Three Essays in Political Methodology

by

Arthur Peter Andrew Spirling

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Curtis Signorino

Department of Political Science

The College

Arts & Sciences

University of Rochester

Rochester, New York

2008

For my mother and for my father.

Curriculum Vitae

The author was born in Carshalton in the London Borough of Sutton, United Kingdom, on August 26, 1979. He attended the London School of Economics from 1997 to 2001 and graduated with a Bachelor of Science degree in 2000. He received a Master of Science degree in 2001 from the same institution, and then attended Nuffield College, Oxford University, from 2001 to 2003. He came to the University of Rochester in the Fall of 2003 and began graduate studies in Political Science. He pursued his research in political methodology under the direction of Professor Signorino and received the Master of Arts degree from the University of Rochester in 2007.

Acknowledgements

Social reformer Sir Thomas Fowell Buxton once modestly observed that "with ordinary talents and extraordinary perseverance, all things are attainable." Those things apparently include my completion of a dissertation in political science at the University of Rochester, although even Sir Thomas would be surprised to learn quite *how* ordinary one's talents can be in practice. With this in mind, there are several individuals who I wish to thank for their contribution to my education, work and emotional well-being over these last several years.

Chronologically, the story begins at Nuffield College, Oxford, where Iain McLean was a sage source of counsel and helped me garner admission to the University of Rochester in 2003. David Firth tried patiently to teach me some statistics, and has been a constant source of helpful advice since that time. Working and coauthoring with both these individuals has taught me much about our related interests, and the travails of getting published.

At Rochester, Curtis Signorino, my advisor, allowed me the flexibility and support to explore new avenues of methods research without the micromanagement that would have pained us both. Preferring to gently steer rather than direct, Curt provided much needed pastoral care when I felt particularly frustrated and anxious—especially in the last year of my dissertation writing, which coincided with the stress of the market. Emphasizing the 'grand plan' as well as the details, Curt has taught me a great deal about how to 'play the game' of our discipline. Also serving on my committee, Tasos Kalandrakis is one of the hardest working men in the business, and his diligence in reading and commenting on my papers is second-to-none. Our projects together have been the most challenging and rewarding experiences of my career to date, and I expect our cooperative efforts will continue for some time yet. Larry Rothenberg stepped forward when I needed an extra body on my committee for bureaucratic reasons, but he has become an important member of my team in the meantime. I thank him for his time and advice. Though not on my committee, I am very grateful to Michael Peress and Tony Lenzo who both supported my work in quite different ways. Michael's deep understanding of econometrics allowed him to provide valuable technical feedback on my papers. Tony, who managed `the star lab` during the writing phase of my dissertation, ran a very tight ship and ensured that both the hardware and software I had access to was in the best possible working condition.

In the broader methods community, I am very grateful to Jeff Gill and Andrew Martin at Wash U, Michael Herron at Dartmouth and Kevin Quinn at Harvard. All these people kindly spent time helping me improve my papers with lengthy and valuable comments.

My comrades in the department were constant sources of amusement and encouragement throughout the writing of this dissertation; in my cohort, I thank Matthew Platt, David Carter, Michael Epstein, Arnd Plagge and Subhashish Ray. Peter Loewen spent an all too short visit at Harkness, during which time we became firm friends. In the early days, Justin Fox, Antoine Yoshinaka and Kris Ramsay were a strong influence on my outlook and behavior as a graduate student.

I wish to convey a very special thanks to my family that loved me whether my research was progressing or stagnating: my mother, my father, my sister and my brother. They were constant in their support, affection and financial assistance; I hope I can, one day, repay these debts. Finally, perhaps most of all, I thank Sonia Tasbas: without her love and kindness, I know only that my outlook—academic or otherwise—would be the poorer. I remain truly grateful that I have found such a wonderful woman to spend my life with.

Abstract

This political methodology dissertation consists of several distinct essays that apply three new estimation techniques to all three subfields of the discipline—international relations, American politics and Comparative politics. The first is an application of reversible jump Markov chain monte carlo, a more general form of MCMC popular for model search problems in statistics. We apply RJMCMC to the current Iraq conflict in order to identify change points in terms of civilian casualty numbers. The second chapter uses the Bradley-Terry model for pairwise contests to estimate the 'power' of actors in legislatures. We apply this generalized linear model to the United States Senate. We use both an unstructured and covariate based model to show that, *inter alia* senators' party identification, ideological tendency, leadership rank, service length and gender all affect their ability to influence others in the chamber. The third chapter shows that British Members of Parliament's voting decisions on roll calls can affect their constituency performance at election time. In particular, 'rebeling' on government bills apparently hurts them, while voting independently on less important matters tends to benefit them. We use a non-parametric, random forest algorithm to estimate the relationship, since the number of parameters (the number of roll calls) far exceeds the number of observations (constituency performances) to be predicted.

Contents

List of Tables

List of Figures

Chapter 1

Introduction: Purpose and Plan

As befits its title, the following dissertation is composed of three self-contained essays within the broad rubric of 'political methodology.' For current purposes, this term refers to the use of *quantitative* techniques that can give scholars insights into political phenomena. The stress on 'quantitative' is important since the chapters below all make claims that attributes of political events and actors can be *measured* and *compared*. In line with this, the essays develop tools that other political scientists can use to analyze the data scenarios they face. Though these themes are consistent threads, the dissertation is eclectic: in terms of the models of inference, the nature of the techniques used, the substantive topics of investigation, and the questions that are actually answered. To this end, one of the papers is avowedly Bayesian, while two are not. Two papers utilize parametric methods, while the third uses non-parametric techniques. The papers cover the three broad areas of substance seen in the discipline: International Relations, American and Comparative politics respectively.

Pursuing such a diverse agenda in one work is not uncontroversial; King (1989), for example, argues for a 'unified' methodology (in practice, maximum likelihood) in political science that would facilitate the application of consistent standards, communication and the cumulation of knowledge within the discipline. With good reason, he warns against techniques which are "imported intact and without adaptation" from other disciplines (King, 1989, 3). The current dissertation is more sanguine about such endeavors: it borrows techniques more commonly seen in bio-statistics, pattern-recognition and market research literatures. The contention is that, if assumptions (and thus limitations) are clearly described, readers may judge for themselves whether or not the model is appropriate. In any case, in reviewing the technical literature in each essay, an argument is made for the *absence* of an existing method to solve the specific problem addressed. Moreover, as will be explored, the particular problems here are neither uniquely rare nor trivial—that is, the chapters are not 'solutions in search of applications'.

A second concern for work in political methodology is that of specifying appropriate 'theory'. Achen (2002, 424), for example, argues that "we have yet to give most of our statistical procedures legitimate theoretical micro-foundations." For Achen (2002, 437), a micro-foundation is "a formal model of the behavior of the political actors under study. The model might emerge from decision theory, game theory, or some other formalism. Then the statistical setup is derived from the model, with no further ad hoc adjustments." In the second chapter, this prescription is broadly followed: a model of human behavior is suggested and an estimator derived directly. In the first and third chapters it is not. That is in part because these other essays are more concerned with data exploration and summary, rather than establishing "reliable empirical generalizations." In statistics, such endeavors are common and their usefulness is in part derived from the fact that they simplify and describe (for the analyst) an otherwise complex picture of information. As a first pass on the data they are helpful: previously hidden regularities emerge and help scholars (including the present author) to think more systematically about what is driving behavior 'under the hood.' Of course, it is a thin line between 'data mining' and 'data dredging' but these essays make a good faith attempt to avoid the latter. In any case, as argued above, introducing new techniques to the discipline where current ones simply cannot be used is a fruitful exercise.

1.1 Plan

The first essay (Chapter 2) uses a reversible jump Markov chain monte carlo approach for examining structural breaks in the current Iraq conflict. The interest is in identifying patterns of violence, especially as they relate to polity building. We find four 'turning points'—all of which correspond to increased civilian casualties. These are arguably congruent with insurgent responses to widely publicized events including the arrest of Saddam and the holding of elections. Thus the chapter makes a contribution to the systematic study of international conflict and comparative political development. Methodologically, the chapter adds to the political science toolkit by suggesting a more general Bayesian approach to change point detection that is philosophically appealing.

The second essay (Chapter 3) suggests a new way of measuring 'power' for actors in structured settings like legislatures and courts. In contrast to *a priori* pivotality indices (of which Shapley-Shubik and Banzhaf are most famous) which have been criticized for their measurement of 'luck' rather than 'power', this essay introduces a data-driven Bradley-Terry approach that proceeds directly from a theoretical model of (pairwise) actor behavior. We apply the model to the 108th Senate. In its unstructured form, we are able to produce a valid rank ordering of senators, with individuals such as John McCain, Bill Frist and Robert Byrd appearing in the power top ten. When a structured, covariate model is used, we show that, empirically, chamber centrists lack power relative to *party* medians. Moreover, we are able to demonstrate that, contrary to some expectations, female senators tend to be as powerful (on average) as males.

The third essay (Chapter 4) uses a non-parametric regression approach to examine the relationship between a member of the British parliament's legislative activity and their constituency electoral performance. In contrast to assumptions of no effect, we show evidence of a subtle reward/punishment dynamic, contingent on the nature of the roll calls in question. This paper combines quantitative and qualitative research on roll calls in the years 1997–2001. Methodologically, we describe a new 'random forests' approach for political scientists interested in problems for which the number of parameters far exceeds the number of observations, and for which the causal story is potentially complex and non-linear.

We conclude in Chapter 5 with a discussion of future applications and extensions to the work here, along with some comments on the role and future of political methodology more generally.

Chapter 2

‘Turning Points’ in the Iraq Conflict: Reversible Jump Markov Chain Monte Carlo in Political Science

President Bush believes that the region is at a true turning point. He believes that the people of the Middle East have a real chance to build a future of peace and freedom and opportunity.

—Condoleezza Rice, National Security Advisor (2003)

2.1 Introduction

The study of *inter*– and *intra*– state conflict is a mainstay of political science. As an *international* conflict that increasingly resembles a *civil* war, the current situation in Iraq provides both a testing ground for theories on the duration and termination of different types of conflicts (e.g. Filson and Werner, 2004; Stam and Bennett, 2006), as well as a rich source of data for empirical work. This is quite separate from its obvious importance as a political, military and economic event in progress. In part due to its contemporaneous nature, political scientists have access to carefully, daily recorded, military and civilian casualty information: an unusual and excitingly fine level of detail. Of course, the utility of any data is only as good as the way it is explored and analyzed. Here, we suggest that a fruitful approach for political scientists lies in examining the time series for (potentially multiple) structural breaks and their effects. For scholars of American politics and public policy, the way that these change points correspond with administration statements on the progress of the war may be particularly intriguing. This notion extends to Comparative institutions scholars interested in the potentially pacifying effect of various post-war 'state-building' activities. In keeping with the increasing acceptance and popularity of Bayesian methods in political science, in undertaking our study we justify and adopt a novel (to political science) approach that uses a more general form of Markov chain monte carlo (MCMC) techniques, well-known to statisticians as 'reversible jump' MCMC (Green, 1995). We do so primarily for computational reasons.

Examining civilian casualty data from the official cessation of hostilities (May 2003) to May 2007, we find evidence of four change points. These breaks are approximately contemporaneous with (1) the capture of Saddam, and the emergence of the *Abu Graib* scandal (late 2003 to Spring 2004); (2) the installation of the Iraqi Interim Government, and the subsequent handover of power to the Iraqi Transitional Government (Summer 2004 to early 2005); (3) the legislative elections for, and negotiations to form, the first full-term Iraqi government (the early months of 2006); (4) the assumption of security and some military responsibilities by the Iraqi government (August/September 2006) . In every case, the frequency with which such incidents occur is *increasing* after the break.

2.2 Background and Data

The United States and allied forces attacked Iraq with aerial bombardments, followed by a land invasion, on March 20th, 2003. By mid-April, Iraq's capital city, Baghdad and Saddam Hussein's home region of Tikrit was under allied control—bringing a *de facto* end to the war. A fortnight later, on May 3rd, 2003, President Bush declared that allied combat operations would now officially cease. As with all conflicts, the war has not been costless. What marks the Second Iraq War though, is the *continued* loss of life *after* the Iraqi army was formally defeated. At the time of writing, some 3,000 coalition force members had died in addition to at least 57,000 civilian fatalities since military operations began (sources are `http://icasualties.org/oif/` and `http://www.iraqbodycount.org/` respectively. Some studies have placed the number of civilian fatalities at a much higher number. For example, Burnham et al. (2006) claim up to 600,000 deaths). Violence has not yet abated despite the passing of some presumably important landmarks in what some characterize as the development of Iraq's polity and stability: for example, the capture of Saddam (December, 2003), the placing of the former dictator on trial from 'crimes against humanity' (July 2004) and his execution (December 2006); the killing of Saddam's sons, Uday and Qusay (July 2003); National Assembly elections (January 2005); the drafting (December 2003–March 2004) and subsequent referendum approval (October 2005) of a constitution; the election of a new president (April 2005) and the forming of a governing coalition (May 2006); the execution of an *Al-Qaeda* ringleader, Abu Musab al-Zarqawi, thought responsible for planning many terrorist attacks (June 2006); the assumption of security responsibility by the Iraqi government (September 2006). We are interested in violence for the post-(official) war period: although we certainly cannot make firm causal claims, our study will enable us, for example, to make statements about the plausibility of various events as “turning points” and allows us to pass some exploratory comments on how new democratic institutions and state apparatus developments are effecting Iraqis. Hence our study focuses on May 3rd 2003 through to the present time of writing (May 2007).

Our data are drawn from `iraqbodycount.org` a (online) data base that records civilian deaths in Iraq “that have resulted from the 2003 military intervention by the USA and its allies. The count includes civilian deaths caused by coalition military action and by military or paramilitary responses to the coalition presence (e.g. insurgent and terrorist attacks)” (Dardagan and Sloboda, 2006). The data in raw form record deaths at the *day* level, from January 2003 through to the present and are compiled from (primarily Western) media reports and other sources. Since uncertainty often exists on precise numbers, especially when different agencies have conflicting figures for the same incident, the data base reports a range of possible death numbers from a ‘minimum’ to a ‘maximum.’ Potential ‘over-counting’ is a concern, so we use the ‘minimum’ and define a ‘casualty incident’ as involving five deaths or more (our findings below are similar when we define the incident threshold at ten or twenty deaths). For the purposes of this paper, we focus on the (changing) *frequency* of attacks, rather than their *size* (above our minimum). In part this is a behavioral assumption: we would contend that, at least initially, terrorists were able to control *how often* they planned to inflict casualties, rather than *how many.* There were 1682 such incidents in our time series, and we graph their occurrence in Figure 2.1; there, the solid line is the cumulative incident count, the solid dots are simply jittered incident occurrences (for which the y-axis is *not* the scale). We also report various dates that may of interest and to give readers a sense of timing perspective. Although univariate time series work is not regularly encountered in political science, it is valuable in the current context as a 'first glance' exploration before covariate information becomes available. We think that such work helps to prompt both theorizing and data gathering for more nuanced and sophisticated analysis.

2.3 Estimation Problem

The *single* change point problem, estimated using Markov chain monte carlo techniques, has been discussed for and by political scientists elsewhere (see Western and Kleykamp, 2004). That treatment is similar to the (hierarchical) presentation given by Carlin, Gelfand and Smith (1992): suppose $\mathbf{y} = (y_1, \ldots, y_T)$ is a vector of observations of the random variable Y (casualty incidents) over time and let f and g be unknown densities in the same parametric family with $y_i \sim f(Y|\lambda_1), i = 1, \ldots, k$, $y_i \sim g(Y|\lambda_2), i = k+1, \ldots T$. We wish to estimate k the (single) change point which takes (discrete) values

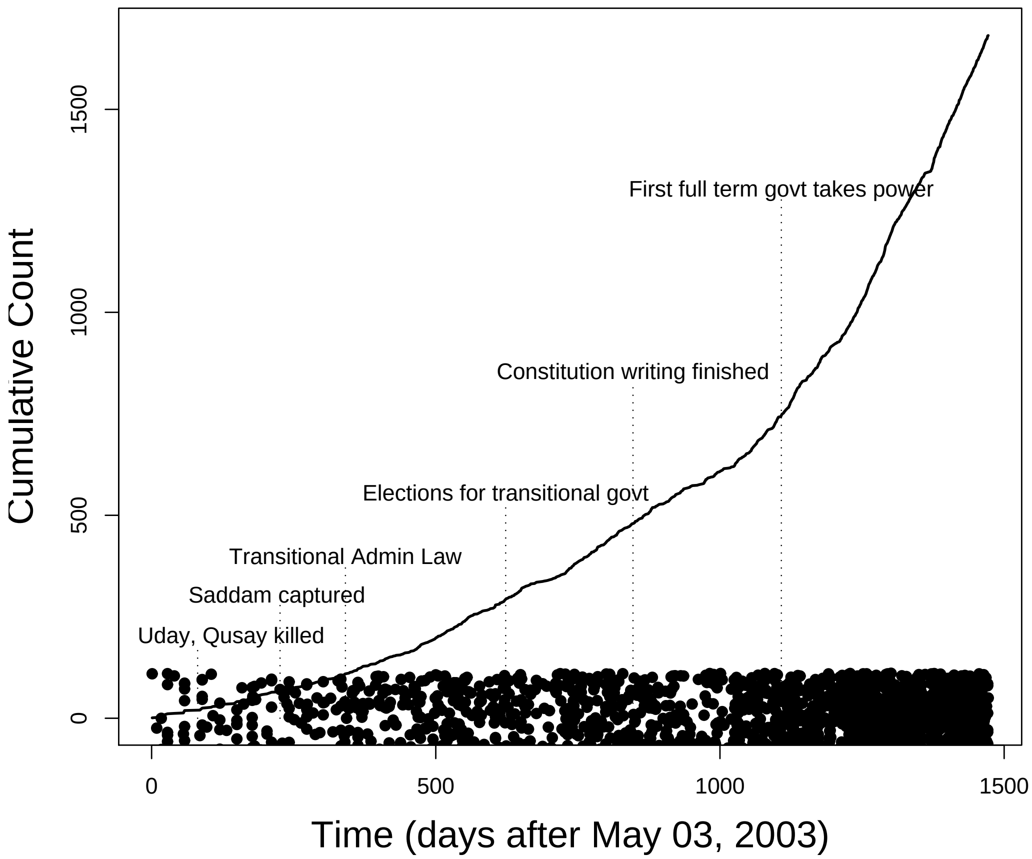

Figure 2.1: Iraq Casualty Incidents, May 2003–May 2007. Solid line represents cumulative counts; solid dots are jittered incident occurrences; various dates of interest demarcated.

in $\{1, 2, \ldots, T\}$. A frequentist approach proceeds by maximizing

$$\mathcal{L}(\mathbf{y}) = \prod_{i=1}^{k} f(y_i|\lambda_1) \prod_{i=k+1}^{T} g(y_i|\lambda_2) \tag{2.1}$$

to obtain k and the parameters λ_1 and λ_2 (which for the count case are arrival rates for a Poisson) if they are of interest. A Bayesian approach proceeds by placing a prior $\tau(k)$ on the change point. There are computational advantages of a Bayesian MCMC approach here since (a) maximizing (2.1) requires optimization in a space that is not continuous (recall that k is discrete) which, say, Gibbs sampling does not; (b) the resultant non-nested models may be straightforwardly compared using Bayes factors (Chib, 1998); (c) missingness in $\mathbf{y}$ is handled systematically. This is quite apart from the philosophical appeal of Bayesian approaches of which political scientists are increasingly aware (see, for example, Gill 2002, 1–6 and Jackman 2004, 486).

