

# The Political Methodologist

NEWSLETTER OF THE POLITICAL METHODOLOGY SECTION  
AMERICAN POLITICAL SCIENCE ASSOCIATION  
VOLUME 19, NUMBER 2, SPRING 2012

## Editors:

JAKE BOWERS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN  
*jubowers@illinois.edu*

WENDY K. TAM CHO, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN  
*wendycho@illinois.edu*

BRIAN J. GAINES, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN  
*bjgaines@illinois.edu*

## Editorial Assistant:

ASHLY ADAM TOWNSEN, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN  
*townsen5@illinois.edu*

## Contents

<b>Notes from the Editors</b>	<b>1</b>
<b>Articles</b>	<b>2</b>
John E. Jackson: <i>On the Origins of the Society</i> . . .	2
David Darmofal: <i>Modeling Spatial Heterogeneity with Geographically Weighted Regressions in R</i>	7
Tsung-han Tsai and Jeff Gill: <i>superdiag: A Comprehensive Test Suite for Markov Chain Non-Convergence</i> . . . . .	12
McKendon Laffleur and John Beiler: <i>President-Parser: Automated Mark-Up Utility</i> . . . . .	18
<b>A Note from Our Section President</b>	<b>21</b>

Welcome to the latest issue of *The Political Methodologist*. We have a nice round up of articles to bring you section news, provide reflections, and supply you with research tools. Our first article, by John Jackson, brings us back to our roots. It recounts the history of the society from the perspective of one of our founding and distinguished mem-

bers. This article allows us a chance to enjoy a look back at our history while also giving us important perspective as we move forward into the future. The next three articles provide software introductions and tutorials. David Darmofal reviews the method of geographically weighted regressions (GWR). He provides an overview of the method that includes a tutorial on implementing these methods in the R `spgwr` package. Tsung-han Tsai and Jeff Gill follow with an introduction to the `superdiag` function in R. This function integrates all of the standard empirical MCMC convergence diagnostics in one command. McKendon Laffleur and John Beiler also introduce software. `PresidentParser` parses speech acts and prepares them for processing in the Profiler Plus environment. Thanks to our contributors for helping us all be more productive and efficient with the software component of our research. We end, as always, with a note from our section president. We welcome Rob Franzese to the post. His note introduces himself, provides a status report for the section, and sets goals for his presidency.

A big thanks to all of our contributors. We appreciate the time and thought you have given to these articles. If you have ideas for future articles in *TPM*, do not hesitate to contact us. Until the next issue, happy reading!

*The Editors*

## Articles

### On the Origins of the Society

**John E. Jackson**

University of Michigan

*jjacksn@umich.edu*

September 3, 1983 will not be commemorated as the next national holiday, yet for some it is a day worthy of a celebration. Nor did it initiate ten days that shook the world, yet it marked the beginning of a quiet but profound revolution in Political Science. On that date, at the urging of a young Berkeley agitator-turned-methodologist named Steven Rosenstone, commenting on an APSA panel the day before, a group of young radicals gathered on the steps in the lobby of Chicago's Palmer House Hotel.<sup>1</sup> Accounts at the time put the number at twelve, but only seven have been positively identified in subsequent documents. Those seven are: Christopher H. Achen, John H. Aldrich, Larry M. Bartels, Henry E. Brady, John E. Jackson, George E. Marcus, Steven J. Rosenstone, and John E. Sullivan. Larry Bartels, the youngest possible member, now denies attending, and claims he was not even in Chicago. The remaining individuals have remained unidentified and unindicted co-conspirators though there are rumors about their identities. It is also true that, much like Babe Ruth's famous Wrigley Field home run in 1932, the number of people who claim to have attended now exceeds the capacity of the lobby and some of those claims are from individuals who, at best, were in a pre-arithmetic state of development.

A subsequent manifesto identifies the cell's purposes as three-fold: 1) To organize a formal meeting to create an agenda for action; 2) To take over a set of panels at the 1984 meeting of the American Political Science Association; and, 3) To institutionalize themselves as a field. There were surely other even more incendiary topics covered, but because the group was very careful not to take minutes so there could be deniability, these are not documented. Though this manifesto hardly had the fiery rhetoric of the Port Huron Statement, its intentions were as profound and its impacts more permanent. It initiated a bloodless revolution that without killing anyone or thing, save a few cross-tabs, some factor analyses, and a path analysis or two, permanently altered the way empirical analysis would be done.

This essay chronicles, with a bit of narrative, how we got to a 25th anniversary celebration of the summer con-

ference of the Society for Political Methodology, aka the APSA organized section on methodology, and earlier just the Political Methodology Group. This first section covers the organizational activities. The next section chronicles the history of the group's formal journals, *Political Methodology* and *Political Analysis*. Historical documents referenced in this essay (noted with an asterisk) along with other documents are stored in the archives, informally know as Jackson's file drawer. Many are also contained in an appendix posted to the PolMeth server. Most attention is paid to the very early years, as there are more and better summaries of later events.

#### Year One: 1983–1984

After the Palmer House Meeting, a steering committee composed of Chris Achen (Chair), John Jackson, Donald Kinder, and Steven Rosenstone was formed to plan and raise money for a several-day conference in the summer of 1984. A proposal written by Stanley Feldman and Steven Rosenstone secured \$4,000 from the NES Board of Overseers. This despite one board member saying he did not understand the importance of discussing "re-fried least squares." With help from Hank Heitowit, Director of the ICPSR Summer Program, additional money was raised from the University of Michigan to host the conference. Heitowit was critical both in the fund raising and in local arrangements for the conference.

The Steering Committee, after an exchange of drafts among themselves and with Henry Brady, suggested four topics to use in initiating the call for papers and participants. These were: 1) Methodological issues raised in analyzing the 1984 NES rolling cross-section design; 2) Aggregate versus individual level analysis, the so-called ecological correlation problem; 3) Quantal choice and survey response models; and, 4) Simultaneous equation modeling. Achen, in his role as First Secretary of the Central Committee, sent out what has come to be known as the April 10th letter\* inviting applications to participate in a four -day conference in Ann Arbor in July, 1984. It boldly suggested that if things worked well this would be the first in a series of annual workshops!

The first summer conference was held July 20–24, 1984 at the Institute for Social Research at the University of Michigan. The list\* of attendees at some or all of the ses-

President, Society for Political Methodology, 1985–1987. Originally prepared for the 25th year celebration of the annual Summer Meetings of the Society for Political Methodology. Ann Arbor, MI. July 12, 2008

<sup>1</sup>The paper Rosenstone was discussing was by John Jackson and Charles Franklin, not Achen as reported in a recent history of political methodology.

sions totals nineteen, including local faculty and graduate students.

Response to the formal meetings and informal discussions was overwhelmingly enthusiastic. Some participants called this the best conference they had ever attended, though given their ages one might wonder what significance to attach to this ( $\alpha = .5$  maybe?). (Several were still waiting for their Ph.D. diploma to arrive in the mail.) One author later referred to the spontaneous combustion of ideas, projects, and strategies, though the near 100 degree temperatures and no air conditioning may have contributed to the combustion.

David John Gow and Gary King were commissioned to begin the process of turning the small cadre of disorganized but committed methodologists into a formally recognized organized section of the American Political Science Association. They circulated a memo\* dated Aug. 1, 1984 on this subject, including a draft set of by-laws, recommendations for an executive committee, and a copy of the petition that required a minimum of 100 signatures in order to be considered for statehood.

The putsch continued with the take over of two panels at the 1984 APSA meetings. One was chaired by Steven Rosenstone and the other by Chris Achen and both featured work and speakers from the Ann Arbor conference. The second session was followed by the first organizational meeting of the still unrecognized movement for a political methodology. The meeting confirmed the recommendation that the central committee become the executive committee and that First Secretary Achen become the first President of the virtual section (though no one then had a clue about what a virtual anything was).

In August, 1984 Steven Rosenstone wrote a long paper\* describing these activities, laying out the intellectual agenda for the next several years, outlining the organizational tasks and preparing the foundation for a proposal to the NSF to fund the summer conferences. This was distilled to a shorter August 27, 1984 paper\* co-authored by Achen, Jackson, and Rosenstone with contributions from Henry Brady, that laid out the intellectual questions and approaches from the summer session and made a point of describing the issues as generic to all Political Science, not just to studies of opinion or elections or to specific types of data. Henry Brady distributed a longer Sept. 10, 1984 document titled, "Steps Towards Improving Political Methodology"\* that expanded on the intellectual and organizational themes of the two previous memos.

### Year Two: 1984–1985

Two significant milestones were reached during the second year of Achen's presidency. With the assistance of Merrill Shanks and the Survey Research Center at UC-Berkeley, along with a second contribution from the NES, funding

was obtained for a second summer conference. Achen sent out a call for papers and participants\* in the spring for a meeting July 25–28, 1985 in Berkeley, CA. (Note that consistent with the behavior of other underground groups of the time, the call is undated and the location of the meeting has relocated far from the initial meetings.) The conference agenda\* covered five broad topics, ranging from models of survey responses to time series and included a "business" meeting.

One very notable aspect of this meeting, in addition to its being held at all, is the way paper presentations and discussions combined work advancing the issues identified in Ann Arbor and in subsequent memos with new methodological questions and procedures. The best example is the evolution of discussions about how to analyze the rolling cross-sections in the 1984 NES study into a much broader discussion of time series and pooled time series and cross-section analysis led by Jim Stimson, Neal Beck, and Mel Hinich (two of the new additions).

The second milestone was Achen's agreeing to become the editor of *Political Methodology* and its adoption as the group's journal for presenting and disseminating new work focused solely on methodological topics. (A fuller discussion of the organization's journal history is a separate section of this paper.)

Lastly, the conference ended with a business meeting of the Political Methodology Society, as the group was named in Achen's call for participation. At this meeting *Political Methodology* was made the official journal for the Society, Gow reported on the ongoing efforts to secure formal statehood within the APSA, plans were discussed for a third conference, and John Jackson was elected the second president. The latter decision proving that the group could accomplish a peaceful transition and regime change.

The accomplishments of the year are very well described in a newsletter\* circulated by David John Gow, the secretary/treasurer (further evidence of the institutionalization of authority), in August, 1985. The newsletter also announced a methods panel at the forthcoming APSA meetings, chaired by Henry Brady, with participation by Douglas Rivers, James Stimson and other, "Political Methodological Society notables."

### Year Three: 1985–1986

Year Three was a banner year for the movement and marked its arrival as a fully public organization at the center of empirical methodology in Political Science. Two signature events brought to fruition the extended efforts of many individuals during the first two years.

On January 31, 1986 David John Gow sent a letter\* to Thomas E. Mann, Executive Director of the American Political Science Association (the seat of all power in the discipline) claiming the right for methodologists to be recognized

formally within the APSA. David politely requested rather than demanded recognition. This claim was buttressed by a list of current officers and 109 signed petitions (106 are still in the archives). It is not clear where, how, or possibly from whom David obtained some of the signatures. It is quite possible of course that some names are an alias for prominent individuals who were sympathetic to the revolution but did not want to be or could not be associated with it. Others may have been bought on the street. The good news is that apparently the APSA did not run background checks on the names, nor did other authorities.

David received a letter from Thomas Mann dated May 12, 1986\* informing us that we were now a formal organization, recognized as the section for methodology within the APSA.<sup>2</sup> No more clandestine activities, and the authorities now had to negotiate directly with the Society's elected leaders. This accomplishment is not merely a symbolic one. It meant the Society, functioning as the organized section, now controlled the planning and content of the methods panels at the annual meeting. This is not a trivial or innocuous authority, as attested to in a 1990 letter\* from John Freeman, the Society President, to Theodore Lowi, the APSA president, exerting the organization's sovereign right to designate Charles Franklin as the organizer for the Methods Sections of the APSA meetings and member of the Program Committee that year. As an exercise of the newly conveyed sovereign authority, Steven Rosenstone was designated the organizer for the Empirical Theory and Methods section of the 1986 APSA meetings.

