depth, and it will provide a foundation for understanding the more technical material in advanced texts. These sections would also make the text a good base for a third semester course in advanced regression, with supplemental material for more detailed treatment of topics selected by the instructor.

Annette Steinacker received her PhD from University of Rochester in Fall, 1995. She is an Assistant Professor in the department of political science at the University of Miami. Her dissertation tested the ability of a bargaining game incorporating intercity competition, the mayor's political risk, and the city's need for economic growth to explain city decisions to grant property tax abatements to new businesses. Her primary research interest is urban economic development.

Review of *Essentials of Econometrics*, Damodar Gujarati (McGraw Hill, 1992, 466 pages).

Kevin Smith
University of Nebraska, Lincoln
kbs@unlinfo.unl.edu

One way to think of *Essentials of Econometrics* is as *Basic Econometrics-lite.* The two books have much in common, and the overlap is particularly noticeable in the introductory treatments of the general linear model.

However, it would be a gross disservice to view *Essentials* as simply a watered-down version of *Basics.* The books are aimed at two different audiences, and for a first year, graduate-level statistics course, *Essentials of Econometrics* may be the better choice. Indeed, this is one of the more user-friendly introductions to the essential foundations of social science methodology that I have come across. This text is currently being used in my department for a graduate level methodology course and has, with few exceptions, been well received. As someone who is in the process of searching for a text book as the basis for an introductory graduate-level methods course, I highly recommend *Essentials.*

In contrast to *Basics*, which begins with single-equation regression models, *Essentials* devotes almost a third of its pages to thorough review of probability and the foundations of statistical inference. This knowledge–critical to a fundamental understanding of the general linear model–is assumed rather than explicated in *Basics* (although an appendix does cover some basic statistical concepts). The coverage in *Essentials* is impressive, both for its comprehensiveness and presentation. The presentation goes beyond a cookbook description, but Gujararti resists bombarding his readers with lengthy mathematical proofs (for those interested they are mostly available in *Basics*).

The second third of *Essentials* is devoted to a thorough explication of the linear regression model, including discussions of functional forms and dummy variables. This is presented as a natural progression from the probability-oriented first third of the book. This section covers five chapters (about 170 pages) and would probably provide the primary focus for a first year regression course. All the appropriate basics are covered in a thorough and easily accessible manner.

The final third of the book covers how to diagnose and correct for the common problems encountered in regression analysis. The latter has particularly good coverage of multicollinearity, heteroscedasticity, autocorrelation and model specification.

Compared to *Basics*, *Essentials* is thus better suited to an audience that has a limited statistical background. The book provides a lucid introduction to the meat and potatoes methodology in the social sciences in a logical and systematic fashion. The book requires no matrix algebra and only a passing acquaintance with calculus.

As an introductory text to this"meat and potatoes methodology" *Essentials* ranks as one of the best. The text is user friendly, and the examples well integrated and illuminating. The problems at the end of each chapter should provide useful assignment material. Although I suspect some students will not appreciate the emphasis a number of the problems give to equation solving, but there is no doubt these exercises are properly aimed at providing a good intuitive understanding of the material. The only complaint in this regard is that the examples are heavily tilted towards economics. While this is understandable, some students may find themselves wishing for a little more that is immediately relevant to their chosen discipline.

The emphasis on providing a thorough grounding in the fundamentals means, of course, that *Essentials* does not compare well to *Basics* on more advanced topics. There is, for example, little on time series econometrics and nothing simultaneous equations models. The treatment of logit and probit techniques in *Essentials* is also fairly rudimentary in comparison with *Basics*. These topics are collapsed into a final "selected topics" chapter. *Basics* is thus much more likely to serve as a useful reference for the practitioner.

However, this should not be construed as a criticism of *Essentials*. While *Basics* is more likely to be taken from my shelf as a reference, I would be hesitant to assign it as a first-year, graduate level stats text. *Basics* begins with the assumption of a statistical background that, rightly or wrongly, a large number of first year political science graduate students do not have.

While I give *Essentials* high marks as an introductory text, I do have some reservations. I think a brief introduction to the matrix approach to regression, at least in an appendix, would be useful. In Chapter 9 of *Basics* Gujarati gives a succinct and compact introduction to matrix applications, and I think a condensed version of this would be helpful. The lack of any real attention to time series is also a little disappointing.

For the most part, though, such criticisms constitute minor nitpicking. Although compared to *Basics*, *Essentials* is, well, pretty basic, it deserves consideration as a strong candidate for a first regression course.

Review of *Regression With Graphics: A Second Course in Applied Statistics*, Lawrence Hamilton (Brooks/Cole, 1992, 363 pages).

Kevin Smith
University of Nebraska, Lincoln
kbs@unlinfo.unl.edu

Regression With Graphics is one of those unabashedly cookbook texts that statisticians hate, but that graduate students in the social sciences can actually understand. As I place a high premium on the latter, I think Hamilton's text should find a useful home in a first year, graduate level regression course.

There's little emphasis here on formal proofs or long treatments of statistical theory. Instead the text focuses on providing students with an intuitive understanding of regression analysis – how to do it, what programs to do it on, how to interpret output, how to figure out when something is wrong, and how to fix it. The text thus places a premium on accessibility and, as the title implies, relies heavily on graphical presentations to achieve this. This proves to be a useful aid. The plots are well integrated with the text and should serve as useful heuristics to math-phobic social science students in a way that algebraic presentations cannot.

This book is best suited to those who are seeking to provide students with a practitioner's grounding in regression analysis. It is primarily designed to answer questions of what method to use where, and how to interpret the results. More technical expositions are sprinkled throughout the book, but these are separated from the main text by boxes. While they are there for the interested student, they are not necessary for understanding the fundamentals of the lesson. Matrix algebra and calculus are options for the interested reader, not requirements.

There is one important caveat to this text: the book has a fairly rudimentary introduction to pre-regression statistics. The first chapter covers measures of central tendency and dispersion, and introduces graphing and transformation techniques. Appendix 1 also covers the basics of population and sampling distributions. However, this book is best suited to students who already a reasonably solid grounding in the fundamentals of probability and hypothesis testing. Chapter 1 and Appendix 1 cover the basics, but these are much better suited for review purposes rather than primary teaching texts. The heart of the book consists of chapters 2-6, which cover the classical regression model and some of its variants. This is where the book really shines. Hamilton introduces residual analysis, power transformations, ANOVA and other techniques right from the beginning in a clear and accessible fashion. Chapter 4 provides one of the more succinct and comprehensive treatments of performing basic regression diagnostics I have seen in an introductory text. This point is worth underlining, as diagnostics often are given short shrift in such textbooks. While econometric theory is in short supply here, practical information is not.

Hamilton uses examples from a broad range of the social sciences to support the text, although most of these tend to have an environmental theme. These are not always relevant to political science (perhaps Cassava production in Central Africa is of burning interest to some, but not to the students I teach). Still, it is refreshing to have a first regression text that is not so heavily tilted towards economics examples.

The problems at the end of the each chapter provide good learning exercises, and are for the most part aimed at developing interpretive skills, rather than math skills.

The last two chapters of the book cover more advanced topics: logit and factor analysis. The logit chapter is well done, and offers a particularly good exposition on how to interpret logit results. Teaching students how to get substantive interpretations of this technique is an exercise that often results in more confusion than illumination. Hamilton manages to avoid this all too common pitfall.

