

# The Political Methodologist

NEWSLETTER OF THE POLITICAL METHODOLOGY SECTION  
AMERICAN POLITICAL SCIENCE ASSOCIATION  
VOLUME 15, NUMBER 2, WINTER 2008

## Editors:

PAUL M. KELLSTEDT, TEXAS A&M UNIVERSITY  
*kellstedt@polisci.tamu.edu*

DAVID A.M. PETERSON, TEXAS A&M UNIVERSITY  
*dave@polisci.tamu.edu*

GUY WHITTEN, TEXAS A&M UNIVERSITY  
*whitten@polisci.tamu.edu*

## Editorial Assistant:

MARK D. RAMIREZ, TEXAS A&M UNIVERSITY  
*mdramir@polisci.tamu.edu*

## Contents

<b>Notes from the Editors</b>	<b>1</b>
<b>Articles</b>	<b>2</b>
James S. Krueger and Michael S. Lewis-Beck: Is OLS Dead? . . . . .	2
Nonna Mayer: Reflection on the methods of Political Science on both sides of the Atlantic . . .	5
<b>Teaching</b>	<b>8</b>
Garrett Glasgow: Interdisciplinary Methods Training In Political Science . . . . .	8
<b>Computing and Software</b>	<b>11</b>
Stephen R. Haptonstahl: Web-Based Data Collection with PHP and MySQL . . . . .	11
<b>The L<sup>A</sup>T<sub>E</sub>X Corner</b>	<b>17</b>
Mark D. Ramirez and James E. Monogan, III: Posters in L <sup>A</sup> T <sub>E</sub> X . . . . .	17
<b>Book Reviews</b>	<b>22</b>
David Siroky: Review of <i>The Elements of Statistical Learning</i> , by Trevor Hastie, Robert Tibshirani, and Jerome Friedman . . . . .	22
<b>Announcements</b>	<b>23</b>
William G. Jacoby: ICPSR Summer Program in Quantitative Methods of Social Research . . .	23

Thomas Plümper: Essex Summer School in Social Science Data Analysis & Collection . . . . .	25
EITM 2008: Washington University in St. Louis . .	25
EITM 2008: Duke University, Call for participants and Mentoring Faculty-in-Residence . . . . .	26
<b>Section Activities</b>	<b>28</b>
A note from our Section President . . . . .	28

## Notes From the Editors

A common perception is that the advancement of political methodology—particularly quantitative methods—is creating a divide between those with advanced quantitative training and those trained in alternative methods. This issue of *The Political Methodologist* shows that this divide is more of a social construction within the discipline than real.

James S. Krueger and Michael Lewis-Beck survey the use of statistical estimators among political science journals assessing the perception that the use of ordinary least squares regression has given way to more complex statistical procedures. Although they find growth in the use of more sophisticated estimators, the use of OLS remains . . . quite ordinary. Their article is more than an assessment of the use of OLS, but a systematic look at the diversity and growth of quantitative methods within the discipline over time. We anticipate that some of our readers may disagree with Krueger and Lewis-Beck's conclusion about OLS. We encourage readers interested in replying to their paper to contact the editors. We are hoping to publish reactions and thoughts on the utility and ubiquity of OLS in the next issue of *TPM*.

Nonna Mayer’s article provides an innovative perspective on the qualitative versus quantitative debate. Instead of opposing forces, Mayer notes that these approaches complement each other in the research of French political scientists. Further, she shows that recasting the typologies around the reactivity of the method and theoretical approach provides new insight into more relevant commonalities among research.

Garret Glasgow provides a portion of the section’s report on the implementation and advantages of interdisciplinary methods training. The report highlights multiple avenues for interdisciplinary methods training for political scientists and increasing formal arrangements among academic departments to engage in training students. Together, these articles speak to possible bridges among methodologists rather than divides.

In the computing and software section, Stephen R.

Haptonstahl provides a detailed description of how to use web-based software for data collection and management. We’re sure this will be of use to many *TPM* readers engaged in both small and large data collection efforts. The “ $\text{\LaTeX}$  Corner” returns this issue with a tutorial on constructing scientific posters. This article should be of interest to students making their first  $\text{\TeX}$  poster and faculty implementing poster sessions in their courses. Next, David Siroky provides a review of *The Elements of Statistical Learning*. According to Siroky, the book should help researchers get more from their data. Finally, there are a list of announcements, section activities, and the first message from the section’s new president, Philip A. Schrodt.

Thanks to all of the contributors of this issue. As always, we look forward to ideas for future issues that relate to teaching, research methods, and political inquiry.

*The Editors*

## Articles

### Is OLS Dead?

**James S. Krueger and Michael S. Lewis-Beck**

University of Iowa

*james-krueger@uiowa.edu and michael-lewis-beck@uiowa.edu*

A graduate student worries aloud: “All I know is ordinary least squares. These results won’t convince anyone.” A colleague posts on his door an “OLS” sign, over it the familiar circle with the negative slash. A leading methodologist praises maximum likelihood estimation, and damns ordinary least squares. They all beg the question: Is OLS dead? But is it? In this article, we discuss the state of the OLS estimator in contemporary political science, offering documentation of its use, and trends in its use.

Regression analysis, in some version, has served as the statistical workhorse in political science. Given the classical linear regression (CLR) assumptions are met, then ordinary least squares (OLS) is an optimal estimator. Interestingly, the listed CLR assumptions vary some from text to text. [See the excellent essay by Larocca (2005), on this point.] For a “classic” statement of the CLR assumptions, see Kmenta (1997). There are those who argue that, against assumption violations, regression analysis is robust, while others argue it is fragile. [Consult the relevant discussions in King (1986); Lewis-Beck (1980, 30; 2004, 935-938).] Supposing the “fragile” perspective, then the use of OLS can be especially problematic. Taking into account these argu-

ments, plus the increasing attention to measurement level issues and newer maximum-likelihood-based statistical procedures, the expectation is that OLS has largely ceased to appear in the leading research outlets.

To test this expectation, we examined a large sample of current published research, content-analyzing the methods employed. In particular, we explored our top three general journals, *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics* (1990-2005), yielding an  $N = 1756$  scientific research papers. (Articles dealing with methods, theory, up-dates, exchanges, communications, workshops, or symposia were excluded).

Table 1 shows the frequency of use for different quantitative techniques. These techniques are classified as ‘more’ or ‘less’ sophisticated than OLS based on whether the material would normally appear before or after OLS in a quantitative methodology course. ( In any classification of this kind, judgment calls are inevitable. For example, it might be argued that ANOVA is a kind of OLS. We kept ANOVA as a separate category, because of its traditional use in experimental designs and because of the generally simpler models it estimates. To take another example, consider the time se-

Table 1: Statistical Methods Employed in Articles in *APSR*, *AJPS*, and *JOP* (1990-2005)

Statistical Method	Raw Frequency	Percent Frequency
<b>Less Sophisticated</b>		
ANOVA	40	1.8
Correlations	89	4.0
Difference Tests	95	4.3
Descriptive Statistics	238	10.7
<b>Subtotal</b>	<b>462</b>	<b>20.8</b>
<b>OLS</b>	<b>684</b>	<b>30.8</b>
<b>More Sophisticated</b>		
Advanced Regression	186	8.4
Time Series	77	3.5
Logit	318	14.3
Probit	216	9.7
Other MLE	97	4.4
Scaling and Measurement	31	1.4
Latent Variables	8	0.4
Simulation	10	0.5
<b>Subtotal</b>	<b>943</b>	<b>42.6</b>
No Method Reported	132	5.8
<b>Total</b>	<b>2221</b>	<b>100.0%</b>

Note: Methods are grouped by degree of sophistication. Different statistical methods employed within the same article are reported separately. For this reason the total N of the table exceeds the total N of the articles in the dataset.

ries classification of an autoregressive distributed lag model, when estimated with OLS using a lagged dependent variable on the right-hand side. We maintain this in the time series category, due to its heavy reliance on the time series nature of the data. It should be mentioned that if either example were placed in the OLS category instead, it would only strengthen further the role of OLS). Note, finally, that Table 1 includes all statistical methods employed in each article.

According to this classification scheme, OLS is by far the most popular method, appearing in nearly 31 percent of the papers. Taken all together, statistics that appear more sophisticated—logit, probit, other MLE, time series, scaling and measurement, latent variables, simulation, advanced regression (e.g., generalized least squares, two-stage least squares, weighted least squares)—appeared 42.6 percent of the time. The remaining quantitative methods are actually less sophisticated than OLS—difference tests, descriptive statistics, correlations, ANOVA, none—and total 20.8 percent of the cases. In sum, OLS, or its lesser cousins, stand as the dominant methodological choice in this blue-ribbon sample.

Of course, one may begin data analysis with OLS, then go on to another method. Thus the question: for those

cases in Table 1 that employed OLS, how many turned to a second method? Just 37.1 percent. In other words, the overwhelming majority stayed with OLS as the principal analysis method. For the minority that did not stay with OLS, what was their next move? Almost half—45.3 percent—went on to a method less sophisticated than OLS (e.g., correlations, difference tests, descriptive statistics); just over half—54.7 percent—did go on to a method more sophisticated than OLS (e.g., an advanced regression or MLE technique). Overall, then, OLS continues to be the dominant method used, even allowing for a second round of analysis. Is there any evidence that, despite its dominance, OLS use is changing? Yes, a bit. For that small minority who step beyond OLS to something more sophisticated, one observes that, over time, the sophisticated choice is becoming somewhat more likely. For example, applying a simple signs test, one observes for seven out of the first eight years of the series, the second choice technique was less sophisticated than OLS; in contrast, for the second eight years of the series, the second choice technique was more sophisticated than OLS in six of the eight years. (See Table 2). There is movement, then, but again it must be mentioned this occurs in a very small group of papers.

Table 2: Statistical Methods Employed in Conjunction with OLS in Articles in *APSR*, *AJPS*, and *JOP* (1990-2005)

Year	Less Sophisticated than OLS*		More Sophisticated than OLS <sup>+</sup>	Total
1990	9	>	6	15
1991	9	>	6	15
1992	6	>	5	11
1993	11	=	11	22
1994	7	<	10	17
1995	9	>	6	15
1996	10	>	7	17
1997	9	>	8	17
1998	11	>	10	21
1999	6	<	9	15
2000	10	>	6	16
2001	5	<	10	15
2002	9	<	13	22
2003	1	<	8	9
2004	1	<	13	14
2005	2	<	11	13
Total	115	<	139	254

Note: Cases are articles which report both OLS and an additional statistical method.

\*Less sophisticated than OLS includes: ANOVA, Correlations, Difference Tests, and Descriptive Statistics.

+More sophisticated than OLS includes: Advanced Regression, Time Series, Logit, Probit, Other MLE, Scaling and Measurement, Latent Variables models, and Simulation.

## Conclusion

OLS is not dead. On the contrary, it remains the principal multivariate technique in use by researchers publishing in our best journals. Scholars should not despair that possession of quantitative skills at an OLS level (or less), bars them publication in these top outlets. In itself, it need not be an impediment, as these data demonstrate.

Why is OLS so entrenched? We can only speculate. However, its putative advantages come to mind. OLS offers common coin, easily exchanged among most scholars in the discipline. It is simple to run, simple to understand. On occasion, it might even stand as a BLUE estimator, when classical linear regression assumptions are met.

## References

- Larocca, Roger. 2005. "Reconciling Conflicting Gauss-Markov Conditions in the Classical Linear Regression Model." *Political Analysis* 13(2):188-207.
- Lewis-Beck, Michael. 1980. *Applied Regression: An Introduction*. Beverly Hills, CA: Sage.
- Lewis-Beck, Michael. 2004. "Regression." In *The Sage Encyclopedia of Social Science Research Methods*, Vol. 3., ed. Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao. Thousand Oaks, CA: Sage, 935-938.
- King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30(3): 666-687.
- Kmenta, Jan. 1997. *Elements of Econometrics*, 2nd edition. Ann Arbor, MI: University of Michigan Press.

## Reflection on the methods of Political Science on both sides of the Atlantic

**Nonna Mayer**

Research director at CNRS-Centre de recherches politiques de Sciences Po, Paris

French Political Science Association

*nonna.mayer@sciences-po.fr*

For its Ninth Congress in Toulouse (5-7 September 2007), the French Political Science Association (AFSP) invited the American Political Science Association (APSA) to hold a joint “table-ronde”<sup>1</sup>, comparing methods on both sides of the Atlantic. It took the form of three consecutive panels, devoted to qualitative and quantitative approaches, to the dimension of time and to contextual and inference problems. During three days, 18 papers were presented, over 60 participants attended, contrasting ways to validate theories and models were discussed at length, illustrated by concrete research examples. The objective here is less to sum up all that was said than to outline the main differences and convergences of our methodologies.