On March 4, 1986 a formal proposal was submitted by the Political Methodology group to the NSF for funding the summer conferences in 1986 to 1988\*. The co-principal investigators acting for the organization were John Jackson (as President), Chris Achen, Henry Brady and Steven Rosenstone, though it reflects the ideas and writing of a larger set of individuals accumulated over the previous two years. Much of the proposal contains sections lifted directly from earlier memos and provides detailed descriptions of the intellectual progress over that period. The NSF notified the Society in May, 1986 that the proposal would be funded, following negotiations over the budget. The final version was approved and signed on June, 24, 1986.

These two events meant the organization had both sovereignty and resources, putting us ahead of some third world countries. There were some concerns about maintaining that sovereignty, however. There had to be at least 100 people willing to contribute real money to the authorities (the APSA, who takes these things very seriously as money is involved) and not just sign a petition.

Through the work of Henry Brady, as organizer, and

James Alt, Harvard University agreed to host the Third Summer Conference. Clearly, we had become the Establishment, even if we were still housed in non-air-conditioned dorms. On May 29, 1986 Jackson sent a memo\* to all Society members announcing both the APSA and the NSF decisions and summarizing a memo sent to department chairs on May 15, 1986 soliciting proposals for papers to be presented at the conference.

The Cambridge meetings were held Aug. 7–10, 1986. A list of participants and the agenda are archived. As in Berkeley the previous year, a number of new and important faces joined the group. Discussions were lively and went well beyond the formal sessions, continuing the tradition started in Ann Arbor that this was methods 24/4 (no one could last let alone afford seven days) and only for the truly committed.

There was an extended discussion at the conference about the relationship with the publishers of *Political Methodology*. After two editions under Achen's editorship it became apparent the publisher and editor disagreed on important issues. The conclusion of the discussion was that the relationship with the publisher was terminated and the Society began a search for a new publisher and new journal.

At the 1986 APSA meetings, in addition to developing a full set of panels under the leadership and guidance of Steven Rosenstone, the Society elected a new slate of officers, except for President as Jackson for forced to serve his full two year sentence. David John Gow decided to remain in Australia, much to our collective dismay. The new officers were Stanley Feldman, Secretary/Treasurer; Henry Brady, Program Director; and Doug Rivers and James Stimson, members of the executive committee. Stimson also agreed to lead the effort to create a new journal and to become its editor.

#### Year Four: 1986–1987

There was now irrefutable evidence that the Society had become the Establishment, a thought frightening to many. Allan Kornberg and Robert Bates approached the Society's leadership about hosting the Fourth Summer Conference, an idea enthusiastically accepted. And, though still in dorms there was a/c for the first time. The letter\* soliciting proposals for papers went out February 25, 1987. The conference was held August 6–9, 1987 in Durham, completing the ABCD cycle.

In addition to the usual lively and stimulating discussions came the discovery that we did not present ourselves as the young radicals from the 1960's and 70's we thought we were. We shared the dining hall with the participants in the cheerleader and athletic camps Duke was running at

<sup>2</sup>Thus began a slightly schizoid nature to the organization that persists. The Society for Political Methodology organizes the summer conferences, maintains the website, and distributes papers. The APSA organized section publishes a journal, newsletter, conducts business meetings, and distributes various awards. Functionally there is little distinction between the two, and one is safe treating them as a single entity though recent leaders have been clever in suggesting ways the distinction can work to the collective's advantage.

the time. On the way into a meal one of these children was overheard to ask who the nerds were. After a lot of denial and finger pointing the consensus was they meant the person with the tie.

Another milestone was passed at the APSA meetings in 1987. The Society elected Stanley Feldman as its third president. A second regime change and a third administration, clearly establishing the Society among the mature democracies of the world and not simply an organization of anarchists as we wished to believe at one point.

### Years Five and on : 1987–

The fifth and sixth summer conferences, organized by Stanley Feldman were held at UCLA and the University of Minnesota, respectively. The Minnesota meeting marked a significant new venture for the Society. The NSF grant included funds specifically for bringing graduate students to the summer meetings. Feldman solicited applications from graduate students in all political science departments, not just those whose faculty were regular attendees. This significant innovation, recognizing the importance of recruitment and socialization for all revolutions, formalized what had previously been an informal practice. Beginning with Charles Franklin's contribution to the Ann Arbor meeting, and continuing with John Williams at Harvard and Liz Gerber at Minnesota, graduate students had participated as co-authors. Other graduate students, such as Stephen Ansolabehere at Harvard, attended as graduate student hosts. All subsequently participated and contributed as faculty. But from Minnesota on, there would be a regular contingent of graduate students invited and expected to participate in the meetings.

By year Five the organization appeared and functioned much like other scholarly organizations. (See Spring, 1988 newsletter\* from Stanley Feldman.) Conferences were held, with papers being regularly published in a variety of journals and the scope of the topics expanding well beyond the initial agenda. In fact some participants have lamented the loss of discussions of a collective intellectual agenda.

Meetings were organized and conducted according to well established rules that permitted a bit of raucous behavior. Formal procedures for nominating and electing officers were in place, despite John Freeman's regular references to the oligarchy, of which he quickly became one when he was elected President in 1989. The concern about not maintaining the APSA's required size were soon alleviated. In short, the revolution had succeeded though in 1988-1989 that wasn't necessarily evident or commonly believed.

The NSF money was husbanded very carefully, in large part because of the generosity of members in using their own funds to attend meetings and in part because of the cagey ability of members to negotiate resources from the host institutions. The initial grant was scheduled to run

through 1988, but was extended until June, 1990. Under Freeman's leadership a new NSF proposal was prepared in January, 1990, though his name does not appear among the co-investigators. Still some need for deniability we presume. The listed investigators are John Jackson, Larry Bartels, Henry Brady, Stanley Feldman, and Gary King. This was funded and with extensions continued to 1993.

### Journals

There is no better signal of the Society's impact and influence than the current status and reputation of the its journal, *Political Analysis* and the wide distribution of *The Political Methodologist*. This was not always the case. There were regular complaints and extensive evidence that existing journals would not publish articles discussing and exploring methodological topics and had difficulty reviewing competently articles that used sophisticated methods. A vibrant, expanding field needed a venue for presenting new work and for stimulating further exploration. This was before the internet and electronic circulation, meaning that print journals were a vitally important part of any scholarly endeavor. Neal Beck stated this need very articulately in a 1986 letter, "...it seems imperative to publish some sort of volume...." Besides, every revolutionary group of the time had its own underground press, visible in every possible above ground location. This publication venue now exists, thanks to the energy and creativity of many individuals.

The first journal, *Political Methodology*, was originated and edited by George Marcus and John Sullivan. Its first volume appeared in 1974. Though the journal was underpromoted and not widely circulated it published some important methodological papers, a few of which are still cited. There were extensive discussions at the early meetings and conferences about the journal question. An agreement was reached in 1984 for *Political Methodology* to be the Society's officially sponsored journal and for Chris Achen to become the editor. Achen announced this in the call for proposals for the 1985 summer conference and encouraged submissions on a broad range of methodological topics.

The arrangement did not satisfy the various parties with an increasing rift developing between the editor and publisher, who owned the copyright for the journal. After an extended discussion at the Cambridge meeting the group decided to terminate the agreement with the publisher. This left the group without any journal, much to everyone's concern. The previous quote from Beck's letter after the decision expressed the collective thoughts as well as his own.

James Stimson, editor designate, and John Jackson, Society president, were designated as a committee to find a publisher for a new journal. Stimson's efforts on behalf of the journal were immense, despite knowing the work facing him if he succeeded. The organization also decided the journal would be an annual, rather than a quarterly. The logic

being that having one rather than four deadlines a year gave the editor more time to recruit and polish manuscripts and created less pressure to publish articles that needed additional work.

A prospectus for the new journal, to be titled *Political Analysis*, was sent to a number of university presses in Dec., 1986. The response was not overwhelming—in fact some presses never responded at all. Surely it was because of the organization's controversial nature, radical agenda, counter cultural views, and clandestine beginning. It was a classic Catch-22. A quality journal was needed to build the intellectual content and size of the organization while university presses wanted evidence of the intellectual products and the size of the market before beginning discussions. One press expressed enough interest to send a possible contract in July, 1987. This contract stipulated a guarantee of \$8000 from the Society, but said, "The press would incur all other financial risk." We kept looking.

The stochastic event frequently associated with successful revolutions occurred when Colin Day was appointed the new director of the University of Michigan Press in Dec., 1987. Colin had been a very entrepreneurial, intellectually aggressive, and successful associate editor at Cambridge, where he had initiated several important series in economics and other quantitative social sciences. His mandate from Michigan was to be all of these things. His agenda included more quantitative work in Political Science. We were more than happy to assist, and promise him access to a very lively and creative group of scholars. His appointment was announced at a meeting on Monday, Dec. 14 and a proposal\* from Stimson and Jackson was sent to the UM Press on Dec. 16, 1987.

After several rounds of discussion, on March 23, 1988 Colin Day sent a letter\* to Stimson and Jackson saying the Press' editorial board had agreed to offer to publish *Political Analysis* for a period of five years, "in the first instance." A contract was then negotiated, but not signed. The issue was the specification of who the Press was contracting with to produce the journal. Apparently given its origins the Press was a bit uncertain about what type of organization Stimson and Jackson were fronting. It was proposed that the contract be with the Methodology Section of the American Political Science Association. The Association, being the rational and perfectly foresighted hegemon, wanted no part of aiding and abetting an underground publication by this subversive group that would be outside their control and that one day might rival their own carefully managed publications. After lengthy negotiations and bargains, and possibly some threats and other behaviors that were never documented a contract was signed. The APSA agreed the section could contract for its own journal, so long as we behaved and did not directly challenge their control of the

discipline.<sup>3</sup> Jim Stimson was then approved as editor by the Press Board\*, and we were in business. Stanley Feldman, the Society's president, was able to announce that we now had a journal again in his 1988 newsletter.

But, the hard work had not even begun. Stimson edited three volumes of the journal, with the first one appearing in 1989 and all appearing in a timely manner. He recruited manuscripts from the leading members of the Society and from non-members. The Society members were strongly committed to the journal, in the manuscripts they submitted and in the reviewing they did. The norm was to provide rigorous and collegial reviewing. The reviewers' task was to advise the editor on the publishability of the paper *and* to provide the author advice on how to make the article a publishable article, or an even better article if it was already acceptable. This effort contributed importantly both to individual scholarship and to the collective advancement of the field. It was understood and practiced that the success of the journal, and of the Society more broadly, depended on the quality and impact of the scholarship and that everyone had to contribute. Stimson established a standard on which subsequent editors have been able to build and to create a journal with very high recognition and impact

### Concluding Remarks

This essay recounts the initial years of the Society for Political Methodology and how it started on its path to becoming the primary place for empirical work in Political Science and an organization that other subfields want to emulate. It only covers the organizational aspects of those years. Others, King (1991), Bartels and Brady (1993), Jackson (1996), and most recently Box-Steffensmeier, Brady and Collier (2008) and Franklin (2008), document the intellectual progress and innovations that occurred along the way. It does not try to relate all the obstacles and impediments to empirical Political Science that in part motivated the group and that were experienced at times during this period. It rather tries to convey the creativity, energy, comradeship, and very often the fun associated with and stimulated by the enterprise.

One should not leave with the impression that success had been achieved at the point this narrative ends. Far from it. It did not seem that way at the time and it wasn't. An extraordinary amount of work, imagination, good fortune, and leadership were required to go from a summer conference with twenty-four attendees to the current mega sessions; to go from an annual journal with limited impact to a top rated journal; and from a section with 109 members to being the second largest in the Association! The largest being a multinational global conglomerate, aka comparative politics. The Society has been blessed with a steady inflow of people with incredible energy, imagination and leadership

<sup>3</sup>Recent editors and Society presidents have placed the Society, the section, and the journal on a formally structured, clearly delineated and legally sound foundation, to the benefit of all.

which has gotten us to where we are.

### Acknowledgements

I have many people to thank, too many to list here. All of the original participants responded to an earlier version with helpful suggestions and corrections. A current activist was also very generous with detailed comments. Any and all errors are my own, but unintentional and not part of a continuing effort to obscure and conceal events and individuals.