The final chapter covers principal components and factor analysis. I found this chapter to be a little out of place. Not that there is anything substantively wrong with it; it would serve as a good introduction to the techniques for any practitioner. But I question why factor analysis gets a whole chapter, while other topics of interest to the practitioner-oriented audience Hamilton obviously seeks, are virtually ignored. For example, there is nothing on time series or simultaneous equations, tools that practitioners are just as likely to need as factor analysis.

Overall, I rate this as a high quality text. While some more math oriented instructors may bemoan the general lack of technical sophistication, I, to the contrary, applaud it. The biggest problem I encounter in teaching students statistics is fear, pure and simple. Hamilton is friendly to the math-phobic, while providing enough of a technical explanation to satisfy the curious (the appendix on the matrix algebra approach to regression, for example, is quite well done).

However, there are some shortcomings with this text. The intricacies of model specification are not covered very well. I found the lack of any real attention to time series disappointing. And it is unlikely that this is a text that will serve as a primary reference for even marginally sophisticated quantitative political scientists. Although, in all fairness I will admit that reading the chapter on factor analysis offered me a concise refresher on a technique I had not used in some time.

The book does succeed in what it sets out to do. As Hamilton says in the preface, "I place more emphasis on practical issues and troubleshooting than on statistical theory." And he is as good as his word.

Reasons to choose this textbook are its accessibility to students, its superior treatment of regression diagnostics and its unabashed focus on the practicalities of regression analysis. It is not suitable (nor designed) for those with no statistical background, nor as a methodological reference that no political scientist should be without.

Kevin B. Smith is an assistant professor of political science at the University of Nebraska Lincoln. His research interests are state-level policy in the substantive areas of education, crime and the environment. His work has appeared in the Journal of Politics, Political Research Quarterly, American Politics Quarterly and Public Administration Review.

Review of *Introduction to the Theory and Practice of Econometrics*, 1st edition Judge, et-al (Wiley, 1982, 839 pages).

George Krause
University of South Carolina
krauseg@garnet.cla.sc.edu

The text informally referred to as *"Baby" Judge et al.* provides a comprehensive overview of basic econometric theory.[1] Although I have not used this text in any of the graduate courses that I have taught at the University of South Carolina, I have used it over the years in my own research dating back to my days as a graduate student in the Department of Economics at West Virginia University. Due to its highly technical treatment of econometrics, those students lacking a solid background in differential calculus and matrix algebra will not find this a friendly book to use, even as a reference source. Moreover, students with technical degrees requiring a background in mathematics but not statistics, will not find much in terms of statistical theory beyond the conventional normal (probability) distribution. Univariate statistics are covered insofar that it pertains to the normal probability distribution (e.g., mean, variance, covariance, standard deviation) . Therefore, it is recommended that students using this book should have a familiarity with probability theory as well as topics related to inference before using it. In most graduate programs in economics, econometrics courses using such a text have a prerequisite of a course in mathematical statistics at the level of Hogg and Tanis's (1983) *Probability and Statistical Inference.*

The assignments at the end of each chapter contain a useful blend of algebraic and numerical exercises that provide advanced students a thorough knowledge of the "theory" and "practice" underlying econometric practice. These "practice" exercises primarily focus on issues of estimation such as alternative forms of statistical estimation for the same given set of data. In doing so, the authors wish to emphasize the importance of estimation choices made by researchers and its subsequent impact on the empirical results that are obtained. Unfortunately, there are no political science examples contained in this text. Moreover, most of these numerical applications-oriented exercises are presented in an arid manner by listing data without any substantive or theoretical context.

The general linear regression model is given ample attention in chapters six and seven (a combined 115 pages). The former chapter introduces students to this statistical model in matrix algebra form. The overview description of the general linear model (pages 111-117) is laid out in a clear, concise manner. Those students with an elementary understanding of matrix algebra will find this presentation to be intuitively appealing. The discussions on point estimation via the least squares method are rather technical, yet quite insightful. Especially useful is the brief section on the covariance matrix and its purpose in tying sampling variability to alternative estimators (pp. 134-136). The presentation of the proofs to demonstrate that least squares estimator (LSE) is the best linear unbiased estimator is rather blurred. I feel that Kmenta's fine econometrics text, *The Elements of Econometrics*(1986), does a much better job of conveying this information.

Whereas chapter six deals with least squares statistical estimation of the general linear model, chapter seven examines the same model yet assumes that the error vector follows a normal distribution and proceeds to employ maximum likelihood estimation as a means to derive estimates of coefficient parameters and standard errors for a given model. Some attention is given to interval estimation within the context of this model. This primarily pertains to the variability of estimated coefficients from their true values. Likelihood ratio tests involving both linear restrictions on the model's parameters and for general specification purposes are presented in a rigorous, yet lucid manner.

Discrete and the more general family of limited dependent variable regression models are given short shrift in this text (only 14 pages - chapter 18). The presentation of the logit and probit model is aptly expressed in matrix algebra terms. This discussion also covers the derivative approach used to generate marginal probabilities from the coefficient estimates employed to assess the relative impacts of independent variables on the discrete dependent variable. The only type of limited dependent variable model presented by Judge et al., is a surface treatment of the conventional Tobit specification (pp. 526 - 528). Other more advanced limited dependent variable models such as special truncated models, multinomial and ordered probit and logit models, Heckit two-stage estimation, and the like are not covered here. I believe that a more preferred advanced general text covering these issues would be William Greene's *Econometric Analysis* (1993) or Jan Kmenta's *Elements of Econometrics* (1986).

One of the major strengths of this book is the discussion of model identification in a systems framework, and its subsequent estimation in chapters 12 and 13. In the former chapter, the authors' present the material in a understandable manner by first discussing the set-up of the simultaneous equation framework before delving into unique

[1] A newer version of this text is available: the 2nd edition was published in 1988. - *Editor*

problems that arise (i.e, least squares bias; estimating reduced form parameters; and converting reduced form parameters into structural ones). The authors' clearly spell out the general types of a priori restrictions that can be made when modeling a simultaneous equation system (p. 350). Model identification is also covered in moderate detail for a general econometrics text (pp. 356-361). This chapter ends with a conventional supply-demand application of simultaneous equation model using the indirect least squares method of estimation. The next chapter provides an in-depth, treatment of the various means of estimating systems of equations and their implications for statistical inference. This comprehensive treatment includes presentation of indirect least squares (ILS), generalized least squares (GLS), two-stage least squares (2SLS), three-stage least squares (3SLS), limited maximum likelihood (LIML), and full information maximum likelihood (FIML) as alternative estimation approaches to simultaneous equation models. Chapter 13 ends with a nice, brief application a simultaneous model of supply and demand that is used for forecasting purposes. This will be of interest as a reference source for those who wish to forecast future outcomes or perform counterfactual analysis within a simultaneous equation framework. While I commend Judge et al. for its thorough and clear treatment of model identification and simultaneous equation models, it may be too technical for students who do not possess a strong background in mathematics. If this is the case, I recommend *Basic Econometrics* (1995) by Damodar Gujarati as a less technical text that does a fine job of addressing these econometric issues on an elementary level.