Here, we are interested in exploring *multiple* change point and such work (Bayesian or otherwise) is much less common in political science. In part this is because, with respect to the logic above, there are profound computational difficulties in generating proposals for situations where we suspect there are more than a couple of change points. One approach, suggested by Chib (1998) and applied to American politics by Park (2006), treats the change point model as a type of time series Markov mixture model, where the observations are (assumed) drawn from latent state variables. Notice that this approach requires *separate* Markov chain monte carlo runs for the different numbers of change points hypothesized (Leonte, Nott and Dunsmuir, 2003).

An alternative solution is to use *reversible jump* Markov chain monte carlo which allows us to complete the computational operations in one 'go' as well as allowing us to be *a priori* agnostic over the number of parameters to be estimated.

Typically when MCMC is used in political science the parameter vector θ has a known number of components, denoted n. For the single change point problem $n = 3$ (these are k, λ_1 and λ_2). Now consider a very different scenario which arises for an unknown number of k change points: for every possible k, we need to estimate $2k+1$ parameters—the change points themselves and then parameters of the densities before, between and after them. That is, we have a set of $\mathbb{M}_k = \{1, \ldots, K\}$ candidate *models* of our data generating process, each with a *different* number of parameters. Otherwise put, the number of parameters is, of itself, a parameter. More formally, the k^{th} model in $\mathbb{M}_k$ has associated parameter vector θ_k which contains n_k parameters such that $\theta_k \in \mathbb{R}^{n_k}$.

Continuing to denote our data vector $\mathbf{y}$, the joint distribution becomes:

$$\begin{aligned} p(k, \theta_k, \mathbf{y}) &= p(\mathbf{y}|k, \theta_k)p(k, \theta_k) \\ &= p(\mathbf{y}|k, \theta_k)p(\theta_k|k)p(k). \qquad (2.2) \end{aligned}$$

Since we have a constant of proportionality we can rearrange and reexpress (2.2) into the more familiar

$$\underbrace{p(k, \theta_k|\mathbf{y})}_{\text{posterior}} \propto \underbrace{p(k)p(\theta_k|k)}_{\text{prior(s)}}\underbrace{p(\mathbf{y}|k, \theta_k)}_{\text{likelihood}}. \tag{2.3}$$

Notice that $p(\mathbf{y}|k, \theta_k)$ is simply the likelihood, while $p(\theta_k|k)$ is the prior for the parameter vector, given a particular data generating process and $p(k)$ is the prior on the model itself. We wish to generate samples from (2.3). Setting up a Markov chain to do this may be difficult though, because it is required not simply to move around the parameter space for any particular θ_k, but to also 'jump' from space to space (from model to model) depending on the k in question.

This type of problem is given a general formulation by Green (1995), known as reversible jump MCMC (RJMCMC), of which standard MCMC algorithms are special cases. Green explicitly discusses a Poisson count change point problem and we followed his approach for our application (though we varied the priors somewhat to ensure that our results were robust to such alternative specifications). Although well known to statisticians, the details somewhat technical, and readers are guided to Brooks (1997) who gives an accessible overview for political scientists.

The *implementation* of RJMCMC, in particular the efficiency of proposals, can be problematic in practice and Hastie (2005) devotes considerable attention to designing a technique to do this. We used his `Automix` sampler (with a maximum of ten possible change points) for our estimation. Though the full details are somewhat technical, drawing on Hastie (2005, 202–203), it is instructive to summarize the way that the model of the data generating process is selected. The first two stages of the sampler produce a Normal mixture distribution for every possible value of k. In the third stage, assuming the Markov chain is currently in state (k, θ_k), `Automix` allocates the parameter vector θ_k to a component l_k of the mixture and uses it to standardize θ_k. Then a new model k' is proposed, along with a commensurate (new) mixture which has component l'_k. To obtain the new state vector $\theta'_{k'}$, the standardized vector is transformed using the mean and the covariance matrix of the mixture component $l'_{k'}$. `Automix` then accepts the proposed state $(k', \theta'_{k'})$ with some specified acceptance probability. A particularly pleasing feature of this software is that issues such as burn in and the requisite number of post-burn iterations are handled automatically.

2.4 Results

There are three sets of (posterior) distributions that interest us here:

1. the posterior of k: this enables us to answer the question "how many change points in the data?" This will have support $k = 1, \ldots, k_{\max}$ where $k_{\max} = 10$.

2. the posterior of change point *positions* conditional on some estimated k. More intuitively, this enables us to answer the question "given a particular number of change points, *when* did they occur in the data?"

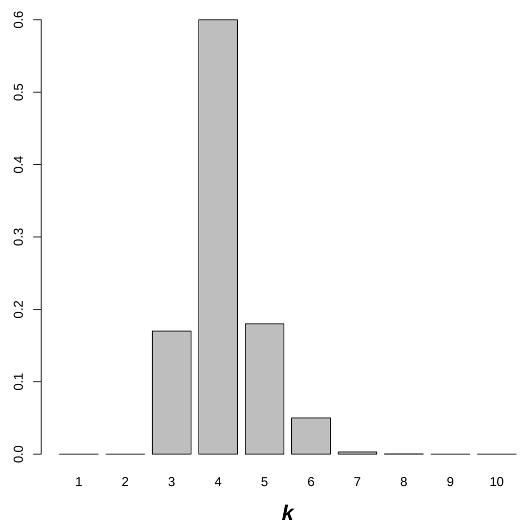

Figure 2.2: Posterior of k: number of change points, Iraq casualty data.

3. the posteriors of the rates for each period, conditional on some estimated k: that is, given the *number* of change points, and *when* they occurred, we can answer "what were the *effects* of the change points?"

In Figure 2.2 we display the posterior for k, the number of change points. The strongest evidence (in the sense of Kass and Raftery (1995)) is for $k = 4$ and we will explore this possibility exclusively. In Table 2.1 we summarize the results for $k = 4$ model in a way that answers questions 2 and 3 above. The first break, in late January 2004 occurs between incidents that may be of import. The first was the capture and arrest of Saddam at a farmhouse near Tikrit in December 2003. The subsequent months saw both an insurgency uprising lead by rebel Shia cleric Muqtada al-Sadr in Baghdad and the diffusion of abusive photographs taken at the *Abu Graib* prison where coalition forces were holding Iraqi detainees. The political fallout of the latter was profound, and criticism of the Bush administration by allied, Arab and other politicians was widespread. This event, arguably, rallied and spurred sectarian hatreds and violence. The break marks a sharp increase in the casualty rate, doubling from one incident every four days, to one every two days.

The second break occurs in August of 2004, a little while after the Iraqi Interim Government assumed power from the Coalition Provisional Authority (in June 2004). This new entity, under the Premiership of Iyad Allawi was subsequently recognized as the legitimate sovereign government of Iraq by both the United Nations and the Arab League (an important regional player). Allawi quickly announced new security measures to tackle insurgency forces and was criticized by some for their draconian nature. As part of this offensive, the Iraqi Interim Government began to censor the critical reports of media outlet *al Jazeera*. Included in the highest posterior density interval for this break is the January 2005 democratic elections for the Iraqi Transitional Government. This change point saw an increase in violence from one incident every two days, to four incidents every five days.

The third break itself, in February 2006, occurs not long after the elections for the first full term Iraqi government (December 2005) and at around the time of the protracted negotiations to form a new coalition government. These talks were deadlocked for some time, lasting from December through to April of the following year. Jawad al-Maliki, leader of the Islamic Dawa

Party would become Prime Minister after the original candidate Ibrahim al-Jaafari proved unacceptable to the Sunni and Kurdish representatives in parliament. Once again, violence surged after this point with, on average, five incidents occurring every three days.

The fourth and final break occurs in September of 2006, a time when the Iraqi government assumed control of national security for approximately 70 percent of the country. The first specific task of the Iraqi Security Forces was, and is, to tame insurgency (with coalition logistical and medical support). By now incident rates were approaching three per day. In Figure 2.3 we summarize our findings in a different way: the open circles represent the median incident rate between the relevant breaks which are demarcated by the broken lines. For reference, we again draw the jittered incidents themselves on the plot.

2.5 Discussion

Our study—to our knowledge the first that uses RJMCMC in a political science context—suggests that violence is increasing and that important state-building activities, like democratic elections, are contemporaneous with upticks in casualties. Apart from this rather grim substantive conclusion, we found that investigating time series on violence to be an interesting and fruitful exercise. If a Bayesian approach is pursued, then reversible jump techniques seem most helpful. We hope that our discussion here will encourage others in political science to consider such methods in future.

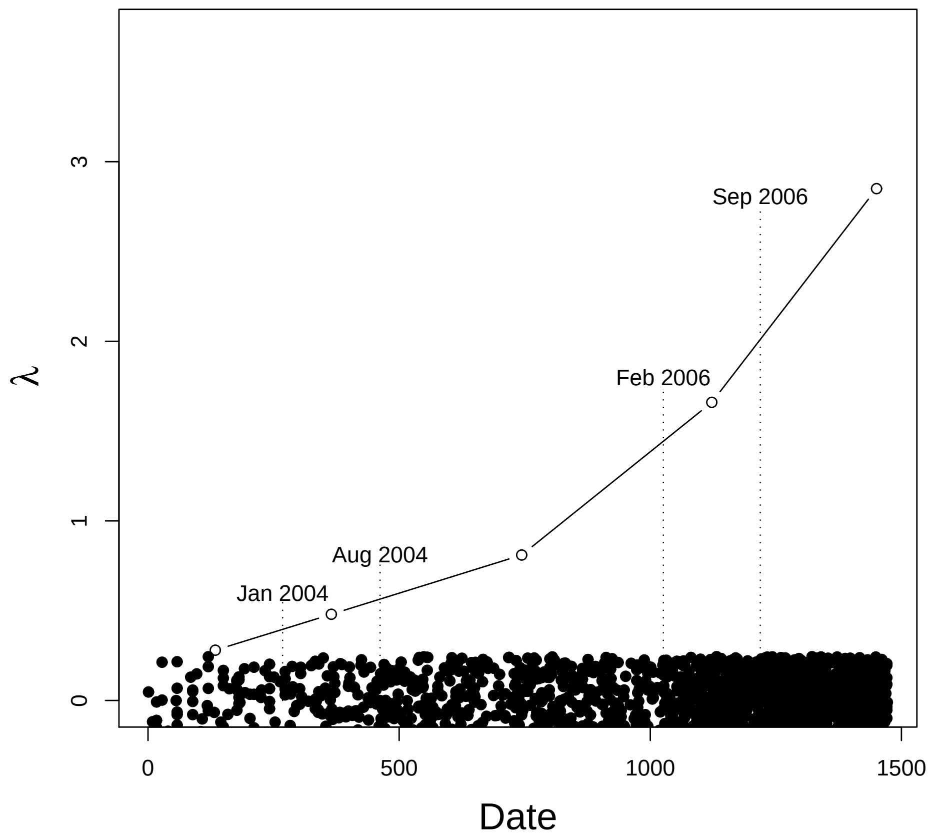

Figure 2.3: Change in rate of casualty incidents over time: rate on y-axis, actual incidents as jittered points. Dates correspond to breaks found.

As noted above, we do not establish causation in any sense: the events we noted were simply occurring at around the same time as the breaks in the time series and it is speculative that they may be of direct importance. This suggests some interesting avenues for future research: for example, one possibility is that increasing violence is a product of an increasingly organized insurgency. On this point, notice that the solid line in Figure 2.1 resembles a exponential curve of form $F(t) = a\mathrm{e}^{rt} + \epsilon$ (we are indebted to an anonymous *TAS* referee for this observation). Modeling—both theoretical and statistical—of this apparent pattern would allow us to think about the development of the conflict in Iraq in a more systematic way. We leave this for future research.

Period	Time	Rate (λ)	Comments
before break 1		0.283	1 incident every 4 days
break 1	Jan 26, 2004 [Jul 18, 2003–Apr 30, 2004]		Capture of Saddam (Dec 2003) *Abu Graib* scandal (Apr 2004)
btwn break 1 and 2		0.476	1 incident every 2 days
break 2	Aug 7, 2004 [Jul 10, 2004–May 4, 2005]		Iraqi Interim Government assumes power (Jun 2004) elections for Iraqi Transitional Govt (Jan 2005)
btwn break 2 and 3		0.814	4 incidents every 5 days
break 3	Feb 22, 2006 [Feb 16, 2006–Mar 24, 2006]		elections for first full term Iraqi government (Dec 2005)
btwn break 3 and 4		1.660	5 incidents every 3 days
break 4	Sept 4, 2006 [Aug 23, 2006–Sept 20, 2006]		Iraqi govt assumes control of counter-insurgency operations for much of country (Sept 2006)
after break 4		2.850	8 incidents every 3 days

Table 2.1: Change point median dates [90% HPD] in Iraq casualty data, with rates between breaks. Comments note political and other activities approximately contemporaneous to break.

Chapter 3

Measuring Power in Political Science: A New Method with Application to the Senate

Power is not revealed by striking hard or often, but by striking true.

—Honoré de Balzac

3.1 Introduction

Scholars have long conceptualized politics as a process of conflict over resources. To determine "who gets what, when and how" (Lasswell, 1936), political scientists have invoked notions of 'power' (Dowding, 1996, 1–8). Early positive political theory made the case that median voters—be they in the electorate or in parliaments—were particularly decisive and hence 'powerful' (Black, 1948; Downs, 1987; Riker, 1962). Later theorists—particularly those associated with the 'Rochester school' (Amadae and Bueno de Mesquita, 1999)—have focused on the role of institutional rules and norms in giving 'players' advantages over one another in settings like Congress. This literature is, by now, enormous, and includes foundational work by Riker (1982, 1986) and others (e.g. Shepsle (1979), Shepsle and Weingast (1987), Krehbiel (1991)). Despite this theoretical work, the *empirical* investigation and measurement of power—even in structured settings like parliaments, committees and courts—has proved difficult and contentious, concentrating on *a priori* metrics that are removed from the data that we have regarding agents' actual actions and decisions (Felsenthal and Machover, 1998, 2004).

This paper is an attempt to address this problem and to bridge a gap between the theoretical treatments of power familiar to positive political theorists and the empirical work of political methodologists. Following Dowding (1996), we suggest an *actor-based*, *data driven* approach. Thus, we treat 'power' as a latent variable or 'ability' possessed to varying extents by actors like Congressmen (Senators in particular). We demonstrate that this trait is straightforwardly uncovered by studying easily available voting records. A strength of this approach is that the definition of power is theoretically removed from factors that influence *how* powerful individuals actually are. Methodologically, our contribution to the discipline is the introduction of a new version of the Bradley-Terry (Bradley and Terry, 1952) model for pairwise comparisons, which includes regressors that are thought to covary with actors' power. We show the utility of both the conceptual and empirical suggestions by studying 'power' in the 108th United States Senate where we show the role that ideology, institutional arrangements, geography and personal factors play in determining the power of its members. To wit, in Section 3.2 we discuss the problems with traditional voting index measures of power; in Section 3.3 we suggest, derive and discuss a new data driven model; in Section 3.4 we apply the model to the 108th Congress and produce a ‘power list’ of Senators therein; in Section 3.5 we show how the power can be predicted by its proposed ‘causes,’ and we discuss the impact of, *inter alia*, party, committee assignment, geographic and personal factors. In Section 3.6 we conclude and suggest future avenues for research.