### References

- Bartels, Larry M. and Henry E. Brady. 1993. "The State of Quantitative Political Methodology." In *Political Science: The State of the Discipline II*, ed. Ada Finifter. Washington, DC: The American Political Science Association.
- Box-Steffensmeier, Janet M., Henry E. Brady and David Collier. 2008. "Political Science Methodology as a Disciplinary Crossroads: An Overview." In *Oxford Handbook of Political Methodology*, ed. Janet Box-Steffensmeier, Henry E. Brady and David Collier. Oxford, UK: Oxford University Press.
- Franklin, Charles H. 2008. "Quantitative Methodology." In *Oxford Handbook of Political Methodology*, ed. Janet Box-Steffensmeier, Henry E. Brady and David Collier. Oxford, UK: Oxford University Press.
- Jackson, John. E. 1996. "Political Methodology: An Overview," In *New Handbook of Political Science*, ed. Robert Goodin and Hans-Dieter Klingemann. Oxford, UK: Oxford University Press.
- King, Gary. 1991. "On Political Methodology." *Political Analysis* 2: 1–29.

## Modeling Spatial Heterogeneity with Geographically Weighted Regressions in R

David Darmofal

University of South Carolina  
darmofal@mailbox.sc.edu

### Introduction

Scholars have long recognized that behavioral relationships may vary across units of observation and thus, covariates may vary in their effects across these units. This recognition has led to an interest in modeling heterogeneity. This is particularly evident in time series analysis of structural breaks, where scholars employ Chow tests to examine whether covariates vary across time. Just as behavioral relationships may vary over time, they may also vary over space. Indeed, many theories in political science predict this spatial heterogeneity. Consider, for example, models of voting behavior, policy diffusion, or comparative political economy where behavioral parameters are hypothesized to vary spatially.

One approach to modeling spatial heterogeneity is to employ spatial Chow tests (Anselin 1990), where now parameters are allowed to vary across distinct spatial subsets of the data, as opposed to the temporal subsets used in time series analysis. Such an approach, however, rests on the critical assumption that parameters do not vary within the spatial subsets of the data (which thus constitute distinct "spatial regimes") and instead vary only across them. This assumption may not be valid for many applications,

where we instead have reason to expect continuous spatial heterogeneity in parameters. Geographically Weighted Regression (GWR) provides a method to model this continuous spatial variation in parameters (Fotheringham, Brunson, and Charlton 2002, Fotheringham, Charlton, and Brunson 1998). Despite the wide applicability of GWR, however, it has seen only limited use within political science (but see Calvo and Escobar 2003, Cho and Gimpel 2009, Cho and Gimpel 2010, Darmofal 2008, Darmofal n.d.).

Happily, GWR models can be easily estimated in the `spgwr` package in R. The current version of this package, version 0.6-13, written by Roger Bivand and Danlin Yu, with contributions by Tomoki Nakaya and Miquel-Angel Garcia-Lopez, includes a variety of GWR models and weighting approaches. In this article, I first examine GWR and some of the central concepts involved with this modeling approach. Next I examine how GWRs can be estimated with the `spgwr` package through a sample application to voting during the New Deal realignment. I conclude by discussing some strengths and weaknesses of the `spgwr` package and possible future extensions for this package.

### Geographically Weighted Regression

In the standard regression model, the effects of covariates are estimated as global parameters that do not vary by unit. Often, however, we will wish to relax this assumption of non-varying parameters and allow the effects of covariates to vary depending upon the spatial locations of the units of observation. GWR departs from the standard regression framework by allowing the estimated parameters to vary geographically. The result is a continuous spatial plane of

parameter values. These parameters are measured at particular observed locations, typically the centroids of polygonal units such as counties, states, or countries (Fotheringham, Charlton, and Brunson 1998, 1907). (In this article's application, for example, I measure the parameters at the centroids of counties in the continental United States). The result is the following geographically weighted regression model with spatially varying parameters:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (1)$$

where, as Fotheringham, Brunson, and Charlton (2002, 52) note, " $u_i, v_i$  denotes the coordinates of the  $i$ th point in space and  $\beta_k(u_i, v_i)$  is a realization of the continuous function  $\beta_k(u, v)$  at point  $i$ ."

In calibrating equation (1), the contribution of other locations to the estimates  $\beta_k(u_i, v_i)$  is determined via a weighting function. The choice of weighting function should ideally reflect substantive theory. However, because scholars frequently lack substantive theory predicting the spatial interactions between units, scholars instead often rely on arbitrary choices of weighting functions instead. The `spgwr` package includes three classes of weighting functions, the bisquare, tricube, and Gaussian approaches, with two alternative Gaussian functions incorporated.

The bisquare weights function, `gwr.bisquare`, takes the form:

$$w_{ij}(g) = (1 - (d_{ij}^2/h^2))^2 \quad (2)$$

where  $d_{ij}$  are the distances between locations,  $h$  is the bandwidth, and  $w_{ij}(g) \neq 0$  if  $d_{ij} \leq h$ .

Alternatively, the tricube weighting function, `gwr.tricube`, takes the form:

$$w_{ij}(g) = (1 - (d_{ij}/h)^3)^3 \quad (3)$$

with  $d_{ij}$ , again, the distances between locations,  $h$  again the bandwidth, and  $w_{ij}(g) \neq 0$  if  $d_{ij} \leq h$ .

The Gaussian weighting function from the earliest versions of `spgwr`, `gwr.gauss`:

$$w(g) = e^{-(d/h)^2} \quad (4)$$

has been joined by a second Gaussian weighting function, `gwr.Gauss`, which is the default weighting function since release 0.5 of `spgwr`:

$$w(g) = e^{-(1/2)(d/h)^2} \quad (5)$$

where  $h$  in both Gaussian weighting functions is the bandwidth, and  $d$  is the distance between units.

Each of these weighting functions includes a bandwidth term, a critical component in any GWR analysis. The bandwidth affects the spatial smoothing of the estimates. As the bandwidth becomes smaller, fewer and more proximate units exert influence on the estimate at location  $i$ . As a consequence, smaller bandwidths produce less spatial

smoothing and greater variance in parameter estimates than do larger bandwidths. As bandwidths become larger, more units, and more spatially distant units exert influence on the estimate at location  $i$  and the GWR estimates come to more closely approximate OLS estimates with an equal weighting scheme (Fotheringham, Brunson, and Charlton 2002, 45).

`Spgwr` supports both fixed and adaptive bandwidths. Fixed bandwidths apply the same bandwidth to all units and are appropriate when the areal units are regularly spaced, as would be the case with a chessboard surface. This article's application is to counties in the continental United States. Counties, of course, are not regularly spaced. Instead, smaller counties are clustered in the Eastern United States and larger counties are located in the Western United States. The application of a fixed bandwidth approach to counties would result in undersmoothed GWR estimates with large standard errors in sparse Western counties and oversmoothed GWR estimates in more densely located Eastern counties. The use of an adaptive bandwidth approach will produce larger bandwidths for Western counties and smaller bandwidths for Eastern counties than will the fixed bandwidth approach. As a consequence, I employ the adaptive bandwidth estimation approach in this article's application.

Where does this bandwidth come from? Given the sensitivity of GWR estimates to the bandwidth, this is a critical question for estimation. Typically there is little substantive theory to guide the choice of bandwidth. As a consequence, scholars instead generally let the bandwidth be estimated from their data.

The current version of `spgwr` provides two methods for calculating bandwidths. The first of these is a cross-validation approach that selects bandwidths by minimizing the root mean square prediction error for the GWRs. If one employed a simple least squares criterion that included all units, each unit would be the best predictor for itself and thus the bandwidth would tend toward zero. Cleveland's (1979) cross-validation approach avoids this problem by dropping unit  $i$  from the calculation of the bandwidth for itself. This drop one cross-validation approach is the method employed in `spgwr`. If one prefers an alternative approach, `spgwr` also provides for bandwidth calculation via minimization of the Akaike Information Criterion (AIC).

## Application to the New Deal Realignment

I examine how spatial heterogeneity can be modeled with GWRs in `spgwr` via an application to voting behavior during the 1932 presidential election. The 1932 election has long been viewed as one of the principal examples of a realigning election. As a consequence, scholars have sought to explain why aggregate voting behavior changed in this election. In my example, I examine changes in county-level



voting for the Democratic presidential candidates between 1928 and 1932. An extended discussion of the New Deal realignment can be found in Darmofal (2008). Here, I focus on demonstrating how `spgwr` can be employed to examine spatial heterogeneity in the sources of changes in county-level Democratic support between 1928 and 1932.

I employ five covariates to model county-level voting changes in the 1932 election. Two covariates measure the size of the Democratic and non-voting populations in the county in the preceding presidential election. The covariate `lagdem` measures the proportion of the county's voting age population in 1928 that voted for Al Smith. The covariate `lagnv` measures the proportion of the county's voting age population that did not vote in the 1928 presidential election. To examine the interactive effect of both populations, I also include the interaction term `lagdem x lagnv`.

Immigrant populations play a large role in accounts of the New Deal realignment. To examine how the presence of immigrants affected aggregate changes in voting in 1932, I include the covariate `forgnpct`, which measures the percentage foreign-born in the county. Finally, to examine how changes in population affected changes in voting behavior, I include the covariate `popch10`, which measures the county-level population change between the 1928 and 1932 elections in ten thousands. The dependent variable in the GWR analysis is `chgdem`, which is the proportion of the county's voting age population that voted for Franklin D. Roosevelt in 1932 minus the proportion of the county's voting age population that voted for Al Smith in 1928.

First, I estimate a standard model with global, non-varying coefficients. The R code for OLS estimation is:

```
dem32tpm.lm <- lm(chgdem ~ lagdem + lagnv + ldlm +
  forgnpct + popch10,
  data=tpm32, weights=elgvtrs)
```

The resulting OLS output is:

```
Residuals:
    Min       1Q   Median       3Q      Max
-103.425  -2.291    2.651    6.644   70.030

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.3497905  0.0087509   39.972 <2e-16 ***
lagdem      -0.6112717  0.0355939  -17.173 <2e-16 ***
lagnv       -0.3294783  0.0143822  -22.909 <2e-16 ***
ldlm        0.7510355  0.0839982    8.941 <2e-16 ***
forgnpct    -0.3933030  0.0154712  -25.422 <2e-16 ***
popch10     0.0018061  0.0001554   11.622 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.014 on 3085 degrees of freedom
Multiple R-squared:  0.3541,    Adjusted R-squared:  0.3531
F-statistic: 338.3 on 5 and 3085 DF,  p-value: < 2.2e-16
```

As can be seen, each of the global estimates reaches statistical significance at a  $p < .001$  level (two-tailed tests).

The standard analysis would employ these global estimates to examine the substantive effects of the covariates, for example by varying the values on `popch10` from one standard deviation below the mean to one standard deviation above and examining the differences in changes in Democratic support across these two types of counties. It is important, however, to probe beyond these global estimates to examine variation at the local level. Does the positively signed global effect of this covariate, for example, hold across all of the continental United States? Or are there subsets of counties that differ in sign from the global result.

A GWR analysis allows researchers to examine how local results differ from global results such as those in the linear model output above. An insignificant global result, for example, may mask countervailing positive and negative effects of a covariate at the local level. Alternatively, even where the global result is statistically significant, one can identify the percentage of units with insignificant effects, or the percentage of units whose effects differ in sign from the global result.

The first step in a GWR analysis is the calculation of the bandwidth. In `spgwr`, this is accomplished by using the `gwr.sel` command to define the bandwidth. As stated earlier, either a fixed or an adaptive bandwidth can be employed. Given that counties are not regularly spaced, it would be inappropriate to use a fixed bandwidth. As a consequence, I instead employ the adaptive bandwidth option by employing the "adapt=TRUE" option. This option finds the proportion of neighboring observations to include in the weighting function as a function of the data in its search for the cross-validation score.

The code for calculating the adaptive bandwidths for this application is:

```
tpm32.bw <- gwr.sel(chgdem ~ lagdem + lagnv + ldlm
  + forgnpct + popch10, data=tpm32,
  coords=cbind(tpm32$x.coord, tpm32$y.coord),
  adapt=TRUE)
```

Here, the regression formula is specified first, the data frame (`tpm32`) is listed, and a set of coordinates is specified (in this case, the coordinates represent the centroids of the counties in the United States in 1932, i.e., today's continental United States). In the calculation of the adaptive bandwidths, I employ both the default `gwr.Gauss` Gaussian weighting function and the default drop one cross-validation approach.