Judge et al.'s treatment of time series econometric models covers approximately 75 pages; however, it does not cover advanced topics such as COMFAC (common factor restrictions) models, unit roots, cointegration, error correction models, frequency domain/spectral methods, and vector autoregression analysis. Where it does succeed is in its thorough treatment of ARIMA (Box-Jenkins) models. Those wishing to find a solid presentation of ARIMA models in a general econometrics book should examine this text. The authors' progress the reader from the basics of time series analysis (e.g., autocovariance and autocorrelation functions and lag operators) through identifying and estimating autoregressive (AR) processes, moving average (MA) processes, and then joint ARIMA processes.

The chapter on bivariate time series (i.e., multiple time series analysis) is sketchy and quite dated. The chapter on distributed lag models (chapter 27) provides a fundamentally sound presentation of the basic formulations for finite (Almon) and infinite (geometric) lag structures. Unfortunately, more recent developments in distributed lag structures such as Shiller's which imposes stochastic lag restrictions and the gamma lag which has an inverted-V functional form are not covered here. In general, more preferable coverage of these time series analysis topics can be found in other general econometric texts such as the highly accessible *Econometric Models and Economic Forecasts* (1991) by Robert S. Pindyck and Daniel L. Rubinfeld, or Kmenta's *Elements of Econometrics.*

The discussion of model specification issues found in chapter 22 is underdeveloped. Although issues centered on nested model specifications are examined largely by comparing various ad hoc variable selection rules such as r-square , adjusted r-square, Mallows Cp , and Amemiya's Prediction Criterion , more powerful "information" criteria (e.g., Akaike criterion, Schwarz criterion, Hannan-Quinn criterion) that can be used to specify empirical models are not examined in this text but are covered in *"Big" Judge et al. (1985)*; – i.e., Judge et al.; *The Theory and Practice of Econometrics.* Moreover, the topic of non-nested (encompassing) model specification tests that are of interest to researchers attempting to test competing theories against one another are not presented here. In addition, general (omnibus) model specification tests such as Ramsey's RESET test for ordinary least squares (OLS) regression models are not explored. The chapters on regression diagnostics focus on the detection and possible solutions to problems relating to heteroskedasticity, autocorrelation, and multicollinearity. Although this treatment is fundamentally sound, it is extremely dated. For instance, alternative robust covariance matrix estimation approaches that have come into vogue in recent years, such as White's method used to address heteroskedasticity of some unknown form and the Newey-West method which serves as a generalization of White's method used to deal with problems arising due to serial correlation, are not covered in this text. A much better general econometrics text for addressing the topic of regression diagnostics and subsequent solutions is William Greene's *Econometric Analysis.*

A unique component to this general text is its coverage of Bayesian estimation and inference (chapter 8). This material presents how both "informative" and "noninformative" priors can be used to facilitate statistical estimation when the researcher has a priori information about the key variables of interest and believes that a parameter can be more accurately summarized in terms of its posterior distribution. Interval estimation is discussed within a Bayesian context by focusing on the highest posterior density (HPD) interval, which is analogous to the conventional confidence interval employed in statistical inference based on sampling theory. Point estimation uses the mean square error matrix as a criterion to compare the posterior mean for the beta matrix to its analog estimated via maximum likelihood estimation. There is also a sub-section on hypothesis testing within a Bayesian context that may be useful for those interested in this approach to statistical inference.

In sum, this book's strengths lie in its application of matrix algebra to both present basic econometric theory and conduct empirical analysis in a rigorous, comprehensive manner. Students using this book will acquire a sound fundamental treatment of the main topics in econometrics. Due to the advanced nature of the presentation and lack of political science applications, most graduate students in political science will not find this book to be a user-friendly text. A course using this book as its primary text requires a proficient knowledge of matrix algebra and differential calculus. Moreover, since this is a general graduate level text that was published 14 years ago, it does not do an adequate job of covering recent developments in the area of econometric methodology. Despite these limitations, this text has significant utility as a reference text for those interested in either elementary Bayesian estimation and inference or an advanced (technical) treatment of elementary econometrics.

References

Greene, William H. 1993. *Econometric Analysis*, 2nd ed. New York: Macmillan.

Gujarati, Damodar N. 1995. *Basic Econometrics*, 3rd ed. New York: McGraw-Hill, Inc.

Hogg, Robert V., and Elliot A. Tanis. 1983. *Probability and Statistical Inference.* Second edition. New York: Macmillan.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee. 1982. *Introduction to the Theory and Practice of Econometrics.* New York: Wiley and Sons.

Judge, George G., William E. Griffiths, R. Carter Hill, and Tsoung -Chao Lee. 1985. *The Theory and Practice of Econometrics.* New York: Wiley and Sons.

Kmenta, Jan. 1986. *Elements of Econometrics.* 2nd edition. New York: Macmillan.

Pindyck, Robert S., and Daniel L. Rubinfeld. 1991. *Econometric Models and Economic Forecasts.* Third edition. New York: McGraw-Hill.

Review of *Elements of Econometrics*, 2nd ed. Jan Kmenta (MacMillan, 1986, 786 pages).

Mohan Penubarti
University of California, Los Angeles
mohan@ucla.edu

A large number of quantitatively oriented graduate students (and researchers) in our field do not have even rudimentary knowledge of probability or distribution theory. Worse yet, a number of students seem to have the impression that Least Squares estimation and statistical analysis are isomorphic. A large part of the problem is probably premature exit from the statistical sequence offered by most political science graduate programs. As a result, students learn Least Squares in the introductory course in regression analysis but do not follow on to learn more. But an equally large part of the problem is the way in which we introduce statistical inference to students. Since the sequence of topics in the textbook typically shapes the syllabus, selection of the appropriate textbook becomes especially important. Fortunately, the choice among introductory texts in econometrics has grown quite considerably. While a number of them are good books for reference, I believe that few compare favorably to Jan Kmenta's *Elements of Econometrics* as a useful text from which to teach political science graduate students.

The typical introductory econometrics textbook begins in one of two ways: (1) classical linear regression model, with all the prerequisites (review of basic algebra, matrix algebra, probability theory, distribution theory) crammed into an appendix (Gujarati's *Basic Econometrics*, a popular choice in econometrics courses, is a good example); or (2) all of these prerequisites are stuffed into an introductory chapter (see Maddala's *Introduction to Econometrics*, for example). These books then proceed with Least Squares, assumptions, violations, potential corrections, multiple regression, and on to advanced topics. In the process, political science graduate students learn very little about probability and distribution theory and thus lack a strong foundation to continue on to advanced courses or self-learn advanced topics. While these books are useful in specific contexts, they are not particularly good textbooks from which to teach principles of statistical inference in a systematic way.

In this respect, *Elements of Econometrics* has few peers. Kmenta begins with a discussion of sampling distribution and shows how one can derive them experimentally using a number of interesting examples. I translate these examples into political science settings and have students run Monte Carlo experiments using computer software to drive home concepts such as a population parameter, estimator, estimate, sampling distribution, as well as some rudimentary properties of an estimator. It is especially useful for them to see through computer experiments that the expected value of the sampling distribution is (close to) the population parameter. Having established that we can in fact develop reasonable guesses of unobserved population parameters, Kmenta provides the motivation for the theoretical derivation of sampling distributions. He then launches into probability theory, probability distributions, estimation, and hypothesis testing. Almost 200 pages are taken up developing basic statistical inference before the classical linear regression model is even considered.