3.2 Problems with Traditional Approaches to Measuring Power

Scholars have been interested in measuring the power of actors in structured settings for over half a century.[1] Beginning with Shapley and Shubik's seminal 1954 contribution, with extensions most famously by Banzhaf (1965), this branch of positive theory has failed to find widespread acceptance in political science for at least three reasons. First, when voters have different weights, different indices yield different results for the same data and, generally, the resulting ambiguity cannot be resolved because there is no objective evidence on the actual distribution of power (Leech, 2002*a*,*b*). Second, for institutions with large numbers of actors, there are computational difficulties with performing the requisite calculations (Leech, 2002*a*,*b*). Third, there

[1] Social scientists have discussed ‘power’ in various ways for well over a century: at least since, for example, Marx and *Das Kapital*. The American Political Science Association felt it was sufficiently important and unresolved even in 2006 to devote their annual meeting to “Power Reconsidered.”

is much sceptism on the question of whether the indices actually measure 'power' at all (Barry, 1991; Dowding, 1996; Riker, 1964).

The third criticism is the central issue here. Barry (1991) argues that, since power indices are based on the probability that an actor is pivotal, they measure something akin to 'luck' or 'decisiveness' rather than power.[2] The point is that any definition or measurement of power must incorporate the notion of getting one's preference in the face of resistance from other actors. Moving on from this notion, Dowding (1996) considers pivotality as a *resource* among many.[3] The implication is that one's power index score is simply one of several independent variables explaining or 'causing'—but theoretically distinct from—an individual's 'power' (which might be thought of as a latent variable). Thinking *this* way enables political scientists to break away from some nonsensical statements such that the power of the Chief Justice in the Supreme Court is $(\frac{1}{9})$, or that the Chairman of the Finance Committee in the United States Senate is is equal in power terms to a freshman senator who has never held a committee post $(\frac{1}{100})$.

In sum: power indices whatever their hue, are not ideal tools for studying what political scientists conventionally think of as 'power.' A new method is needed which, *inter alia*,

[2] A natural defence is that power indices are intended for *a priori* analysis; that is, power indices tell us about the distribution of power *before* we consider the preferences of actors or their ability to set an agenda (see Felsenthal and Machover (1998) for a comprehensive review. See also: Lane and Berg (1999) and Holler and Widgren (1999) in their response to Garrett and Tsebelis (1999)).

[3] See also Krehbiel (1998) for a development of 'pivotality' with respect to US legislatures.

1. defines power as the capacity to get what an actor wants in the face of resistance...

2. treats power as a (latent) capability...

3. estimates an actor's power as a function of independent variables...

4. incorporates commensurate statements of uncertainty and...

5. allows for explicit comparison of the effects of different predictors.

In the next section, we show one way to achieve these aims.

3.3 Statistical Theory and Model

'Power,' as considered here, is a *latent* variable. Since we cannot measure it directly, we need a statistical model that takes *observable* data, estimates an unobservable trait, and outputs a metric. We made the case above that, in the broadest sense, power is an actor's ability to obtain his preferred outcome. In the case of legislatures, which are our concern here, members require majorities to pass bills and an actor's power in these settings will turn on his capability to form legislative coalitions for the issues he cares most about. There is resistance, in the form of actors who vote against the bill, precisely because they stand to lose from its passage. After introducing a little notation, we now provide a way to systematically measure actors —in fact, senators'—capabilities to achieve their preferences in such settings

Let $i = 1, \ldots, I$ index the senators and $j = 1, \ldots, J$ index the bills on which they vote. For any particular j, there exists a senator who proposes the bill, denoted i^p_j. In any session of the Senate, there will be not more than J such individuals, since all bills are proposed by someone, but not all senators propose. The proposer of bill j, i^p_j, is assumed to seek a coalition to vote for the bill and we denote that coalition $c_{i^p_j}$ which has $|c_{i^p_j}|$ members (one of whom is the proposer). It is natural to assume that i^p_j prefers a coalition that is a majority of all senators voting on j, but this is not strictly required here.[4] For now, we suppose that the benefits of passing j—be they pork, credit, news headlines—will accrue only to i^p_j and that other legislators pay some cost in supporting the bill, perhaps because the total size of the pork or publicity 'pie' available is now decreased for them and their home districts. Else, there is some simple opportunity cost of the time they take to vote. In attempting to form the coalition, we say that i^p_j 'convinces' a senator to join $c_{i^p_j}$ if she subsequently votes in favor of the bill (and hence in favor of i^p_j's ideal point relative to the status quo). The nature of the convincing could take many forms: the proposer may attempt to reason, argue, threaten or bargain with his colleagues explicitly, or else the inducement may be implicit. In subsequent votes, say $j+1$, the situation may well be reversed with the previous proposer now joining a coalition $c_{i+1^p_{j+1}}$ backing a senator $i+1$ who had been a member of $c_{i^p_j}$. We say that two senators 'interact' in a particular session of Congress if one proposes a bill and the other joins its backing coalition. Notice that for any particular bill j proposed by i, there are $c_{i^p_j} - 1$ interactions.

[4]We assume rather that a proposer is coalition size maximizing: his utility is increasing in $|c_{i^p_j}|$. We could think of this as increasing the *probability* that j is passed.

For each senator, the statistical model we discuss below will determine a score λ_i such that if we know only that Senator A and Senator B are involved in an interaction, the probability that it is A convincing B (i.e. B is backing A's proposal) rather than B convincing A, is the difference of the scores $\lambda_A - \lambda_B$:

$$\text{logodds}(A \text{ convinces } B|A \text{ and } B \text{ interact }) = \lambda_A - \lambda_B. \tag{3.1}$$

The interpretation is then straightforward: the greater the value of λ_i relative to other senators, the more powerful we hold that senator to be.

The derivation of the estimator begins by assuming that there exists a latent and hence unobservable power for each senator denoted α_i. For any given interaction (any particular bill) between a pair of senators, let $\pi_{AB} \in (0, 1)$ be the probability that the interaction involves Senator A convincing Senator B. Since an interaction must involve either A convincing B or B convincing A it must be true that $\pi_{AB} + \pi_{BA} = 1$.[5] Write the odds that A convinces B as a function of their latent powers, such that

$$\frac{\pi_{AB}}{\pi_{BA}} = \frac{\pi_{AB}}{1 - \pi_{AB}} = \frac{\alpha_A}{\alpha_B}. \tag{3.2}$$

[5]An objection may be that party or some other factor compels senators to vote one way or the other, in which case this equality will not hold. By studying *personal amendments* in Section 3.4 we ameliorate these concerns somewhat but see also our 'order effect' specification below.

This formulation has the obvious consequence that, if $\alpha_A > \alpha_B$, *in any given interaction*, $\pi_{AB} > \pi_{BA}$. That is, for any given interaction between A and B, A is more likely to be *forming* the coalition (and gaining his preferred outcome) than B is.

The problem of *estimating* α_i can be approached via a logistic regression. To see how, first let $\alpha_i = \exp(\lambda_i)$. Then, some rearrangement yields[6]

$$\pi_{AB} = \frac{\exp(\lambda_A)}{\exp(\lambda_A) + \exp(\lambda_B)}. \tag{3.3}$$

Suppose that A and B interact a total of N_{AB} times. Of these interactions, let n_{AB}—where $n_{AB} \leq N_{AB}$— be the number of times that A convinces B. Then, so long as the N_{AB} interactions are independent of one another, and the same probability π_{AB} applies to each interaction, n_{AB} has a binomial (N_{AB}, π_{AB}) distribution. If we are also willing to assume that all the other interactions between the other senators are also independent, we have a logit model that can be estimated via maximum likelihood:

$$\text{logit}[\Pr(A \text{ convinces } B)] = \lambda_A - \lambda_B. \tag{3.4}$$

We can then compare the relative size of λ_A and λ_B to see which senator is more powerful.[7]

[6] From Equation (3.2), it is obvious that $\pi_{AB}(1 + \frac{\alpha_A}{\alpha_B}) = \frac{\alpha_A}{\alpha_B}$; substitute $\frac{\alpha_B}{\alpha_B}$ for 1 and note that $\alpha_i = \exp(\lambda_i)$ to obtain (3.3).

[7] Importantly, the model does not require that the matrix of interactions is 'complete' in the sense that every senator interacts at some point with every other. The model implicitly assumes transitivity: if Senator A is more powerful than B, and B is more powerful than C, then A is also more powerful than C.

A concern with this formulation might be that, in fact, party pressures and thus *a priori* similar preferences between senators, obviates the need to 'convince' other legislators in any real sense. Otherwise put, some senators are disposed to supporting others like them, whatever the proposers power. We will tackle this issue head-on by estimating a version of our logistic regression as

$$\text{logit}[\Pr(A \text{ convinces } B)] = \lambda_A - \lambda_B + \delta w \tag{3.5}$$

where w is an indicator taking the value 1 if the proposer's party is the same as that of an individual he convinces and -1 otherwise. Hence, δ will capture the 'natural advantage' of proposing to someone in a senator's party, and the λ_i will reflect the ability *absent* this advantage.

Notice that our method here is very different to that of previous endeavors to measure power. Unlike index methods, it is an *a posteriori* in the sense that we infer 'power' *after* observing actor's *actual* decisions—rather than before. Hence the terms 'data driven' and 'actor based.' Second, unlike subjective rankings it is objective and relies on a clearly defined definition of power. This means, in contrast to 'panel of expert' surveys, that our findings (below) are exactly replicable by any other political scientist who chooses so to do. Moreover, it is much cheaper and faster to calculate.

This estimator is not new to this paper; it is the Bradley-Terry (Bradley and Terry, 1952) pairwise comparison (with equation 3.5 representing an 'order effect' specification) method in a novel setting. This approach is well known and well studied by statisticians working in fields as diverse as sport team rankings (Agresti, 2002), journal citation patterns (Stigler, 1994) and competition for mates in the biological sciences (Stuart-Fox et al., 2006).[8] This is not the first paper in the discipline to suggest thinking of some political phenomena as pairwise interactions: for example, Groseclose and Stewart (1998) study the value of Congressional committee assignments using (dyadic) transfers of representatives.[9] The paper is perhaps closest in spirit (though not in execution) to work by Wawro (2000) who studies legislative entrepreneurship in the House. Designing a novel "entrepreneurship scales score" based on five observables of behavior, Wawro shows that party-based career prospects are strongly linked to a representative's records of actively introducing legislation. Below, we build on this work by presenting congruent findings—though with a different causal direction: in particular, that party and committee advancement aids senators in their quest to pass (potentially controversial) legislation. Also, like Wawro we provide scholars with a metric that may be used for further research.

[8]Such models can be fitted with many standard statistical packages. For this paper, `R` (R Development Core Team, 2006) was the environment of choice in conjunction with the `BradleyTerry` library (Firth, 2005).

[9]See also King (2001) who discusses some possible extensions of standard international relations models that assume dyadic interactions, and Fowler (2006) who studies the 'connectedness' of legislators in both chambers—though not in an explicitly pairwise way.

3.4 Application: 108th Senate

We use data from the 108th Senate, that met between January 2003 and December 2004. The universe of cases for the present analysis is the 255 amendments, by 74 different senators, proposed and voted upon during this time. Amendments are preferred to bills as a whole since they "tend to reflect more specific changes to a bill that are less susceptible to deviations from the sponsor's original intent" (Fowler, 2006, 9). That is, they are more amenable to the notion of *personal* coalitions formed to achieve an individual's goals. Studying the Senate in particular has the further benefit that, subject to some constraints negotiated via Unanimous Consent Agreements, its members may propose amendments at essentially any time, in any order, without permission from a Committee of the Whole. This means that, unlike the House, the status quo is afforded less protection—an ongoing concern for those studying power (see Bachrach and Baratz, 1962, for example). There are 101 actors in the current data set: 74 proposing senators, 26 non-proposing but voting senators and the President (where his views on the amendments are known). Of the amendments, 112 were winning, 139 were losing, and four were tied votes. Recall that the number of observations is the number of interactions, and is thus well in excess of the number of amendments: otherwise put, each amendment corresponds to multiple interactions and hence multiple observations.

One concern might be that senators vote *strategically*: against their first preferences for an alternative that is *a priori* less preferred. They might also *propose* strategically, to ensure their preferred alternative is selected by majority rule (Riker, 1982, 1986). However, it is not immediately apparent that strategic (let alone 'killer') amendments are common. Second, strategic voting on *amendments*, as opposed to bills, should be relatively rare: we can generally assume that those voting for an amendment *want* it to pass. Third, strategic voting is presumably a product of the organizational structure—like the timetabling procedures—of the Senate: but this is precisely the sort of consideration that ought to be *included* in the calculation of our metric. Lastly, votes in the Senate are part of the public record, so we expect senators to treat their vote seriously and think through its consequences.[10]

In Table 3.1, we report a selection of the power estimates for this specification. Recall that there is no sense in which the power measure is absolute, and all the coefficients are scaled relative to zero which is assigned to President Bush.

Examining the table, we note that the long serving (since 1979) John Warner of Virginia (senior senator) is ranked first. Second is Mitch McConnell of Kentucky (senior senator). In the 108th Senate he was elected as the majority whip by the Republicans and, at the time of writing, was the leader of Republicans in the Senate. Judd Gregg, the senior (Republican) sena-

[10]Notice that simply proposing (and forming coalitions) for more amendments cannot in and of itself increase a senator's power in this model: see Appendix 3.7 and Appendix 3.9.

tor from New Hampshire is at third. Gregg chaired the Health, Education, Labor and Pensions Committee in the 108th Congress, and became chair of the Budget Committee in the 109th. Robert Byrd, the longest serving senator in history is the first Democrat on this power list. From West Virginia, Byrd is, at the time of writing, the President *Pro Tempore*—a role he had previously held in the 1980s and 1990s. John McCain of Arizona, then chair of the Commerce committee, is followed by Bill Frist of Tennessee, the Senate Majority leader for the Congress in question. Thad Cochran and Charles Grassley, chairs of the Appropriations and Finance Committees respectively (in the 109th Congress) hold the 7th and 8th spots. Barbara Boxer, the junior senator from California is the first woman to feature on the list. At the bottom of the ranking, we see both senators from Hawaii (Daniel Akaka and Daniel Inouye) along with Jim Jeffords (Republican to 2001, Independent thereafter).[11] Fitting an 'order effect' model such that the 'natural advantage' of proposing to like-minded partisans is controlled for—as specified in equation 3.5—makes little appreciable difference to this rank order: see Appendix 3.8 for more details.

Subjective rankings—published in magazines such as *Time* and by consultancy groups like *Knowlegis*—reach very similar conclusions to ours. The *Knowlegis* power list for 2005 for example, one session after the 108th, places Thad Cochran at 1 (7 in our ranking), Mitch McConnell at 4 (2), Charles Grassley at 7 (8), John McCain at 8 (5), Bill Frist at 10 (6), Arlen Spec-

[11]Importantly, the rank ordering is very different to that which might be garnered from looking solely at the number of amendments proposed by each senator. See Appendix 3.9 for *this* rank ordering.

tor at 9 (15) and Orrin Hatch at 2 (19). This similarity to our ranking is remarkable given that (a) they are discussing a different Congress and that (b) our approach is based on a simple, model-based voting metric that is relatively straightforward to compute. The *Knowlegis* list, by contrast, is based upon a vast and expensive survey incorporating what is theoretically much more information.[12]

As noted above, a pleasing feature of the current estimator is that the power estimates λ_i have meaning outside of a simple rank ordering. Recall that $\lambda_A - \lambda_B$ is the (anti-logged) probability that, conditioned on two senators interacting, it is A that convinces B to back his amendment rather than the other way round. Consider, for example, John McCain and Russ Feingold (Wisconsin), coauthors of the Bipartisan Campaign Reform Act of 2002. The probability that, in any interaction between these two, it is McCain proposing the bill and Feingold backing it is

$$\frac{\exp(\lambda_{\text{McCain}})}{\exp(\lambda_{\text{McCain}}) + \exp(\lambda_{\text{Feingold}})} = \frac{\exp(21.58)}{(\exp 21.58) + \exp(19.85)} \approx 0.85.$$

By contrast, the same calculation for McCain and Rick Santorum yields a probability of (very close to) 1. This is hardly surprising given that the senator from Pennsylvania did not offer an amendment in the 108th Congress,

[12]According to their website, "*Knowlegis* staff carefully researched, sorted and considered thousands of data points to determine . . . power. . . reviewed thousands of media articles, hundreds of bills that passed out of committee . . . over a thousand amendments. . . We collected data on the leadership, committee, and caucus positions of each Member . . . researched relevant campaign contributions, and considered any characteristic or action that could contribute to their Sizzle-Fizzle factor. . . there are more than 10,000 data points and variables that were considered in the 2006 Knowlegis Power Rankings" source:`http://www.congress.org/congressorg/power_rankings/backgrounder.tt`

but it lends some validity to the estimator.

A further factor in favor of the new approach concerns the goodness of fit: in contrast to standard methods this is at least meaningful (what is the goodness of fit for a subjective survey?) and, in fact, respectable, at some 85% of interactions correctly predicted. This figure is within the ballpark of similar statistics for industry standards like `NOMINATE` (see Poole and Rosenthal, 1997), though given the structure of the underlying statistical model it is calculated somewhat differently. Appendix 3.10 reports more details.

In sum, Table 3.1 seems to be a reasonable 'influence list' both in terms of its underlying statistical model and its actual contents; it tells us little, however, about the *causes* of power, a subject to which we now turn.

3.5 Structured Modeling

We usually have theories about what explains the power of different individuals; indeed, sometimes we treat characteristics that are causes of power as if they were synonymous with power itself: consider, for example, the notion that 'the rich are powerful.' A pleasing feature of the current approach and estimator is that we can *separate* these notions and explicitly incorporate individual specific explanatory variables as predictors of power. We estimate

$$\lambda_i = \sum_{r=1}^{p} \beta_r x_{ir} \tag{3.6}$$

and thus we predict the power of each senator as a (linear) function of explanatory variables $x_{i1}, x_{i2}, \ldots x_{ip}$ with coefficients $\beta_1, \beta_2 \ldots, \beta_p$ (see Firth (2005) and Springall (1973) for details). Since the approach here is essentially a logistic regression, we can interpret coefficient estimates as positive or negative, in terms of their marginal effects on power, and we have commensurate standard errors. We can thus make uncertainty statements about our predictors.