The default, verbose, setting reports the search for the adaptive bandwidths using the cross-validation approach:

```
Adaptive q: 0.381966 CV score: 11.72391
Adaptive q: 0.618034 CV score: 12.70606
Adaptive q: 0.236068 CV score: 10.99102
Adaptive q: 0.145898 CV score: 10.40008
Adaptive q: 0.09016994 CV score: 9.967558
```

```

Adaptive q: 0.05572809 CV score: 9.698851
Adaptive q: 0.03444185 CV score: 9.429194
Adaptive q: 0.02128624 CV score: 8.980359
Adaptive q: 0.01315562 CV score: 8.753248
Adaptive q: 0.008130619 CV score: 8.567631
Adaptive q: 0.005024999 CV score: 8.336769
Adaptive q: 0.00310562 CV score: 8.350944
Adaptive q: 0.004292369 CV score: 8.295654
Adaptive q: 0.004134318 CV score: 8.280937
Adaptive q: 0.00374139 CV score: 8.269902
Adaptive q: 0.003818894 CV score: 8.270277
Adaptive q: 0.0037007 CV score: 8.270073
Adaptive q: 0.00374139 CV score: 8.269902

```

The bandwidths calculated by the `gwr.sel` function are then employed in the GWR regression estimation, using the `gwr` function:

```

tpm32.gauss <- gwr(chgdem ~ lagdem + lagnv +
ldln + forgnpct + popch10,
data=tpm32, coords=cbind(tpm32$x.coord,
tpm32$y.coord),
adapt=tpm32.bw, weights=elgvtrs, hatmatrix=TRUE)

```

Here, the regression formula is again specified first, followed by the data frame, the coordinates, and the bandwidths calculated via the `gwr.sel` function, in this case `tpm32.bw`. Next, optional weights are specified. In this case, given the variation in the eligible voting populations across the counties, I weight the data by the number of eligible voters in the county (`elgvtrs`). Finally, I employ the `hatmatrix=TRUE` option.

The resulting GWR output is:

```

Kernel function: gwr.Gauss
Adaptive quantile: 0.00374139 (about 11 of 3091)
Summary of GWR coefficient estimates:
      Min.   1st Qu.   Median   3rd Qu.   Max.   Global
X.Intercept. -0.308800  0.225200  0.350400  0.487100  1.702000  0.3498
lagdem       -5.411000 -0.811200 -0.400800 -0.013690  2.505000 -0.6113
lagnv        -1.687000 -0.510900 -0.294500 -0.077540  1.074000 -0.3295
ldln         -4.997000 -0.660900  0.285200  1.056000  6.130000  0.7510
forgnpct     -7.156000 -0.817900 -0.260600  0.166300  2.830000 -0.3933
popch10      -0.327800 -0.031150 -0.009888  0.001131  0.183400  0.0018
Number of data points: 3091
Effective number of parameters (residual: 2traceS - traceS'S): 536.501
Effective degrees of freedom (residual: 2traceS - traceS'S): 2554.499
Sigma (residual: 2traceS - traceS'S): 0.04451834
Effective number of parameters (model: traceS): 499.0506
Effective degrees of freedom (model: traceS): 2591.949
Sigma (model: traceS): 0.04419555
Sigma (ML): 0.04047085
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): -9861.288
AIC (GWR p. 96, eq. 4.22): -10555.82
Residual sum of squares: 5.062716
Quasi-global R2: 0.6849732

```

First, as can be seen, the global parameter estimates (included again in the last column of output) mask a great deal of variation at the local level. For example, while the global parameter estimate for `forgnpct` is -0.3933, the parameter estimates at the local level range from -7.156 to

2.83. Where the global estimate for `popch10` is 0.0018, the local parameter estimates range from -0.3278 to 0.1834.

We can probe further by examining the incidence of significant parameter estimates at the local level. Table 1 presents the percentage of statistically significant estimates at the local level. The first column in the table lists the covariate, the second column lists the percentage of all units with statistically significant estimates that are positively signed, and the third column lists the percentage with statistically significant estimates that are negatively signed ( $p < .01$ , two-tailed tests).

Table 1: GWR Significant Estimates

Covariate	% Sig. +	% Sig. -
lagdem	0.68	10.45
lagnv	0.16	20.58
ldln	3.88	1.04
forgnpct	0.74	7.64
popch10	0.13	6.31
intercept	53.93	0.07

Several results at the local level are of particular note. For example, the interaction term, `ldln`, has a positively signed global effect, but less than 5 percent of all counties have statistically significant effects. Next, 3.9 percent of all counties have positively signed significant effects for this covariate, while 1 percent have negatively signed significant effects. A little over 8 percent of all counties have statistically significant effects for the `forgnpct` covariate. The negatively signed global effect is matched by 7.64 percent of counties at the local level while slightly less than one percent of counties have positively signed local effects. Interestingly, while the global estimate for `popch10` is positively signed, more counties (6.31 percent) have negatively signed local effects than have positively signed local effects (0.13). This demonstrates how global parameter estimates can mask effects at the local level.

One of the advantages of a form of spatial analysis such as GWR is that these approaches lend themselves naturally to the mapping of results. Rather than relying on a global estimate, as standard modeling approaches do, we can map the GWR estimates to see where statistically significant local effects are located. Figure 1 provides such a mapping for the GWR estimates for the `forgnpct` covariate.

In Figure 1, counties with estimates that do not reach statistical significance at a  $p < 0.01$  level (two-tailed test) are mapped in white. Counties with positively signed statistically significant effects are shaded in brown. The remaining counties with statistically significant effects are shaded in the colors denoted in the legend, depending on their coefficient estimates for the `forgnpct` covariate.

What is perhaps most notable from this geographic map-

ping of the results is how few counties across the country exhibited statistically significant effects for this covariate, despite the significant global estimate. Although similar information is contained in Table 1 above, the mapping of the local estimates makes this even clearer. The few counties with positively signed statistically significant effects can be found in Georgia, Michigan, Montana, New York, North Dakota, and Pennsylvania. Counties with negatively signed statistically significant effects are located in Arkansas, Colorado, Delaware, Florida, Georgia, Idaho, Illinois, Iowa, Kansas, Maryland, Michigan, Missouri, Montana, New Mexico, North Carolina, Oklahoma, Pennsylvania, South Carolina, Texas, Utah, Virginia, Wisconsin, and Wyoming. Such mapping of the local parameter estimates can aid the researcher in identifying additional covariates that may account for the spatial heterogeneity in the modeled effects.

### Evaluating `spgwr`

Estimation time in `spgwr` is not quick, but it is manageable. The selection of adaptive bandwidths took 3 minutes and 48 seconds in 64-bit R on a Dell Precision T3500 64-bit Quad Core machine. The estimation of the GWR model took 38 minutes and 20 seconds on the same machine.

Scholars familiar with R will find `spgwr` an easy package to apply. For scholars less familiar with R, the learning curve for `spgwr` is not steep. Scholars interested in recent advances in GWR analysis will, however, find that the current version of `spgwr` has some limitations. Perhaps most importantly, the current version of `spgwr` does not contain functions for the estimation of GWR models that incorporate spatial dependence. Spatial heterogeneity and spatial dependence can co-exist. In such cases, it would be preferable to model spatial lag dependence or spatial error dependence while also modeling spatial heterogeneity. The current version of `spgwr` does not afford this possibility.

Spatial count models, currently absent from the functionality of `spgwr` would also be a useful addition. Similarly, as work advances on spatial GLM's, stronger functionality for the estimation of generalized GWR models would be welcome. Although the current version of `spgwr` includes functions both for finding a bandwidth for a generalized GWR model (`ggwr.sel`) and a function for estimating generalized GWR models (`ggwr`), the documentation for `spgwr` warns against reliance on these functions. Much additional work remains to be done in this regard.

Finally, helpful additional extensions would include tests for spatial nonstationarity of the covariates. For example, the `gwr` functions developed by Mark S. Pearce for `Stata` include as the default option a test for spatial nonstationarity in which the standard deviations of the parameter estimates are compared against those from a Monte Carlo simulation that mimics spatial randomness (Pearce 1998, 20-21). Alternatively, Leung et al.'s (2000) test for spatial nonsta-

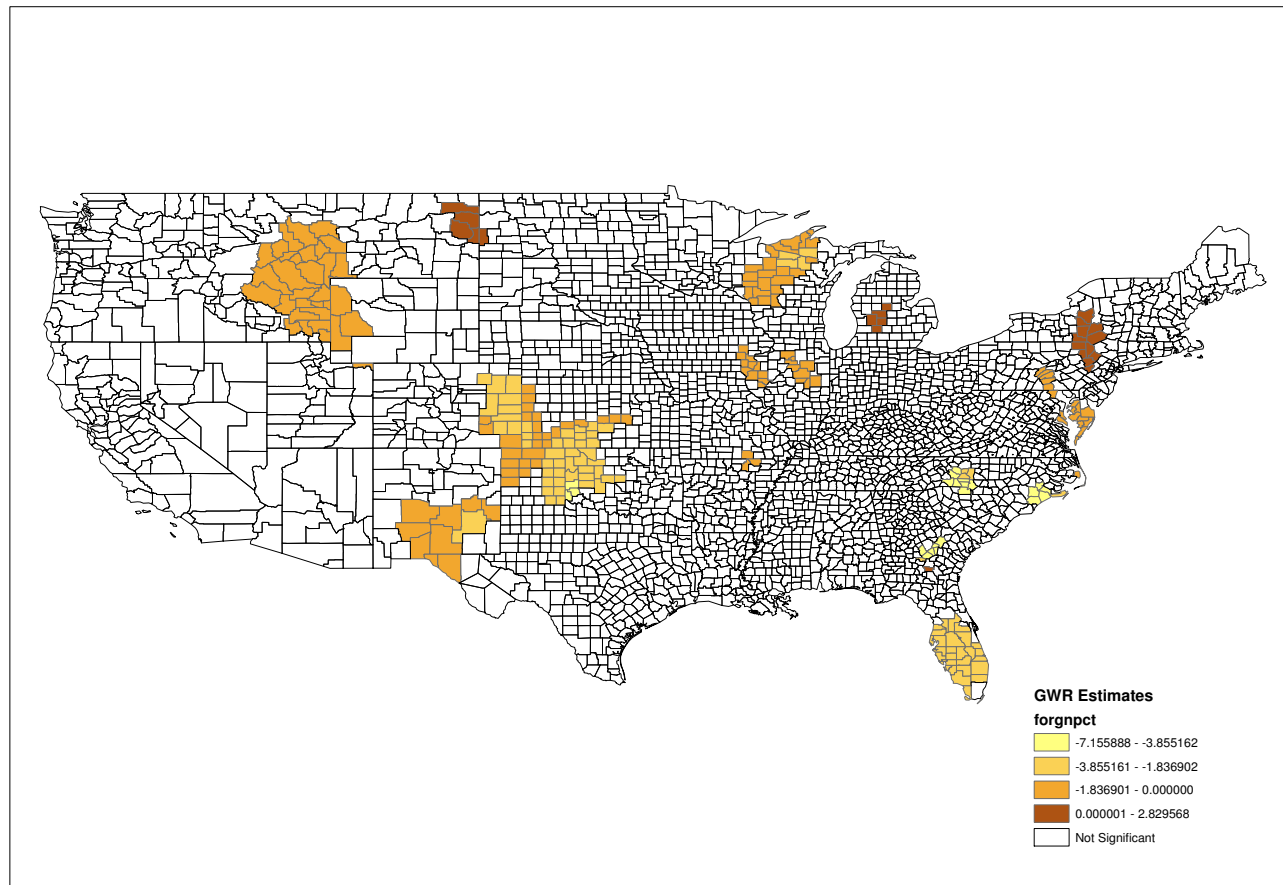
tionarity (see Fotheringham, Brunson, and Charlton 2002, 93-94, 214) could be added. Both sets of tests are included in Fotheringham, Brunson, and Charlton's dedicated GWR software. `Spgwr` is currently lacking in comparison to both `Stata` and GWR in tests for spatial nonstationarity.