When the regression model is eventually introduced, students have a much broader grasp of the link between $\hat{\beta}_1$ and $\hat{\beta}_2$ of the sample regression function and β_1 and β_2 of the population regression function and realize that Least Squares is just one estimator (or formula) for generating guesses of β_1 and β_2. Consequently, they are less fixated by Least Squares but also have a better foundation in statistical theory to explore more advanced topics. Kmenta spends about 150 pages on regression analysis, covering the usual sequence of assumptions, violations, and corrections. He then considers in some depth a topic which is extremely important in empirical research but gets little attention in most statistical texts: estimation with deficient data. The particular deficiencies he deals with are: (1) errors of measurement; (2) grouped data (especially problematic for research using survey data); and (3) missing observations.

The rest of *Elements of Econometrics* contains topics that are probably outside the scope of a first year course in regression analysis but which make the book useful for a more advanced course or as a reference text: models with binary regressors, restricted coefficients, nonlinear functional forms, distributed lags, qualitative dependent variables, limited dependent variables, varying coefficients, unobserved variables; the Generalized Linear Regression model; and simultaneous equations. There is no treatment of time-series analysis on somewhat dogmatic grounds. In Kmenta's words, "these models have no economic content and their use for modeling exogenous variables is not theoretically justified."

In my *Introduction to Probability and Statistical Theory* course at UCLA, I recommend Goldberger's *A Course in Econometrics* as an alternative text and my lecture notes are largely based on that text. However, the level of technical difficulty in Goldberger is too high for many students in the course. In addition, Goldberger contains almost no numerical examples. In contrast, *Elements of Econometrics* assumes only basic mathematics, very little calculus, and does not introduce matrix notation until multiple regression on page 392. Further, it has a number of examples based on real world data along with exercise problems of varying difficulty. Like other econometrics texts, alas, there are few or no political science examples. If there is a drawback to *Elements of Econometrics*, it is a failure to tie the examples to computer assignments (a criteria on which Gujarati's *Basic Econometrics* does quite well). A better integration of econometric principles and techniques with computer assignments using actual data is likely to bridge the gap between coursework and practical research for many graduate students.

Overall, if you are looking for a well written and carefully thought out introduction to statistical inference that is appropriate for political science graduate students, I highly recommend Jan Kmenta's *Elements of Econometrics.*

Mohan Penubarti (mohan@ucla.edu) is an assistant professor at UCLA. His research interests focus on international trade and competition issues. His most recent publications have appeared in Journal of International Economics and International Interactions.

Review of *Econometric Models and Economic Forecasts*, Robert S. Pindyck and Daniel L. Rubinfeld (McGraw Hill, 1976, 576 pages), and *Applied Linear Regression*, Sanford Weisberg (Wiley, 1980, 283 pages).

Andrew Skalaban
University of California, Davis
ajskalaban@ucdavis.edu

Econometric Models and Economic Forecasts by Pindyck and Rubinfeld (*P and R*) is an applied introductory econometrics text. As an applied text and reference it is excellent, and I recommend it highly. For those interested in complete expositions of the latest theoretical developments in the field, you would do well to look elsewhere.

I use *P and R* in the second course of a required two quarter methods sequence at U.C. Davis. I try and cover most of the material in the book's first 11 chapters. This covers everything from a quick review of basic statistics to simple simultaneous equation estimation. The last third of the book deals mostly with univariate time-series models, and while some of our students may find this useful later as a reference, it is really beyond the scope of our second methods course. I teach the book in conjunction with Hamilton's *Statistics* with STATA. Student enthusiasm for the book is, to be generous, restrained. However, I do not take this as being a bad sign. Like all econometric texts, *P and R* is expensive; and I'm not sure students ever get over the initial shock. After all, they could probably buy 10 or more copies of *The Prince* for what *P and R* costs. Furthermore, while the quality of our recent incoming graduate students has been quite high overall, the typical beginning Davis graduate student does not have a great deal of background in math or statistics. For these, and probably most political science graduate students I would think, *P and R* is a challenge; but it is a challenge that can be met by most of them.

P and R does not require calculus or matrix algebra. Derivations using calculus or matrix algebra appear in appendices at the end of every chapter. This is a good mix. Students without much math background can get through the main body of each chapter without getting thoroughly lost, while the more mathematically inclined do not get short-changed. The book does a very good job of explaining and showing the derivations of the basics–least squares estimation, generalized least squares, maximum likelihood estimation of the logit model, etc. It spends about six chapters on OLS estimation, interpretation, diagnostics, and the extension to GLS. *P and R*'s chapter on models with qualitative dependent variables is very good. Its juxaposition of probit and logit maximum likelihood models with a variation on weighted least-squares is an excellent pedagogical choice.

Econometric Models and Economic Forecasts also has its weaknesses. It provides a solid introduction to pooled cross-sectional time-series models, but its coverage could be more comprehensive. For example, it does not discuss the differences between models that are cross-sectionally or time-serially dominant. *P and R* also does not cover count models. While this probably goes beyond what I could reasonably cover in a ten week course, I would prefer a chapter devoted to these models to which the students could later refer. These are somewhat serious deficiencies; however, for me, they are more than outweighed by *P and R*'s strength in applications.

The reason that I have gone back to using *P and R* is for its examples. Each chapter usually contains several good examples of a real applications. Even though these examples are overwhelmingly drawn from economics research, they usually have a substantive policy focus, e.g. housing, that makes them accessible to virtually all social science students. *P and R* integrates theory and applications better than any other econometrics text that I have seen. The book also reprints an unusually large number of interesting data sets which can be used in exercises.

Econometric Models and Economics Forecasts' approach to the regression model is pretty standard econometrics. A nice introductory discussion of consistency, unbiasedness, and efficiency form a base out of which modelling choices are developed. Problems associated with measurement error and the practice of model specification receive, by the authors' own admission, less attention. *P and R* has improved its treatment of OLS regression diagnostics over previous editions, but the discussion is not up to the quality of other topics covered in the book. Since I consider regression diagnostics an absolutely essential topic, my lectures often diverge from *P and R* and concentrate more heavily on the treatment of regression diagnostics and model building found in Sanford Weisberg's *Applied Linear Regression*.

Weisberg is one of the leading statisticians in the field of regression diagnostics (especially outliers and influential cases). His text on the regression model, *Applied Linear Regression*, has a slightly different flavor than that found in most econometrics texts. Weisberg emphasizes the importance of checking assumptions and the use of that information in model building. Whereas *P and R* includes a number of tests for heteroskedasticity, e.g. whether there is constant variance across the range of values of an independent variable, Weisberg emphasizes graphical analysis of residuals to check for homoskedasticity and as an aid in determining functional form. He even goes as far to suggest that added variable plots be used as an aid in model specification and discusses procedures such as the Box and Cox method for deriving the functional form of a regression equation based on the data. Clearly this deviates from standard line of argument which holds that specifications flows directly, and only, from theory. As such, I find it provides a nice pluralist supplement to an econometrics text and is valuable in helping students with practical data analysis problems. Moreover, the subject of case influence is an exceedingly important one and Weisberg's approach is invaluable in helping researchers understand the usefulness of subjects like robust regression.

Applied Linear Regression does require some calculus and matrix algebra in many sections. It is practically impossible to do justice to the discussion of case influence without matrix algebra. As a general statistics text, it draws its illustrations and examples from a wide array of disciplines–from biology to public policy. This makes for an interesting read, however, it also makes the text somewhat less attractive for a graduate methods class in political science since the basic rationale for teaching our own discipline-centered courses is to tie the statistical applications to political research. Thus I find Weisberg's text an excellent supplement and/or reference.