### The quali-quantitative debate

It is a fact that in France quantitative approaches are less developed than in the States, where even qualitative researchers have received a basic formation in statistics, and know how to read an equation, a regression line, a factor analysis. In France one tends to give more importance to the historical and philosophical positioning of a problem, training in statistical methods is offered by fewer institutions, rational choice models are not popular (Billordo 2005b, 2006), and quantitative analysis forms a small minority of the articles published in the main reviews (one third of all articles published in *French Political Science Review* between 1970 and 2004 according to Billordo 2005a). The borders between quali and quanti approaches was the issue addressed by the first panel. Where the Americans tended to see distinct epistemologies, different conceptions of causality, “two cultures” (Mahoney and Goertz 2006), the French on the contrary insisted on the necessity to go beyond this opposition, questioning what basically differentiates the two approaches. Is it the fact of counting, opposing those who count to those who give account, in French “ceux qui comptent” vs “ceux qui racontent”? Is it a problem of arithmetic, mathematics, statistics? Is it the number of cases studied, small or big-n? Are survey research and comprehensive interviews, case and variable oriented approaches so antagonist? Where should one put the QCA (Qualitative Comparative Approach) developed by Charles Ragin, based on Boolean logic, which

does not actually count, but puts a phenomenon into an equation according to the presence or absence of certain elements, and the way they combine?

On the whole the divide between qualitative and quantitative methods seems far more institutionalized in the States, where it is embodied in distinct academic departments and recruitment procedures, and is represented by two different methodological standing groups at APSA. But precisely because the separation is less rigid in France, it seems more natural to combine the two approaches, as shown by most of the French papers for the table-ronde. This could be an asset, at a time where all over Europe mixed methods designs, triangulation, and combining comprehensive and explicative approaches, are becoming fashionable (Moses, Rihoux and Kittel 2006).<sup>2</sup>

### Assessing time

The second panel explored the time dimension. The papers apprehended it in many different ways, time as period, as process, as event, as series of sequences, as interval, time as the present moment and time as the past and its memories. The advantages and limitations of several methods were compared with sophisticated models such as survival analysis, optimal matching analysis, protest event analysis. But time is also the specific time of the interview or of the observation, when it takes place, how long it lasts, what relationship settles between interviewer and interviewee. Most participants insisted on the limits of the “one shot” interview to grasp the subjects with their contradictions, their evolutions, and their interactions, for quantitative as well as for qualitative approaches.

### Assessing context

The third session enlarged the notion of context. At first we had in mind ecological analysis and the classical problems of inference. But some understood it also as the subjective context, the way people interviewed feel about their surroundings. Others dwelt on how experimentation can manipulate the context in order to test the effect of the variables, in or out the laboratory. Context was also taken in the sense of the scale of analysis selected, and the mul-

<sup>1</sup>Co-organised by Nonna Mayer (AFSP) and Andrew Appleton (APSA/French Politics Group). A special thanks to the chairs of APSA’s two methodological sections, Janet Box-Steffensmeier and James Mahoney, who greeted me at their business meeting in APSA’s 2006 Congress in Philadelphia and enthusiastically supported this project, and to the French Politics group whose mediation was essential. The table ronde’s programme and papers summaries (French and English) are available on AFSP’s website, <http://www.congres-afsp.fr/>.

<sup>2</sup>See for instance the dynamic Standing group on Political Methodology at ECPR (European Consortium for Political Research) chaired by Benoît Rihoux, Jonathon Moses and Bernhard Kittel (ECPR) and the workshop they propose at the coming ECPR session on “Methodological Pluralism? Consolidating Political Science Methodology” (Rennes, 11-16 April 2008).

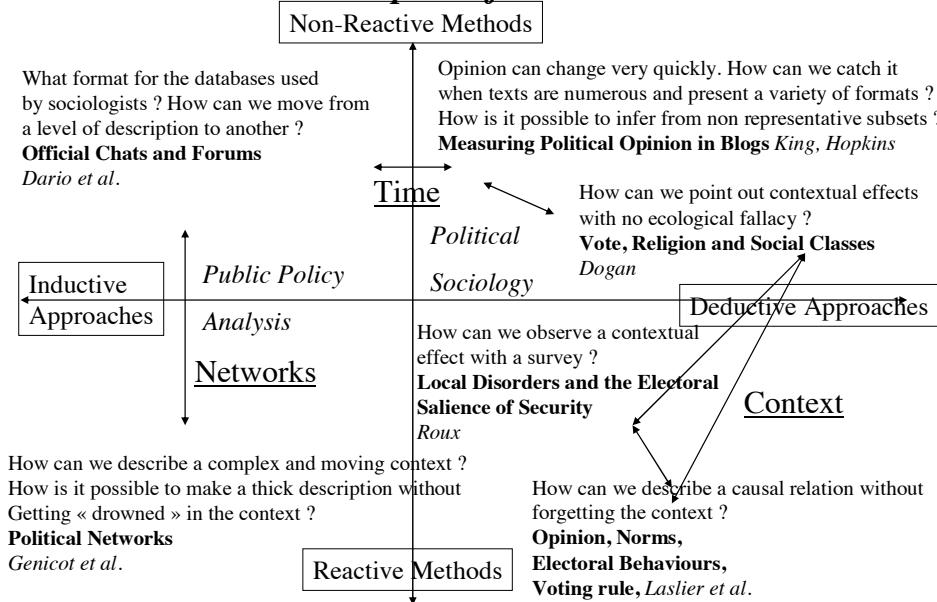
tiple levels—in this case local, national and European—at which on can grasp the relations between actors and the dynamic of their opinions, both in the instant and in the long run. Finally the debate focused on the new types of context in constant transformation generated by the development of Internet (blogs, mailing lists) and the challenge they represent to the traditional quali and quanti methods.

**A common space of discussion**

Many questions were asked, many research tracks opened during these three days. If obvious differences appeared in the methods discussed, yet there also was a common space of discussion between qualitativists and quantitativists, which Mathieu Brugidou, chair of the last session, attempted to map in the following graph based on the 6 papers he discussed.

The vertical axis opposes inductive and deductive approaches, those which move from theories and hypotheses to their empirical validation and those which on the contrary, prefer to start by observation and immersion in the field and move up from there. The second axis opposes reactive to non reactive methods. The former deal with tests, surveys, interviews, getting a reaction from the actors observed, the latter deal with a given object already there that they do not influence. For each paper is specified (in boldface) the topic and the methodological issue. The arrows show the possible lines of discussion connecting papers, the objects they have in common are underlined and in italic appears the sub discipline concerned. To fully understand the graph one must go back to the papers, available on the AFSP’s website. Yet even without doing so, it shows that the quali quanti methodological divide is not the only, nor necessarily the most relevant one.

***TR1(s3) :Inference, context, new approaches : a common space of discussion ?***



The paper by Genicot et al., about public policies actors in Europe, is positioned in the reactive/inductive quadrant, lower left. It shares with Dario et al., (upper left), who study companies forums and chats, a same object: networks, and a similar inductive approach, considering the configuration of the network is not given before hand, it will emerge from the analysis. Yet Genicot and her colleagues have opted for a purely qualitative approach by interviews, while Dario et al., offer a quantitative approach of non reactive data (email lists) to make sense of the evolving configuration of the networks. King et Hopkins who follow the evolution of political opinions expressed in millions of blogs, share with Dario and his colleagues a common moving object, the Net, and the use of sophisticated statistical models. But they are in the upper right quadrant because they give preference to deductive methods, starting with a predefined categorisation of the political opinions they code. Dogan's paper, which offers an ecological analysis of votes, is in the same quadrant and faces with King and Hopkins the common problem of inference. But Dogan also shares a common preoccupation, the effect of context, with Roux who is interested in the subjective perceptions of context by the voters and Laslier and his colleagues who perform electoral experimentations, artificially manipulating context, both situated in the lower right quadrant (deductive-reactive).

The Franco-American table-ronde was but a first step

to confront and exchange our methodological know-how, see how close and how different we are, and overcome the gap between the so called qualitative and quantitative research. We hope it will be followed by many others.

## References

- Billordo, L. 2005a. "Publishing in French Political science journals: an inventory of methods and sub-fields." *French Politics* 3(2):178-186.
- Billordo, L. 2005b. "Methods Training in French Political Science." *French Politics* 3(3):352-357.
- Billordo, L. and A. Dumitru. 2006. "French Political Science: Institutional Structures in teaching and research." *French Politics* 4(1):124-134.
- Mahoney, J. and G. Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14(3): 227-249.
- Moses, J.A. and B. Rihoux, and B. Kittel. 2005. "Mapping political methodology: reflection on a European perspective." *EPS* 2005(4): 55-68.

## Teaching

### Interdisciplinary Methods Training In Political Science<sup>1</sup>

**Garrett Glasgow**

University of California, Santa Barbara

*glasgow@polsci.ucsb.edu*

Although interdisciplinarity is frequently cited as a positive goal in academia, it generally remains more of a buzzword than a practice, and some are skeptical that an interdisciplinary approach will ever catch on in political science (e.g., Moran 2006, McKenzie 2007).

One possible exception to this is interdisciplinary methods training in political science. Interdisciplinary methods training in political science has existed in an informal way for decades—political scientists have a long history of adopting methods from other academic disciplines and applying them to substantive political science problems (Beck 2000). In recent years a growing number of political science departments have shown an interest in more formal arrangements for interdisciplinary methods training.

We will first examine how these interdisciplinary methods training programs are implemented. We will then study the strengths and weaknesses of an interdisciplinary approach to political science methods training. Finally, we will make a set of recommendations for those departments that are considering adopting an interdisciplinary approach to methods training.

#### Implementation of Interdisciplinary Methods Training

We are aware of five primary approaches to interdisciplinary methods training in political science, each of which represents a different level of commitment to interdisciplinary methods training. We discuss each of these approaches here in order from the strongest to the weakest commitment by departments to interdisciplinary methods training.

The first approach is represented by those departments or programs set up explicitly as interdisciplinary units who produce students that can then go on to join political science departments or otherwise be regarded as political scientists. Examples of this approach are the Quantitative Methods in the Social Sciences (QMSS) MA program at Columbia, the Social Science department at Caltech, and the Political Science department at the University of California, Irvine. Although the methods training in these pro-

grams certainly qualifies as interdisciplinary, we will not focus on them in this article, as our primary goal here is to provide information to existing political science departments considering an interdisciplinary approach to methods training. Adopting this method of interdisciplinary methods training would most likely involve combining or reorganizing existing departments, which would be a formidable and largely unnecessary task for a political science department simply seeking to upgrade its methods teaching.

A second approach is to establish graduate fellowships for a select group of students with an interest in interdisciplinary quantitative methods. We are aware of two examples of this approach to interdisciplinary methods training, the Mellon Interdisciplinary Graduate Fellow Program offered through the Institute for Social and Economic Research and Policy (ISERP) at Columbia University, and the Pre-doctoral Fellowships offered by the Quantitative Social Science Initiative (QuaSSI) at Pennsylvania State University. Both of these fellowships are offered through organizations that stand apart from the political science departments at these institutions. While not as daunting a task as reorganizing a department along interdisciplinary lines, this approach is still demanding in terms of resources and administration, although some of this burden could be reduced if a department were to work to establish a fellowship program in an existing interdisciplinary institute on campus (assuming one exists).

A third approach to interdisciplinary methods training in political science is the “certification” program. Under these programs students in political science programs who wish to pursue advanced methodological training can join a “methods certification” program. These students remain political science students, but take methodological classes in other social science departments beyond those available in their home department. Students that complete these classes gain a certification on their transcript that they have advanced methods skills.

To the best of our knowledge, 10 political science departments in the U.S. offer their students the option of joining a certification program in advanced methods. Infor-

<sup>1</sup>This article originally appeared as a section of the 2007 Undergraduate and Graduate Methodology Committee Report. I thank Lonna Atkeson, Paul Gronke, Dean Lacy, and Alan Zuckerman for their helpful comments.



mation on each of these certification programs was gathered through internet searches and email correspondence, and is available on the authors website.