In all, however, `spgwr` is a useful package for examining spatial heterogeneity. While continued extensions of the package are needed, even the current version will be attractive for scholars who have not been accustomed to modeling spatial heterogeneity. Rather than assuming that global parameter estimates hold for all locations, spatial heterogeneity should be modeled when it exists. `Spgwr` presents an opportunity for such modeling, and a particularly attractive one for scholars who use R for their research.

### References

- Anselin, Luc. 1990. "Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis." *Journal of Regional Science* 30(2): 185-207.
- Calvo, Ernesto and Marcelo Escobar. 2003. "The Local Voter: A Geographically Weighted Approach to Ecological Inference." *American Journal of Political Science* 47(1): 189-204.
- Cho, Wendy K. Tam and James G. Gimpel. 2009. "Presidential Voting and the Local Variability of Economic Hardship." *The Forum* 7(1): 1-21.
- Cho, Wendy K. Tam and James G. Gimpel. 2010. "Rough Terrain: Spatial Variation in Campaign Contributing and Volunteerism." *American Journal of Political Science* 54(1): 74-89.
- Cleveland, W.S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74: 829-836.
- Darmofal, David. 2008. "The Political Geography of the New Deal Realignment." *American Politics Research* 36(6): 934-961.
- Darmofal, David. Forthcoming. *Spatial Analysis for the Social Sciences*. Book manuscript, under contract at Cambridge Press University.
- Darmofal, David and Peter F. Nardulli. 2010. "The Dynamics of Realignment: An Analysis Across Time and Space." *Political Behavior* 32(2): 255-283.
- Fotheringham, A.S., C. Brunson and M.E. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, UK: John Wiley & Sons.
- Fotheringham, A.S., M.E. Charlton and C. Brunson. 1998. "Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis." *Environment and Planning* 30: 1905-1927.
- Pearce, Mark S. 1998. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Stata Technical Bulletin* 46: 20-24.

Figure 1: GWR Estimates



## superdiag: A Comprehensive Test Suite for Markov Chain Non-Convergence

Tsung-han Tsai and Jeff Gill  
Washington University in St. Louis  
*t.tsai@wustl.edu* and *jjgill@wustl.edu*

### Overview

The use of Markov chain Monte Carlo to solve difficult estimation problems, both Bayesian and non-Bayesian, is now quite common in political science and related fields. This is because it is increasingly easy to set-up and run a Gibbs sampler or Metropolis-Hastings kernel to produce marginal

posterior (sampling distribution) summaries for standard regression table construction. Programs like WinBUGS and JAGS automate the construction of the actual sampler from modeling statements dictated by the user. Both of these programs can also be called in multiple ways from R, which greatly simplifies the data-handling process. Yet we have noticed from conference papers, papers to review, and even published work, that the issue of Markov chain convergence is often poorly addressed.

A Markov chain that is not in its stationary (target) distribution does not produce valid empirical draws for inferential purposes. Therefore it is imperative that researchers perform sufficient analysis to assure themselves and their readers that the ergodic process is complete. This should involve multiple empirical diagnostics as well as graphical approaches to look for evidence of non-convergence. Unfor-

unately many authors provide no discussion of such checking or perhaps use only one tool. See Gill (2008) for specific advice and a discussion of various related problems.

We provide here an easy-to-use R function that integrates all of the standard empirical MCMC convergence diagnostics with one simple command. Users can simultaneously run functions supplied in menu form by the R suites `boa` and `coda`. Furthermore, the function automatically alters parameter choices, e.g. window definitions in the Geweke diagnostic, which are important features of the tests. Our aim is to help improve convergence testing in political science by making the mechanics much easier.

## The Commonly Used Convergence Diagnostics

Four tests dominate practice: Geweke, Gelman-Rubin, Raftery-Lewis, and Heidelberger-Welch, all named after their authors. The Geweke (1992) test compares some proportion of the early era of the chain after the burn-in period with some nonoverlapping proportion of the late era of the chain. A formal difference of means test, based on an asymptotic standard normal statistic, is performed with the idea that in stationarity there should not be an appreciable difference in the means of the two periods. The default uses the 0.0 to 0.1 and 0.5 to 1.0 periods. Gelman and Rubin's (1992) convergence diagnostic compares a small set of independently run chains with different starting points that are overdispersed relative to the target distribution. This is based on normal or Student's-t theory approximations to the marginal posteriors using an ANOVA-based test. The core part of the Heidelberger and Welch (1983) diagnostic uses a Brownian bridge assumption to produce a comparison of a sum from an early part of the chain to an appropriately scaled sum from the full length of the chain, both after the burn-in period. There is a second part of the test, the half-width comparison, that is less important and is often ignored. The Raftery and Lewis (1992) diagnostic gives a rough indication of convergence for a currently running chain. It takes thresholds from the user and estimates the chain length that provides a satisfactory result based on setting up a parallel (but not Markov) chain during a pilot run, where the iterations are a binary series according to whether the generated value in the primary chain at that stage is less than a chosen quantile. These four diagnostics are described in detail in Gill (2007, Chapter 11) along with other approaches.

Note that only the Gelman and Rubin diagnostic requires the user to run multiple chains for the convergence analysis. We think that there is more to be learned by running the other diagnostics on multiple chains. Furthermore, when there was controversy about one long run versus multiple short runs, computer resources were more limited (see the discussion in Kass, et al. [1998]). Running multi-

ple chains allows the user the ability to start from different points in the multidimensional sample space, change the random seed, and alter parameters of the diagnostic. However, running multiple chains, saving/processing them, and moving up and down the menu systems of `boa` and `coda` can add to the overhead of running MCMC, thus providing a disincentive. This extra work may explain the insufficient attention to convergence issues that we have observed in recent papers. Our goal here is to provide a function that reduces this administrative burden thus motivating researchers towards a more robust practice.

## Introducing `superdiag`

The R function `superdiag` is mostly a wrapper that calls each of the four popular formal diagnostics from the underlying `coda` routines without the `coda` menu structure. The output very closely resembles standard output from these functions and can be automatically dumped to a text file, rather than the screen, with the `sink()` function. In several instances described below we manipulate the diagnostics to exploit the provision of multiple chains. The code for `superdiag` is in the Appendix here and can also be downloaded at <http://jgill.wustl.edu/computing.html>.

### Geweke Diagnostic

Our extension of the Geweke diagnostic automatically alters the window specification. There are two parameters: the first gives the proportion of the chain to use advancing upward from the starting value, and the second gives the proportion of the chain to use descending from the final value downward. The default in `boa` and `coda` uses Geweke's (1992) defaults of 0.1 and 0.5: the first 10% of the values and the last 50% of the values. If a user of `superdiag` gives only one chain to analyze, then we use these defaults. However, if multiple chains are provided, only the first uses the defaults and all other chain analyses get random non-overlapping proportions up from the start of the chain and down from the end of the chain.

### Heidelberger and Welch Diagnostic

For the Heidelberger and Welch diagnostic we manipulate two parameters. The halfwidth part of the test calculates a  $(1 - \alpha)\%$  credible interval around the sample mean for each parameter dimension, where the estimated asymptotic standard error is the square root of spectral density divided by the non-discarded sample size,  $s(0)/n^*$ . The default in `boa` and `coda` is  $\alpha = 0.05$ . If the mean divided by this halfwidth is lower than  $\epsilon$  (defaulted to 0.1), then the halfwidth test is passed for this dimension. We sample with replacement common  $\alpha$  values for each chain to alter the size of the credible interval that creates the halfwidth value. This adds robustness to the effects of (necessarily) arbitrary choices of

the  $\alpha$  parameter. Secondly, we modify the default value for  $\epsilon$  by separately sampling uniformly in the interval  $[0.01 : 0.2]$  to provide a richer range of test criteria for the ratio. If a single chain is analyzed then the defaults are used, but with multiple chains the user will see differing combinations of these parameters.

### Raftery and Lewis Diagnostic

For each parameter (separately), the Raftery and Lewis is run as a pre-processing algorithm to determine a potential chain length to use after burn-in. First select the posterior tail threshold of interest, defaulted to  $q = 0.025$ , then determine a “tolerance” for this quantile, defaulted to  $r = 0.0005$ , followed by the desired probability of being within that tolerance, defaulted to  $s = 0.95$ . So generically we get a 95% probability of being in the interval  $[0.0245 : 0.0255]$ , which is the default here when only one chain is supplied to `superdiag`. Also, we need a convergence tolerance value  $\epsilon$ , which is used to determine a stopping point based on a parallel chain process (defaulted to 0.001). The diagnostic then runs a pilot sampler whose length is determined as if there is no autocorrelation in the parallel chain, given by rounding:  $n_{\text{pilot}} = \left\lceil \Phi^{-1} \left( \frac{s+1}{2} \right) \frac{\sqrt{q(1-q)}}{r} \right\rceil^2$ , where  $\Phi^{-1}()$  is the inverse of the normal CDF. The process will then return: the length of the burn-in period ( $M$ ), the estimated number of post burn-in iterations required to meet these goals ( $N$ ), and a specified thinning interval ( $k$ ). For each of these four parameters we sample from a vector (changeable by users) of values around the defaults (larger and smaller) to provide a reasonable range of alternatives.

### Example Model

This example comes from Gill (2007, Chapter 11), which is an extension of Norrander (2000), using tobit models to account for 15 states that do not have a death penalty on the books and therefore cannot express public support through a count of executions. The usual questions center on ideological, racial and religious makeup, political culture, and urbanization as causal effects of state-level executions (1993-1994 in the data here). If  $\mathbf{z}$  is a latent outcome variable in this context with  $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$  and  $z_i \sim \mathcal{N}(x_i\boldsymbol{\beta}, \sigma^2)$ , then the observed outcome variable is produced according to:  $y_i = z_i$  if  $z_i > 0$ , and  $y_i = 0$ , if  $z_i \leq 0$ . This gives the likelihood function:

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = \prod_{y_i=0} \left[ 1 - \Phi \left( \frac{x_i\boldsymbol{\beta}}{\sigma} \right) \right] \times \prod_{y_i>0} (\sigma^{-1}) \exp \left[ -\frac{1}{2\sigma^2} (y_i - x_i\boldsymbol{\beta})^2 \right]. \quad (1)$$

A flexible parameterization for the priors is given by Gawande (1998):

$$\boldsymbol{\beta} | \sigma^2 \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{I}\sigma^2 B_0^{-1}), \quad \sigma^2 \sim \mathcal{IG} \left( \frac{\gamma_0}{2}, \frac{\gamma_1}{2} \right) \quad (2)$$

with vector hyperparameter  $\boldsymbol{\beta}_0$ , scalar hyperparameters  $B_0, \gamma_0 > 2, \gamma_1 > 0$ , and an identity matrix  $\mathbf{I}$ . The resulting full conditional distributions for Gibbs sampling are given for the  $\boldsymbol{\beta}$  block,  $\sigma^2$ , and the  $z_i | y_i = 0$  as:

$$\begin{aligned} \boldsymbol{\beta} | \sigma^2, \mathbf{z}, \mathbf{y}, \mathbf{X} &\sim \mathcal{N} \left( (B_0 + \mathbf{X}'\mathbf{X})^{-1} (\boldsymbol{\beta}_0 B_0 + \mathbf{X}'\mathbf{z}), \right. \\ &\quad \left. (\sigma^{-2} B_0 + \sigma^{-2} \mathbf{X}'\mathbf{X})^{-1} \right) \\ \sigma^2 | \boldsymbol{\beta}, \mathbf{z}, \mathbf{y}, \mathbf{X} &\sim \mathcal{IG} \left( \frac{\gamma_0 + n}{2}, \right. \\ &\quad \left. \frac{\gamma_1 + (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})}{2} \right) \\ z_i | y_i = 0, \boldsymbol{\beta}, \sigma, \mathbf{X} &\sim \mathcal{TN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2) I_{(-\infty, 0)}, \end{aligned} \quad (3)$$

where  $\mathcal{TN}()$  is the truncated normal and the indicator function  $I_{(-\infty, 0)}$  gives the truncation bounds. The prior parameters for the inverse gamma distribution are designed to give a diffuse form:  $\gamma_0 = 300, \gamma_1 = 100, B_0 = 0.02$ , and  $\boldsymbol{\beta} = \mathbf{0}$ .