Taken together, *Econometric Models and Economic Forecasts* and *Applied Linear Regression* allow me to do what I believe I need to do in our second-quarter required methods class. I can cover an introduction to the regression model rigorously without needlessly alienating students who come into our program with little math/stats background or who are skeptical of the usefulness of statistics for their research. I think that I can even claim a few converts among this latter group. The basic grounding of the course in *P and R*'s econometric approach helps students navigate their way into the contemporary work being done in political methodology which is increasingly drawing its focus from econometrics. Finally, integrating the material from Weisberg allows me to cover some topics I find essential for training students to use the regression model in an informed way.

Andrew Skalaban is currently an Assistant Professor of Political Science at U.C. Davis. His research focuses mainly on public opinion and political economy. Professor Skalaban's work has appeared in the American Journal of Political Science, Journal of Politics, Political Methodology, and other journals.

Review of *Introductory Statistics for Business and Economics*, 4th edition, Thomas H. Wonnacott and Ronald J. Wonnacott. (Wiley, 1990, 815 pages).

Janet M. Box-Steffensmeier
Ohio State University
jboxstef@magnus.acs.ohio-state.edu

Wonnacott and Wonnacott's *Introductory Statistics for Business and Economics* is an appropriate text for a course that is the first comprehensive introduction to statistics. I have used this book to teach the second quarter graduate class of the quarter sequence that all Ph.D. students at Ohio State are required to take, and it has worked well. The first class is primarily research design and uses a variety of texts, including King, Keohane, and Verba (1994). The third class focuses on the violations of assumptions and regression extensions, such as logit/probit, and usually Gujarati, 3rd ed. (1995), is used. The students taking the first and second courses, in contrast to those skipping to the third course or to our advanced methods courses, have very little mathematics or statistics background. Thus, we need to start at ground zero. Wonnacott and Wonnacott only assume high school algebra and so their text fits the needs of the students enrolled in this class well. Those students who know calculus are encouraged to study the optional appendices at the end of the book, which contains proofs, and some of this material is presented in class. No matrix algebra is required to understand the text, nor is matrix algebra offered as an option (matrix algebra is used extensively in the third course of the sequence at Ohio State University).

Wonnacott and Wonnacott cover probability and univariate statistics, which is one reason this text is applicable for the first course in a statistics sequence. Their text complements more advanced econometrics texts, such as Gujarati, because it covers probability and regression thoroughly, and then one can turn to Gujarati for more technical coverage of regression, violations of assumptions, and diagnostics (many of the more advanced texts do not cover probability). It is helpful for students to have overlapping presentations of statistical material, particularly of such a fundamental topic like regression, which can also be used to demonstrate related techniques such as correlation and analysis of variance. These texts work well together in a quarter system. I think that they would make a powerful combination on a semester system also. Wonnacott and Wonnacott's text is not likely to be as satisfactory for a second course in a two-semester statistics sequence (for example, there is no systematic discussion of the violations of assumptions) nor the best text for an undergraduate political science methodology course (there is probably a need for a slower paced introduction with a lot more political science examples). The first fifteen chapters of the text present basic classical statistics and include topics such as descriptive statistics, probability, sampling, point estimation and confidence intervals, hypothesis testing, ANOVA, regression, and some regression extensions. This ordering of topics has worked well. The last ten chapters include important topics such as nonparametric statistics, Bayesian inference, and time series. This text offers instructors great flexibility. Instructors can choose from these last ten chapters as time allows without extreme concern about the order. Also, some parts of the early chapters can be abbreviated or taught later with little interruption of continuity.

Some more advanced topics, such as simultaneous equations and time series analysis, are also included in Wonnacott and Wonnacott. These two chapters, as well as three others, are included in Wonnacott and Wonnacott's *Introductory Statistics for Business and Economics*, 4th ed., but not in their *Introductory Statistics*, 5th ed. Their presentation in these chapters will give students a rough idea of these topics only, not an in-depth examination. Given the rationale of being able to read more than one presentation of any particular topic, I would opt for having students buy the former rather than the latter. For the additional money, it is useful to have these extra chapters. The text is more of a teaching text than a reference text. There are not enough subtleties presented or enough in-depth analysis of, for example, regression diagnostics or simultaneous equations for the text to be useful as a reference for practicing political scientists.

One of the advantages of Wonnacott and Wonnacott's text is that it shows the logical relationships between what are often treated as separate topics or chapters in other texts, for example, equivalence of confidence intervals and hypothesis testing. That is, the logical connections are more explicit in Wonnacott and Wonnacott than other texts. They generally emphasize intuition before theory by presenting an example before the equations. They also give attention to interpretation and have a lot of practical examples and problems. There are some political science examples, although instructors would probably want to further supplement these in the lecture. There are more general social science examples and an index of examples and problems, which instructors considering this text may find useful. A big plus of Wonnacott and Wonnacott is that there are problems, problems, and more problems. Many students need extensive practice working problems so this is clearly an advantage. The answers to odd-numbered problems are in the back of the book. Providing answers to the numerous odd-numbered problems means that there are enough problems for students to work on their own to check their understanding and plenty of problems left to assign for evaluating students. A student workbook contains the complete solutions to all the odd-numbered problems as well as more than 250 review problems. Another add-on available to instructors is a computerized test bank.

Wonnacott and Wonnacott do provide some examples of computer output from Minitab but more of this is needed. Ideally, computer data sets would be available like those for Greene's *Econometric Analysis* (1993) or Dougherty's (1992) *Introduction to Econometrics*, the latter of which contains data and manuals for the programs used. Obviously with a first year required course, most instructors are going to want students using a computer package to conduct data analysis. Compared to other texts, Wonnacott and Wonnacott does not provide much linkage between their text and a computer package. Nor are there instructor add-ons such as overheads of tables or graphs, which can be very useful and convenient. The instructor's manual does contain solutions to all the exercises but no comments about the material is provided. Unfortunately, like many other statistics books, the sparse index leaves something to be desired. Finally, the arrangement of problems can be distracting because some problems are interspersed between the different sections of the chapter. This arrangement means that students often do the problems by concentrating only on the examples presented in the previous section rather than determining more generally how to solve a problem (the review problems at the end do help with this issue).

In sum, this text provides a solid introduction for the first course in statistics at the graduate level and can be used quite well with a more advanced econometrics text. The good balance maintained between intuition and theory and the numerous problems available are some of the major advantages.

References

Dougherty, Christopher. 1992. *Introduction to Econometrics.* New York: Oxford University Press.

Greene, William H. 1993. *Econometric Analysis*, 2nd ed. New York: Macmillan.

Gujarati, Damodar N. 1995. *Basic Econometrics*, 3rd ed. New York: McGraw-Hill, Inc.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry.* Princeton: Princeton University Press.

Janet M. Box-Steffensmeier is an assistant professor at Ohio State University. Her research interests center on American politics, specifically legislative politics and public opinion, and on methodology, specifically time series and duration analysis. Her work has appeared in the American Political Science Review, American Journal of Political Science, and Political Methodology.

Review of *Introduction to Econometrics*, Christopher Dougherty (Oxford University Press, 1992, 399 pages).