We used institutional and personal information to explain power in the Senate. Collecting such data is straightforward: the Senate itself, the Government Printing Office and the United States Census Bureau provided all the relevant variables below in electronic form. We break the findings into four subsections dealing with 'Party and Ideology,' 'Committees and Agenda Control,' 'Geographic Factors' and 'Career Factors.' The intention here is not to provide an exhaustive account of power, but to demonstrate the strength of the approach and possible avenues for future research. Before describing the results, notice that the nature of 'power' now being considered is altered somewhat. In the previous section, power was an ability to be ascribed to individual senators—it was a 'personal' characteristic. Now though, power is a function of variables and is being treated in an 'institutional' sense, separate from the individuals who wield it. Otherwise put, it is now a maintained assumption that institutional (and other) characteristics *make* individuals powerful.[13]

[13]Notice that a potential endogeneity concern—that individuals *already* endowed with latent power *ipso facto* obtain important institutional advantages—is being avoided by construction.

3.5.1 Party and Ideology

In contrast to the 'textbook Congress' of Fenno (1973), there is increasing evidence that Congressional voting is ideological and party driven (McCarty, Poole and Rosenthal, 2006).[14] This is more true of the House than the Senate, but nonetheless it suggests some testable hypotheses. As noted above, senators must assemble majorities to obtain their own preferences. If a majority party exists (the Republicans for the 108th Congress) we might expect those from the majority party to be more powerful than those from the minority (Democrats). A refinement on this theme is that we expect *leaders* of Congressional parties to be especially powerful: the Senate Majority Leader (Bill Frist for the 108th Congress), for example, has the ability to schedule debate. Other officials—which we refer to as 'junior leaders'—such as Policy Committee Chairs have powers to design and execute policy ideas. In Table 3.2 we give the results of the model for a `Majority` dummy variable and `SeniorLeadership`, a dummy that denotes either the senior Senator for each party, or the whips for each party. We also interact these variables. The positive and significant effects of being a Republican and being part of the Senate's senior leadership are evident from Table 3.2. From the interaction term it is evident that being a Republican *and* a leader adds an extra fillip to one's power. To be clearer here and recalling equation (3.6), consider the power of some senator A who is a member the majority party

[14] See also Huitt (1957) who discusses cross cutting tensions of ideology and party as it applies to senators.

and also a senior leader. By Table 3.2,

$$\lambda_A = 0.286 \times 1 + 0.205 \times 1 + 0.754 \times 1 = 1.245.$$

The power of Senator B who is a rank-and-file Republican is

$$\lambda_B = 0.286 \times 1 + 0.205 \times 0 + 0.754 \times 0 = 0.286.$$

The probability that, if these two interact, it is the Republican leader proposing and B supporting is

$$\frac{\exp(1.245)}{\exp(1.245) + \exp(0.286)} = 0.72.$$

If B is a rank-and-file Democrat, then this probability rises to

$$\frac{\exp(1.245)}{\exp(1.245) + \exp(0)} = 0.78.$$

In Table 3.3 we estimate the same model with `AllLeadership`, a dummy that includes all senior leaders (as in `Leadership`) in addition several other categories: Conference Chairs, Party Committee Chairs, Conference Secretaries and Senatorial Campaign Chairs. This model fits the data as well as the previous model (note the similar values of the Akaike Information Criterion (AIC)). Interestingly though, general leadership status does not confer the power that senior leadership does—notice that the coefficients for `AllLeadership` and the interaction are now smaller than the commensurate ones in Table 3.2.

To demonstrate how American politics has become increasingly polarized, McCarty, Poole and Rosenthal show that legislators' ideological positions, and the median positions for parties, have become increasingly disparate along a liberal-conservative dimension. As politics becomes more polarized, we might have several conflicting expectations. On the one hand, senators who occupy the center ground—are close to the legislative median— may find it easier to broker deals with others to their left and right, and hence will be more powerful. On the other hand, if voting is strongly party based, then perhaps the most ideologically extreme in each party will be able to motivate their 'core' supporters into backing their preferred positions. This will be *a fortiori* true for relatively radical senators from the *majority* party. At base, this is one version of the 'party versus floor median' debate well discussed in literature elsewhere (Cox and McCubbins, 1993; Rohde, 1991, *cf.* Krehbiel, 1998).

We can measure ideological extremism via senators' `NOMINATE` scores (Poole and Rosenthal, 1997).[15] In particular, the score for the senator in one dimension is treated as their 'conservatism' (we label this '`Conservatism`' and the higher the score, the more conservative and less liberal they are). 'Extremism' is a different concept and can be ascertained by taking the absolute value of this score (`Extremism`): a very high score now implies a senator very far to the left *or* to the right, while a low score reflects a senator in the center of the chamber. We add `Distance from Median` that records the absolute

[15] In particular, `DW-NOMINATE` scores, available from `http://voteview.com/dwnomin.htm`

value of senators' distances from their party median in terms of `NOMINATE` scores. In Table 3.4 we report the effects of conservatism (controlling for majority party membership and distance from the median) and in Table 3.5 we report the effects of extremism.

The lessons from Tables 3.4and 3.5 are interesting. Once we control for majority party membership, conservatism and distance from the median have a negative impact on power. Consider, for example, Senator A who is a moderate Republican with a `NOMINATE` score of 0.2 (the scale runs -1 through 1). By contrast Senator B is deeply conservative with a `NOMINATE` score of 0.8. Their powers are

$$\lambda_A = -0.591 \times 0.2 + 0.844 \times 1 + \times - 0.702 \times |0.2 - 0.441| = 0.557$$

and

$$\lambda_B = -0.591 \times 0.8 + 0.844 \times 1 + -0.702 \times |0.8 - 0.441| = 0.119.$$

In any interaction, the probability that A proposes some coalition, while B joins it, is around 0.61. To see more of the interplay between ideology and power, consider Figure 3.1: here, we plot the Senator's power estimates against their `NOMINATE` scores, and then impose a solid loess curve. The top graphic displays the plot for all senators, and the one below is the same graphic for senators who proposed amendments in the 108th Senate.

Notice that, in the top panel of Figure 3.1, the extremes of the Senate are rewarded in power terms: the loess dips slightly as it crosses the middle of the ideological spectrum. But, interestingly, in the lower panel (which represents the most powerful senators), we see two things: first, majority party status boosts one's power—notice that the loess rises as it moves right. *Within* the majority party (the Republicans) though, the most powerful senators are not drawn from the far right wing: notice the high λ_i recorded for those with a `NOMINATE` score around 0.4. In Table 3.5, the positive coefficient on `Extremism` suggests once again that senators towards the middle of the chamber lack power. Interestingly, it is also evident while being in the majority party is beneficial, it does not pay to be a right wing Republican: rather, the powerful are from the median of the party. Figure 3.2 confirms this idea: in the first panel, the power ratings for all senators are shown, and in the bottom panel, the analysis is restricted to the proposers only. Notice that the bulk of the mass occurs around 0.4 for both parties, with the most powerful senators of both parties occurring just above and below this scaling. Otherwise put, it does not pay to be the median of the chamber, but it does pay to be the median of your party.

3.5.2 Committee and Agenda Control

Positive political theorists, especially those of the 'Rochester school' (Amadae and Bueno de Mesquita, 1999), have suggested that it is the organization of Congress in terms of its *committees* that confers power on actors. Indeed, in a series of articles Shepsle and Weingast (1987) and Krehbiel (1991) discussed precisely *why* committees are powerful. Of course, not all commit-

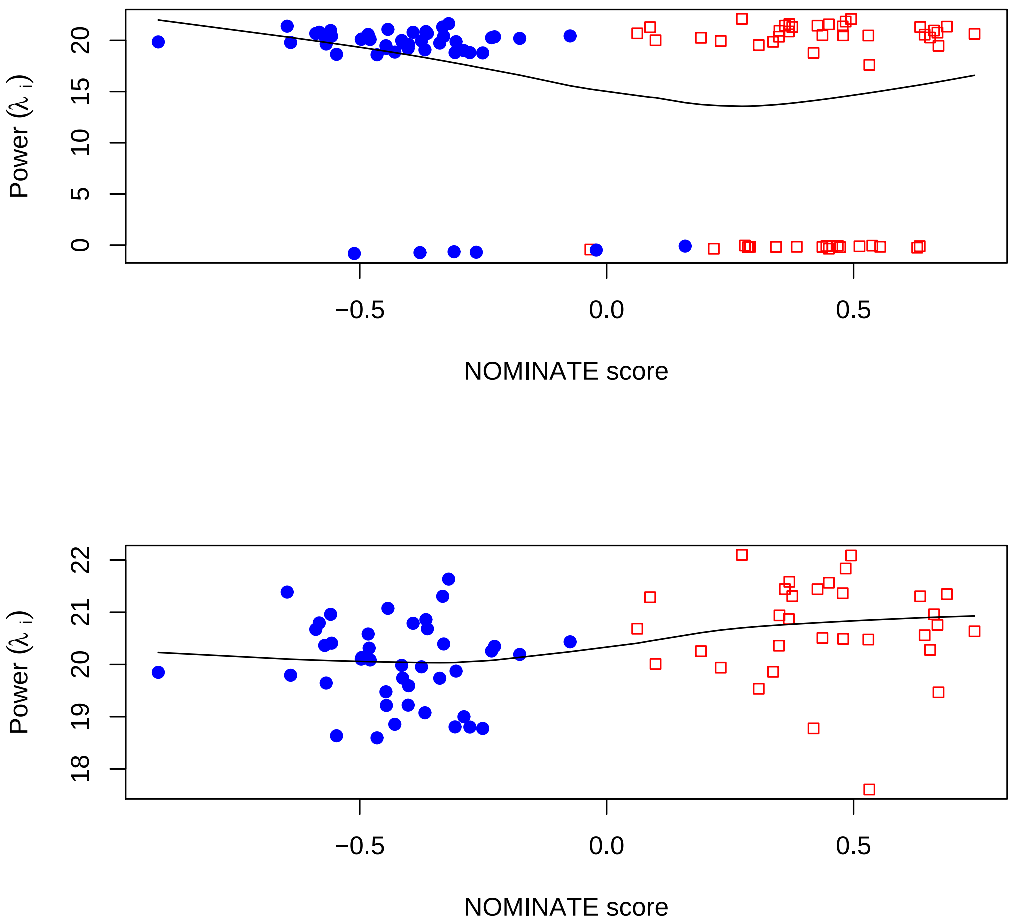

Figure 3.1: NOMINATE score versus power (λ_i); open squares are Republicans, closed circles are Democrats; solid line is loess. Top figure is for *all* senators; bottom figure is for all *proposing* senators.

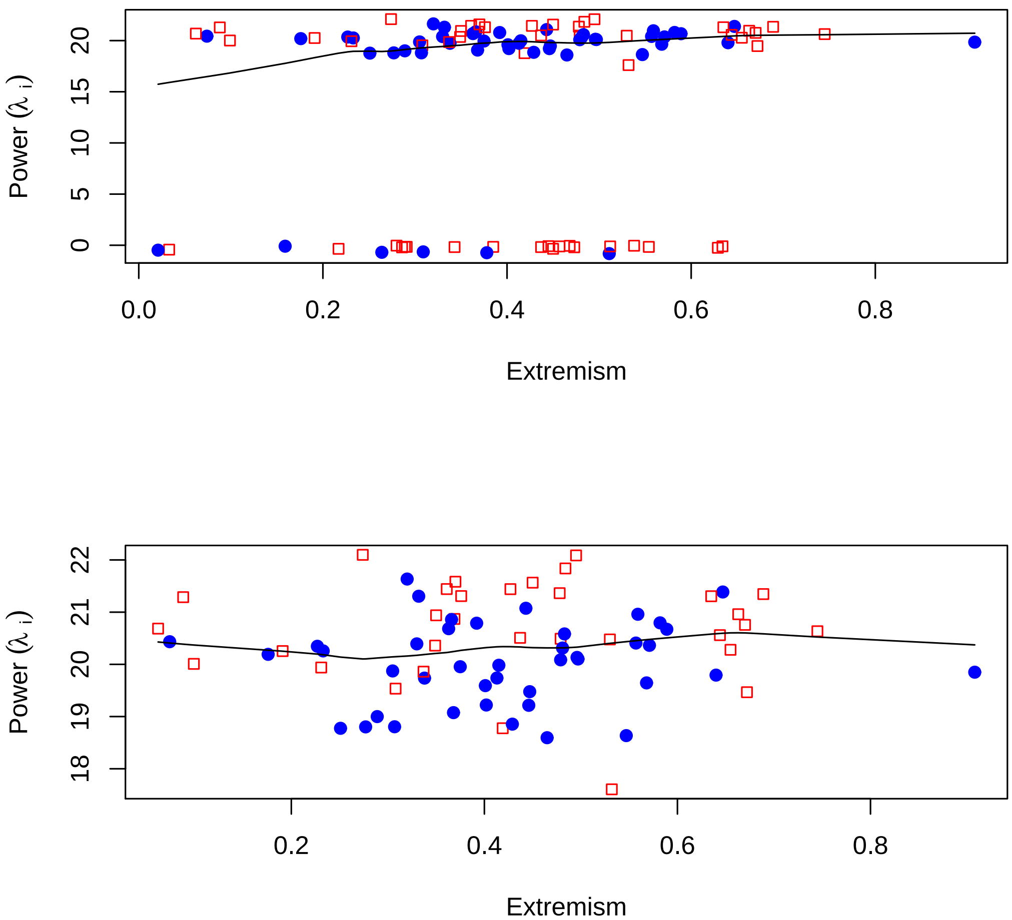

Figure 3.2: Extremism versus power (λ_i); open squares are Republicans, closed circles are Democrats; solid line is loess. Top figure is for *all* senators; bottom figure is for all *proposing* senators.

tees are created equal, and they attract different memberships with different motivations (Fenno, 1973). Nonetheless, there is general agreement that financial committee positions—those that have 'power of the purse' tend to be the most coveted spots for ambitious senators. This is not least because such positions enable members to channel substantial pork to their home states. For example, Charles Grassley, Chairman of the Senate's Appropriations Committee, used the 2004 appropriations bill to steer some $50m to Iowa's visionary complex of biomes, *Earthpark* (*The Economist*, 2006).

Committees themselves are, of course, hierarchical structures. As mentioned above, Grassley was the *Chair* of Appropriations, not simply a member. Chairs have several *de jure* responsibilities and rights pertaining to timetabling, hearings and the selection of bills to be considered. Other than the Chair, who must be a member of the majority party, committees are constructed of minority and majority party members. Above, we discussed reasons why majority party senators might be powerful, and presumably this goes *mutandis mutatis* for majority party committee members.

Since space is limited, we only discuss some of the committees and their members here. In particular, in Table 3.6 we report regression coefficients for the Appropriations, Armed Services, Commerce, Finance and Rules and Administration committee all denoted with these names. We also include a majority party interaction term, and a term for Chairs (of any committee) denoted `Chair`. Where Table (3.6) reports positive coefficients, the Committee assignment increases the power of a senator relative to one *not* on the that committee. Majority party members get an extra boost in power on the Armed Services, Appropriations and Finance committees, but not on the Rules and Administration or Commerce committees. Caution is required in interpreting the findings for some committees, such as Rules and Administration, which are notoriously weak and generally involved in less interesting work. Indeed, senators may accept a role on such committees as the 'price' for serving on more powerful committees elsewhere. Hence, the magnitude of the coefficient maybe somewhat misleading: senators are powerful because of their 'good' committee assignments which are *correlated* with weak assignments, not *caused* by them.[16]

From the perspective of a senator, the most powerful position is a role on the Finance committee (notice that the addition of the `Finance` coefficient and that for `Majority`×`Finance` is a larger number than for any other committee). Perhaps unsurprisingly, Chairs are more powerful than rank-and-file members; in Figure 3.3 we compare chairs to Democrats and non-chairing Republicans who propose amendments. Notice that the median power of chairs clearly exceeds that of the other groups and, in fact, their entire inter-quartile range is more powerful than that of Democrats.

3.5.3 Geographic Factors

One of the results of the Constitutional Convention of 1787 was the so called 'Connecticut Compromise.' Dealing with the creation of the United States' legislative bodies, the Compromise proposed two houses: a lower

[16] We are grateful to Antoine Yoshinaka for this suggestion.

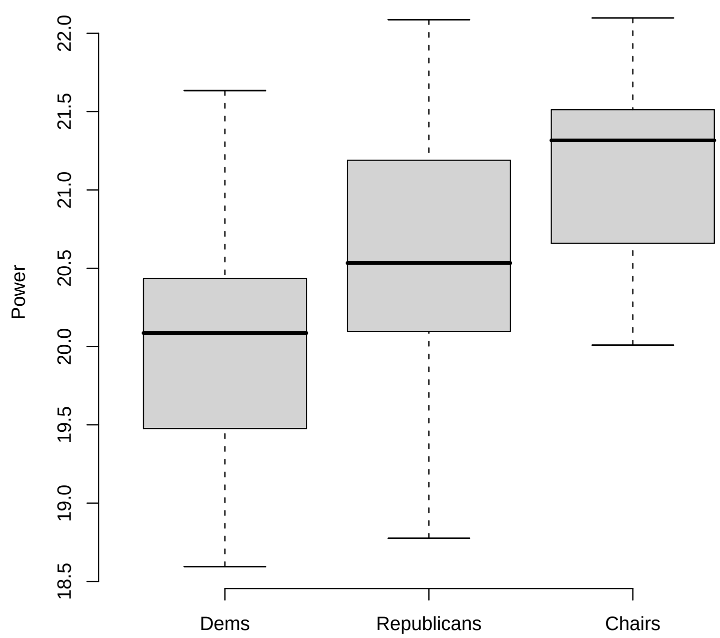

Figure 3.3: Boxplot showing relative power of committee chairs, non-chairing Republicans and Democrats who propose amendments.

house elected in proportion to population, and a senate, in which each state would have two representatives, regardless of its population. The fear of large state tyranny, though admonished as unlikely and illogical by the likes of Madison and Hamilton, motivated small states—like Delaware, Maryland, New Jersey and Connecticut—to seek institutional protection.