We programmed a Gibbs sampler in R to provide marginal distributions from the full conditional distributions, running it 50,000 times and disposing of the first 40,000 iterations. The results are summarized in the table below and the R code for the sampler is given at <http://jgill.wustl.edu/computing.html>.

	Mean	SE	95% HPD
Constant	-14.545	3.672	[-21.766:-7.350]
Past Rates	171.146	8.048	[155.200:186.608]
Political Culture	0.346	0.145	[0.060:0.622]
Current Opinion	3.974	1.067	[1.858:6.022]
Ideology	3.142	1.111	[0.973:5.315]
Murder Rate	0.009	0.080	[-0.157:0.159]

### Running superdiag

As noted, `superdiag` is designed to reduce the time it takes to manipulate and test multiple chains with the standard diagnostics. The default parameter values for the four tests remain defaults for the first chain and then we systematically manipulate them for robustness. Sometimes, particularly for Raftery and Lewis, this causes an unusual combination that gives a very pessimistic view of convergence. Users are cautioned to pay attention to such combinations and possibly discount the result.

The input object definition provided to `superdiag` is flexible. The four formal functions of diagnostics in `coda` accept only an `mcmc` or `mcmc.list` object as input. Conversely,

the input supplied to `superdiag` can be a list object, or either `mcmc` and `mcmc.list` objects, regardless of the number of chains. Therefore, researchers who generate MCMC samples on their own rather than by programs like `WinBUGS` or `JAGS`, can easily analyze data by using `superdiag` without necessary conversion. Since our Gibbs sampler output was created with an R function and for each chain the last 10,000 values out of 50,000 were saved to text files, we loaded the MCMC output and built objects according to:

```
tobit.list <- list()
for (i in 1:5) {
  tobit.list[[i]] <- mcmc(read.table(paste(
    "tobit.mcmc.out",i,sep=""),header=TRUE))
}
tobit.list <- as.mcmc.list(tobit.list)
```

However, chains produced by `WinBUGS` or `JAGS` can be read in directly as `mcmc` objects with the function `read.coda`. Users of `R2jags`, `Rjags`, `runjags`, `BRugs`, `rbugs`, `R2WinBUGS`, and `MCMCpack` are already working with appropriately formatted data objects for `superdiag`.

Running the diagnostic with our newly created list is done with one simple command, producing:

```
superdiag(tobit.list,burnin=0)

Number of chains = 5
Number of iterations = 10000 per chain before discarding
                        the burn-in period
The burn-in period = 0 per chain
Sample size in total = 50000

***** The Geweke diagnostic: *****
Z-scores:
      chain 1  chain 2  chain 3
Constant    0.19006 -1.19822 -1.97097
Past.Rates  0.43421  0.35479 -0.10703
Political.Culture 0.09149  0.23044  1.18154
Current.Opinion 1.04999 -0.26106  2.11368
Ideology    -0.70922  1.08343  1.09934
Murder.Rate 1.12295  0.41441 -0.30343
Window From Start 0.10000  0.03854  0.62622
Window From Stop 0.50000  0.92317  0.28442

      chain 4  chain 5
Constant    -0.77899 -0.14759
Past.Rates  0.81997  0.25356
Political.Culture -1.67556 -0.27691
Current.Opinion -2.47919 -0.15100
Ideology     1.51378  0.38307
Murder.Rate  -0.53252 -0.67923
Window From Start 0.25000  0.40163
Window From Stop 0.28153  0.12059
```

```
***** The Gelman-Rubin diagnostic: *****
Potential scale reduction factors:
      Point est. Upper C.I.
Constant          1          1
Past.Rates        1          1
Political.Culture 1          1
Current.Opinion   1          1
Ideology          1          1
```

```
Murder.Rate          1          1

Multivariate psrf
1

***** The Heidelberger-Welch diagnostic: *****

Chain 1, epsilon=0.1, alpha=0.05
      Stationarity start      p-value
      test      iteration
Constant      passed      1      0.275
Past.Rates    passed      1      0.303
Political.Culture passed    1      0.968
Current.Opinion passed    1      0.267
Ideology      passed    1      0.432
Murder.Rate   passed    1      0.925

      Halfwidth Mean      Halfwidth
      test
Constant      passed    -14.5451 0.09711
Past.Rates    passed    171.1460 0.16613
Political.Culture passed    0.3461 0.00381
Current.Opinion passed    3.9738 0.02737
Ideology      passed    3.1423 0.02991
Murder.Rate   failed     0.0088 0.00198

Chain 2, epsilon=0.032, alpha=0.005
      Stationarity start      p-value
      test      iteration
Constant      passed    1      0.871
Past.Rates    passed    1      0.660
Political.Culture passed    1      0.882
Current.Opinion passed    1      0.616
Ideology      passed    1      0.770
Murder.Rate   passed    1      0.932

      Halfwidth Mean      Halfwidth
      test
Constant      passed    -14.57477 0.08874
Past.Rates    passed    171.28755 0.16882
Political.Culture passed    0.34533 0.00342
Current.Opinion passed    3.97391 0.03253
Ideology      passed    3.14973 0.02591
Murder.Rate   failed     0.00947 0.00206

Chain 3, epsilon=0.133, alpha=0.01
      Stationarity start      p-value
      test      iteration
Constant      passed    1      0.1009
Past.Rates    passed    1      0.2762
Political.Culture passed    1      0.6218
Current.Opinion passed    1      0.0683
Ideology      passed    1      0.4488
Murder.Rate   passed    1      0.8134

      Halfwidth Mean      Halfwidth
      test
Constant      passed    -14.58034 0.07927
Past.Rates    passed    171.09983 0.18120
Political.Culture passed    0.34438 0.00319
Current.Opinion passed    3.96566 0.03392
Ideology      passed    3.15523 0.02413
Murder.Rate   failed     0.00937 0.00181

Chain 4, epsilon=0.152, alpha=0.025
      Stationarity start      p-value
      test      iteration
```

```
Constant      passed      1      0.7090
Past.Rates    passed      1      0.4077
Political.Culture passed    1      0.1044
Current.Opinion passed    1      0.0286
Ideology      passed      1      0.2394
Murder.Rate   passed      1      0.6763
```

```
Halfwidth Mean      Halfwidth
test
Constant      passed    -14.5491 0.07792
Past.Rates    passed    171.1474 0.15642
Political.Culture passed    0.3435 0.00371
Current.Opinion passed    3.9472 0.03155
Ideology      passed    3.1478 0.02464
Murder.Rate   failed     0.0105 0.00221
```

Chain 5, epsilon=0.082, alpha=0.01

```
Stationarity start      p-value
test      iteration
Constant      passed      1      0.181
Past.Rates    passed      1      0.592
Political.Culture passed    1      0.335
Current.Opinion passed    1      0.409
Ideology      passed      1      0.349
Murder.Rate   passed      1      0.197
```

```
Halfwidth Mean      Halfwidth
test
Constant      passed    -14.5786 0.08361
Past.Rates    passed    171.1698 0.15530
Political.Culture passed    0.3430 0.00352
Current.Opinion passed    3.9430 0.02890
Ideology      passed    3.1574 0.02409
Murder.Rate   failed     0.0105 0.00208
```

\*\*\*\*\* The Raftery-Lewis diagnostic: \*\*\*\*\*

```
Chain 1, converge.eps = 0.001
Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
Constant	3	4061	3746	1.08
Past.Rates	2	3802	3746	1.01
Political.Culture	3	4028	3746	1.08
Current.Opinion	3	4061	3746	1.08
Ideology	3	4061	3746	1.08
Murder.Rate	3	4129	3746	1.10

```
Chain 2, converge.eps = 0.001
Quantile (q) = 0.001
Accuracy (r) = +/- 0.0025
Probability (s) = 0.99
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
Constant	2	1061	1061	1.00
Past.Rates	2	1061	1061	1.00
Political.Culture	2	1061	1061	1.00
Current.Opinion	2	1061	1061	1.00
Ideology	3	1297	1061	1.22
Murder.Rate	3	1297	1061	1.22

```
Chain 3, converge.eps = 2e-04
```

```
Quantile (q) = 0.001
Accuracy (r) = +/- 0.005
Probability (s) = 0.999
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
Constant	2	434	433	1
Past.Rates	2	434	433	1
Political.Culture	2	434	433	1
Current.Opinion	2	434	433	1
Ideology	2	434	433	1
Murder.Rate	2	434	433	1

```
Chain 4, converge.eps = 0.005
Quantile (q) = 0.1
Accuracy (r) = +/- 0.0025
Probability (s) = 0.99
```

You need a sample size of at least 95543 with these values of q, r and s

```
Chain 5, converge.eps = 0.005
Quantile (q) = 0.001
Accuracy (r) = +/- 0.005
Probability (s) = 0.975
```

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
Constant	1	202	201	1
Past.Rates	1	202	201	1
Political.Culture	1	202	201	1
Current.Opinion	1	202	201	1
Ideology	1	202	201	1
Murder.Rate	1	202	201	1

Running the function also generated the following warning:

**Warning message:**  
**In superdiag(tobit.list, burnin = 0) :**  
**The burn-in period is negative or zero**

which is not material to us since we have already removed the proposed burn-in values. Even though this chain appears to be in convergence, note that there are differences resulting from parameter changes. For example, Geweke test for Current Opinion across the five chains returns:

```
1.04999 -1.283986 0.81387 -2.98846 0.5089598
```

gives one value typically outside of the normal range for expected convergence. It is critical, however, to keep in mind that testing at  $\alpha = 0.05$  means that even in actual convergence, about 1 out of 20 values will be in the tails. Therefore it is important to look at the complete picture.

### Concluding Remarks

It is important to remember that these convergence diagnostics are actually indicators of *nonconvergence* rather than evidence of convergence. Failing to find evidence of nonconvergence with these procedures is comforting but not decisive. Careful practitioners should treat positive results from one test with continued skepticism, and run multiple diagnostics on any single Markov chain, any one of which can



provide sufficient evidence of failure. Our goal here is simply to encourage such caution by making the process easier, therefore improving standard practice in the literature.

## References

- Brooks, Stephen. P. and Gareth O. Roberts. 1998. "Convergence Assessment Techniques for Markov Chain Monte Carlo." *Statistics and Computing* 8: 319–335.
- Gawande, Kishore. 1998. "Comparing Theories of Endogenous Protection: Bayesian Comparison of Tobit Models Using Gibbs Sampling Output." *Review of Economics and Statistics* 80(1): 128–140.
- Gelman, Andrew and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4): 457–511.
- Geweke, J. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In *Bayesian Statistics 4*, ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger. Oxford, UK: Oxford University Press, 169–193.
- Gill, Jeff. 2007. *Bayesian Methods for the Social and Behavioral Sciences*. Second Edition. New York, NY: Chapman & Hall.
- Gill, Jeff. 2008. "Is Partial-Dimension Convergence a Problem for Inferences from MCMC Algorithms?" *Political Analysis* 16(2): 153–178.
- Heidelberger, Philip and Peter D. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31(6): 1109–1144.
- Kass, Robert E., Bradley P. Carlin, Andrew Gelman, and Radford M. Neal. 1998. "Markov Chain Monte Carlo in Practice: A Roundtable Discussion." *The American Statistician* 52(2): 93–100.
- Norrander, Barbara. 2000. "The Multi-Layered Impact of Public Opinion on Capital Punishment Implementation in the American States." *Political Research Quarterly* 53(4): 771–793.
- Raftery, A. E. and Lewis, S. M. 1992. "How Many Iterations in the Gibbs Sampler?" In *Bayesian Statistics 4*, ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger. Oxford, UK: Oxford University Press, 763–773.

## Appendix

This appendix gives our R code for the `superdiag` function.