All of these certification programs require additional courses of students beyond those necessary for a political science degree without the certificate. Typically students seeking the certificate are expected to take an additional 4 to 7 courses beyond those required for their degree.

Four of these certificate programs focus on survey research methods, while the others are focused on quantitative social science methods more generally. Five of the certificate programs are housed in interdisciplinary research centers (Washington, Duke, North Carolina, Ohio State, and Michigan), two are housed in specific academic departments (Statistics for Portland State and Political Science for Indiana University—Purdue University Indianapolis, and three are set up as interdisciplinary units without a specific home in a department or research center (UCSB, Florida, and Stanford).

As these certificate programs are relatively new, little is known about the job market value of the certification on the transcript above and beyond the classes needed to attain it, although there is some anecdotal evidence that some employers find them valuable.

A fourth approach is to offer methods as a field within the political science department, and allow or even expect students to take some of the additional methods courses required to complete the field outside of the political science department. In many ways this approach is a more flexible, less formal version of a certificate program, and may carry less of an administrative burden for the department. However, without the administrative framework and formal agreement between departments represented in a certificate program some students might have difficulty enrolling in methods classes in other departments. We are uncertain of how many departments have adopted this strategy.

Finally, a fifth approach is to send students to a summer methods program. The best known of these programs is the Summer Program in Quantitative Methods of Social Research at the Inter-University Consortium for Political and Social Research (ICPSR), which focuses on general methods training. The ICPSR Summer Program offers three instructional tracks in quantitative methods (beginning, intermediate, and advanced), as well a variety of short courses on advanced methods and applied topics. In 2007 the fee for an 8-week course was \$2000 for students from ICPSR member institutions. The Essex Summer School in Social Science Data Analysis & Collection is another option for a summer program offering interdisciplinary methods training. The Summer Institute in Survey Research Techniques at University of Michigan also accepts students for summer courses from other institutions, and is focused on teaching survey research methods.

Obviously this approach is the least demanding in terms of administrative work for the department, although it still has a resource cost. Many departments send students to summer methods programs on a regular basis, making this approach the most common approach to interdisciplinary methods training by far.

Thus, our research reveals a tiny number of departments or programs that produce political scientists that are explicitly interdisciplinary, while a few other political science departments have made formal arrangements to allow students to further their methods training in other departments or allow students to satisfy some methods requirements in other departments. Many more departments offer students a relatively limited opportunity to pursue interdisciplinary methods training through attendance at summer programs such as ICPSR.

Of course, there may be a significant amount of informal interdisciplinary methods training that is difficult to observe (such as a student taking methods classes outside of the political science department on his or her own initiative or on the advice of an advisor), but it is clear that formal interdisciplinary methods training is rare in political science.

However, interdisciplinary methods training in political science appears to be gaining some momentum—all 10 of the certification programs we studied, as well as the QMSS program at Columbia, were established within the last 10 years. There is also some evidence that the ICPSR summer program is moving in a more interdisciplinary direction. While an interdisciplinary approach to methods training is still rare, it appears to be an idea that has gained at least some momentum. Thus, a clear accounting of the strengths and weaknesses of an interdisciplinary approach to methods teaching would be helpful to those departments considering such a move.

### **Strengths and Weaknesses of an Interdisciplinary Approach**

As formal interdisciplinary methods training is a recent arrival in political science, we have very limited experience with interdisciplinary methods training. Nevertheless, we have obtained enough information to construct a preliminary list of the strengths and weaknesses of the interdisciplinary approach:

#### *Strengths*

1. Most obviously, interdisciplinary methods training gives students additional opportunities to improve their methods skills and to develop a distinctive methodological skill set that will distinguish them from their peers on the job market and in their subsequent career.

2. Interdisciplinary methods programs allow political science departments to offer a wider variety of methods classes to students than is available within the department. This is most likely to be helpful to smaller departments that have relatively few methodologists on the faculty.
3. Interdisciplinary methods programs also give students an opportunity to forge relationships in related fields that may be useful. For instance, a student interested in political psychology would not only benefit from the methodological training available in the psychology department, but would also benefit from meeting potential collaborators and colleagues among the psychology graduate students and faculty.
4. Many methodological innovations in political science are adapted from other fields (Beck 2000). Interdisciplinary methods programs are one conduit by which new methods and practices can be introduced to the field of political science.

#### *Weaknesses*

1. In some cases interdisciplinary training might increase the time to completing the degree. This can have negative consequences not only for the student (who must find a way to finance graduate education for some additional time), but also for the department (as this will increase the average time to degree).
2. Although additional methodological training is clearly valued in political science, it is not clear if interdisciplinary approaches carry the same weight. Some potential employers may view interdisciplinary training as a fad or irrelevant, and wish to see students focused more on “mainstream” political science. The relative lack of job listings that cross over even our traditional subfield distinctions (American, Comparative, etc.) suggests that many departments will struggle to see how students with an interdisciplinary focus will fit into their faculty. Although interdisciplinary students might be ideal for more general job listings such as “individual behavior,” such listings are rarely seen.
3. Interdisciplinary training also reduces the amount of time students spend in their home department, which means fewer opportunities to network with their fellow students and faculty.
4. Participation in an interdisciplinary methods program will also likely increase the number of students from outside the political science department taking political science methods classes. While this may be a positive in some aspects, it can also increase the administrative burden on the department and have effects on pedagogy (as some students will not have a political science background).
5. In some cases allowing political science students to participate in interdisciplinary methods training may also have a negative impact on enrollments in some political science classes as students substitute methods classes in other departments for classes in the political science department. This in turn may have budgetary or other implications for the department.
6. Participation in the interdisciplinary program may impose an administrative burden on the department, particularly on methods faculty, who may not only be teaching larger classes, but may be asked to participate in other administrative tasks related to the interdisciplinary program (such as tracking student progress, designing curriculum, etc.). As interdisciplinary programs are not exclusive to any one department, it may be unclear who will compensate faculty for taking on these additional duties.
7. In the long run, interdisciplinary programs may lead some departments to view methods training as something that can be accomplished by sending students to other departments. This in turn could reduce demand for hiring political methodologists, and also produce graduate students who are not well versed in political science approaches to methodology.

#### **Recommendations**

Departments considering an interdisciplinary approach to methods training should first consider what goals they hope to accomplish with such a move. While interdisciplinarity is often commented favorably upon in academia, students that take a truly interdisciplinary approach are often disadvantaged when seeking jobs (McKenzie 2007). Thus, an interdisciplinary methods approach is more likely to be successful and useful to students if it is used to extend and enhance the methodological offerings of the political science department rather than create truly interdisciplinary students.

These departments should also consider what administrative and resource burdens an interdisciplinary approach to teaching methods group would involve. While such an approach can provide a lot of “bang for the buck” by giving political science students access to a large number of advanced social science methods classes, it is not without cost. Potential increases or decreases in class enrollments and demands on faculty time should be considered. There are likely to be tradeoffs—for instance, requiring additional methods classes above and beyond the required political science classes will preserve political science enrollment, but may increase the average time to degree in the department.

Obviously, few departments will seek to reorganize themselves along interdisciplinary lines. Establishing interdisciplinary institutes or fellowships at interdisciplinary institutes is more likely to be feasible, but would still be a formidable task. Thus, political science departments seeking to move their methods training in an interdisciplinary direction would most likely either adopt the certification approach, send students to other departments to complete a political science methods requirement, or establish a regular “pipeline” for sending students to summer methods programs.

For departments that decide to set up a certification program for interdisciplinary methods training, begin by contacting those departments that offer classes that might be useful to political science students and proposing an interdisciplinary unit. Other social science departments (such as anthropology, communication, economics, education, geography, and sociology), the statistics department, and even the mathematics department are good candidates. Devising a general curriculum that is acceptable to all participating departments (such as 4 additional methods classes) is the next step. There is also likely to be some administrative work in getting the certification program approved above the departmental level. Establishing a clear framework for the responsibilities of each participating academic unit at this time is important. Issues such as resources and course relief for administrative work related to the program should be clearly worked out early on to avoid problems and misunderstandings down the road.

Sending students to summer methods programs on a regular basis is a comparatively low cost way to pursue an interdisciplinary approach to methods training. However, even at this relatively low level of commitment to interdisciplinary methods training there might be policies a department can establish to most effectively integrate these summer programs into a larger methods program. For instance, should departments send any students who are interested to a summer methods program, only the methods students, or only those that are deemed the best in some kind of internal competition? Should students attend at the end of the departments methods sequence, or only once the student has a clear dissertation topic? As with implementing any of the other approaches to interdisciplinary methods training outlined here, a clear accounting of goals, costs, and benefits is vital.

## References

- Beck, Neal. 2000. ““Political Methodology” A Welcoming Discipline.” *Journal of the American Statistical Association* 95(450):651-654.
- McKenzie, Kenneth. 2007. “Comment on Moran’s ‘Interdisciplinarity and Political Science’.” *Politics* 27(2):119-122.
- Moran, Michael. 2006. “Interdisciplinarity and Political Science.” *Politics* 26(2):73-83.

## Computing and Software

### Web-Based Data Collection with PHP and MySQL

**Stephen R. Haptonstahl**

Washington University in St. Louis  
*srhapton@wustl.edu*

#### Introduction

Computer-based forms can greatly increase the efficiency and accuracy of data collection. Collecting the data in digital form makes it easier to import it to statistical software and facilitates easy backups. Many researchers reach first for a spreadsheet like MS Excel because, as a free form tool, it is easy to get started without any programming skills and no Internet connection is needed. However, Web-based forms provide some clear advantages: more than one person can enter data at a time without fear of writing over each other’s work; the data is stored on a server where it

is (almost certainly) safer; the researcher can impose more structure on the data entered using pull-down menus, check boxes, and radio buttons. Two caveats: an Internet connection and some technical skills are required. However, if those entering the data will have Internet access and if you can create and upload a simple Web page, then this article fills in the rest, providing the tools you need to set up your own Web-based data collection form.

Here’s how it works. Talk to the administrator of your Web server to get a MySQL account and make sure phpMyAdmin is installed. Using the code book you wrote

for your data set, modify the template downloaded from my Web site (<http://haptonstahl.org/srh/>) by adding boxes called **controls** in the template for each piece of data you want to collect; this Web page you create is called the **front end** of the form. Upload this file and two other files to a folder on the Web server. Within your Web browser use phpMyAdmin to set up the database with fields corresponding to the controls in the Web form; this database is the **back end** of the form. Once set up, test everything. Those entering the data use any Internet-connected computers to enter the data simultaneously, which is then stored safely in a database on the server. Once the data is entered, use phpMyAdmin to view the data and export for use in your statistical software.

There are some skill and system prerequisites for this project. You should be comfortable editing a simple Web page using a text editor like Notepad or WinEdt, uploading it to a server, and viewing the page using your Web browser. If not, you can acquire an overview of the process for creating a Web page from <http://tinyurl.com/e5uwo> and learn basic HTML in a few minutes from <http://tinyurl.com/h922b>. The server you use must have PHP, MySQL, and phpMyAdmin installed. If you want to know what each of these programs do, I recommend starting with the Wikipedia entries on each. These are free and probably are already on your Web server, but the server administrator will know for sure. To use MySQL, you need another user name and password, possibly different from the ones you use to upload files to the server, and the name of the database assigned by the administrator. Save this information for later; a useful form for this is available at <http://tinyurl.com/m4cf4>.

## Setting Up the Front End

The front end is the Web page itself, and all you have to do is modify a template I have used with other projects. Create a *working folder* where you will work on the site; this can be a folder on your local machine, in which case you upload the files to the Web server after each edition, or you can work directly on the server. Download the zip file containing the template files (<http://tinyurl.com/egvvt>) and extract the contents to this working folder. Two files, `add_record.php` and `show_records.php`, contain PHP code that you may never need to modify; leave those alone. The file you will edit for your project is `index.php`, but I will start by dissecting a simpler file, `simple-index.php`.