Christopher J.W. Zorn
Ohio State University
czorn@magnus.acs.ohio-state.edu

Like many texts used in first-year graduate programs in political science (e.g. Maddala, Gujarati), Christopher Dougherty's *Introduction to Econometrics* is intended for use in a year-long undergraduate course in econometrics for economics majors. Beginning with a review of rudimentary probability and sampling theory, Dougherty covers basic least-squares regression in the familiar assumptions-violations-remedies fashion. Other chapters discuss transformations of variables, use of dummy variables, dynamic processes, and simultaneous equations.

Unlike other texts, however, *Introduction to Econometrics* was specifically written to keep the level of mathematical background needed to a minimum, while still providing thorough coverage of the issues. The author endeavors to achieve these sometimes conflicting goals by a number of means. For example, there is no discussion of matrix algebra, nor is it used in or needed for understanding the text. Dougherty also eliminates summation notation in formulas, instead using the variance/covariance terminology familiar to readers of Hanushek and Jackson's *Statistical*

Methods for Social Scientists. With two exceptions, the text also does not require use of calculus; basic differential calculus is used in deriving expressions for α and β in regression equation, and an appendix of the review section uses integrals in discussing expected values and variance in continuous variables. Both of these examples are optional, and contained in separate sections which may be skipped with no real loss in continuity. Finally, the text does not discuss the general linear model or familiar qualitative-variable (e.g. probit/logit) models. The result is a text which, while covering much of the same ground as similar texts, does so in a much less technical manner.

The strengths of the book are several. The introductory review section is especially good, covering unbiasedness, efficiency and consistency as well as probability and sampling in a clear, concise 30 pages. The discussion of variable transformations is also useful; Dougherty's discussion of interpreting coefficients in log-linear models as elasticities is especially intuitive and comprehensible. The book has a good discussion of specification issues, both in terms of the standard specification-error discussion, and in the chapters on dummy variables, dynamic processes and simultaneous equations. In addition, the book comes well "packaged". Student editions contain a floppy diskette containing data for analysis in MicroTSP or Microfit, as well as manuals for these programs (the data are also provided as Lotus 1-2-3 spreadsheet files for instructors using other software). The Instructor's Manual has extensive comments covering the material presented, tables and graphs for use as overheads, and reviews of all exercises. Finally, a series of sample graphics are also included on an instructor's disk, for use with computer projection screens becoming more and more common in such courses. All computer materials are amply documented.

Yet despite these strengths, Dougherty's book is not the optimal choice for teaching econometrics in a political science context. As one would expect from a text aimed at economists, the book has no political science examples to speak of. Beyond that, however, most of the examples in economics are of especially limited interest to most political scientists. For example, demonstrations used throughout the text revolve around estimation of demand functions for various commodities. This is also true of the exercises at the end of each section, and the data provided to students with the text. Additionally, while many of the examples are classics in economics (e.g. Friedman's discussion of measurement error in estimating consumption functions), they often require a level of knowledge of economic theory above that of the traditional political science student. A smaller point is the book's lack of any discussion of univariate statistics or nonparametric measures of association, which are often covered in the opening weeks of introductory-level methods courses.

My sense is, however, that most instructors will see the biggest liability of Dougherty's *Introduction to Econometrics* in its lack of treatment of mathematical issues. In particular, the omission of matrix algebra limits its usefulness in any course sequence which goes on to discuss general linear models, maximum likelihood methods, or other more advanced topics. Because of these drawbacks, the usefulness of this text in graduate-level political methodology courses is restricted. The book might form a good basis for a less rigorous,"second-track" methods sequence, but there are others that would be even better, and in any case few departments have such separate arrangements.

Despite these limitations, however, there is much to like about the book. The author's presentation is clearer and more intuitive than most in this area; this strikes me as one positive result of the less-rigorous presentation style. This fact alone makes the book worth a look as a possible (instructor's) supplement to an assigned text; one might certainly assign a different text and teach out of this one. For the working political scientist, the book would be as good to have around for reference and review as most in its class, though here again the level of presentation will be seen as a boon to some and a bane to others.

Christopher J. W. Zorn is a doctoral candidate in Political Science at The Ohio State University. His research interests include judicial politics, American political institutions, and political methodology. He is currently writing his dissertation on federal government litigation strategies in the U.S. Supreme Court.

Syllabus for Research Design

Larry Bartels
Princeton University
lbartels@wws.princeton.edu

The following syllabus is for a Ph.D. seminar in the Woodrow Wilson School at Princeton. It is designed for a diverse group of students with interests in politics, economics, demography, and technology; most but not all of them already have some statistics under their belts. The readings are correspondingly diverse, though some are taken from the reading lists of similar courses taught by Herb Weisberg at Ohio State (Watson), David Collier at Berkeley (Lakoff), and Mark Hansen at Chicago (Durkheim). In addition to the books listed, the reading for each section includes works on similar topics using different methodologies to provide some basis for comparison. For example, along with Fenno's *Homestyle* we do Stokes on the nationalization of electoral forces and Cain, Ferejohn, and Fiorina on the personal vote; along with Card and Krueger's *Myth and Measurement* we do papers on the New Jersey Income Maintenance Experiment; and so on.

I would assign somewhat different books if I was offering this course either to political science Ph.D. students or to undergraduates, though the fundamental aims would be similar in those cases.

This course is designed to provide an introduction to a variety of empirical research methods used by social scientists and policy analysts, including case study, experimental, historical, observational, survey-based, and statistical methods. The emphasis is on the logic of research design and the inferential strengths and weaknesses of various research methods. The primary aims of the course are to make students sophisticated consumers of diverse empirical research and to inform their subsequent investment in more specialized methodological skills. Our reading will be roughly evenly divided between methodological works and notable applications. Our discussions will focus both upon the readings and upon course participants' own research, with the relative emphasis to be determined by students' backgrounds and interests.

1. Theories (week 1)
    - Thomas S. Kuhn, *The Structure of Scientific Revolutions* 2nd ed.
    - James D. Watson, *The Double Helix*
2. Concepts and Categories (week 2)
    - George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*
3. Observation and Measurement (weeks 3 and 4)
    - Richard F. Fenno, Jr., *Home Style: House Members in their Districts*
    - Norman M. Bradburn and Seymour Sudman, *Polls and Surveys: What they Tell Us*
4. Experiments and Quasi-Experiments (weeks 5 and 6)
    - Thomas H. Cook and Donald T. Campbell, *Quasi-Experimentation*
    - David Card and Alan Krueger, *Myth and Measurement: The New Economics of the Minimum Wage*
5. Historical Analysis (weeks 7 and 8)
    - David H. Fischer, *Historians' Fallacies: Toward a Logic of Historical Thought*
    - Richard Neustadt and Ernest R. May, *Thinking in Time: The Uses of History for Decision Makers*
6. Statistical Analysis (weeks 9 and 10)
    - Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*
    - Christopher H. Achen, *Interpreting and Using Regression*
    - Emile Durkheim, *Suicide*
7. Benefit-Cost Analysis (weeks 11 and 12)
    - Edward M. Gramlich, *A Guide to Benefit-Cost Analysis*, 2nd ed.
    - Stephen Breyer, *Breaking the Vicious Circle: Toward Effective Risk Regulation*

Larry M. Bartels teaches in the Politics Department and Woodrow Wilson School of Public and International Affairs at Princeton University. He is a past president of the Methods Section of APSA.