Of course, as argued with respect to the power indices approach above, the fact that voting resources are *de jure* equal across states should not imply that all senators are equally powerful. One way to examine this notion more formally is to use the present model with geographic factors as explanatory variables. We do not have particularly strong priors about the effect of geographic factors on a senator's power, but some vague ideas might be as follows.

Historically, representatives from the South were very powerful actors. Up until the 1960s, Southern states were solidly Democratic, and possibly dissenting voices—from blacks and poor whites—were excluded from voting (Key, 1949). As a result, Southern senators faced few challenges in their home states and could use the committee seniority system—that rewarded long service irrespective of party affiliation—to obtain powerful chairmanships. In the 'post-reform' period, this systematic concentration of power in Southern hands was much reduced Rohde (1991). Moreover, the South is no longer under hegemonic Democratic control as demonstrated by George Bush with victory in every Southern state in his 2004 Presidential reelection. Nonetheless, we might still expect, for historical or other reasons, that senators from the South—Texas, Louisiana, Arkansas, Mississippi, Alabama, Florida, Georgia, North Carolina, South Carolina, Virginia and Tennessee—wield disproportionate power. We use a `South` dummy taking the value of 1 for senators from these states.

The expected effect of a state's wealth on a senator's power is arguably ambiguous. On the one hand, senators from rich states may be able to procure greater 'home-grown' funding for their campaigns and causes, rendering them more influential. On the other hand, senators from poorer places may have more sway in Washington because they can point to underfunded public services and crumbling infrastructures in their home states as evidence that they have a more urgent claim to the nation's resources. Combined with suitably deployed rhetorical skill, we could imagine poverty may boost a state's representative's powers. We measure wealth use the Census Bureau's Median Household Income statistics for each state in dollars (`Median Income`).

We have similarly vague priors *viz* the effect of population density on power. On the one hand, small, primarily urban, densely populated states have senators who literally represent 'more' citizens (we measure this with Census Bureau's population estimate for 2003, `Population`), which may aid rhetorical appeals. On the other, sparsely inhabited, primarily agricultural states may increase the power of senators who represent them, in part because, for historical reasons, farm-based financial aid—an important component of rural states' federal funding—is easier to deliver than other types of support. We use the land area of the states and denote this variable `Land Area`.

The significant, negative coefficients in Table 3.7 suggest that poorer, smaller, more densely populated, non-Southern states will yield senators with more power. Of course, this might not correspond to any particular state. To help interpretation, in Figure 3.4, we color a map of the contiguous United States according to the predicted λ_i that a senator from that state based on the coefficients of Table 3.7. Based only on geographical factors, the most powerful state—in terms of its Senators—is California and it is shaded lightest. New York and West Virginia are similarly light colored. 'Weaker' states are dark colored. There are no particular regional patterns discernable from Figure 3.4, except perhaps a band of states from New York west through Missouri which appear disproportionately light (and thus powerful) relative to their neighbors.

3.5.4 Career Factors

For most politicians, a position in the Senate is a career ambition (Brace, 1984; Rohde, 1979). But, once attained, senators have strong incentives to seek reelection (Mayhew, 1974). As implied above, this is in part because long service is linked to promotion in terms of committee and other assignments. Here we calculated the years of service since first entering the Senate through 2004 and denoted this variable `Service`. Though not formally associated with greater rights or responsibilities, longevity of service *for a particular state* makes a senator the 'senior' representative of his constituency. We use a `Senior` dummy to check for any extra power effect that

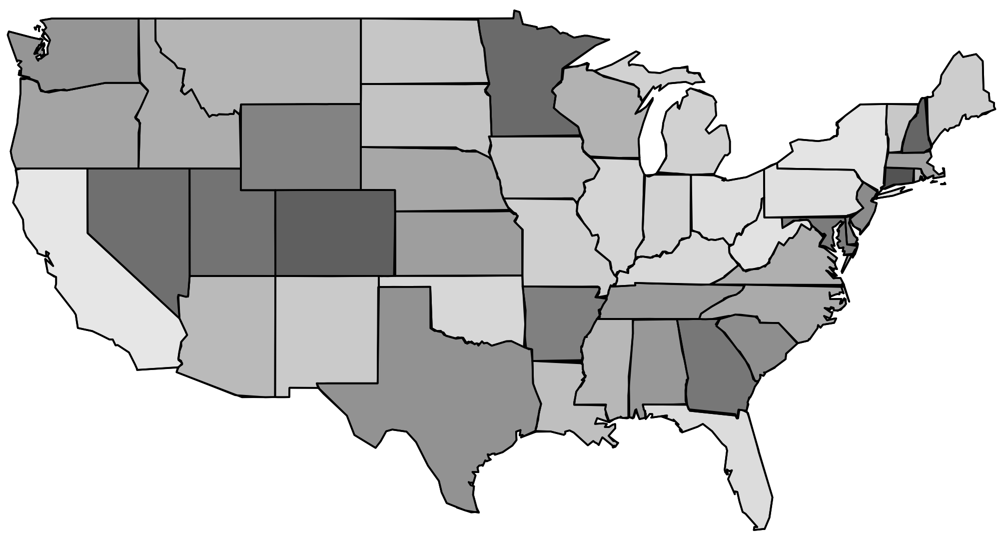

Figure 3.4: Map of contiguous United States with shading proportional to ‘power’ given by coefficients in Table 3.7: lighter states are associated with more powerful senators.

such status confers.

From a career perspective, sociologists have long argued that being male strengthens advantages one in the work place, and this is no less true in politics. Consider, for example, the comment and excitement drawn by Nancy Pelosi's ascension to the Speakership of the House in 2006.[17] For this reason, we add a dummy variable for `Male` here.

The coefficients for `Service` and `Senior` status are much as we might anticipate in Table 3.8: longer time served in the Senate, as well as seniority makes for a more powerful senator. The coefficient for sex though is perhaps not as expected. Being male is actually associated with a *lower* power than being a female. In Figure 3.5 we report a boxplot for males and females, in terms of their power. Interestingly, though the median power of the sexes is approximately equivalent, the distributions are very different: while male senators are counted among the most powerful, they are also some of the weaker members of the Senate. Females, by contrast, are heavily concentrated in the upper power ranges.

3.5.5 Summary of Findings

In summary, a senator is more powerful if the senator is:

- a member of the majority party, and has a leadership position within

[17] Pelosi herself seemed well aware of her exceptionism and, in her acceptance speech noted that "[i]t is an historic moment for the Congress, and an historic moment for the women of this country... For our daughters and granddaughters, today we have broken the marble ceiling. For our daughters and our granddaughters, the sky is the limit, anything is possible for them."

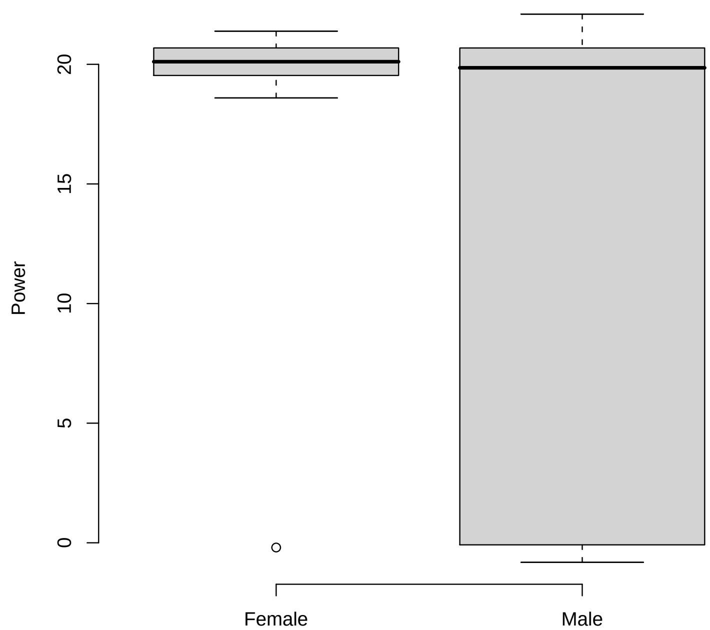

Figure 3.5: Boxplot showing relative power of male and female senators.

the party;

- from the center of the ideological distribution of her party;
- a majority party member of the Finance and Appropriations committees;
- chair of a committee;
- is from a relatively poor, densely populated non-Southern state;
- is female, long serving and is the Senior Senator from her state.

3.6 Discussion

This chapter proposed a new way to measure power in structured settings. Rather than relying on *a priori* metrics that arguably measure something akin to 'luck' or 'pivotality' rather than power, we suggested an approach that assumes the powerful have a greater capability to form coalitions in order to pass bills that they care about. This method is straightforward to implement, and allows the separation of 'power' from its 'causes' in a standard generalized linear model framework. Thus, we can talk of predicted probabilities and make statements of uncertainty regarding the 'effects' of certain variables. We applied the statistical model to the 108th Senate, using (personal) amendments as the bills around which Senators attempt to form coalitions and we found that *inter alia* majority party membership, moderation within a party, chairing of committees, long service records and seniority all make for more powerful senators. These findings in and of themselves are perhaps not shocking, but they do demonstrate the strength and flexibility of the approach, along with establishing some validity of the method.

In terms of future avenues of research, there are obvious ‘cross section’ and ‘time series’ extensions to this work. Here, we choose to study the Senate: its 100 members are a manageable number of observations for which to ascertain biographical and other data. There is no reason why, with more time, political scientists could not execute a similar model for the House of Representatives. Temporally, we studied one Congress and thus the results here are something of a ‘snapshot.’ Studying actors’ power over time may allow a more complete picture: certainly we could imagine that changing rules in Congress, along with the prevailing political climate, might alter the power of say, chairs and party leaders.

Taking the model outside of the United States Congress is a possibility too, though any such extension requires knowledge of actors’ preferences in interactions. The United Kingdom House of Commons, for example, may provide another test ground for the model via private members bills. Unfortunately, due to strong party whipping and strategic voting, deciding actors’ preference in this circumstance is not always straightforward. There is a similar caveat for the US Supreme Court, though in that case it is simply unclear who ‘leads’ a coalition. Outside of American and Comparative politics, International Relations with its concentration on specifically dyadic interactions may be amenable to such an approach. We leave this for future work.

	Senator	Power (λ_i)
1	Warner, John	22.10
2	McConnell, Mitch	22.09
3	Gregg, Judd	21.84
4	Byrd, Robert	21.63
5	McCain, John	21.58
6	Frist, Bill	21.57
7	Cochran, Thad	21.44
8	Grassley, Charles	21.44
9	Boxer, Barbara	21.39
10	Graham, Lindsey	21.36
⋮	⋮	⋮
22	Kennedy, Edward	20.79
23	Daschle, Tom	20.79
⋮	⋮	⋮
29	Kerry, John	20.58
⋮	⋮	⋮
33	Lott, Trent	20.48
⋮	⋮	⋮
45	Clinton, Hillary	20.13
⋮	⋮	⋮
54	Feingold, Russ	19.85
57	Lieberman, Joe	19.74
⋮	⋮	⋮
93	Santorum, Rick	-0.34
⋮	⋮	⋮
99	Inouye, Daniel	-0.73
100	Jeffords, Jim	-0.74
101	Akaka, Daniel	-0.82

Table 3.1: Baseline results for model of power for US Senators in the 108th Senate.

	Estimate	Std. Error
`Majority`	0.286^{***}	0.034
`SeniorLeadership`	0.205^{***}	0.072
`Majority`$\times$`SeniorLeadership`	0.754^{***}	0.104

Table 3.2: Effect of majority party and senior leadership status on power in the Senate. Asterisked coefficients (***) imply $p < 0.01$. AIC: 11052.

	Estimate	Std. Error
`Majority`	0.304^{***}	0.035
`AllLeadership`	0.118^{**}	0.047
`Majority`$\times$`AllLeadership`	0.373^{***}	0.073

Table 3.3: Effect of majority party membership and any leadership status on power in the Senate. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 11152.

	Estimate	Std. Error
`Conservatism`	-0.591^{***}	0.080
`Majority`	0.844^{***}	0.073
`Distance from Median`	-0.722^{***}	0.122

Table 3.4: Effect of conservatism, distance from median and majority status on power in the Senate. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 11169.

	Estimate	Std. Error
`Extremism`	1.14^{***}	0.107
`Majority`	0.783^{***}	0.075
`Majority`$\times$`Extremism`	-0.9636^{***}	0.168
`Distance from Median`	-0.702^{***}	0.123

Table 3.5: Effect of extremism, distance from median and majority status on power in the Senate. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 11136.

	Estimate	Std. Error
`Chair`	0.596^{***}	0.042
`Majority`	-1.055^{***}	0.089
`Appropriations`	0.245^{***}	0.040
`Majority`$\times$`Appropriations`	1.376^{***}	0.100
`Armed Services`	0.189^{***}	0.042
`Majority`$\times$`Armed Services`	1.353^{***}	0.102
`Commerce`	0.264^{***}	0.040
`Majority`$\times$`Commerce`	-0.737^{***}	0.068
`Finance`	0.482^{***}	0.046
`Majority`$\times$`Finance`	1.445^{***}	0.104
`Rules Admin`	0.440^{***}	0.039
`Majority`$\times$`Rules Admin`	0.312^{***}	0.068

Table 3.6: Effect of committee membership, majority party membership and chair status on power in the Senate. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 9434.9.

	Estimate	Std. Error
`Median Income`	$-2.199 \times 10^{-5***}$	2.003×10^{-6}
`South`	-0.256^{***}	3.390×10^{-2}
`Land Area`	$-1.668 \times 10^{-6***}$	2.135×10^{-7}
`Population`	$2.488 \times 10^{-8***}$	1.723×10^{-9}

Table 3.7: Effect of incomes, southern state representation, land area (state size) and state population on senator's power. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 10954.

	Estimate	Std. Error
`Service`	0.029^{***}	0.001
`Senior`	0.543^{***}	0.029
`Male`	-0.521^{***}	0.035

Table 3.8: Effect of incomes, southern state representation, land area (state size) and state population on senator's power. Asterisked coefficients imply $p < 0.01$ (***), $p < 0.05$ (**). AIC: 10205.

Appendix to Chapter 3

3.7 The "Busy" Senator Problem

The concern is that a senator could increase his influence score by simply proposing *more* amendments. That is, by being 'busy' in a legislative sense, he would appear more powerful. Here we show this to be untrue.

First, consider three proposing senators A, B and C. Suppose that A and B decide to somehow combine their efforts such that only *one* of them will propose and form coalitions for all amendments jointly that they formerly worked on separately. Whether A will have B do all the proposing and coalition forming, or whether B will delegate his work to A, write the senator who does the proposing and forming as S_{AB}.

Clearly, S_{AB} is busier—in that he now proposes more amendments—than either A or B. Importantly, though, the probability that any randomly drawn interaction involving S_{AB} and C has C convincing S_{AB}, will be simply the weighted average of the former probability that C convinces A and C convinces B: there is no fillip from proposing more amendments. Hence, 'business' alone cannot yield a higher power rating. To show this, let "$C\text{con}S_{AB}$" be the event that C convinces S_{AB}. The probabilities for the constituent senators are:

$$\overset{A}{\Pr} = \frac{\Pr(A\text{con}C)}{\Pr(C\text{con}A)}$$

and

$$\overset{B}{\Pr} = \frac{\Pr(B\text{con}C)}{\Pr(C\text{con}B)}.$$

The probability for the joint, 'busy' senator is:

$$\begin{aligned}
\overset{S_{AB}}{\Pr} &= \frac{\Pr(S_{AB}\text{con}C)}{\Pr(C\text{con}S_{AB})} \\
&= \frac{\Pr(A\text{con}C) + \Pr(B\text{con}C)}{\Pr(C\text{con}S_{AB})} \\
&= \frac{\Pr(C\text{con}A)\frac{\Pr(A\text{con}C)}{\Pr(C\text{con}A)} + \Pr(C\text{con}B)\frac{\Pr(B\text{con}C)}{\Pr(C\text{con}B)}}{\Pr(C\text{con}S_{AB})} \\
&= \gamma \overset{A}{\Pr} + (1-\gamma) \overset{B}{\Pr}.
\end{aligned}$$

where $\gamma = \frac{\Pr(C\text{con}A)}{\Pr(C\text{con}S_{AB})}$.

Thus, business cannot itself increase a senator's influence.

3.8 Logit Model with Order Effect

We estimate the 'order effect' model as described in equation 3.5 and present the results in Table 3.9. Inspection suggests that the rank order is closely similar to that of the model without an order effect. In fact, performing a

Spearman rank correlation test between the sorted lists yields a $\rho = 0.144$ with a p-value (on the null hypothesis that $\rho \leq 0$) of 0.08. We estimate δ—the advantage of proposing to someone in one's own party—at 0.52 (and is significant at the 1% level). That is, there is a 'home advantage', but it does not disturb our earlier findings greatly when accounted for.

3.9 Amendment Proposers

Table 3.10 gives the rank ordering of senators in accordance with the number of amendments they proposed. Clearly, simply *proposing* more amendments does not make you more powerful: for example, Grassley proposes only 3 times yet is ranked in the top 10. For completeness, Figure 3.6 displays the frequency information from Table 3.10.