`simple-index.php`

```
<html>
<head>
  <title>Project Title Here - Data Entry Form</title>
<?php
  $server_host =      'localhost';
  $mysql_user_id =   'mylogin';
```

```
  $mysql_password =  'mypassword';
  $mysql_database =  'mydatabase';
  $mysql_table =     'mytable';
  $mysql_primary_key = 'id';
  include('add_record.php');
  ?>
</head>
<body>
<center>
<h2>Project Title Here</h2>
<h3>Data Entry Form</h3>
<form action="index.php"
  method='post'>
  <input type='reset'
    value='Reset the form (does not upload info)' />
  <table border cellpadding='2' align='center'>

    <tr>
      <td><b>#</b></td>
      <td><b>Variable</b></td>
      <td><b>Value</b></td>
      <td><b>Description</b></td>
      <td><b>Coding Notes</b></td>
    </tr>

    <tr>
      <td>1</td>
      <td>Test Text</td>
      <td><input type='text' name='testText'
        size='40' maxlength='255' /></td>
      <td>Plain text field</td>
      <td>Details about what to enter go here.</td>
    </tr>

  </table>
  <input type='hidden' name='action' value='add' />
  <input type='submit'
    value='Click to upload this record' />
</form>
</center>
<?php
  $number_of_records_to_show = 5;
  include('show_records.php');
  ?>
</body>
</html>
```

The two sections marked “<?php ... ?>” are PHP code, not HTML, so be careful in modifying these sections. Note that each line of PHP ends in a semicolon. The first PHP section at the top is where you put the MySQL user name, password, and database given to you by the server administrator. You can choose the name of the table, and like any “variable” the table name should contain no spaces or punctuation. Note that a table in MySQL corresponds to a worksheet in MS Excel, so you will need exactly one table for all of the data entered using a single Web-form. I like to name the table something like a one-word version of the project title. In that block you should only have to set those four values.

At the bottom of the file is a small PHP block where you can specify how many of the previously entered records will be displayed at the bottom of the screen. You can set

this to any integer 0 or higher, or to 'all'. It is useful for those doing data entry to be able to see records recently entered. Even though they will not be able to modify those records, if they see an error they can reenter the corrected record with an note to you that the erroneous record should be removed.

Between the two PHP sections is an HTML form. This is the section that you modify to set up the front end interface for the user, which appears in Figure 1.

**Project Title Here**

**Data Entry Form**

Reset the form (this does not upload typed info)

#	Variable	Value	Description	Coding Notes
1	Test Text	<input type="text"/>	Plain text field	Details about what to enter can go here.

Click when complete to upload this record

**Figure 1:** Simple version of the “Front End” Form

Each block of `<tr> ... </tr>` corresponds to a row in the displayed HTML table. Rows in the HTML table (front end) roughly correspond to columns (fields) in the database table (back end) but this is only approximately true; don't confuse the two kinds of tables. There are two HTML rows in the table above. The first is the header row. The second defines one control, a text box, where a piece (field) of data “Test Text” can be entered. If you want a box where up to 255 characters (letters or numbers) can be entered, copy this block. There are other controls (radio buttons, check boxes, text areas) I will discuss later.

Within each row, there are five columns delimited by `<td> ... </td>`. The first is the item number specified in your data code book (because you numbered them, right?) The second is the short title of the field. The third is a bit of HTML “form” code (cf. HTML forms) creating a control in which users can enter data; for text boxes just copy this block but rename the “name”, which in this example is 'testText'. **Each field in the database must have a distinct name that matches exactly the name of the control in the HTML form.** The fourth and fifth columns you can use as you like, however on most computers the fourth column will be completely visible but reading the fifth column may require the user to scroll right, so it is a less convenient reference.

There are three other kinds of controls that are useful instead of simple text boxes. The following sample comes from the template file `index.php` and produces the controls that appear in Figure 2.

#### Section from `index.php`

```
<tr>
  <td>3</td>
  <td>Test Check Boxes (check all that apply)</td>
  <td>
    <input type='checkbox' name='testCheckbox1' />
      Check this one <br/>
    <input type='checkbox' name='testCheckbox2' />
      You can check this one, too <br/>
  </td>
  <td>Check boxes</td>
  <td>Each check box is a separate field that is empty
    if not checked and set to 'value' if checked. If you
    do not specify a value, it simply gets the value 'on'
    when checked.</td>
</tr>

<tr>
  <td>4</td>
  <td>Test Radio 2<br/>(choose one)</td>
  <td>
    <input type='radio' name='testRadio2' value='1' />
      I can spell R <br/>
    <input type='radio' name='testRadio2' value='2' />
      I use R <br/>
    <input type='radio' name='testRadio2' value='3' />
      People ask me how to use R <br/>
    <input type='radio' name='testRadio2' value='88' />
      I write R packages (which ones?)
    <input type='text' name='testRadio2_notes' size='15'
      maxlength='255' /><br/>
  </td>
  <td>Radio buttons</td>
  <td>This combines the radio buttons with a text box.
    Note that the text box has a different name and gets
    stored in a different column of the database.</td>
</tr>

<tr valign='top'>
  <td>5</td>
  <td>Test Text Area</td>
  <td>
    <textarea name='testTextArea' rows='15'
      cols='45'>Default text here (leave blank if you
      like)</textarea><br/>
  </td>
  <td>Text area</td>
  <td>Type lots if you like. In phpMyAdmin I created a
    "TEXT" field for this, which can be very, very long.</td>
</tr>
```

The little square boxes above are **check boxes**. Each is a separate piece bit of data with its own name, `testCheckbox1` and `testCheckbox2`; more than one may be checked. The little circles are **radio buttons**. All of those in a group have the same name, `testRadio2`; only one at a time can be checked (just like buttons on old-style car radios.) The large box at the bottom is a **text area** which is suitable for entering text more than 255 characters.

Radio buttons are good for categorical data as long as the categories are distinct. Check boxes are good for collecting “dummy variable” type data as long as there is no missing data; if missing data is a possibility, use a radio

3	Test Check Boxes (check all that apply)	<input type="checkbox"/> Check this one <input type="checkbox"/> You can check this one, too	Check boxes	Each check box is a separate field that is empty if not checked and set to 'value' if checked. If you do not specify a value, it simply gets the value 'on' when checked.
4	Test Radio 2 (choose one)	<input type="radio"/> I can spell R <input type="radio"/> I use R <input type="radio"/> People ask me how to use R <input type="radio"/> I write R packages (which ones?) <input type="text"/>	Radio buttons	This combines the radio buttons with a text box. Note that the text box has a different name and gets stored in a different column of the database.
5	Test Text Area	Default text here (leave blank if you like)	Text area	Type lots if you like. In phpMyAdmin I created a "TEXT" field for this, which can be very, very long.

Figure 2: Check Box, Radio Button, and Text Area Controls

button with “missing” as one of three choices. I recommend including at least one text area at the end of the form for any comments the coder may have about the data. It is also a good idea to have the coder enter their name somewhere on the form if there is more than one person entering data.

You now have examples of four different controls. To create your own, find the right type of control in the template and copy everything in the `<td> . . . </td>` block surrounding it. Then customize it by setting the item number, control name, and comments. For radio buttons, you can change the values associated with each choice by setting the `value` and you can have more or fewer choices by copying or deleting a `<input type='radio' . . . />` line. Radio buttons are “grouped” simply by giving them the same name; if they have the same name, only one of them can be selected at a time.

You will have to choose names for each of the controls, and these control names must agree *exactly* with the fields you create in the database. Start by reading Jonathan Nagler’s (1995) article on “Coding Style and Good Computing Practices.” Then go through your code book and indicate what kind of control you want to use (text box, radio button, check box, or text area) and name each control (or group of radio buttons.) Keep this list handy for setting up the database.

The template is fairly simple, so you might decide to spend some time customizing it. I recommend some caution here. If you make the page “prettier” you may inadvertently make it harder to use. If you add features like validation using JavaScript or other “client-side” code you may change the form so that it does not run on all browsers. You will have to make a decision on these tradeoffs; if you make such changes, make sure you test the form on every browser (Internet Explorer, Firefox, Safari, Opera, etc.) you can before you start data collection.

### Setting Up the Back End

Now that you have the front end set up to collect the data, you need a back end database to store the data. The front end uses PHP to send data to the database using `add_record.php`, but to set things up you will use a

free Web-based application called phpMyAdmin. MySQL is the database program you will use, but it only has a command line interface built in. phpMyAdmin provides an easier graphic user interface. You do not install phpMyAdmin on your machine; your administrator installs it (if it is not there already) on the Web and MySQL server. You use it through your Web browser by going to a Web address provided by your server administrator. Log in using the user name and password specifically provided for MySQL – the same ones you put in the top PHP block of the Web form.

Once logged in you will create the table that will hold the data. Click on “Databases” and then on the name of the database your administrator provided. You should see a prompt “Create new table on database” (Figure 3). Enter the name of the table that you chose earlier when setting up the front end. For the number of fields, enter one more than the number of field names you chose when going through your code book. For example, the template `index.php` needs 8 distinct data fields, so the table will need 9 fields.

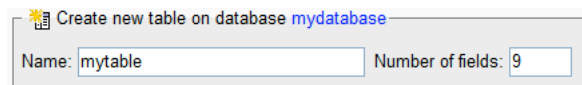


Figure 3: Creating a Table using phpMyAdmin

The next screen will allow you to create the database fields (Figure 4). There are a lot of options here, but you only need three kinds of fields to collect all of your data. First, add a field named ‘id’ (without quotes). This field will be a unique id for each record, even if all of the data fields are repeated across records. Set the type to ‘INT’, the ‘Extra’ to ‘auto\_increment’ and click on the first radio button (under the key) to designate this field as the primary key of the table.

Now enter the names of all of your data fields down the first column. If you may be typing a paragraph for a field and you chose a ‘text area’ control then set the type to ‘TEXT’ which can handle about 8 full typed pages of text. For all other controls (text box, radio button, check box) leave the type set to ‘VARCHAR’ and set the length

Field	Type	Length/Values	Attributes	Null	Default	Extra					
id	INT			not null		auto_increment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
testText	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testRadio1	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testCheckbox1	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testCheckbox2	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testCheckbox3	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testRadio2	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testRadio2_note	VARCHAR	255		not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
testTextArea	TEXT			not null			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Figure 4: Creating Fields in a Table using phpMyAdmin

to the maximum of 255 characters. This does mean that you are limited to 255 characters in text boxes, but if you are typing anywhere near that, you should probably be using a text area control and ‘TEXT’ field type. To create the table corresponding to the template file `index.php` the screen should look like Figure 4. Check the spelling of all of

the field names, including capitalization. Not matching exactly the field names between the form and the database is the most common error; this will cause data entered in that control to ‘just disappear’ and not appear in corresponding field of the database, or possibly appear in the wrong field. Click ‘Save’ to create the table.

If you have a lot of fields (10+) you might choose to enter ten or so at a time. To do this, enter ‘10’ for the number of fields when creating the database and enter the ‘id’ and first nine data fields. To add fields, click on the table name on the left side of the screen, which will show the structure of the table, then enter the number of fields you want to add and click ‘Go’ (Figure 5).

When you are finished testing, clear the data from the database. You can’t do this from the Web form, and neither can anyone else – it’s not a bug, it’s a feature. To clear the table of all data, log into phpMyAdmin, browse to the table structure, and look for the ‘Empty’ tab (Figure 6). Click it.

field(s)
  At End of Table
  At Beginning of Table
  After

Figure 5: Adding Fields to a Table

Figure 6: ‘Empty’ is Good; ‘Drop’ is Bad

### Testing

After you create the front end page and back end database, upload the files all to a single folder on the Web server and view the Web page. With luck, everything looks great, but this is unlikely. Check everything thoroughly, from the spelling of page title to the positioning of each control. Try to add some data. Do you get an error when uploading the data? Does the data appear down at the bottom of the entry form? If you set `$number_of_records_to_show = 0`, you may want to set it to 10 while testing the form to make it easy to see whether the database receives correctly the data you enter.

**Important: Do not click the ‘Drop’ tab; that will delete the table completely and you will have to reenter all of the fields.** If you changed the `$number_of_records_to_show` setting for testing, now is the time to change it back.

A great way to make sure that a data entry screen is well-designed is for the designer of the form to use it. Go ahead and do some of the real data entry yourself. I recommend entering at least 10-20 records. This will catch more problems and help you anticipate how best to train the people who will be doing the rest of the data entry.

### What Goes In Must Come Out

You may want to inspect the data while it is being entered. Log into phpMyAdmin, click on the table name, then click on the ‘Browse’ tab to view the data. If you need to make corrections in the data here you can scroll to the record you want to correct and click the pencil icon in the same row. You can also delete records (by clicking the X on

One easy method for testing the connection between the form and the database is to enter one record with data only in the first field, then another record with data only in the second field, etc. Remember to check each radio button and each check box separately. Checking radio buttons in order makes it easy to verify that you didn’t forget to change the ‘value’ from one button to the next.

the row) and you can delete all of the data or the entire table. The software will ask you for verification before deleting anything, but be careful—once something is deleted, there is no ‘undo’.