```
## CREATE THE DIAGNOSTIC FUNCTION, TSUNG-HAN TSAI AND JEFF GILL
superdiag <- function(mcmcoutput, burnin=10000, confidence.gr=0.95,
  frac1.gw=0.1, frac2.gw=0.5, eps.hw=0.1,
  pvalue.hw=0.05, q.rl=0.025, r.rl=0.005, s.rl=0.95,
  eps.rl=0.001) {

  # mcmcoutput: input chains from jags, bugs, etc.
  # confidence.gr: 1-alpha for testing with the Gelman and Rubin test
  # frac1.gw: frac1 for the Geweke Test
  # frac2.gw: frac2 for the Geweke Test
```

```
# eps.hw: epsilon for the Heidelberger and Welch test
# pvalue.hw: p-value for the Heidelberger and Welch test
# q.rl: q-parameter for the Raftery and Lewis Test
# r.rl: r-parameter for the Raftery and Lewis Test
# s.rl: s-parameter for the Raftery and Lewis Test
# eps.rl: convergence epsilon for the Raftery and Lewis Test

# CREATE A FUNCTION TO DISCARD THE BURN-IN PERIOD
burn <- function(input.matrix, burnin) {
  out <- input.matrix[-(1:burnin),];
  return(out);
}

# THE INPUT SHOULD BE AN "mcmc", "mcmc.list", OR "list" OBJECT
if (class(mcmcoutput) != "mcmc" & class(mcmcoutput) !=
  "mcmc.list" & class(mcmcoutput) != "list")
  stop("The inputs have to be mcmc, mcmc.list, or list objects.");
# CONVERT "mcmc" INTO "mcmc.list"
if (class(mcmcoutput) == "mcmc") {
  mcmcoutput <- as.mcmc.list(mcmcoutput);
}

para.names <- dimnames(mcmcoutput[[1]])[[2]]; # PARAMETERS NAMES
n.chains <- length(mcmcoutput); # THE NUMBER OF CHAINS
dim.chain <- sapply(mcmcoutput, dim); # THE DIMENSION OF EACH CHAIN
# THE NUMBER OF ITERATIONS BEFORE DELETING THE BURN-IN PERIOD
t.iter <- dim.chain[1];
diff.dim <- dim.chain - dim.chain;
if (sum(diff.dim != 0) != 0) stop("The number of iterations or
  variables is not equal for all chains.");

# DISCARD THE BURN-IN PERIOD
if (burnin <= 0) {
  warning("The burn-in period is negative or zero");
  mcmcburnin <- mcmcoutput;
} else {
  mcmcburnin <- lapply(mcmcoutput, burn, burnin=burnin);
}

# SAVE THE SAMPLES AS A MCMC LIST AFTER DISCARDING THE BURN-IN PERIOD
mcmcurnin.list <- vector("list", n.chains);
for (i in 1:n.chains) {
  mcmcurnin.list[[i]] <- as.mcmc(mcmcurnin[[i]]);
}
mcmcurnin.mcmc.list <- as.mcmc.list(mcmcurnin.list);

# THE TOTAL SAMPLES FOR ALL CHAINS
t.samples <- as.matrix(mcmcurnin.mcmc.list);

# REARRANGE THE PRINTED RESULTS
geweke.chains <- matrix(NA, nrow=n.chains, ncol=dim.chain[2]);
heidel.list <- vector("list", n.chains);
raftery.list <- vector("list", n.chains);

# SETUP DIFFERENT WINDOW SPECIFICATIONS FOR GEWEKE
geweke.windows <- matrix(c(frac1.gw,frac2.gw),ncol=2)
for (i in 2:n.chains) {
  win1 <- runif(1,0,0.99); win2 <- 1-runif(1,win1,1)
  geweke.windows <- rbind(geweke.windows,c(win1,win2))
}

# SETUP DIFFERENT PARAMETER SPECIFICATIONS FOR HEIDELBERGER AND WELCH
heidel.params <- matrix(c(eps.hw,pvalue.hw),ncol=2)
pvals <- c(0.1,0.05,0.025,0.01,0.005)
for (i in 2:n.chains) {
  param1 <- runif(1,0.01,0.2); param2 <- sample(x=pvals,size=1)
  heidel.params <- rbind(heidel.params,c(param1,param2))
}

# SETUP DIFFERENT PARAMETER SPECIFICATIONS FOR RAFTERY AND LEWIS
raft.params <- matrix(c(q.rl, r.rl, s.rl, eps.rl),ncol=4)
qvals <-c(0.25,0.1,0.05,0.01,0.001)
rvals <- c(0.001,0.0025,0.0005,0.001,0.005)
svals <- c(0.9,0.95,0.975,0.99,0.999)
evals <- c(0.005,0.0025,0.001,0.0005,0.0002)
for (i in 2:n.chains) {
  param1 <- sample(x=qvals,size=1); param2 <- sample(x=rvals,size=1)
  param3 <- sample(x=svals,size=1); param4 <- sample(x=evals,size=1)
  raft.params <- rbind(raft.params,c(param1,param2,param3,param4))
}
```

```

# RUN DIAGNOSTICS BY CHAIN IN ORDER TO AVOID ONE CHAIN RUINS ALL CHAINS
for (i in 1:n.chains) {
geweke <- suppressWarnings(try(geweke.diag(mcmcburnin.mcmclist[[i]],
geweke.windows[i,1], geweke.windows[i,2]),
silent=TRUE));
if (class(geweke) == "geweke.diag")
geweke.chains[i,] <- t(geweke[1]$z);
heidel.list[[i]] <- suppressWarnings(try(heidel.diag(
mcmcburnin.mcmclist[[i]], heidel.params[i,1],
heidel.params[i,2]), silent=TRUE));
raftery.list[[i]] <- suppressWarnings(try(raftery.diag(
mcmcburnin.mcmclist[[i]], q=raft.params[i,1],
r=raft.params[i,2], s=raft.params[i,3],
converge.eps=raft.params[i,4]), silent=TRUE));
}
colnames(geweke.chains) <- para.names;

# PROVIDE THE BASIC INFORMATION OF MCMC SAMPLES
cat(paste("Number of chains =", n.chains, "\n"));
cat(paste("Number of iterations =", t.iter, "per chain before
discarding the burn-in period\n"));
cat(paste("The burn-in period =", burnin, "per chain\n"));
cat(paste("Sample size in total =", dim(t.samples)[1], "\n"));
cat("\n");

# REPORT RESULTS OF DIAGNOSTICS OF CONVERGENCE
if (n.chains < 2) {
chain.name <- "chain 1";
rownames(geweke.chains) <- chain.name;
cat("The Gelman-Rubin diagnostic is not reported since
the number of chains is less than 2.\n");
cat("\n");
cat("The Geweke diagnostic:\n");
cat(paste("Fraction in 1st window =", frac1.gw, "\n"));
cat(paste("Fraction in 2nd window =", frac2.gw, "\n"));
cat(paste("Z-scores:\n"));
print(t(geweke.chains));
cat("\n");
cat("The Heidelberger-Welch diagnostic:\n");
cat("\n");
print(heidel.list);
cat("\n");

cat("The Raftery-Lewis diagnostic:\n");
cat("\n");
print(raftery.list);
cat("\n");
} else {
chain.name <- "chain1";
for (i in 2:n.chains) {
namei <- paste("chain ", i, sep="");
chain.name <- c(chain.name, namei);
}
rownames(geweke.chains) <- chain.name;

cat("***** The Geweke diagnostic: *****\n");
dimnames(geweke.windows)[[2]] <- c("Window From Start", "Window From Stop");
cat(paste("Z-scores:\n"));
print(rbind( t(geweke.chains), t(round(geweke.windows,5))));
cat("\n");
cat("***** The Gelman-Rubin diagnostic: *****\n");
print(gelman.diag(mcmcburnin.mcmclist, confidence.gr));
cat("\n");
cat("***** The Heidelberger-Welch diagnostic: *****\n");
cat("\n");
for (i in 1:n.chains) {
cat(paste("Chain ",i," epsilon=",round(heidel.params
[i,1],3),"
alpha=",round(heidel.params
[i,2],3),sep=""));
print(heidel.list[[i]]);
cat("\n");
}
cat("***** The Raftery-Lewis diagnostic: *****\n");
cat("\n");
for (i in 1:n.chains) {
cat(paste("Chain ",i," converge.eps = ",round(raft.params[i,4],4),sep=""));
print(raftery.list[[i]]);
cat("\n");
}
cat("\n");
}
# RETURN THE MCMC SAMPLES WITH THE BURN-IN DISCARDED
superdiag.chains <- list(mcmc.samples = mcmcburnin.mcmclist)

```

## PresidentParser: Automated Mark-Up Utility

McKendon Lafleur and John Beielor  
Louisiana State University  
*McKendon@acm.org and jbeiel1@lsu.edu.*

### PresidentParser

As any researcher dealing with text-based data knows, the task of manipulating the text into a useable form is often the most time consuming aspect of the research process. This holds true especially for the study of leader psychology in international relations. This task is what the *PresidentParser* program aims to alleviate.

In short, *PresidentParser* parses speech acts and prepares them for processing in the *Profiler Plus* environment<sup>1</sup>. Since research into the role of elite psychology in international relations is focused on at-a-distance analysis of speech

acts to derive quantitative measures of that leader's psychology, a researcher must often modify those speech acts so that only the verbal material, and thus psychology, of the leader is captured. For instance, many studies of leader psychology rely on leader responses to question and answer sessions. One does not, however, wish to include the verbal material of a reporter in the psychological profile of a leader. Luckily, *Profiler Plus* allows for an easy way to remove this unwanted verbal material while still retaining the integrity and logic of a speech act. *Profiler Plus* includes the ability to insert the XML tags `<ignore></ignore>` into a speech act, which, as the name implies, tells the program to ignore the lines surrounded by the tags. Traditionally, this task has been performed by hand, and anyone who has been involved in a research project of this type can attest that it is often a long, monotonous task. Additionally, with the advent of automated webscraping scripts, the number of speech acts to be processed has begun to grow exponentially, often into the tens of thousands.

<sup>1</sup>*Profiler Plus* can be obtained from [Social Science Automation](#)

The very monotony and repetition of “marking-up” a speech act is what lends this activity to automation. At its heart, “marking-up” is merely an act in pattern recognition. An individual opens a speech act, identifies what type of phrases signal the material of interest (such as “The President:”) and the material to be ignored (such as “Q:”), and then proceeds to insert the `<ignore>` tags in the appropriate places. If there is one task that computers are *extremely* skilled at, it is repetitive pattern recognition. We decided that there had to be a better way to complete this task, which led to a search for a solution. This search resulted in the creation of `PresidentParser`.

This article serves as more than just an introduction to `PresidentParser`, however. The program makes use of state patterns in the programming language Java, which are a fairly common method of creating a text parser. `PresidentParser`, and the description contained within this article, can serve as an example how this fairly straightforward programming method can serve as a useful middle ground in political science. This middle ground is the space between preparing data by hand, and attempting to modify, or create, more full-featured and complex language processors such as TABARI or Profiler Plus. We hope that `PresidentParser` and this article can serve as motivation for the development of fur-

ther tools to aid in the data gathering and preparation process. Towards this end, the program can be downloaded at [parser.johnbeieleer.org](http://parser.johnbeieleer.org) and the source can be viewed, and modified, at <https://github.com/Voomer/Profiler-Plus-Presidential-News-Conference-Parser>.

## PresidentParser: The Guts

Since we wanted the program to be easily portable to different operating systems, the parser used to process text is written in Java. The basic design of the program is a state machine; a state machine can have many possible states, but it can only be in one state at a given time. The program can change from one state to another when certain conditions are met. These properties of a state machine make it ideal for parsing leader interviews and other text documents, since it is possible to model the speakers as states with a change in speaker signaling the transition to another state. The interviews generally have specific signals for the change in a speaker (e.g. “The President.” “Voices.” “Reporter.”) Because the capitalization and punctuation do not occur naturally in English, it is possible to use them to define a state transition. Specifically for `PresidentParser`, there is a state for the recording of text, a state for the text to be ignored, and a preprocessing state to handle some special conditions.

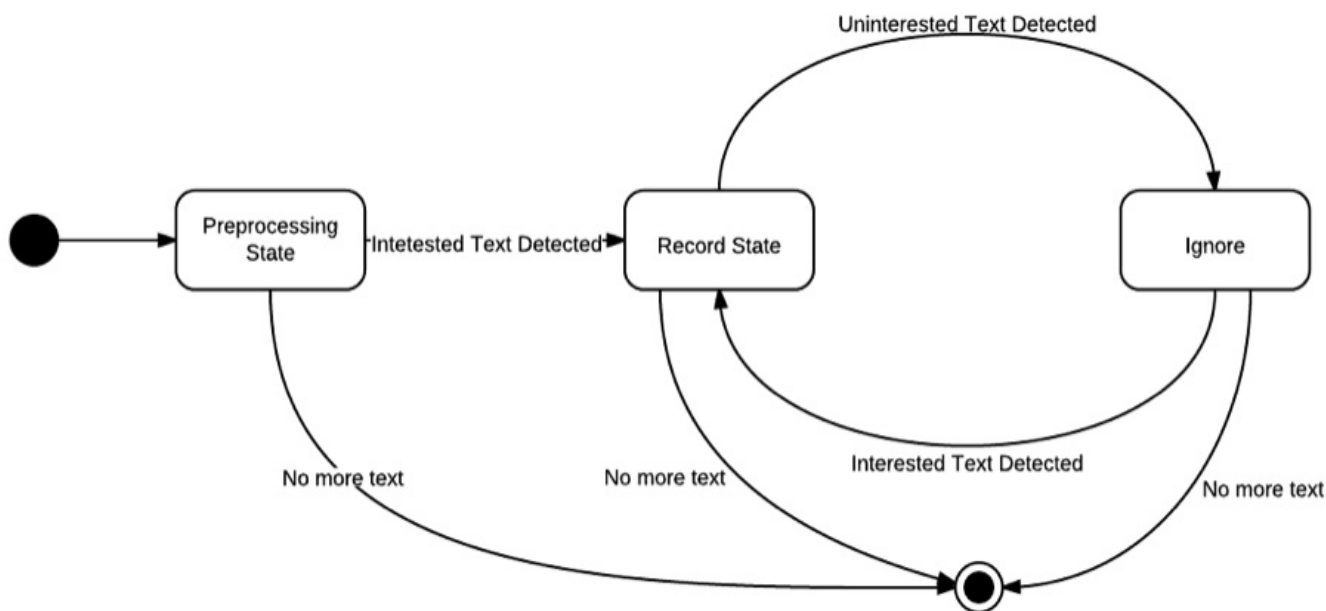


Figure 1: Concept Map for `PresidentParser`

Figure 1 shows a visual depiction of how `PresidentParser` works. The record state reads through the text, looking for signals that the speaker has changed to someone who should be ignored as defined by the text dictionaries. Once such a signal is detected, the ignore tag is appended to the speaker and the program transitions to the ignore state. The ignore state then searches for text that is to be recorded. Once this signal is found, the ignore state appends the closing ignore tag and switches to the record state. This process repeats until there is no more text. For the needs of our particular situation, namely news conferences focusing on the leaders of states, there is some additional logic to handling the peculiarities in the documents such as headers, text in parentheses, etc., but these additions do not change the overall model. The preprocessing state is added because, again specifically relating to the news conferences, the first instance of someone speaking is usually not part of the interview. If the document consists of only one speaker it is not considered an interview, so the program terminates if it reaches the end of the document without finding both states.

The general idea behind `PresidentParser`, and state machines in general, is that if a situation can be modeled inductively, then states and transitions can be used to model the problem. This leads to the more general applicability of state machines. `PresidentParser` was developed for a single situation, but the logic and methods implemented can be used in many different situations that can be modeled in a similar manner.

## How-To Guide

Now that it is clear how the program works we will provide a brief overview of how to actually use the program. `PresidentParser` is run from the terminal or command line and takes four arguments: input directory, output directory, record dictionary, and the ignore dictionary. This leads to a command that has the following form<sup>2</sup>:

```
java -jar /Users/username/parser.jar "/Users/username/input"
"/Users/username/output" "/Users/username/Record.txt"
"/Users/username/Ignore.txt"
```

The operation of the program is that simple.<sup>3</sup> The most difficult part of operating the program is the formulation of the dictionaries. There are two types of dictionaries that can be used. The first is a basic list of signals for the ignore or record states. The second makes use of regular expressions (regex) to make use of pattern matching to signal the transitions. For example, one source of speech acts was [The American Presidency Project](#). These files were formatted in a fairly consistent manner, which allowed us to create a record dictionary that included signals such as “The President:” or “The Presi-

dent.”. The ignore dictionary includes signals such as “Q.” or “Q:”. This is fairly simple; the only difficulty comes in identifying a pattern specific to a document.

For more complex documents, regular expressions can make life much simpler. For example, another focus for our project was the daily briefings given by the U.S. State Department. The problem with these documents, however, was a high level of variability in the speakers such as “Ms. Doe:” “Mr. Doe:” “Mr. Smith:” “Mrs. Davis:”, etc. This provided a problem for creating a comprehensive list of speakers to signal text that should be recorded. We were able, however, to use a regular expression since the speakers did follow the pattern of an uppercase m, followed by some letters, followed by a period and whitespace, followed by other letters, and ending in a colon. The regular expression used in this instance was “M.+?\.\s\w\*\:”. Overall, `PresidentParser` has allowed us to mark up over 20,000 speech acts. This is a task that would have taken an enormous amount of time to do by hand. The program is highly accurate in straightforward cases, such as the existence of only one speaker of interest in a document. In more complex situations, such as the existence of two speakers in addition to reporters and the person of interest, the program becomes less accurate due to the complexity and variability of the text structure.

## Other Text-Parsing Programs

There are, of course, other programs and methods for text parsing other than `PresidentParser` and state machines. In addition to the aforementioned Profiler Plus, is the `TABARI` event-data program by Schrodte. There are problems with both of these programs for the purposes that interested us. First, Profiler Plus is closed source, which means that it cannot be modified to work for different purposes. While `TABARI` is open source, it is a natural-language processor and is much too complex to be modified for the comparatively simple problems we were interested in. The same problem would apply to Profiler Plus if the source was freely available. This brings to light the beauty of state machines. They allow a person to use automated methods in a straightforward manner, without having to delve into the complexity of natural-language processing. As with marking up speech acts, it is not always necessary to bring the full power of natural-language processing to bear on a problem; it is sometimes more convenient to model a problem in an inductive manner, which is conducive to the use of state machines.

One final problem with `TABARI` is that it is written in C++. We think Java is easier to learn than C++. This allows for a lower start-up cost, which causes the creation of

<sup>2</sup>The command is formatted for OS X or Linux, appropriate adjustments should be made for Windows.

<sup>3</sup>For convenience sake the download of `PresidentParser` contains a text file that includes the terminal command that is able to be copy and pasted.

special purpose software programs to be relatively pain-free.

## Conclusion

`PresidentParser` was created to answer a specific need: automate the mark up of thousands of speech acts. The development of this program has shown the ability of state

machines to occupy a happy medium between highly complex natural-language processing programs and dealing with text manually. We hope that our work on this project will serve as an example for how state machines can be helpful in the type of work a political scientist may have to do, and as a jumping-off point for other projects.

## A Note from Our Section President

**Robert (Rob) J. Franzese, Jr.**  
University of Michigan  
*franzese@umich.edu*

### Greetings, Gratitude, and Plaudits

I am deeply honored and sincerely humbled to be writing in the capacity of President of the Society for Political Methodology and the APSA Section on Political Methodology. As the Society approaches its 30<sup>th</sup> birthday (the first summer meeting was in 1983), and with all thanks due to the tremendous academic work, the inspired and insightful leadership, and the prodigious administrative efforts of generations of political methodologists, we remain, as the previous Society president, Jeff Gill, wrote in his inaugural message: “one of the most, if not the most, accomplished groups in political science[:]. We have the top journal in the field for two years in a row [*Political Analysis*, of course, under the outstanding editorship (Thank You!) of Jonathan Katz and Michael Alvarez; oh, and it’s going on 5 years running now...], our Summer Meeting is the envy of every other subfield, and our students do extremely well on the job market... These are achievements that we should both be justifiably proud of and appreciative for the excellent leadership that we have enjoyed over the last quarter century.” Allow me to, and join me here if you will in, extending our sincere and great gratitude to Jeff for his outstanding leadership and mammoth efforts as President. Just a few of his many noteworthy accomplishments as President include aligning Summer Meeting hosts and venues through 2014 (!!), achieving incorporation and 501(c) status for the Society, initiating international partnerships at various levels and in diverse ways with the European Political Science Association (EPSA) and the European Consortium of Political Research (ECPR) and similar initiatives negotiating with other international political-science organizations with

interests in developing and furthering political methodology across the globe. Extremely importantly, Jeff spearheaded the NSF application which successfully renewed the Society’s NSF grant (with Jeff and me as co-PI’s, the grant has been renewed for two years until Fall 2013: this was by no means a formality under the conditions then or currently), so that we can continue to do the important work that this support has allowed: namely, to support graduate students and political methodologists of under-represented groups to attend the summer meetings, the “Visions in Methodology (VIM)” meetings for women in political methodology, a pair of small topically focused conferences, and other small initiatives.

### Status Report

Preparations for the 2012 Summer Meeting, being chaired by Tom Carsey and hosted at UNC-Chapel Hill and Duke, are on schedule and well underway (and the call for applications has already gone out to the PolMeth email listserve as I hope you’ll have noticed!). Regional meetings, SLAMM and NEMP, are continuing strong, and the call for proposals for the first of the small topical conferences under the new grant will be coming soon. All of our (MANY!) committees have been renewed, and refreshed as necessitated by our rules on such, for 2012 and going forward. Several of those committees have sent calls for nominations (to which I encourage you to respond). We continue to grow in numbers, and in depth and breadth of intellectual exploration, within U.S. political science and related fields, and across the globe. Our finances remain quite comfortably strong and protected. In sum, the state of the Society, my fellow methodologists, is strong indeed. Or, again as Jeff wrote in his inaugural message: “...we are in a very healthy position, admired by our colleagues outside of the subfield, financially secure within the subfield, and able to take on new challenges.” Jeff did indeed take on many new chal-

allenges, and to great success, as noted above, and once more please join me in expressing our thanks and appreciation.

### **Goals, Plans, and a Call**

Accordingly, the goals I have set for my own term are, first and foremost, to sustain this across-the-board health, strength, and vigor of the Society and all our endeavors. I would like also to advance and strengthen especially some

specific dimensions of our laudable endeavors. Most importantly, I would greatly welcome and hereby call for ideas, proposals, or suggestions on new initiatives, or areas for redoubled efforts in, our work to further political methodology and political methodologists in under-represented groups. Please feel free to contact me with any suggestions or comments that you might have on that or any other dimension of our work as a Society. I look forward to working with you over the next two years.

University of Illinois at Urbana-Champaign  
Department of Political Science  
420 David Kinley Hall  
1407 W. Gregory Drive  
Urbana, IL 61801

*The Political Methodologist* is the newsletter of the Political Methodology Section of the American Political Science Association. Copyright 2012, American Political Science Association. All rights reserved. The support of the Department of Political Science at the University of Illinois in helping to defray the editorial and production costs of the newsletter is gratefully acknowledged.

Subscriptions to *TPM* are free for members of the APSA's Methodology Section. Please contact APSA (202-483-2512) if you are interested in joining 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 may be sent to any of the editors, by e-mail if possible. Alternatively, submissions can be made on diskette as plain ascii files sent to Wendy K. Tam Cho, 420 David Kinley Hall, 1407 W. Gregory Drive, Urbana, IL 61801.  $\text{\LaTeX}$  format files are especially encouraged.

*TPM* was produced using  $\text{\LaTeX}$ .



**President: Robert Franzese**  
University of Michigan  
[franzese@umich.edu](mailto:franzese@umich.edu)

**Vice President: Kevin Quinn**  
University of California at Berkeley, School of Law  
[kquinn@law.berkeley.edu](mailto:kquinn@law.berkeley.edu)

**Treasurer: Luke Keele**  
Pennsylvania State University  
[lj20@psu.edu](mailto:lj20@psu.edu)

**Member-at-Large: Fred Boehmke**  
University of Iowa  
[frederick-boehmke@uiowa.edu](mailto:frederick-boehmke@uiowa.edu)

***Political Analysis* Editors:**  
**Michael Alvarez and Jonathan Katz**  
California Institute of Technology  
[rma@hss.caltech.edu](mailto:rma@hss.caltech.edu) and [jkatz@caltech.edu](mailto:jkatz@caltech.edu)