3.10 Goodness-of-Fit

Recall that the method calculates a (maximum likelihood) value of λ_A and λ_B for all senator pairs who interact. Above, we defined an 'interaction' as an amendment in which either A proposed and B supported or B proposed and A supported. Then, as in 3.3, the λ_i have the interpretation that

$$\Pr(A \text{ convinces } B | A \text{ and } B \text{ interact}) = \frac{\exp(\lambda_A)}{\exp(\lambda_A) + \exp(\lambda_B)}. \tag{3.7}$$

The conditioning of the probability in (3.3) implies that we need not consider the counterfactual of non-proposing senators as proposers. Hence, the matrix of votes to be predicted denoted $\mathbf{Y_{act}}$, is of dimension $I^p \times I$ where

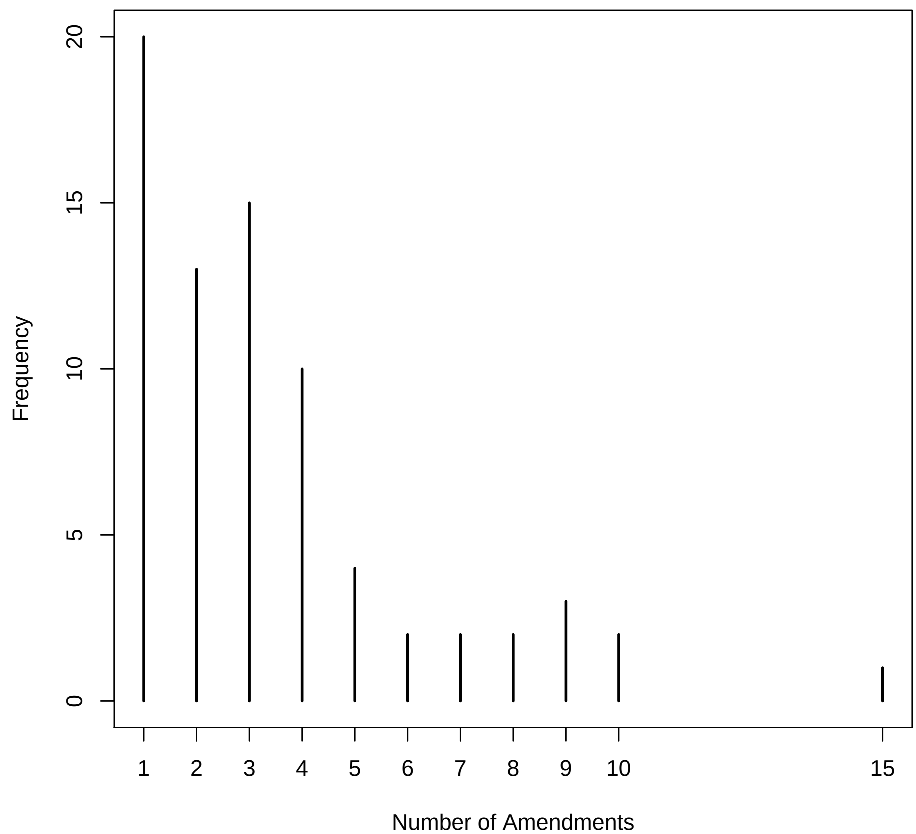

Figure 3.6: Amendment numbers: distribution.

$|I^p|$ is the number of proposing senators (74 for this case) and $|I|$ is the number of senators in total (here, 101, since we included President Bush to aid interpretation). $\mathbf{Y_{act}}$ then, is the actual interactions, and is of the following form:

	Akaka	Alexander	Allard	...
Allard	1	1	0	...
Baucus	3	0	0	...
Bayh	1	0	0	...
Biden	4	2	1	...
Bingamen	9	4	2	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

where the proposers lie to the left and all the voting senators form the columns. So, for example, Joe Biden proposed and was backed four times by Daniel Akaka, twice by Lamar Alexander and once by Wayne Allard. Notice also that senators voting for their own amendments are not, in and of themselves, counted as backers.

For any particular cell of $\mathbf{Y_{act}}$, there exists an associated total number of *possible* interactions (in either direction) for the senators. For example, Joe Biden and Wayne Allard interacted a total of 3 times: once when Allard was the proposer and Biden backed him, twice when Biden was the proposer and Allard backed him. This matrix has symmetric form when proposers as compared to voting senators who were also proposers, but not otherwise.

We denote this matrix $\mathbf{Y_{poss}}$ and a similar segment to that above appears as:

	Akaka	Alexander	Allard	...
Allard	1	1	0	...
Baucus	3	0	1	...
Bayh	1	0	0	...
Biden	4	2	2	...
Bingamen	9	4	3	...
⋮	⋮	⋮	⋮	⋱

From the estimates of Table 3.1 via Equation (3.3), we can obtain predicted probabilities of 'convincing' for any particular pair of senators A and B. For example, consider the following subset of the matrix:

	Akaka	Alexander	Allard	...
Allard	1.00	1.00	0.50	...
Baucus	1.00	1.00	0.48	...
Bayh	1.00	1.00	0.21	...
Biden	1.00	1.00	0.52	...
Bingamen	1.00	1.00	0.74	...
⋮	⋮	⋮	⋮	⋱

where the incidence of `1` implies that the proposer on the left is the one convincing the voter in the column to back the amendment (rather than the other way round) essentially with certainty. We denote this matrix $\mathbf{O_{pred}}$. Since we have direct estimates of the underlying latent abilities, we do not dichotomize this matrix to zeros and ones as may be seen in standard logistic (and probit) regression predicted probability contexts.

The expected votes matrix is the element-by-element multiplication $\mathbf{Y_{poss}} \times \mathbf{O_{pred}}$ yielding another matrix of dimensions $I^p \times I$. Note that every element of $\mathbf{Y_{poss}} \times \mathbf{O_{pred}}$ is non-zero valued. Subtracting this matrix from $\mathbf{Y_{act}}$ yields a matrix of both positive (the model over predicts) and negative entries (the model under predicts). The absolute sum of these columns yields the total number of misclassifications which is (when rounded) 2062 of some 13,869 total interactions. Hence, the percentage of interactions correctly predicted is $1 - \frac{2062}{13{,}869} = 0.85$.

	Senator	Power (λ_i)
1	McConnell, Mitch	19.63
2	Warner, John	19.59
3	Gregg, Judd	19.36
4	Frist, Bill	19.10
5	McCain, John	19.09
6	Cochran, Thad	18.98
7	Byrd, Robert	18.97
8	Graham, Lindsey	18.97
9	Grassley, Charles	18.96
10	Nickels, Don	18.87
11	Boxer, Barbara	18.84
⋮	⋮	⋮
21	Kennedy, Edward	18.30
⋮	⋮	⋮
25	Daschle, Tom	18.19
⋮	⋮	⋮
30	Kerry, John	18.10
⋮	⋮	⋮
33	Lott, Trent	17.97
⋮	⋮	⋮
46	Clinton, Hillary	17.49
⋮	⋮	⋮
52	Feingold, Russ	17.30
53	Lieberman, Joe	17.25
⋮	⋮	⋮
94	Santorum, Rick	-0.89
⋮	⋮	⋮
98	Akaka, Daniel	-1.20
99	Jeffords, Jim	-1.34
100	Inouye, Daniel	-1.51
101	Johnson, Tim	-1.53

Table 3.9: Baseline results for model of power for US Senators in the 108th Senate, logit with ‘order effect’ (party advantage) estimated.

Byrd	Bingaman	Boxer	Dodd	Feinstein	Lautenberg	Kennedy
15	10	10	9	9	9	8
McCain	Daschle	Warner	Dorgan	Harkin	Durbin	Gregg
8	7	7	6	6	5	5
Leahy	McConnell	Biden	Breaux	Cantwell	Clinton	Frist
5	5	4	4	4	4	4
Hutchison	Levin	Mikulski	Reed	Specter	Baucus	Cochran
4	4	4	4	4	3	3
Conrad	Corzine	Edwards	Feingold	Graham (FL)	Graham (SC)	Grassley
3	3	3	3	3	3	3
Kerry	Kyl	Lincoln	Nickles	Schumer	Thomas	Bond
3	3	3	3	3	3	2
Bunning	Dayton	Hatch	Hollings	Inhofe	Landrieu	Lieberman
2	2	2	2	2	2	2
Lott	Murray	Reid	Rockefeller	Snowe	Allard	Bayh
2	2	2	2	2	1	1
Brownback	Campbell	Carper	Collins	DeWine	Ensign	Enzi
1	1	1	1	1	1	1
Hagel	Kohl	Murkowski	Nelson (FL)	Roberts	Sarbanes	Sessions
1	1	1	1	1	1	1
Smith	Stabenow	Voinovich	Wyden			
1	1	1	1			

Table 3.10: Number of amendments per senator.

Chapter 4

Rebels with a Cause? Legislative Activity and the Personal Vote in Britain, 1997–2005

> *The first duty of a member of Parliament is to do what he thinks in his faithful and disinterested judgement is right and necessary for the honour and safety of Great Britain. His second duty is to his constituents, of whom he is the representative but not the delegate. Burke's famous declaration on this subject is well known. It is only in the third place that his duty to party organization or programme takes rank.*
>
> —Sir Winston Churchill, 1954

4.1 Introduction

At least since the time of Burke and his *Speech to the Electors of Bristol* (1774/1975), scholars of politics have debated the proper 'model' for elected representation. While much of the discussion has been normative in nature, country-comparativists are able to distinguish quite different practices across polities. At one extreme of these stylized facts, House members in the United States are perhaps closest to Burke's 'delegates': primarily concerned with delivering (federal spending) 'pork' to their constituents and, indeed, consciously acting in line with the general ideological preferences of their district (e.g. Fiorina, 1974; Mayhew, 1974). The nation, or even the party with which they identify, is thought to exert a relatively weak pull on their decision calculus (Jacobson, 1989). By contrast, akin to a 'trustee' conception, members of parliament in the United Kingdom House of Commons are elected on a party ticket, are thought to have a negligible 'personal' vote, and make policy according to some notion of 'national interest', broadly construed. In particular, members of the governing party generally support the executive's bills, regardless of any constituency or personal concern.

In that they hold representatives to account, elections are identical mechanisms in all systems (Riker, 1982). But contingent on the model of representation in place, the policy-making behavior for which citizens will reward or punish their legislators should vary markedly. Analysts of American politics have devoted considerable attention to this issue and there is a voluminous theoretical and empirical literature that investigates the legislative determinants of individual office-holders' electoral performances (e.g. Canes-Wrone, Brady and Cogan, 2002). Congruent with the 'delegate' assumption, the findings generally suggest that more 'maverick' incumbents—those who vote with their party least often—tend to do better come polling time.

Despite its Burkean vintage though, equivalent research for the United Kingdom House of Commons—or any 'Westminster' system—has been much less developed. That the study of British politics has generally neglected the fundamental link between constituents, their expectations of behavior and their elected agents is unfortunate, but unsurprising. It is unfortunate because ignoring Britain (and Westminster systems more generally) has resulted in an asymmetrical development of political science understanding of representation and accountability. While we know much about the United States where parties are weak, we can say little concerning polities where parties and party voting are strong forces.

The lacuna is unsurprising: with some important exceptions (Kam, 2007; Pattie, Fieldhouse and Johnson, 1994), scholars have assumed away the potential importance of legislative activity, and looked elsewhere for personal vote effects (see, for example, Cain, Ferejohn and Fiorina, 1987; Norton and Wood, 1993). In any case, with over a thousand votes to consider in a given parliamentary session, and no convenient ideal point summary available (see Spirling and McLean, 2007), scholars have not been able to conduct systematic studies that analyze MP legislative behavior *as a whole*—that is, every decision on every vote. It has not helped that public opinion studies point to seemingly inconsistent preferences and contradictory expectations, with citizens claiming that both party unity *and* 'independent-mindedness' matter (Johnson and Rosenblatt, 2007).

We seek to rectify this theoretical and empirical deficit in the current chapter: using a new data set that incorporates some 2200 parliamentary votes cast by around 200 backbench Labour MPs between 1997 and 2005, with commensurate constituency controls, we explain ballot box performance as a function of personal voting decisions in the House of Commons. We show that, all else equal, MPs that disrupt important government business are punished by their constituency voters while those that show freedom of thought and action on less important matters tend to be rewarded for their independence. This effect is worth up to 2.5 percentage points at election time. In pursuing our study we suggest a novel theoretical model of partisan voter behavior in Britain and new non-parametric, statistical methods as yet unseen in political science. Our findings extend the growing literature that emphasizes the importance of seeing MPs as individual and accountable representatives, at least partly responsible for their own electoral fates.

We proceed as follows: in Section 4.2 we give the substantive background for our study, and report on previous efforts in the literature. In Section 4.3, we give a simple formal expression describing the actions of (partisan, Labour) voters in response to different actions on various roll call types at Westminster. In Section 4.4 we describe and use a "random forest" classification technique that allows us to include essentially *every* vote as a possible independent variable, along with suitable controls, for predicting subsequent electoral performance. In Section 4.5 we report results consistent with our model, and show that several bills have a small but significant effect on performance with signs as predicted. Section 4.6 concludes.

4.2 Members, Constituents and the Personal Vote in Britain

Party voting by the electorate, either based on national Cabinet—i.e. executive—policy positions and performance (see Budge, 1999; Crewe, 1983; Norris, 2001) or on cruder partisan loyalties (see Butler and Stokes, 1974) is thought to be the norm for Westminster systems like Britain. Congruent with this principal, almost all legislative business is directly controlled by the executive, and party pressure or 'whipping' is commonly used by the government to herd its 'backbench' MPs into supporting the Cabinet's programmatic agenda when required (e.g. Cowley, 2002, 3–7).[1] By conventional wisdom then, since voters don't (and have no reason to) care about parliamentary behavior, there is little an individual backbencher has to offer her constituents aside of her party label at election time.

Interestingly though, when the public was recently asked to name the qualities that an MP ought to possess, while two-fifths claim that a representative should "be loyal" to her party, some 60 percent also claim that she should "be independent-minded" (Johnson and Rosenblatt, 2007, 165–167). In Figure 4.1 we show the changing nature of public opinion on this subject: clearly freedom of thought seems increasingly important, while loyalty is falling away as a requirement. This is *a fortiori* true for the period in which Labour has held large majorities of seats in government i.e. post-1997.

[1] Whipping typically includes offers or threats related to ministerial promotion: see Kam (2006) for a discussion of the Canadian case.

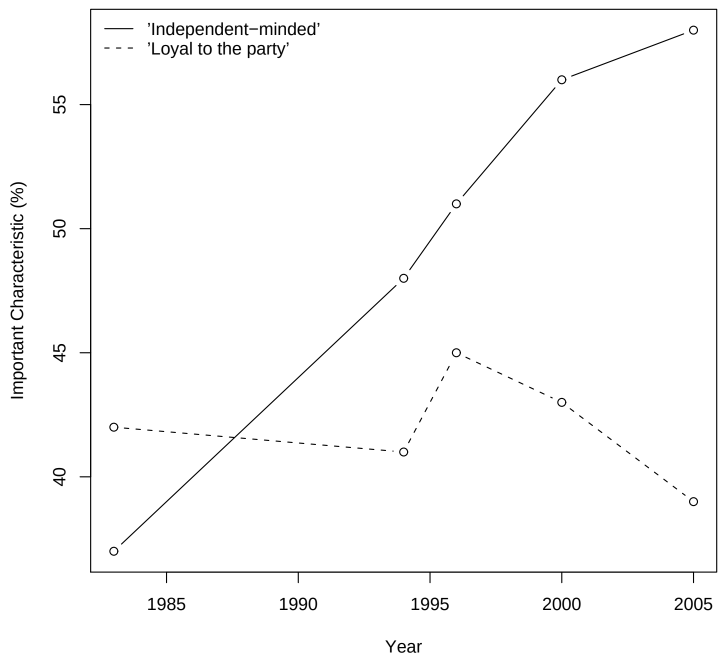

Figure 4.1: Percentage of respondents naming different qualities as "important for an MP to have": solid line is percentage naming 'independent minded', broken line is percentage naming 'loyal to party'. Multiple choice, more than one choice permissible; survey dates are 1983, 1994, 1996, 2000, 2005. Source: Johnson and Rosenblatt (2007, 166, Table 1).

On the one hand, such a diversity of views might imply that voters fail to understand the nature of parliamentary government and the importance of party dominance and discipline therein (e.g. Paltzelt, 2000). On the other hand, it is not impossible to rationalize both preferences, especially for voters who lean towards the governing party: united governments are more efficient, and able to respond rapidly to crises (see, e.g. Alesina and Drazen, 1991; Tsebelis, 2002); they also exhibit greater longevity and thus are more able to deliver benefits to their supporters. On the other hand, perhaps 'maverick' representatives are more able and willing to impose their constituents' views and concerns on the government at large when the attention of national politics—and national parties—is elsewhere (see, e.g., Canes-Wrone, Brady and Cogan, 2002; Pattie, Fieldhouse and Johnson, 1994), or perhaps they avoid dangerous 'group think' (in the sense of Janis, 1972) behind the scenes. But whether different voters hold different preferences, or voters are internally divided, it is not difficult to see the potentially tricky balancing act this presents for MPs involved in Commons roll calls.

Despite—or perhaps because of—the complicated dynamics this suggests, previous work has tended to examine other MP-voter links. A series of studies, for example, examine constituency service as both a goal of members and its effects at election time (see Cain, Ferejohn and Fiorina, 1987; Norton and Wood, 1993; Searing, 1994). Others look to local campaigning activity (e.g. Denver and Hands, 1997; Pattie, Johnston and Fieldhouse, 1995), with some suggesting that, whatever the truth may be, MPs increasingly *believe* they enjoy a personal vote dependent on their appearances, statements and legislative behavior (Cowley, 2005).