When the data has all been entered, it is easy to export the data to a more usable format. Log into phpMyAdmin and click on the table name to view the structure of the table. Click on the ‘Export’ tab. Here you will have several options for the format. To back up your data in a format that you could read back into MySQL, use the default ‘SQL’ settings. Other file formats include  $\text{\LaTeX}$ , Excel, Word, comma-separated versions (CSV), and XML.

To get your data into R choose the “CSV for MS Excel” format. Check the boxes to “Put field names in the first row” and “Save as file”, then click ‘Go’. The resulting downloaded file can be read into R using the default settings for `read.delim()`. One way to get your data to STATA is to save it from R using `write.dta(foreign)`.

## Final Thoughts

How safe is this? Most likely, your administrator performs nightly backups of all MySQL databases on the server, so your data is reasonably safe from hardware failure. Another potential source of problems stems from the fact that the Web form is on the Web: theoretically, anyone can get to it. However, if nobody creates an explicit link to the form, a would-be saboteur would have to know the address to get to the page, and even then all they could do is enter extra records. As long as you have a field for the name of the person entering data, you would be able to filter out bogus data. They could possibly fill the server with data, although, unless automated, this would take a *very* long time. They could only see the last few records (however many are displayed at the bottom of the form) but they could not delete records. All of this is very unlikely, but possible. If this “security through obscurity” is not sufficient for your project, you can arrange with your server administrator to “secure” the entire folder containing the Web form so that a password is required for any access.

Collecting data via the Web takes a little more time to set up initially, but the advantages are significant. One recent project using this method used a long form to collect 180 fields of data for 1100 court cases using half a dozen coders, but this would be useful for smaller or larger projects. Go to <http://tinyurl.com/kyavr> to try both of the forms discussed in this article, download the files, and find other useful links.

## References

- Nagler, Jonathan. 1995. “Coding Style and Good Computing Practices.” *The Political Methodologist* 6(2):2-8.



## The L<sup>A</sup>T<sub>E</sub>X Corner

### Posters in L<sup>A</sup>T<sub>E</sub>X

**Mark D. Ramirez and James E. Monogan, III**

Texas A&M University and University of North Carolina, Chapel Hill

*mdramir@politics.tamu.edu monogan@email.unc.edu*

Political scientists are no longer leaving the poster session to the natural sciences. The graduate student poster session is a time-honored tradition of the section's annual meeting and the success of the faculty poster session at the section's last meeting should propel the poster format into greater prominence. Poster sessions can also be a useful teaching tool for students to show-off their course-related research within their home institution (Bohemke 2002, Damon 2003). So how do we go about making professional looking scientific posters?

WYSWYG software (e.g., Powerpoint) lacks the precision and control most L<sup>A</sup>T<sub>E</sub>X users prefer. And for many of us, creating a poster using software such as Adobe Illustrator means learning a new software package. Creating posters in L<sup>A</sup>T<sub>E</sub>X allows us to easily integrate the same tables, figures, and equations from our T<sub>E</sub>X documents into a poster format. Figures, tables, and equations that look great in your manuscript will also look great in your poster. L<sup>A</sup>T<sub>E</sub>X posters also allow the user to focus on that all important content, while implementing and changing the design of the poster can be saved until after the real work is finished. So how do we go about turning our L<sup>A</sup>T<sub>E</sub>X manuscripts into scientific posters?

In this installment of the L<sup>A</sup>T<sub>E</sub>X Corner, we focus on how to make professional scientific posters using two L<sup>A</sup>T<sub>E</sub>X classes: the `sciposter.cls` and `scrartcl.cls`. Both are available from [www.ctan.org](http://www.ctan.org).<sup>1</sup> The flexibility of L<sup>A</sup>T<sub>E</sub>X allows multiple ways to create posters. We chose these two classes due to their ease of use and ability to suit the needs of political science research. The focus of this article will be on construction within L<sup>A</sup>T<sub>E</sub>X rather than issues related to creativity and presentation style—topics covered elsewhere (e.g., Block 1996, Moore, Augspurger, King, and Proffitt 2001).

#### Making a poster using `sciposter.cls`

The first step in creating a poster using the `sciposter.cls` is to set up the preamble of the document. The preamble should start with the document class com-

mand:

```
\documentclass{sciposter}
```

while the orientation of the poster can be set to either `landscape` or `portrait` (the default) as an option within the document class command

```
\documentclass[landscape]{sciposter}
```

Other document class options include changing the poster size and normal font size. The default paper-size is a0 (83.96 cm x 118.82 cm) with a default normal font size of 25pt. However, it is best to change the font and paper size after including the content of the poster—a topic covered in the next section.

After the document class is set up, it is possible to add whatever packages are going to be useful in the body of the poster. The `multicol` is the easiest way to create columns within the poster so it should always be included in the preamble. The `amsmath` and `graphicx` packages are also recommended for including mathematic formulas and managing figures. Packages are included in the same way as in other L<sup>A</sup>T<sub>E</sub>X documents:

```
\usepackage{multicol}
\usepackage{amsmath}
\usepackage{graphicx}
```

Adding the title of the project `\title`, author or coauthors `\author`, institutional affiliation(s) `\institute`, and an e-mail address(es) `\email` is strait-forward in the `sciposter.cls`. It is even possible to add an institutional insignia to the left and right of the title with the `\leftlogo` and `\rightlogo` arguments. The `sciposter.cls`, however, retains the space available for a logo even when the researcher is not using a logo. The `\nologo` command widens the space available at the top of the poster for the title when the user is not including a logo. It is also possible to change the size of the logo as an option to the argument. The code below provides an example of these components.

```
\leftlogo[0.5]{A&Mlogo}
```

<sup>1</sup>The `sciposter.cls` was created by Michael H. F. Wilkinson, Institute for Mathematics and Computing Science, University of Groningen. `scrartcl.cls` is part of the `koma-script` package. Example posters and associated files are available from Monogan's website: <http://www.unc.edu/~monogan/computing/latex/>.

```
\title{Estimating policy responsiveness
with state-space modeling}
\author{Mark D. Ramirez}
\institute{Texas A\&M University}
\email{mdramir@politics.tamu.edu}
```

It is also possible to change the default font sizes for the title, author, and institutional affiliation by redefining the commands `\titlesize`, `\authorsize`, and `\instsize`. For instance, placing the following command in the preamble increases the font size of the title:

```
\renewcommand{\titlesize}{\Large}
```

The preamble is also the place to pay homage to the conference host on your poster by including a footer with the `\conference` command noting the conference name, location, and date.

```
\conference{name, location, date}
```

The final command in the preamble should be `\maketitle`, which generates the title and related components of the poster similar to the `article` document class. Now that the basics are set up, it is time to focus on the content of the poster.

#### *The body of the poster*

The `multicol` package is the workhorse of the `sciposter.cls`. It organizes and sets the columns in the body of the poster. The `multicol` environment implements multiple columns of text, figures, tables, or equations and balances the length of the final columns. Initially, one should focus on putting together the content of the poster and worry about the column balancing and spacing later. Against this advice, we will first describe the `multicol` environment then the basics of including the poster content.

To use the environment one simply types:

```
\begin{multicols}{4}
```

where 4 specifies the number of columns and can be any integer between 1 and 10. To end the `multicol` environment simply type:

```
\end{multicols}
```

It is possible to include text both within and outside of the `multicol` environment. Text outside of the environment will span the length of the poster—minus the page margins. The abstract of the poster shown in Figure 1 is made by placing the text of the abstract between the `\maketitle` and `\begin{multicols}` commands. The font of the abstract text is increased by placing the text of the abstract within curly brackets—to identify a group—and by using the `\huge` command in front of the text. For

instance, the code for the body of the poster in Figure 1 started with:

```
\maketitle
{\huge Although there is evidence...}
\begin{multicols}{4}
```

This places greater emphasis on the abstract and allows colleagues to read it from a greater distance. Another alternative is to place any references after the `multicols` environment allowing them to span the entire bottom of the poster. Experimentation and good taste are key.

The `multicols` environment will automatically format or balance the text and other items within each column. The default column format setting in the environment is `\flushcolumns`, which aligns the top and bottom baselines of each column. However, this automatic balancing reduces the control over formatting valued by  $\LaTeX$  users. There are several ways to change the auto-column formatting. A rather crude alternative is to use the `\raggedcolumns` command, which balances the text within the `multicol` environment without using the top and bottom baseline for alignment. Another alternative is to simply have  $\TeX$  place all of the white space at the end of the last column by adding an asterisk (\*) to the environment command such as:

```
\begin{multicols*}
text
\end{multicols*}
```

otherwise the space is hidden within interline spacing, which is usually preferable. Neither of these alternatives are particularly satisfying.

Some of the usual  $\TeX$  commands are also ineffective in this environment. In an `article` document class, for instance, the `\pagebreak` command could be used to tell  $\TeX$  to break or end a column after a particular line. However, the `\pagebreak` command does not work in the `multicols` environment. Instead, the `\columnbreak` command tells  $\TeX$  where to end a column. Text or items following the `\columnbreak` command will appear in a new column. Columns can also be adjusted vertically using the `\vspace` command. It is also handy to adjust the size of the lines between columns. The width of the lines between columns can be controlled with the command:

```
\setlength{\columnseprule}{0pt}
```

The lines are omitted when the parameter is set to 0pt.

Prior to formatting with the `multicol` environment, the content of the poster should be included in the document. The content of a poster (e.g., text, figures, tables, equations) uses the same  $\LaTeX$  commands as an `article` document class. One of the great things about making a poster in  $\LaTeX$  is the ability to export (e.g., cut and

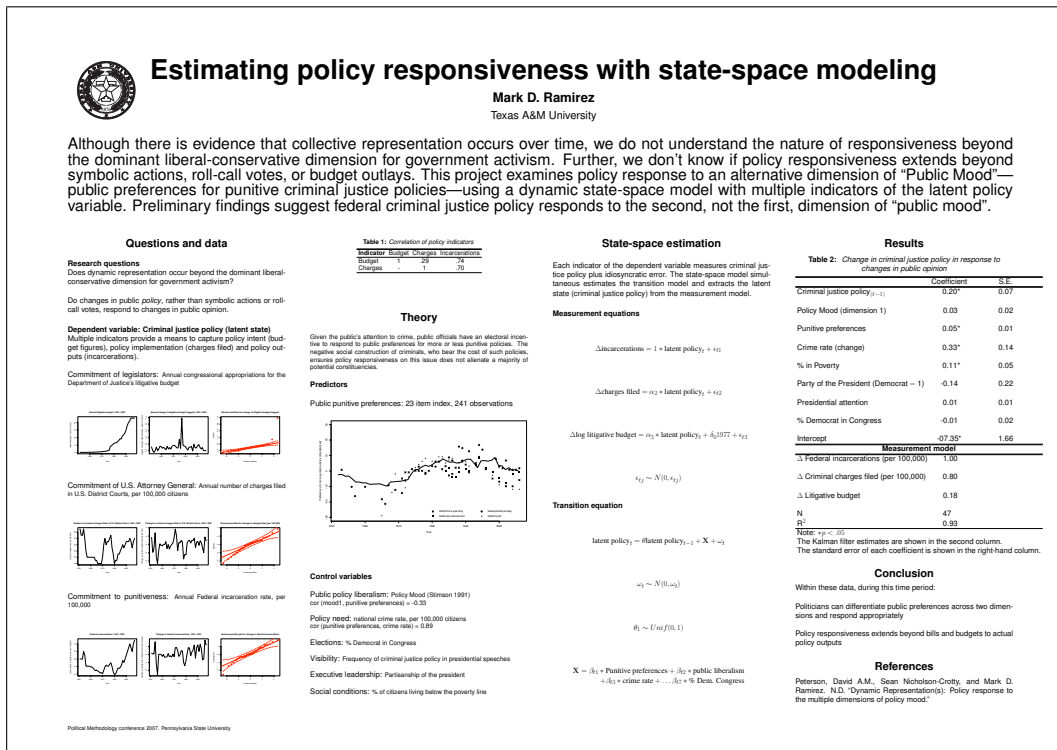


Figure 1: Example of a poster using the sciposter.cls

paste) tables, figures, and equations from a working TeX manuscript into the poster document. To insert a figure, for instance, first invoke the figure environment. Then tell TeX to center the figure within the environment. It's also possible to include an option to re-scale the image—courtesy of the graphicx package.

```
\begin{figure}
\centering
\includegraphics[scale=.5]{figure.eps}
\end{figure}
```

Other options include adding a caption and figure numbering by using the \caption command as is typical in an article document class. Since titles are already apart of the figure, the example poster does not include this option.