Political Methodology 1996 Summer Meeting

Nancy Burns
University of Michigan
nburns@umich.edu

The Summer Political Methodology Meetings will be July 17th through July 21st in Ann Arbor, Michigan.

Some Highlights...

- Public Roundtable, co-sponsored by the ICPSR Summer Program, on Bayesian Approaches to Political Science, Thursday evening (Chris Achen, Larry Bartels, Simon Jackman, Gary King, and Renee Smith)
- Continuous Graduate Student Poster Sessions at the Center for Political Studies

Preliminary Program

Time Series

Unit Roots and Causal Inference in Political Science

- John Freeman (Minnesota)
- Paul Kellstedt (Brown)
- John Williams (Indiana)

- **Discussant:** James Granato (Michigan State)

Estimating Time Series Cross Section Models with Heterogeneous Coefficients

- Nathaniel Beck (UC-San Diego)
- Jonathan Katz (Cal Tech)
- **Discussant:** Jonathan Nagler (UC-Riverside)

Measurement Models for Time Series Analysis: Estimating Dynamic Linear Errors-in-Variable Models

- Gregory McAvoy (North Carolina-Greensboro)
- **Discussant:** Suzanna De Boef (Penn State)

Macropartisanship

Dynamic Models of Macropartisanship

- Donald Green (Yale)
- Bradley Palmquist (Harvard)
- Eric Schickler (Yale)
- **Discussant:** Renee Smith (Rochester)

Party I.D. and Macropartisanship: Resolving the Paradox of Micro-level Stability and Macro-level Volatility

- Robert Erikson (Houston)
- James Stimson (Minnesota)
- Michael MacKuen (Missouri-St. Louis)
- **Discussant:** Christopher Achen (Michigan)

Spatial Model

Estimating Preferred Points in Small Legislatures: The Case of Chilean Senate Committees

- John Londregan (Princeton)
- **Discussant:** Henry Brady (UC-Berkeley)

Markov Chains, Gibbs Samplers, Bootstrap

Bayesian Tools for Social Scientists

- Simon Jackman (Chicago)
- **Discussant:** Gary King (Harvard)

Bootstrap Methods for Non-nested Hypothesis Tests

- Walter Mebane (Cornell)
- Jas Sekhon (Cornell)
- **Discussant:** Douglas Rivers (Stanford)

Polarization and Political Violence

- Mohan Penubarti (UCLA)
- **Discussant:** Charles Franklin (Wisconsin-Madison)

Congressional Support for NAFTA

- Fiona McGillivray (Washington University)
- Alastair Smith (Washington University)
- **Discussant:** Bradford Jones (Arizona)

Political Economy

A Methodology for Estimating the Impact of Partisan Competition on the Economy

- Michael Herron (Stanford)
- **Discussant:** Melvin Hinich (Texas-Austin)

PUBLIC ROUNDTABLE

- Lawrence Bartels (Princeton)
- Simon Jackman (Chicago)
- Renee Smith (Rochester)
- Christopher Achen (Michigan)
- Gary King (Harvard)
- Nancy Burns (Michigan)

Other Participants

- Michael Alvarez (Cal Tech)
- Glenn Beamer (Virginia)
- Rob Franseze (Michigan)
- John Jackson (Michigan)
- Robert Luskin (Texas-Austin)
- Kenneth Meier (Wisconsin-Milwaukee)
- J. David Singer (Michigan)
- Dan Wood (Texas A and M)

Political Analysis

Walter Mebane
Cornell University
wrm1@macht.arts.cornell.edu

Political Analysis publishes full-length articles on topics generally in the area of political methodology, including formal or positive models of political phenomena. The journal follows editorial procedures in general use in political science. It adheres to the Policy on Statistical Reporting, Archiving and Replication of the Political Methodology Section of the American Political Science Association.

For a statement of the policy see the section's newsletter, *The Political Methodologist*, 6, no. 1 (1994): 18–19. All manuscripts are anonymously refereed by subject matter experts. No manuscripts may be reviewed if currently under consideration elsewhere.

Manuscript format and preparation: Manuscripts to be considered for publication should be typed, double-spaced on one side of the paper. Of the four copies to be submitted, three should be free of any author identification on the cover and throughout the text of the manuscript. A brief abstract of the contents of the proposed article should be included in each copy. Electronic submission of LaTeX or LaTeX 2_ε source files is also welcomed and indeed is encouraged (please get in touch with the editor at wrm1@cornell.edu to work out details). The journal uses as guide to matters of form and style the *Style Manual for Political Science*, a publication of the American Political Science Association's Committee on Publications. It is available from the Association's office at 1527 New Hampshire Avenue, NW, Washington, DC 20036. Printed copies of manuscripts for review are to be submitted to:

Walter R. Mebane, Jr.
Political Analysis
Department of Government
125 McGraw Hall
Cornell University
Ithaca, New York 14853-4601

Political Analysis - Volume 5 (John R. Freeman, Editor) Contents:

Henry E. Brady - *Knowledge, Strategy, and Momentum in Presidential Primaries.*

John E. Jackson - *Attitudes, No Opinions, and Guesses.*

William G. Jacoby - *Testing the Effects of Paired Issue Statements on the Seven-Point Issue Scales.*

Paul Kellstedt, Gregory E. McAvoy, and James A. Stimson - *Dynamic Analysis with Latent Constructs.*

Eric Schickler and Donald Green - *Issues and the Dynamics of Party Identification: A Methodological Critique.*

Nathaniel Wilcox and Christopher Wlezien - *The Contamination of Responses to Survey Items: Economic Perceptions and Political Judgments.*

Rainer Winkelmann, Curtis S. Signorino, and Gary King - *A Correction for an Underdispersed Event Count Probability Distribution.*

1996 APSA Political Methodology Panels; Section Chair - Elizabeth Gerber, UCSD

1. Time Series and Panel Models of Partisanship [**Thu 10:45 am**]
2. Testing Formal Theories of Legislatures [**Sun 10:45 am**]
3. New Approaches to Implementing Experimental Research [**Fri 10:45 am**]
4. Modeling the Timing of Political Events [**Fri 3:30 pm**]
5. Models of Political Choice [**Sat 8:45 am**]
6. Dimensionality and Functional Form [**Thu 3:30 pm**]
7. Ecological Inference and Missing Data [**Fri 8:45 am**]
8. Testing Bargaining Models [**Sat 10:45 am**]

6-1: TIME SERIES AND PANEL MODELS OF PARTISANSHIP

- Chair: Bradley Palmquist
- Party Identification and Macropartisanship: Micro-Macro Connections (Michael MacKuen, Robert Erikson).
- Heterogeneity and Individual Party Identification (Janet Box-Steffensmeier, Renee Smith).
- Partisanship and Ideology: An Over Time Analysis of Subgroups (Suzanna De Boef).
- Reassessing Macropartisan Changes (Donald Green, Bradley Palmquist).
- Discussants: John Freeman, Neal Beck.

6-6: DIMENSIONALITY AND FUNCTIONAL FORM

- Chair: John Jackson.
- Computational Political Economy (Ken Kollman, Scott Page).
- Materialist Low Dimensionality versus Structure-Induced Equilibrium in the U.S. House (Jasjeet Sekhon, Walter Mebane).
- Classification and Prediction with Neural Network Models (Langche Zeng).
- Simulating Political Behavior in Chaotic Environments (Thad Brown).
- Discussant: Diana Richards.