When scholars *have* examined parliamentary votes as a predictor of performance, they have tended to be limited in scope: in the case of Pattie, Fieldhouse and Johnson (1994), they deal with just six 'key' votes with various levels of attendant whipping, while Kam (2007) uses 'dissent' on all whipped bills (in the aggregate) to predict voter familiarity with their MP's name.[2] These strategies are not unreasonable given the specific questions that were to be addressed, and the studies are important in that they are rare breaks with the orthodox thinking on voting in Britain. Nonetheless, if we could be agnostic about the potential bills of interest, and thus include *all* votes as possible predictors, we would garner the best of both investigations: we would have a survey of parliamentary behavior that is *deeper* in terms of both the number of votes in the sample and *broader* in terms of their content and the pressure applied in each case. We would thus capture the totality of legislative behavior, rather than a potentially misleading small sample. Before doing so, we need to be clear about what voters desire and expect from their representatives. We outline our intuitions in the next section.

4.3 Formalizing Intuitions

As with other scholars in the field, our work here will concentrate on the behavior of *government* (i.e. Labour) party MPs: the 'power' of the government relative to other parties in a Westminster system means that these representatives receive more media attention, and are more likely to be held accountable for the state of the polity at election time. Thus these are the members for whom the stakes are highest—for whom party unity and/or independence matter most.

[2]Though for New Zealand, rather than Britain, Kam (2007, 135) does show that dissent is an effective vote-winning strategy.

Our intuition is that on almost all issues (whipped or unwhipped) debated in the House of Commons, voters are uninformed. The bills and proposals are too technical and it is too time-consuming to review and understand their contents. Thus, voters rely on heuristics. For the reasons cited above, party unity is prized on issues that are most important to the government's legislative agenda. Notice that this effect will be strongest for bills which exhibit "government-versus-opposition" voting: that is, on roll calls where the non-government parties attempt to foil the executive for the sake of opposition, rather than because they actually have some different and sincere policy preference (see Spirling and McLean, 2007). The reason is that these votes most starkly test the Cabinet's ability to govern; indeed losing on such a roll call would typically trigger a vote of confidence.[3]

By contrast, when votes are relatively free of party pressure, precisely because they do not endanger a government agenda, maverick, independent-minded voting is welcomed. Indeed here, 'dissent' from the MP's party as a whole is a 'good thing.' However, on the rare occasion that free votes arise whose contents and consequences are simple to grasp for the public at large, voters reward MPs on a spatial basis. That is, left-leaning partisan voters want their MPs to plump for the more 'socialist' option while right partisans prefer the more 'conservative' choice.

[3]Such votes dangerously undermine Prime Ministerial authority: John Major called such a vote in 1993 in response to sustained rebellion over a series of European integration votes. Major was compelled to resign his position as party leader—though he subsequently won reelection—in response to the continual sniping. Fifteen years earlier, James Callaghan was forced to call a general election (which he subsequently lost) after suffering defeat in Commons in March 1979.

More formally, suppose there are $1, \ldots, i, \ldots N$ voters in constituency j, for which MP$_j$ is the incumbent governing party member of parliament. For any particular bill (or proposal, or amendment) b that is to be voted upon, there are two possible responses by MP$_j$—'aye' or 'no'. We will write the choice of MP$_j$ as a point o_j in p-dimensional space: $o_j \in \mathbb{R}^p$. For a bill that is essential to government business, denote the Cabinet's preferred outcome as o_g. There are q_1 such bills of a total of B that come to be voted upon in House of Commons. Since the government essentially never loses on parliamentary business it cares about, in a slight abuse of notation, we will use o_g to also refer to the outcome voted for by the *majority* of the governing party. This extends to the $B - q_1$ bills that are not essential to government business. Finally, denote the median voter of constituency j as having an outcome preference on any particular roll call as $o_i \in \mathbb{R}^p$. Notice that we assume she voted for the incumbent at the previous election. We can now write her utility from MP$_j$'s voting record in the Commons as

$$U_i(\cdot) = \underbrace{-w_1 \overbrace{\left(\sum_{b=1}^{q_1} d(o_j, o_g)\right)}^{\text{whipped}} + w_2 \overbrace{\left(\sum_{q_1+1}^{q_2} d(o_j, o_g)\right)}^{\text{unwhipped}}}_{\text{voter uninformed}} - \underbrace{(1 - w_1 - w_2) \overbrace{\left(\sum_{q_2+1}^{B} d(o_j, o_i)\right)}^{\text{unwhipped}}}_{\text{voter informed}} \tag{4.1}$$

where $0 \leq w_1, w_2 \leq 1$ refer to the 'weights' that voter i places on each legislative activity component and $d(\cdot, \cdot)$ is the Euclidean distance between two points in p dimensions.[4] With no loss of generality we assume that voter i compares the utility from (4.1) with some constant c—say, zero—and casts her ballot for MP_j at the subsequent election if $U_i(\cdot) \geq c$, else she abstains or votes for a competitor. Several simple comparative statics, consistent with our reasoning above, follow:

- for (heavily) whipped government business in Equation (4.1), so long as $w_1 > 0$, $\frac{\partial U_i(\cdot)}{\partial d(o_j, o_g)} < 0$. In words: for bills that deal with important government business, (partisan, government party) voters will disapprove of fractious, rebellious MP behavior. Moreover, the further away the MP's vote (o_j) places him relative to the government (o_g), the worse he does.

- for unwhipped roll calls on which voters are *uninformed*, so long as $w_2 > 0$, $\frac{\partial U_i(\cdot)}{\partial d(o_j, o_g)} > 0$. In words: on these bills, voters reward 'independent thinking' insofar as it manifests itself as choosing options different (far from) the party majority option.

- for unwhipped roll calls for which voters are *informed*, so long as $w_1 + w_2 < 1$, $\frac{\partial U_i(\cdot)}{\partial d(o_j, o_i)} < 0$. That is, the voter's utility is decreasing the further the MP's choice is from her ideal (outcome) point.

For reasons that we will discuss below, it is not possible to 'test' this theory directly. Nonetheless, this model will give us a structured way to think about our findings. Before leaving the theoretical discussion here, it is helpful to

[4] Recall that the Euclidean distance between two points $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ is simply $\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$. It is thus always positive, and caters for points in any number of dimensions in space.

clarify three issues. First, this is an *as if* model: we are not contending that most voters directly examine voting records to ascertain their representative's performance. Rather, they receive media and elite reports summarizing these patterns. What matters then is the *types* of MPs and the *types* of voting behavior in a broad sense. Second, there is a potential contradiction in the voter calculus: punishing a rebellious MP also hurts the party by potentially denying them an admittedly unreliable seat in the Commons. We do not doubt that such reasoning occurs, but we will assume that at *some* point an MP's actions are sufficient to merit voting against them, even if it superficially damages the national party in the short term.[5] Third, to the extent that a time trend exists, Figure 4.1 implies that independent thought ought to be increasingly rewarded, in addition to (or perhaps in place of) loyalty. Otherwise put: (relatively) free votes should be better predictors of performance in 2005 than in 2001.

4.4 Data and Method

Ideally, we would test our predictions from Section 4.3 with constituency level data recording citizens' preferences for their MP's specific legislative actions, in addition to partisan, government party voters' views on the desirability of party unity and favored positions on various free-vote issues. We would compare these predilections to their representative's actual record

[5] As a case in point, consider the fate of Neil Hamilton, Conservative MP for Tatton in Cheshire. Accused by *The Guardian* newspaper of taking bribes to ask questions in the House of Commons, Hamilton proclaimed his innocence, sued for libel and went on to fight for his seat in the 1997 General Election, the then 5th safest Tory district in the country with a previous majority of 16,000 votes (27.6 percent). Hamilton lost to an independent challenger, by some 11,000 votes on a 18 percent swing away.

and calculate the effect that this divergence had on the MP's subsequent electoral performance. No such data exist so, instead, we will try to infer micro-behavior from aggregate level findings.

In the Americanist context, scholars have summarized Congress members' legislative behavior using ideal point estimation, and then contrasted this with metrics purportedly capturing the underlying political orientation of citizens in their home district. Thus 'divergence' and the degree to which a member is 'out-of-step' with his constituents may be assessed (e.g. Canes-Wrone, Brady and Cogan, 2002). But such a strategy is fraught with difficulty for the United Kingdom case: the sheer preponderance of strategic, 'government-versus-opposition' voting in the House of Commons makes interpreting industry-standard tools for summary very difficult (Spirling and McLean, 2007).

The analysis of British parliamentary voting in the context of our theory is hampered by two further considerations. First, it is not publicly obvious how much whipping is applied on individual roll calls. While parties certainly acknowledge that strong pressure is present at Westminster, they do not typically publish details of its intensity and, in any case, whipping may take the form of informal threats and offers.[6] Moreover, the fact that a roll call is ostensibly 'free' is no guarantee that it is completely undirected.[7] So,

[6]In practice, party whips contact party members with a schedule concerning upcoming divisions. They literally underline the requirement that members attend and vote appropriately according to the importance the party places on the roll call in question.

[7]See McLean, Spirling and Russell (2003) for a specific incidence of this phenomenon.

though we will use terms 'whipped' and 'unwhipped' below, they are meant only as broad-brush shorthand for the general environment in which the bill is being considered: this is especially true when the parties differ in their application of their whip. The second problem is that, within the parliamentary record, *Hansard*, missingness is endemic and unexplained. Since there is no way to formerly record absences, there is no way of knowing if a missed vote was due to indifference, principled abstention, illness, ministerial business, etc.

4.4.1 Data: 1997–2001, 2001–2005

For each data set, our dependent variable is the difference in electoral success (vote percentage in their home constituency) for each of the MPs in the study between the respective elections. As noted above, we restrict the analysis to government party—i.e. Labour party—MPs for whom we think the theory predicts the strongest effect. Further, we deal solely with MPs who hold seats in England and who did not hold ministerial office at any time during the parliamentary terms. The location restriction avoids including MPs whose electoral fates may be the product of nationalist party dynamics that are difficult to control for. The next avoids the potential endogeneity problems inherent in 'good behavior', front-bench promotion and constituency popularity. Simply put, MPs are given ministerial responsibilities contingent on toeing the party line,[8] but that same status makes them much more recognizable in their home districts, increasing the possibility

[8]See Benedetto and Hix (2007) for a discussion of this issue and its consequences for discipline.

their subsequent vote share is influenced by factors other than their roll call record. For obvious reasons, we also drop MPs who choose not to fight the subsequent election, or who joined the parliament mid-way through via by-elections (and thus for whom we are missing some dependent variable information). These data cuts leave some 225 backbench Labour MPs for 1997–2001, and 194 for 2001–2005.

Our approach is to predict, i.e. to regress, the vote shares on *every single roll call vote* the MP casts in the relevant parliament, data which was compiled from (online) *Hansard* reports.[9] This way, we know that the totality of MP legislative behavior is being captured. We have also collected numerous (standard) controls which include the difference in electoral spending by Labour between the elections (a reasonable proxy for campaign 'effort'—see Pattie, Johnston and Fieldhouse (1995)), the difference in turnout, and the proportion of constituency residents falling into several traditionally marginalized groups that might be disproportionately sympathetic towards a Labour government.[10]

4.4.2 Estimation Problem: $p > n$

The dependent variable of interest is (essentially) continuous. Hence, ordinary least squares (OLS) might seem attractive as an estimation procedure.

[9] See Firth and Spirling (2006) for a discussion of this data. We impose the requirement that at least 50 percent of all the MPs in the sample voted on that roll call for it to enter the data. This avoids including divisions where the majority of MPs were missing.

[10] Compiled from the Census 2001 Report for Parliamentary Constituencies. In particular, the percentage of pensioners, of non-whites, of individuals living with a limiting long term illness (which includes disability), of unemployed, and the proportion of households headed by a lone parent.

Unfortunately, on the 'right hand side' there are around 1300 variables—predominantly parliamentary divisions—which we wish to use as regressors (or 'predictors'). That is, the number of parameters to be estimated, p far exceeds the number of observations, n: we have negative 'degrees of freedom', $p > n$. The consequence is that the standard OLS solution,

$$\boldsymbol{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}, \tag{4.2}$$

is unavailable; moreover, the equivalent of Equation (4.2) for essentially any parametric model—discussed in more detail below—is elusive.

One unsatisfactory solution is to re-specify the regression problem with fewer variables, but this requires an entirely arbitrary *a priori* restriction on the model. A better option is to employ systematic exploration methods that uncover structure and reduce the data down to its 'important' component variables. Moreover, because our theory is somewhat vague, we would like a flexible approach that assumes as little as possible about the precise functional form of the relationship between the $\mathbf{y}$ and $\mathbf{x}$ we alluded to above.

4.4.3 Tree-based Regression and Random Forests

OLS is a special case of 'regression' which we can write more generally as $\mathbf{y} = m(\mathbf{x}) + \boldsymbol{\epsilon}$ where $m()$ describes an unknown model for the data we observe. Depending on the assumptions we make about $m()$—in particular, what we believe we 'know' about the function that it represents—we commit to a *parametric* or *nonparametric* approach. In parametric models, the functional form of the relationship between the dependent and independent variables is assumed known, but it contains parameters whose values are unknown and to be estimated. Political scientists are most familiar with generalized linear models (GLM) where $\mathbf{y}$ is linearly associated with $\mathbf{x}$ profiles but in which the functional relationship between them can be non-linear.[11] Examples include logit or Poisson regression. When analysts have a strong grasp of the underlying physical causal process (and thus, for example, the potentially subtle ways that variables interact), parametric models are a good choice. They are of limited use, however, when the data are not well understood, or theory is not easily forthcoming—especially when the number of independent variables is large and thus the causal story complex. This is precisely because the assumptions that make parametric models straightforward to fit also restrict the variety of relationships that they can adequately describe: and if the model is 'incorrect' for the data (e.g. linearity is assumed when non-linear relationships are more apt), results will be less than satisfactory. Notice that the goodness of fit criteria may be misleading in these circumstances. In any case, for problems in which we have $p > n$, standard parametric models cannot be fit to the data.

In contrast to parametric models, *nonparametric* models make very few assumptions about the functional form between the dependent and independent variables. While parameters of the model are estimated, they do not necessarily have a clear interpretation in terms of the physical causal

[11] That is, $\mathrm{E}(\mathbf{y}|\mathbf{x}) = \mathbf{G}(\boldsymbol{\beta}\mathbf{x})$ where E() is the expectations operator, $\mathbf{G}()$ is some function and $\boldsymbol{\beta}$ are the parameters to be estimated.

process thought to underlie the data: instead the concern is with *function* estimation and good prediction.

A simple nonparametric method, popular in biostatistics, is that of *regression trees* (Ripley, 1996, 213). Within this literature, the task of optimizing predictive performance is performed by *classifying* cases according to their independent variable values via a series of logical splits. Typically, at each split, or 'node', the algorithms divide one independent variable into (binary) disjoint subsets into which the cases can be placed. Further splits are then proposed for any cases that are not in terminal nodes (known as 'leaves' of the tree). There are lots of different rules that may be used to 'cut' the independent variables (Ripley, 1996, 216–220), and hence several different ways to assess how well the tree classifies the problem in question. The general idea is to split the data so that cases within the subsets are as like each other as possible; in regression contexts, this literally means choosing to split a 'parent' node such that the resulting two 'child' nodes have the minimum possible within node variance in terms of their values of the dependent variable.[12] By seeing which variables—when split—best predict the cases, we obtain some notion of 'importance' in a regression context.

To see how this process might work, consider the following toy example: we have data on the 2004 Presidential election, in terms of George Bush's percentage of the vote in any given state. We also have 'explanatory' variables for each state, including its population density, its median household

[12]More details may be found in Appendix 4.7.

income, and the proportion of its population that are black. In Figure 4.2 we report a tree grown for this data using the split criterion defined above.[13] At the top of the tree, we see the first split was on population density, with those having a value of lower than 206.8 sent 'to the left' and those that are more dense than this cutoff sent 'to the right.' On the left, these relatively rural states were once again split on population, with the least dense (< 35.2) sent to the left. The next split for those states occurs in terms of their black population. If it was less than 1.5 percent (0.015 to the left of the figure), they were dropped in the bin of states with a mean value of 65 percent for Bush. Literally, these are large, rural, predominantly white states such as Utah and Wyoming. Each leaf (end value) in the figure corresponds with a different bin, and states are placed there contingent on their variable values.

Several comments are in order here. First, the variables are not all of the same 'importance': different variables at different nodes do better in reducing the variation in the subsequent (split) data sets than others (here, population density does particularly well). Second, there is no guarantee that 'important' variables will be statistically significant in a regression context or *vice versa*: classification and/or tree-regression are not synonymous with OLS. Third, we made a judgement call as to when the tree would stop growing. Clearly, there are 50 states, but only 8 bins as terminal nodes—65.00, 57.17, 46.00, 51.00, 59.78, 54.60, 48.20, 42.29. In this case, the tree

[13]Implemented in the `R` (R Development Core Team, 2006) statistical language and environment using the `tree` package: see Ripley (1996).