A poster generally includes the same sections as an article manuscript: an abstract, research question, theory, data, methods, results, and conclusion. The sciposter.cls uses the same \section and \subsection heading commands as the more familiar article document class. The example poster above contains five sections: Questions and data, theory, estimation, results, and conclusions. These headings are made using the \section\* command. The subheadings (e.g., Predictors, measurement equations) are made by using the bold command \textbf simply as a matter of personal taste.

The inclusion of text is the same as in other TeX documents. Changing the style and fonts of text is also the same. The default font for an a0 poster is 25pt. The 25pt font is large enough for most eyes to read from about four feet away. To change the normal font size, simply indicate the desired size in the document class options:

```
\documentclass[landscape,30pt]{sciposter}
```

This will set the normal size font to 30pt and adjust the other font sizes (e.g., \small, \Large) accordingly. Having multiple font sizes within the body of the poster is accomplished via the same 10 font sizes most TeX users are already familiar with: \tiny, \small, \Large, \huge .... For instance, it might be useful to use a smaller font to describe the details of variables or technical aspects of the research project and a larger font to highlight key points of the research. With a normal font of 25pt, the \small font size (21pt) is still readable from a few feet away. The \small font is used in the example poster to describe the details of each of the variables. Changes in font sizes and text styles, however, should be used cautiously.

*Adjusting the style*

After the content is laid out and the columns are adjusted, it is now time to go back and play with colors and style of the poster. All of these options should be placed in the preamble of the document, but utilizing them is best

saved until after the content is complete. These commands will allow the user to change the look of section headers, the color scheme of the poster, and poster format.

It is possible to differentiate sections within the poster by enclosing section headers in boxes or underlining them. The document class command provides these options. Section headers can be within shaded boxes of color `boxedsections`, underlined `ruledsections`, or set as plain text `plainsections`. For example, the code for the plain text section headers in the example poster is:

```
\documentclass[landscape,plainsections]{sciposter}
```

The user may also want to change or add colors to the poster. The color scheme within the poster is defined by four commands:

```
\definecolor{mainCol}{rgb}{1,1,1}
\definecolor{TextCol}{rgb}{0,0,0}
\definecolor{SectionCol}{rgb}{0,0,0}
\definecolor{BoxCol}{rgb}{0.9,0.9,1}
```

The `mainCol` option sets the background color of the poster with the default background being white. The user may also change the default color of the text or section headers—black—using the `TextCol` or `SectionCol` options respectively. If the poster contains boxes—typically for section headers—than the box color can be set using the `BoxCol` argument. The color codes (i.e. 1, 1, 1 for white) are available in “Chroma: a reference book of L<sup>A</sup>T<sub>E</sub>X colors” available at [www.ukern.de/tex/chroma.html](http://www.ukern.de/tex/chroma.html).

Finally, the default setting in the `sciposter.cls` does not indent paragraphs. To override this default, add `\setlength{\parindent}{2cm}` in the preamble of the document. Of course, the user can set the paragraph indention at any value they like. It is also possible to change the margins of the poster using the `\setmargins{}` command in the preamble. These changes should allow the user to format the poster in a manner that best communicates their ideas.

The final command in the poster is the same as other L<sup>A</sup>T<sub>E</sub>X documents:

```
\end{document}
```

## Making posters using `scartcl.cls`

An alternative approach to making posters is to use `scartcl.cls` to format a small page that scales correctly for a poster size document. Though this document is compact at 390mm x 319.2mm, blowing it up by a factor of 2.828 easily turns the poster into an A0 size document, while blowing it up by a factor of 2.0 creates an A1 size document.

To format a poster using these directions, a user needs to have the `koma-script` package installed from CTAN, and the style file `scrpos.sty`. Once these files are loaded, the preamble of a poster should read as follows:

```
\documentclass[12pt]{scartcl}
\usepackage{scrpos}
```

Any special packages you need to use can also be loaded, but all packages essential to the file structure (i.e., `multicol`) are already loaded by the style file.

Afterward, several commands pass parameters to the style file to format your poster correctly. Please note that spaces between lines of code should be avoided in this environment as this will be interpreted as a new paragraph and disturb the placement of objects on the page. Thus, this code cannot be separated.

```
\begin{document}
\background{3}
\titleEPS{Letter1.eps}{The University of North
Carolina at Chapel Hill}{Robustness of Quantal
Response Models with Omitted Variables}
{James E. Monogan III}
{\frontmatter}
```

In the `background` command, the user should enter the number of columns the poster will include so they can be spaced equally. To create the title, the user has two options: `\titleEPS` creates a header with an institutional logo in `.eps` format; the command expects the logo filename first, followed by the institution name, poster title, and finally author name. Alternatively, `\titleNO` creates a header without the logo, expecting all parameters except the postscript filename. Finally, the command `\frontmatter` makes some necessary spacing adjustments.

Content in `scartcl.cls` needs to be placed in individual boxes within columns. This may offer an advantage in that it forces the author to think about a presentation style that can be quickly and easily followed by a reader. To start a new column, simply type `\begin{pcol}` to start the poster column environment and place `\end{pcol}` at the end of that column. An odd feature of this environment is that empty lines outside of the column environment can distort the presentation. Hence, the end of one poster column needs to be followed by the beginning of the next poster column environment *on the very next line*, just like the code at the start of the document. Comments might be especially useful for handling this. For example, starting a new column with code such as “`\begin{pcol}%Column 1`” could help the author keep track of the file structure. Within the poster column environment, place several boxes to include your content, starting with `\begin{pbox}` and ending with `\end{pbox}`. The poster box environment is flexible and

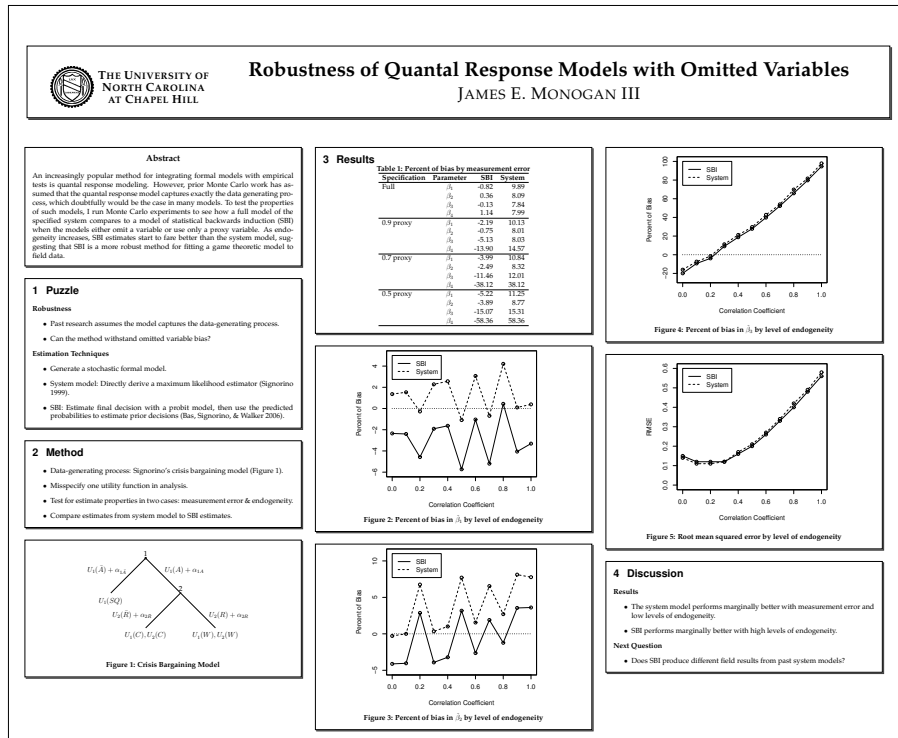


Figure 2: Example of a poster using the scartcl.cls

allows nearly all of the features of regular  $\text{\TeX}$ -processing, including paragraph text, bulleted lists, tables, and figures.

Although `scartcl.cls` accommodates tables, the table environment itself does not work. Hence, to incorporate a table from a paper, use the exact same code, except delete the table environment and write-in the title with regular text. For example, the following code produced the table in the poster featured in Figure 2:

```
\begin{center}
\textbf{Table 1: Percent of bias by measurement
error}
\begin{tabular}{lrrr}
%INSERTED TABULAR CODE FROM PAPER
\end{tabular}
\end{center}
```

A nice feature of this class is that it can incorporate postscript figures, which is what many people use in their articles. In fact, the poster featured in Figure 2 includes a  $\text{\LaTeX}$ -produced game tree using the `egameps` package, which produces postscript figures that are not automatically accommodated by PDF- $\text{\LaTeX}$ . More commonly, though, a writer will want to include a postscript figure created from an outside source, just as in a paper. The following code produced a figure in the working example:

```
\begin{center}
\includegraphics[width=4.5in]{b1.ps}
\vspace{-.1in}\textbf{Figure 2: Percent of bias
in  $\hat{\beta}_1$  by level of endogeneity}
\end{center}
```

Finally, at the end of the document, include the following code:

```
}\backmatter
\end{document}
```

When finished, the writer can create the document by applying  $\text{\TeX}$ ify, using DVIPS, and finally using `ps2pdf`. This three-step process allows a user to create a poster with proper spacing that also incorporates postscript files.

### Conclusion

So which package should a user choose? The `sciposter.cls` is essentially a column-based format, while the `scartcl.cls` relies on placing the content in boxes. However, the differences are more than just a matter of taste. On the one hand, `sciposter.cls` requires less start-up time and coding, while `scartcl.cls` can create bold-looking posters conveniently once the style file `scrpos.sty` is installed. The `scartcl.cls` also seems well-suited to handle a wider range of graphic sizes, while the

`sciposter.cls` user will have to size figures within the limits of the column size. Both packages undoubtedly offer a sharper mode of poster publication that is better suited for scientific communication than any of the commercial alternatives.  $\TeX$  also allows users the flexibility to use other poster styles (i.e., `a0poster.cls`) or create their own. Regardless, these are two alternatives that should have most *TPM* readers constructing professional posters in no time.

## References

- Block, Steven M. "Do's and Don'ts of Poster Presentation." *Biophysical Journal* 71(December):3527-3529.
- Boehmke, Frederick J. 2002. "Teaching Advance Graduate Methods: Using Poster Sessions." *The Political Methodologist* 11(1):2-3.
- Damron, Danny. 2003. "Facilitating Critical Feedback on Capstone Papers through Student Poster Sessions." *PS: Political Science and Politics* 36(4):777-780.
- Moore, Linda Weaver, Phyllis Augspurger, Margaret O'Brien King, and Charlotte Proffitt. "Insights on the Poster Preparation and Presentation Process." *Applied Nursing Research* 14(2):100-104.

## Book Review

### Review of Trevor Hastie, Robert Tibshirani, and Jerome Friedman The Elements of Statistical Learning

David Siroky

Duke University

*d.siroky@duke.edu*

**The Elements of Statistical Learning.** Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Springer Series in Statistics, 2001, 533 pages. \$89.95, ISBN 978-0-387-95284-0 (hardcover).

This is a statistically significant book. It seeks nothing less than to provide a statistical framework for a wide range of methods and techniques that have been developed through the contributions of several disciplines, including artificial intelligence, machine learning, pattern recognition, statistics and computer science. This book tries to show, in my view convincingly, that these developments can be seen as a less or more coherent whole through the lens of statistical theory (cf. Breiman 2001).

Political methodologists and political scientists more generally will find these methods of interest as substitutes or complements in most of the usual classification, censored survival or regression problems that we encounter in studies of voting, conflict, trade, regime change, etc. Whether one is interested in prediction, inference or description, the techniques and algorithms reviewed in this book are worth exploring in addition to, and for some objectives instead of, the more conventional approaches. One way these methods differ from more familiar tools is the explicit focus on three issues—the curse of dimensionality, the tradeoff between simplicity and accuracy, and the multiplicity of good

models—associated, respectively, with Bellman, Occam and Rashomon (Breiman 2001: 200, 206-209). These three problems are philosophical, at one level, yet they influence how we analyze data, and what we claim to have demonstrated with our analysis of those data, each and every day.