6-7: ECOLOGICAL INFERENCE AND MISSING DATA

- Chair: Wendy Tam.
- Missing Data in Political Analysis (Charles Franklin).

- A Consistent Estimator for the Transition Model (Ken McCue).
- Ambivalence and Uncertainty in the Ecology of Race (John Brehm).
- Referenda, Roll Calls and Representation (Jeffrey Lewis).
- Discussants: Gary King, Wendy Tam.

6-3: NEW APPROACHES TO IMPLEMENTING EXPERIMENTAL RESEARCH

- Chair: Kent Portney.
- The Experimental Studio (Shanto Iyengar).
- Innovations in Experimental Design in General Population Attitude Surveys (Paul Sniderman).
- Optimizing Experimental Design for Testing Formal Theory (Rebecca Morton).
- Using the Applied Mulitmedia-Based Experimental Research (AMBER) in Analysis of Criminal Court Sentencing (Jerry Goldman).
- Discussants: Arthur Lupia, James Druckman.

6-4 : MODELING THE TIMING OF POLITICAL EVENTS

- Chair: Jonathan Nagler.
- Does Uncontested Mean Uninteresting? A Unified Model Of Legislative Elections (Jonathan Katz).
- Split-Population Survival Models: An Application to the Timing of PAC Contributions (Janet Box-Steffensmeier, Peter Radcliffe).
- Using Cluster Analysis to Derive Early Warning Indicators for the Middle East, 1980-1996 (Philip A. Schrodt, Deborah Gerner).
- A Remedy for Some Common Methodological Problems Associated with Dyadic Conflict Analyses (Neal Beck, Richard Tucker).
- Discussants: Brad Jones, Gretchen Hower.

6-5 : MODELS OF POLITICAL CHOICE

- Chair: Alana Northrop.
- Choice Models in Theory and Practice: Independence of Irrelevant Alternatives and the Decision to Litigate (Christopher J. Zorn).
- Explaining Voter Choice in Western European Democracies (Andrew Martin, Kevin Quinn, Andrew Whitford).
- Economic Voting in a Unified Model of Vote Choice (William Morgan).
- Can People Explain Their Own Vote (Andre Blais, Richard Nadeau).
- Discussants: Lanny Martin, R. Michael Alvarez.

6-8: TESTING BARGAINING MODELS

- Chair: Scott Page.
- What Gets Vetoed? An Empirical Analysis of Microlevel Data 1945-1992 (Charles Cameron, John Lapinski, Charles Reiman).
- The Rationality of US-China Relations Before and After Tiananmen (Catherine Langlois, Jean-Pierre Langlois).
- Legislative Entrepreneurship and Campaign Finance (Gregory Wawro).
- Experimental Elections: Efficiency and Information Aggregation (Sugato Dasgupta, James Granato).
- Discussant: David Epstein.

6-2: TESTING FORMAL THEORIES OF LEGISLATURES

- Chair: Jacob T. Levy.
- A Linear Latent Factor Model of Voting (James Snyder, James Heckman).
- Computing the Value of Committee Seats (Tim Groseclose, Charles Stewart).
- Estimating Ideal Points in Small Legislatures (John Londregan).
- Measuring Legislative Instability: A Test for Cycling in the Former Russian Parliament (Jo Andrews).
- Discussants: Keith Krehbiel, Mike Munger.

Instructions For Using the Political Methodology Electronic Paper Archive

The current version of this document is available at: **http://www.ucr.edu/polmeth/instruct.html**

Introduction

To facilitate the distribution of papers of interest to political methodologists an electronic archive has been established at UC Riverside. This document describes the procedure for submitting papers and retrieving papers. Papers can be retrieved via anonymous-ftp or web-browsing software such as Netscape. Please read these instructions carefully. By following these instructions you help to make it as easy as possible for others to retrieve your work.

References used here are assuming you are submitting a paper for the 1996 Summer Political Methodology conference. For other conferences, the term 'methods96' would be replaced by the appropriate conference (i.e., 'apsa96', 'midwest96', 'southern96', etc) name. If you are submitting a working paper, you would substitute 'working96' for 'methods96.' [You should modify these names 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.

First, paper authors are asked to produce postscript and/or into 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. You can read a separate document explaining how to do this with Microsoft Word at the methodology web-site (http://wizard.ucr.edu/polmeth/msword.html).

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.

Second, include a plain-text (ascii) version of the abstract. It is generally a good idea to avoid lines of text longer than 80 columns in the abstract. The abstract can take any form you would like. However, the first 4 lines of the file must be the following:

```
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"
conference = "conference-code"
```

Following these 4 lines, **include a blank line before beginning the text of the abstract**.

The conference-code should be taken from the following list. If the paper was presented at a conference not listed below, choose what you think makes sense. If the paper you are submitting was not presented in its submitted state at a conference, choose 'working96'.

- apsa96 = American Political Science Association Annual Meeting.
- methods96 = Summer Political Methodology Meeting.
- midwest96 = Midwest Political Science Association Annual Meeting.
- southern96 = Southern Political Science Association Annual Meeting.
- western96 = Western Political Science Association Annual Meeting.
- working96 = No Conference Associated With This Version of Paper.

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"
conference = "methods96"
```

File 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 of Files

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).

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 '96'. 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 '96b'.
- 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 '96'. 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 '96b'.
- 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 '96'. Again, if the author has another paper on the archive already using this name, then use '96b'.

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.

Single 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 alvar96.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 alvar96.zip:

- alvar96.abs : abstract of Alvarez's paper.
- alvar96.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.
- alvar96.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.
- alvar96.txt : Text of Alvarez's paper, plain-text (ascii) format.
- README : File explaining what is in each of the above files.

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

```
/pub/polmeth/working_papers96/incoming
```

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

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

References used here are assuming you are retrieving a paper from 1996. *Note: Prior to 1996 there were separate directories for several different conferences.*

If you subscribe to the political methodology list-server (polmeth@wizard.ucr.edu) you should receive abstracts of the 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.]

Anonymous ftp Instructions

Alternatively, You can use anonymous ftp to connect to wizard.ucr.edu. You then go to the /pub/polmeth/working_papers96 and get the appropriate .zip file (be sure to use a binary get).

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/working_papers96
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 (on a PC, you might use PKunzip rather than unzip) as follows:

```
% unzip 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:
```

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.

Last updated on April 29, 1996 by Jonathan Nagler.

Please send suggestions, comments, or trouble-reports to: nagler@wizard.ucr.edu or www@wizard.ucr.edu.

The Political Methodologist
Department of Political Science
University of California
Riverside, 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 College of Humanities, Arts, and Social Sciences of the University of California, Riverside in helping to defray the editorial and production costs of the newsletter.

Subscriptions to *TPM* are free to members of the APSA's Methodology Section. Please contact APSA (212 483-2512) to join the section. Dues are $8.00 per year.

Submissions to *TPM* are welcome. Articles should be sent to the editor by e-mail (nagler@wizard.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, CA 92521-0118. LaTeX format files are especially encouraged; but the editorial staff of *TPM* is able to handle almost any format. The deadline for submissions for the next issue is November 1, 1996.

TPM was produced using LaTeX on a PC running Linux and a Hewlett-Packard workstation.