In the author's words, this is a book about *learning from data*. The book differentiates supervised and unsupervised learning, although the former receives the lion's share of attention (the first 13 of 14 chapters). Both supervised and unsupervised learning require a training and test set of data: the difference lies in what one knows and does not know about the problem. In the supervised case, one knows the output; in the unsupervised case, the outcome is unknown. Knowing the outcome allows one to create a *learner* that predicts the output of 'unseen' objects from a test set of data given the inputs. The problem is called supervised because the known outcome guides the learning process. The quality of the learner is typically judged on the basis of its ability to predict the output values or classes of new objects. Examples of methods for the supervised learning problem include feed-forward neural networks, nearest neighbor methods and support vector machines. The final chapter is devoted to the unsupervised (or unlabeled) learning problem. The objective here is to depict the patterns and relationships among the predictors and thus to uncover any clusters in the data. Examples of this form

of learning problem include principal components analysis, multidimensional scaling and cluster analysis.

The book covers a lot of ground and the reader will become well acquainted with a number of methods and the theory behind them. After a theoretical overview, the book begins with linear methods for regression and classification (Chapters 3-4) and covers a range of topics, including Basis Expansions and Regularization Methods (5), Kernel Methods (6), Model Assessment, Selection, Inference and Averaging (7-8), Additive Models, Boosting, Neural Networks, Support Vector Machines, Flexible Discriminants and Prototype Methods (9-13). The final chapter covers unsupervised learning problem tools such as self-organizing maps, association rules, exploratory projection pursuit (14). Many of these methods, especially the Lasso, Classification and Regression Trees (CART), Projection pursuit regression models (PPR), Multivariate Adaptive Regression Splines (MARS) and Boosting methods have been fruitfully applied to diverse areas from finance to face recognition. Many of the techniques have obvious relevance to high-dimensional political science problems across the sub-fields, but have not yet been much explored.

Four qualifications are in order. First, those searching for a data analysis book will be disappointed. Although the book comes with a number of data sets, the focus is on theory. Not much is offered (at least to students) in the way of code to reproduce the figures or apply the methods mentioned. The reader searching for this will not have to look far, however: **Matlab** and **R** have some excellent versions

for most of these methods, and sometimes several versions. Second, the book is both long and deep, thus requiring a significant time investment. The level is probably appropriate for an intermediate-advanced student of political methodology who has completed several courses in theory and a few data analysis projects. Third, although the authors distinguish between unsupervised and supervised learning problems, they make no mention of semi-supervised learning, which uses both identified and unidentified (or labeled and unlabeled) data and has a number of practical applications. Fourth, the book is also a bit biased toward those learning methods on which the three authors have worked in their past work, but this is not necessarily bad since it makes their presentation of that material compelling and insightful.

Despite these comparatively minor complaints, some of which might be easily enough incorporated in the second edition, the price (89.95 USD from Springer, 59.95 from Amazon) is reasonable for a book with over 200 (color!) figures. Political methodologists have largely ignored developments in statistical learning. *Elements of Statistical Learning* is a well written and illustrated text that should help redress that problem, hopefully encouraging a fruitful cross-pollination.

## References

- Breiman, Leo. 2001 "Statistical Modeling: The Two Cultures," *Statistical Science* 16(3):199-231.

## Announcements

### ICPSR Summer Program in Quantitative Methods of Social Research 23 June - 15 August 2008

**William G. Jacoby**

Director, ICPSR Summer Program

[jacoby@msu.edu](mailto:jacoby@msu.edu)

The Inter-University Consortium for Political and Social Research (ICPSR) is pleased to announce the 2008 Summer Program in Quantitative Methods of Social Research. The main component of the Summer Program is held on the campus of the University of Michigan, in Ann Arbor. Lectures and workshops on a wide variety of topics in research design, quantitative reasoning, statistical methods, and data processing are presented in two four-week sessions. The first session runs from June 23, 2008 until July 18, 2008. The second session runs from July 21, 2008 until August

15, 2008. The contents of the two sessions are largely independent of each other, although some second-session workshops do assume that participants are familiar with material from first-session courses. The 2008 ICPSR Summer Program will also offer a number of three- to five-day workshops throughout the summer. These shorter workshops are held in a variety of locations: Amherst, MA; Ann Arbor, MI; Bloomington, IN; Chapel Hill, NC; and New Haven, CT.

Further information about the ICPSR Summer Program in Quantitative Methods of Social Research is available on our

web site: <http://icpsr.umich.edu/sumprog/>. The complete set of course descriptions will be posted by February 8, 2008 and we will begin accepting applications on February 20, 2008. Please feel free to e-mail us with any further questions, at [sumprog@umich.edu](mailto:sumprog@umich.edu).

### 2008 ICPSR Summer Program Course List

All classes are held in Ann Arbor, MI, unless otherwise noted.

#### First Session (June 23 - July 18)

##### *Lectures*

Introduction to Computing  
 Mathematics for Social Scientists, I  
 Mathematics for Social Scientists, II  
 Data Mining  
 Statistical Computing Using R/S  
 Statistical Graphics for Visualizing Data

##### *Workshops*

Introduction to Statistics and Data Analysis, I  
 Quantitative Analysis of Crime and Criminal Justice  
 Quantitative Research on Race and Ethnicity  
 Quantitative Historical Analysis  
 Rational Choice Theories of Politics and Society  
 Game Theory: Basic and Advanced Topics  
 Regression Analysis I: Introduction  
 Regression Analysis II: Linear Models  
 Regression Analysis III: Advanced Methods  
 Multivariate Statistical Analysis  
 Maximum Likelihood Estimation for Generalized Linear Models  
 Introduction to Applied Bayesian Modeling for the Social Sciences

#### Second Session (July 21 - August 15)

##### *Lectures*

Introduction to Computing  
 Matrix Algebra  
 Complex Systems Models in the Social Sciences  
 Missing Data: Statistical Analysis of Data with Incomplete Observations

##### *Workshops*

Introduction to Statistics and Data Analysis, II  
 Regression Analysis II: Linear models  
 Categorical Analysis  
 Time Series Analysis  
 Simultaneous Equation Models  
 "LISREL" Models: General Structural Equations with La-

tent Variables  
 Longitudinal Analysis  
 Advanced Topics in Maximum Likelihood Estimation  
 Advanced Topics in Bayesian Methods

##### *Shorter Workshops*

Scaling Methods for Survey Data (June 2-13, Co-sponsored by the Survey Research Center Summer Institute in Survey Research Techniques)  
 Integrating Biomarkers into Population-Based Research (June 2-6, Chapel Hill, NC)  
 Panel Data Analysis using Stata (June 2-6)  
 Analyzing Developmental Trajectories (June 9-12, Amherst, MA)  
 Categorical Data Analysis (June 9-13)  
 Network Analysis: An Accelerated Introduction (June 13-15)  
 Collaborative Psychiatric Epidemiology Surveys (June 16-20)  
 Project on Human Development in Chicago Neighborhoods (June 16-20)  
 Latent Trajectory Growth Curve Analysis (June 16-20, Chapel Hill, NC)  
 Introduction to Applied Bayesian Statistics for Social Scientists (June 23-27, Chapel Hill, NC)  
 Hierarchical Linear Models: Introduction (June 23-27, Amherst, MA)  
 Analyzing Multilevel and Mixed Models using Stata (June 23-27)  
 Hierarchical Linear Models II (July 7-11)  
 Designing, Conducting, and Analyzing Field Experiments (July 9-11, New Haven, CT)  
 Introduction to Spatial Regression Analysis (July 14-18, Bloomington, IN)  
 Welfare, Children, and Families: A Three-City Study (July 21-23)  
 Using Secondary Data for Analysis of Marriage and Family (July 24-25)  
 Network Analysis: An Introduction (July 21-25)  
 Structural Equation Models (July 21-25)  
 Providing Social Science Data Services: Strategies for Design and Operation. (August 11-15)  
 Introduction to Multilevel Models Using SAS (August 11-15, Chapel Hill, NC)

### Contact Information for the ICPSR Summer Program:

Mail: ICPSR Summer Program, P.O. Box 1248, Ann Arbor, MI 48106-1248, USA.

Web: <http://icpsr.umich.edu/sumprog/>

E-mail: [sumprog@umich.edu](mailto:sumprog@umich.edu)

Voice: (734) 763-7400

Fax: (734) 647-9100.



## Essex Summer School in Social Science Data Analysis & Collection 7 July - 15 August 2008

**Thomas Plümper**  
University of Essex  
*tpluem@essex.ac.uk*

The Essex Summer School in Social Science Data Analysis has taken place annually at the University of Essex for over 40 years. The summer school is designed to develop the skills of social scientists in various aspects of statistical, mathematical and analytical skills which are increasingly required in modern research. The first Summer School at the University of Essex, held in 1967, was sponsored by UNESCO. Subsequent schools have been sponsored by the Nuffield Foundation, and, latterly, by the European Consortium for Political Research. With over 700 participants, the summer school is not only the oldest but also one of the biggest of its kind.

The Summer School provides an opportunity for social scientists to make contact with professional social scientists from other countries. Besides classroom instruction and daily laboratory work, there are evening seminars on research topics of special interest. The 2008 Summer School offers courses on Logit and Probit Analysis (Instructor: Marco Steenbergen), Structural Equation Models (Pe-

ter Schmidt), Duration Models (Matt Golder), Time Series Analysis (Harold Clarke), Bayesian Econometrics (Simon Jackman), Causality (Nathaniel Beck), Qualitative Comparative Methods (James Mahoney), among others. In addition, the 2008 summer school presents courses on formal theory and modelling and on discourse theory.

For further information please see the website at <http://www.essex.ac.uk/methods/> or contact the summer school office at [sumsch@essex.ac.uk](mailto:sumsch@essex.ac.uk). If you wish to suggest an additional course please contact Thomas Plümper at [tpluem@essex.ac.uk](mailto:tpluem@essex.ac.uk).

### For further details:

Mail: Essex Summer School, University of Essex, Colchester CO4 3SQ, UK

E-mail: [sumsch@essex.ac.uk](mailto:sumsch@essex.ac.uk)

Web: [www.essex.ac.uk/methods/](http://www.essex.ac.uk/methods/)

## The Washington University Summer Institute on the Empirical Implications of Theoretical Models June 9 - 26, 2008

**Steven Smith**  
Washington University in St. Louis  
*smith@WUSTL.EDU*

Washington University's Weidenbaum Center and Department of Political Science sponsor a summer institute on the problems of testing theoretical models of politics. The institutes are designed for advanced graduate students and junior faculty whose research and teaching would benefit from training seminars on the link between methods of empirical analysis and theoretical models. The 2008 EITM Summer Institute is comprised of five seminars over a three-week period, a basic seminar and four advanced seminars (the last week will provide students with an option to take one of two seminars).

Participants selected for this program will receive a \$1,000 stipend. This will be used to cover their travel, room,

and board. Up to 25 subsidies are available for full-time participants. Participants shall be responsible for any and all costs of living expenses in St. Louis for the three-week long institute. Funding is provided by the National Science Foundation.

### *Application information*

Application deadline: February 15, 2008

Applicants should submit the following:

1. A curriculum vita with name, contact information, current location, and position

2. A short, 1-2 page statement of why you would like to participate in the program
3. A transcript (optional)
4. 1 or 2 letters of endorsement (optional, but strongly recommended for graduate students)

Materials can be sent via email to [eitm@artsci.wustl.edu](mailto:eitm@artsci.wustl.edu) in either Word or PDF format. If you would prefer to mail them, the address is EITM Application, Weidenbaum Center, Campus Box 1027, Washington University, St. Louis, MO, 63130-4899.

#### *2008 Seminars*

- June 9-12  
Theoretical and Methodological Foundations  
Professors Randall Calvert and Andrew Martin  
Washington University in St. Louis
  - June 13-14  
Random Utility Models and Strategic Choice  
Professor Mark Fey  
University of Rochester
  - June 16-18  
Operationalizing the Spatial Model  
Simon Jackman  
Stanford University
  - June 19-21 , 23  
Experimental Applications  
Professor Rick Wilson  
Rice University
  - June 24-26  
Issues in Testing Positive Theories of Legislative Politics  
Keith Krehbiel  
Stanford University
- or
- International Relations Applications  
Robert Walker  
Washington University in St. Louis

## EITM at Duke University June 16 - July 11, 2008 Call for Participants and Mentoring Faculty-in-Residence

**John Aldrich**  
Duke University  
[aldrich@duke.edu](mailto:aldrich@duke.edu)

Duke will host the seventh annual Summer Institute on **EITM: Empirical Implications of Theoretical Models** this summer, June 16th through July 11th, 2008. Funded by the National Science Foundation (NSF), this program seeks to leverage the complementarity between formal models and empirical methods. EITM is training a new generation of scholars to integrate theoretical models more closely, effectively, and productively with empirical evaluation of those models. The Summer Institutes are highly interactive training programs for advanced graduate students and junior faculty. They are led by teams of scholars from across the discipline who are working at the forefront of such empirical-theoretical integration.

### Call for Participants

We welcome applications from advanced graduate students who have passed all qualifying exams, preferable with a completed dissertation prospectus or plan but not yet at writing-up stages. Graduate students will benefit most from the program if they are committed to using both theoretical models and empirical data in their dissertations. They should have some training in both formal methodology and quantitative analysis, and advanced training in at least one of these areas. We also welcome applications from junior faculty looking to improve their defended dissertation in a direction that incorporates EITM, or who are embarking on an EITM-style post-dissertation project. We will base admission substantially on the quality and potential of research proposed in the application. We intend to accept about 25 participants. Applicants will be notified of admission status by email by March 31.

#### *Application information*

Application deadline: February 29th

A complete application consists of the following four components:

1. **Curriculum Vita** with name and contact information, current location and position. If you are a student, the CV should indicate your current status in graduate school (year in program, whether you've passed qualifying exams, whether you've defended a dissertation proposal).

2. Description of your **EITM research proposal** (5-10 pages). We will base admission substantially on the quality and potential of this proposal—particularly its integration of theoretical modeling and empirical testing.
3. Brief (1-2 page) **statement of interest** and purpose in applying for the summer program.
4. **Two letters of recommendation** sent as email attachments to [eitm@duke.edu](mailto:eitm@duke.edu). Please ask your letter writers to place your name and “EITM” in the e-mail message’s subject heading and to email the letters directly to us.

Please submit application materials as **PDF** or **MS-WORD attachments** via e-mail to [eitm@duke.edu](mailto:eitm@duke.edu).

Applicants will be notified of our acceptance decisions (by e-mail) by March 31.

#### *Financial Support*

There are no fees or tuition. Dormitory lodging, meals and domestic travel expenses will be provided.

For more information, go to [www.poli.duke.edu/eitm/](http://www.poli.duke.edu/eitm/).

### **Call for Mentoring Faculty-in-Residence**

An important feature of this year’s program is that the efforts of our regular Lecturing Faculty will be augmented by a team of Mentoring Faculty-in-Residence (MFR).

**Responsibilities:** Each MFR will have a mentoring group, consisting of a small number of EITM participants. We expect there will be six MFRs, each assigned no more than five mentees. A primary responsibility of each MFR is to work closely with his/her mentees, helping them integrate ideas and methods from the Institute into their own projects. Secondly, MFRs will work closely with lecturing faculty to develop a set of teaching materials for a semester-length course reflecting EITM principles. These course materials will be made available on-line for use by instructors throughout the discipline. MFRs will also give presentations of their own current research. MFRs must commit to participating in the entire four-week Institute.

**Qualifications:** We expect that MFRs will be drawn from the ranks of tenure-track or recently tenured political science faculty who use EITM methods in their research. The most important qualification for the Mentoring Faculty-in-Residence is experience with developing, executing and publishing EITM research projects. Participation in past Summer Institutes (either our rotating Institute or our sister program at Washington University in St. Louis) is helpful but not required. Experience with graduate teaching in formal and quantitative methods is similarly helpful, but also

not essential (although MFRs should have the appropriate preparation for teaching in these areas.) Applications from faculty whose home institutions do not currently offer EITM training are particularly welcome. We hope that participating in the Summer Institute will enable the Mentoring Faculty to develop their teaching and mentoring expertise, expand the number of universities that teach EITM methods, and deepen course offerings where EITM is already part of the curriculum.

**Deadline:** The application deadline is February 29, 2008.

**Application:** Complete applications consist of the following components:

1. **Curriculum Vita** with name and contact information, current location and position, and names of two people we may contact as references if needed.
2. Brief (1-3 page) **statement of interest and purpose** in applying for a position as Mentoring Faculty-in-Residence. Please indicate any skills and background that would be particularly useful as an MFR.
3. Brief (1-3 page) **description of current research**.

Application materials should be sent as **PDF** or **MS-WORD attachments** via e-mail to [eitm@duke.edu](mailto:eitm@duke.edu). Please indicate “MFR application” in the subject line.

**Financial support:** Domestic travel expenses, dormitory housing and meals are covered, along with stipend of approximately \$7,725.

**Child Care:** We intend to offer child care for MFRs. Details will be made available in the near future.

**Notification:** Application decisions will be made by March 31, 2008.

For more information, go to [www.poli.duke.edu/eitm/](http://www.poli.duke.edu/eitm/).

### **Content of the EITM Summer Institute**

EITM Summer Institutes are organized into 4 week-long modules, each with a different substantive and methodological focus. This year’s fourth week will be split between a mini-module and participant presentations. This year’s EITM program and faculty (as so-far committed) are:

- Week One (June 16 - June 21): Institutions and Institutional Analysis  
Lead Lecturers: John Aldrich, Duke University and Arthur Lupia, University of Michigan

This unit explores Empirical Implications of Institutional Models. It traces the origins, successful development, and potentially problematic aspects of the New Institutionalism literature, combining lectures and innovative class activities to understand modern studies of the causes and consequences of institutional choices. Activities use examples of bureaucratic performance, voter competence, Congressional organization, election laws, separation of powers, coalition bargaining, jury decision-making, political development, etc. The week also addresses (a) some constructive debates on the appropriateness to political contexts of the modern proliferation of equilibrium concepts and statistical-estimation procedures, (b) how incomplete information affects institutional efficacy, and (c) innovative data-collection methods. Past work teaches critical lessons, but this week aims to improve the scientific and social value of new research, helping to shape the new *new institutionalism*.

- Week Two (June 23 - June 27): Experimentation in the Social and Behavioral Sciences  
Lead Lecturer: Wendy Wood, Duke University, and a series of special lecturers.

This week will be composed of a series of presentations and projects about experimentation in the social sciences. The week will begin with an overview of experimentation and research design and will then consider the use of experimentation in political science, social psychology, experimental economics, and political psychology. We will take advantage of the re-

sources at Duke, including training in the use of software commonly used for the design and implementation of experiments and the running of experiments in Duke's DIISP lab, exposure to psycho-physical lab techniques, and training in neuron-experimentation and the use of field experiments.

- Week Three (June 30 - July 4): Complexity: Computational Models and Social Networks  
Lead Lecturers: Scott de Marchi, Duke University and James Fowler, University of California, San Diego

This week will provide a practical and hands-on introduction to using computational methods, focusing on how they relate to closed-form analytical models and empirical tests. As a way of grounding the key topics in computational modeling, the module will cover social network theory and the techniques used to analyze politically-relevant networks (with a substantive focus on problems such as congressional cosponsorships and judicial citations). A key feature of this treatment will be to demonstrate how one connects the analysis of social networks with specific hypotheses and tests on observed data. Finally, the module will also provide one additional substantive unit based on the interests of guest faculty. In previous years, this has included computational models of elections, international conflict, and bargaining.

- July 7 - July 11  
Participant Project Workshop and Mini-Module  
TBA

## Section Activities

### A note from our Section President

In late February, a contingent from the Society for Political Methodology, including several members of our Undergraduate and Graduate Methodology Committee, our past president Jan Box-Steffensmeier, and myself, will be attending the APSA's fifth "Teaching and Learning Conference" in San Jose. We plan both to listen to what others are doing in the area of quantitative methods pedagogy, and also to present some of our own ideas on how SPM can create and share material to improve instruction and, perhaps, reduce the number of times one must figure out how to put  $\hat{\beta} = [X'X]^{-1}X'y$  onto a projector slide.

This initiative represents something of a departure for SPM, which has historically focused primarily on advancing the state of the art in research techniques. Yet, with the exception of those in our membership who are outside the academic environment and those who have yet to

begin academic positions, virtually all of us spend a great deal of effort working on methods instruction, often ranging from the most elementary introductions to the interpretation of statistical tables, to post-doctoral training in venues such as ICPSR and EITM.

In many cases, this instruction is done very well: each year, when the APSA issues its list of teaching award winners, I am pleased to see how many names I recognize from our ranks. But the level of effort required is substantial—no skimming the readings ten minutes before class and walking in to say "So, what does X have to say?" (not that anyone would ever do that...). Meticulously proof-read handouts, realistic problem sets, commented statistical package scripts and yes, seemingly endless projector slides are all part of a thorough classroom presentation.

It is nice to do this well, nice to get the occasional recognition, particularly when it comes from students

or peers who are not teaching quantitative methods, but we also do it because it serves a vital collective function: preparing future methodologists. Ours is a cumulative and technical discipline and consequently cannot simply be re-invented or re-discovered by each new generation, nor can it be picked up by reading a couple of books and then filling in the rest through observation and common sense. Instead, like engineering, mathematics and the natural sciences, it has to be taught gradually, in systematic layers of ever increasing complexity, and, ideally, starting fairly early in the undergraduate curriculum. Only by doing this, and doing it well, will we have students we can push to the cutting edge of our craft in the short period of time available for graduate education.

In October I met with several people at the National Science Foundation to discuss whether funding might be available for these initiatives, and received a great deal of encouragement and suggestions. There are multiple ways

we might do this, and in discussions at the “Teaching and Learning Conference” and elsewhere we intend to formulate a strategy. We will keep you posted in future issues of TPM and the PolMeth listserv.

One item of business: One of the things for which the President is responsible is appointing committees. We’ve got a *lot* of committees! I’ve still got a small number of vacancies to fill, and in any case the whole process starts again in August. We’re trying very much to expand participation, so if you would be interested in working on one of these (most do their work exclusively by email), please send me a note at [schrodt@ku.edu](mailto:schrodt@ku.edu). A list of the Society’s standing committees can be found on the <http://polmeth.wustl.edu> web site.

Cheers,

Philip A. Schrodt  
*University of Kansas*

Department of Political Science  
2010 Allen Building  
Texas A&M University  
Mail Stop 4348  
College Station, TX 77843-4348

*The Political Methodologist* is the newsletter of the Political Methodology Section of the American Political Science Association. Copyright 2007, American Political Science Association. All rights reserved. The support of the Department of Political Science at Texas A&M in helping to defray the editorial and production costs of the newsletter is gratefully acknowledged.

Subscriptions to *TPM* are free to members of the APSA's Methodology Section. Please contact APSA ([http://www.apsanet.org/section\\_71.cfm](http://www.apsanet.org/section_71.cfm)) to join the section. Dues are \$25.00 per year and include a free subscription to *Political Analysis*, the quarterly journal of the section.

Submissions to *TPM* are always welcome. Articles should be sent to the editors by e-mail ([tpm@polisci.tamu.edu](mailto:tpm@polisci.tamu.edu)) if possible. Alternatively, submissions can be made on diskette as plain ascii files sent to Paul Kellstedt, Department of Political Science, 4348 TAMU, College Station, TX 77843-4348. See the *TPM* web-site, <http://polmeth.wustl.edu/tpm.html>, for the latest information and for downloadable versions of previous issues of *The Political Methodologist*.

*TPM* was produced using L<sup>A</sup>T<sub>E</sub>X on a Mac OS X running iTeXMac.



**President: Philip A. Schrodt**

University of Kansas  
[schrodt@ku.edu](mailto:schrodt@ku.edu)

**Vice President: Jeff Gill**

Washington University in St. Louis  
[jgill@wustl.edu](mailto:jgill@wustl.edu)

**Treasurer: Jonathan Katz**

California Institute of Technology  
[jkatz@hss.caltech.edu](mailto:jkatz@hss.caltech.edu)

**Member-at-Large: Robert Franzese**

University of Michigan  
[franzese@umich.edu](mailto:franzese@umich.edu)

**Political Analysis Editor: Christopher Zorn**

Pennsylvania State University  
[politicalanalysis@la.psu.edu](mailto:politicalanalysis@la.psu.edu)