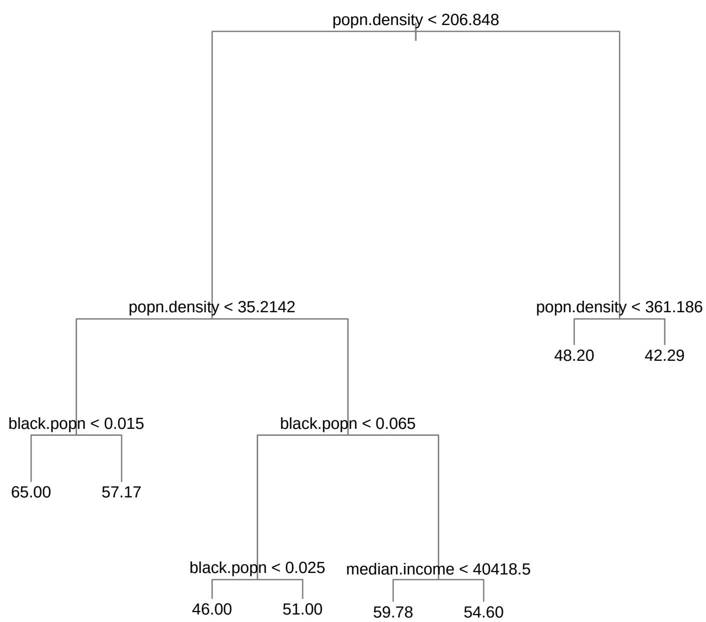

Figure 4.2: Example of a classification tree for American states, predicting (categorizing) the percentage of vote for George Bush in the 2004 Presidential election.

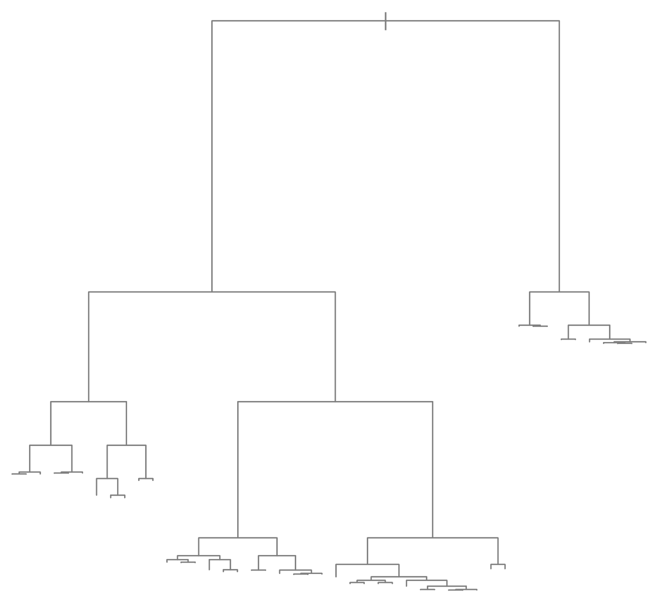

Figure 4.3: Example of a tree that uniquely classifies American states, in terms the percentage of vote for George Bush in the 2004 Presidential election. Node labels removed for clarity of presentation.

is grown until no more than 10 cases reside in a terminal node. In this toy instance, we could have, in fact, insisted on a perfect fit to the data: 50 terminal nodes with one for each state. Such a tree is given—without the distraction of node labels—in Figure 4.3. Its complexity should be obvious.

For most purposes, fitting a large data set with lots of independent variables in a way that resembles Figure 4.3 won't be helpful: we will 'overfit.' That is, we will essentially reproduce the complexity of the data by fitting its random ('noisy') features which are not part of the causal process in which we are interested. Such trees have high variance, in the sense that altering the data on which they were grown (by, perhaps, adding new points out of the sample) would result in very different findings. This is undesirable given that we want to know about the *general* structure of the data at hand. To reduce this variance, we can 'prune' the tree back after it is grown, and there are various methods for doing so (Ripley, 1996, 221–231). Alternatively, we can 'stop' growing at some pre-defined point. But less complex trees introduce a new problem of bias: simpler models may not be good enough to correctly represent the subtleties of the data we are interested in. Though we can make the selection of 'right-sized' trees automatic using various cross-validation procedures, there may be a better way to proceed.

One solution is to grow *multiple* trees—each potentially classifying the cases differently—and find some way to aggregate the results from all of them. A popular and powerful way to do this is Breiman's (2001) random forests algorithm.[14] The key idea is that each tree will be unpruned and thus have low bias. And, while the individual trees will have high variance, for reasons noted below, there will be a low correlation between them. Combining the trees will give the best of both worlds: low bias and low variance.

Random forests actually builds on an earlier technique—also proposed by

[14] Implemented in `R` by Liaw and Wiener (2002).

Breiman (1996)—called ‘bagging’.[15] Bagging is an abbreviation of *b*ootstrap *agg*regat*ing* and, like most tree methods, it begins by splitting the data into a ‘training’ and ‘test’ set. In random forests, we do not need to make this split. We proceed as follows:

1. take a *bootstrap sample* from the data, D. In practice, this means taking a sample of size N—with replacement—from D (see Efron and Tibshirani, 1993, for a general discussion). Denote this sample as D_k and notice that the *same* observations can therefore appear multiple times in D_k, and others from D might not appear at all. In fact, approximately one third of the cases will not appear in the sample (called ‘out of bag’).

2. train the predictor using D_k . That is, grow one tree from that sample until the cases reside in sufficiently small nodes (commonly, five cases). These trees are not pruned.

3. repeat steps 1 and 2 $1, \ldots, b, \ldots B$ (say, 500) times.

4. aggregate the results of the B trees. In a regression context, we average across the trees.

Random forests get their name from the randomness employed at each node in each tree: rather than employing the best split among *all* variables to continue to grow the tree, a random forest splits a *randomly chosen subset* of all the predictors; in practice this subset is approximately one-third of the total number of independent variables (around 350 here). This strategy results in

[15] See Ruger et al. (2004) for a political science application of an earlier contribution by Breiman et al. (1984).

low correlation between the trees, and yet the predictive performance generally exceeds that of related techniques—such as neural nets (see Beck, King and Zeng, 2000; de Marchi, Gelpi and Grynaviski, 2004)—without overfitting. A schematic diagram describing random forests is given in Figure 4.4.

We will be interested in several quantities from our models: first, the goodness of fit, which is calculated in a similar way to the R^2 for a standard regression with the difference that we explicitly attempt to predict data that was not involved in fitting the trees. Second, we want to know how 'important' different variables are; to do this, we will inspect a residual sum of squares measure for trees where the variable is involved, relative to those where it is not. Third, the marginal effect of any given variable on $\mathbf{y}$ will be a concern. Broadly, we obtain this by comparing the predicted $\mathbf{y}$ when the variable is in the (aggregate) random forest model, to when it is not. Appendix 4.8 gives more technical details. Notice that because random forests are not attempting to estimate regression coefficients in the usual, GLM, way there is no requirement that $n > p$.

If random forest classification is successful, we expect observations which are 'alike' to end up in the same terminal nodes across trees. We can use this fact to impute missing observations, and Appendix 4.9 explains how.

4.4.4 Estimation Plan

Given the above explication, our investigation now proceeds in several steps, which combine both quantitative and qualitative methods:

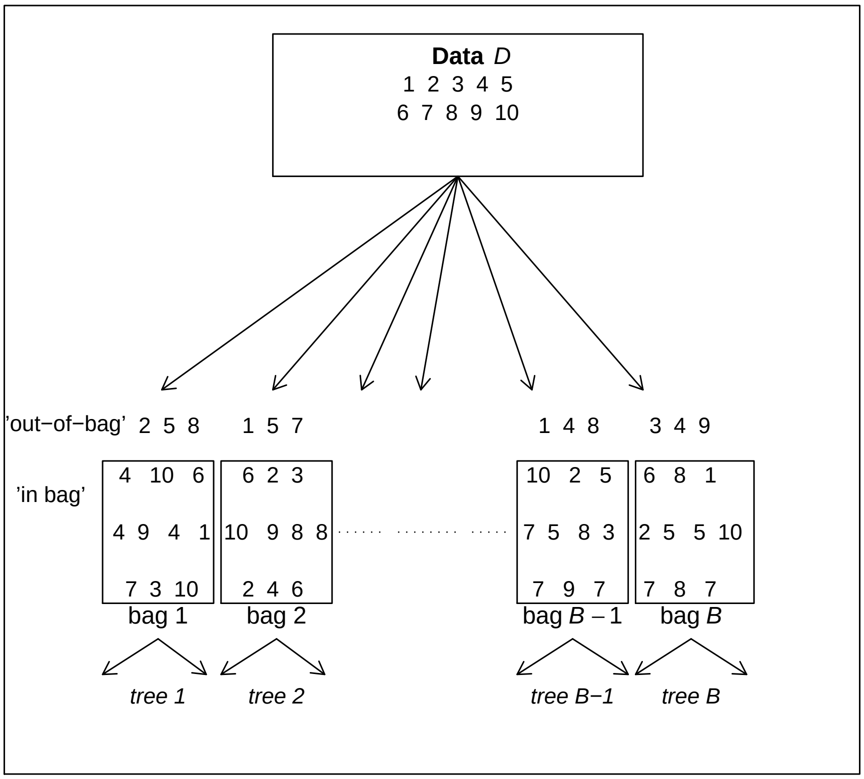

Figure 4.4: Random forest schematic. Here the data set consists of ten observations labeled 1 through 10. These are 'bootstrap sampled' into bags (1 thorough B). For each bag, some observations are 'left out.' A tree is grown from each bag, and the results aggregated.

1. impute values for the missingness in the divisions using a random forest classification

2. use imputed data and random forest to calculate relative 'importance' of different controls and roll calls in predicting electoral success.

3. using 2 as a guide, obtain and report the (random forest) marginal effect of some important variables (roll calls)

4. use narrative histories (such as Cowley (2002) and Cowley (2005)) to establish nature of bills—whipped/unwhipped, etc.

As a further comparison we will also report OLS coefficients for the variables from 2 (using each variable with all controls to predict performance). To reiterate, we will use data classification techniques to find the most 'important' variables (which will involve a judgement call on our part) and *then* see if and how they fit our theory. Because the method fits by maximizing prediction, concern will focus on the *types* of roll calls that matter ('heavily whipped', 'free', 'issue of conscience', 'programmatic concern', etc.), rather than their precise content. In other words, we will *identify* those bills that best predict the dependent variable in question and then *describe* their broad characteristics in terms of our interest in party discipline versus independence. Because essentially all votes are used, there is thus no danger that we are looking in the 'wrong place' when predicting electoral performance from legislative behavior.

4.5 Results

The results are in two parts, first for 1997–2001 and then, separately, for 2001–2005. Of course, for both data sets, we follow the stages set out in the previous section.

4.5.1 First term: 1997–2001

For our first data set, recall that the dependent variable is the difference in vote share that the backbench Labour MP's garnered in their home constituency for the 1997 versus 2001 election. In fact, both events were landslides for the Labour party as a whole, and Labour lost just six seats (of a 179 seat majority) in the latter. Nonetheless, there was quite significant variation in terms of the performance differences of individual MPs with a mean change of -1.59 percent (variance of 14.77), and a maximum loss of -12.5 percent recorded by Khabra Piara (Ealing Southhall), and a maximum gain of 12.7 percent recorded by Michael Jabez Foster (Hastings and Rye). In Figure 4.5, we graph the most important 30 variables from the random forest regression where, x-axis scale refers to the decrease in the average residual sum of squares when the variable is in the tree, as opposed to when it is left out. The (mean) pseudo-R^2 goodness of fit measure for the model was around 0.06.

Note first that the controls are seemingly essential and all contribute more than any roll call in accounting for differences in the dependent variable. In particular, the percentage of citizens in a district with a debilitating illness

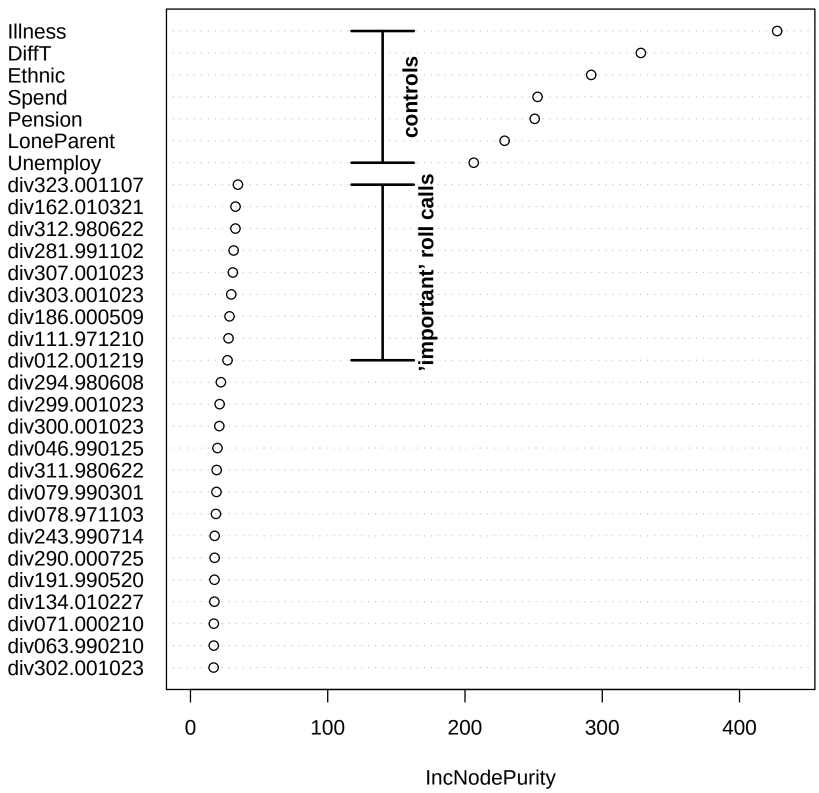

Figure 4.5: Variable importance plot for 1997–2001 data. Controls and 'important' roll calls used in subsequent analysis shown as subsets of all variables. x-axis gives the increase in the regression performance from including each variable.

(`Illness` in the graphic) is most important, followed by turnout differential (`DiffT`), the proportion of non-whites (`Ethnic`), the proportion of Pensioners (`Pension`), the proportion of lone parent households (`LoneParent`), the proportion of citizens unemployed (`Unemploy`) and then the electoral spending differential (`Spend`). The following points on the graphic refer to the importance of the roll calls and have labels beginning with `div`. Deciding which of these bill votes are worthy of further inspection is essentially a judgement call, and here we consider the top 9, below which there is a slight drop off in importance.

In Table 4.1 we report our results: for each variable, we give the marginal effect from the random forest: literally, this is the difference in the predicted constituency performance when an MP votes with the minority of his party as opposed to when he votes with the majority. We also report an OLS coefficient (with p-value) from including that predictor in a regression—with all controls. In both cases, a positive number means the MP did better at election time while a negative one means he was worse off. Since the controls are not the focus of the study, they are left out of the table. As noted above, it can be quite difficult to know for sure how much party pressure was applied in the various votes, but column 3 gives a casual author-based coding of the discipline imposed for that vote, ascertained from a qualitative reading of the relevant House of Commons debate and surrounding literature; the whipping is one of four ordinal levels: ‘free’ (essentially no whipping), ‘weak’ (little whipping, perhaps timetabling concerns), ‘moderate’ (some pressure, government has preference for outcome), ‘strong’ (tight whipping, government has strong, programmatic concerns).

Beginning first with bills which Cowley (2002) discusses as central to the Government's post-1997 legislative agenda, we note that the signs are in the expected direction. That is, rebeling on the Transport Bill (row 7) leads to MPs being punished at election time—i.e. a negative coefficient—that is statistically significant. This roll call dealt with the Government's attempts to privatize the National Air Traffic System. The rebels on this roll call were backing an (non-Government) amendment that would have substantially delayed the denationalization.

The vote on the Crime and Disorder Bill (rows 3) was concerned with lowering the age of homosexual consent from 18 to 16, such that it would be on a parity with heterosexual consent. They were ostensibly unwhipped, yet this belies some potential party pressure on MPs. For this bill, voting *for* an amendment to enforce a minimum age of at least 18 in certain circumstances was punished by the electorate. Within the realms of the model we advanced above this suggests two possibilities: first, that voters were sympathetic to such social equality issues, and the MPs that vote for them. For gay rights, such preferences seem to be the case: some 93% of British respondents in a recent survey supported laws to protect gay people from discrimination (Cowan, 2007), while the legalization of same-sex civil partnerships appears to have occasioned much less heated debate in the UK than their potential introduction in the US. Certainly, those taking the more conservative position here did worse. Second, note that though the government was *formally*

Code	Name	Discipline	'Rebel' Effect (RF)	'Rebel' Effect (OLS)	*p*-value
div323.001107	Programming of Bills	weak	+0.179	+1.144	0.038
div162.010321	Weights and Measures (S.I.)	weak	+0.291	+0.827	0.256
div312.980622	Crime and Disorder Bill	free	-0.011	-0.391	0.460
div281.991102	City of London (Ward Elections) Bill	moderate	-0.176	-0.845	0.092
div307.001023	Election of Speaker	free	+0.213	+1.743	0.006
div303.001023	Election of Speaker	free	+0.142	+1.165	0.075
div186.000509	Transport Bill	strong	-0.271	-1.287	0.031
div111.971210	Doctor Assisted Dying	free	-0.147	-0.472	0.360
div012.001219	Human Fertilisation and Embryology	free	+0.171	+0.822	0.132

Table 4.1: 'Most important' parliamentary votes for predicting electoral performance by backbench Labour MPs in 2001 general election. First column is data code for the division with date after point: so, div323.001107 is division 323 taking place on November 7, 2000. Name of bill in column 2; random forest 'coefficient' in column 3 for being a 'rebel'(in the minority and/or frustrating the Government's ambitions); OLS coefficient in column 6; *p*-value for that coefficient in column 7.

uncommitted to the amendment, it allowed it time for consideration and, when the House of Lords rejected it, the government reintroduced it via the separate Sexual Offences (Amendment) Bill. Indeed, when the Lords once again refused its assent, the government threatened to enact it directly over the wishes of the upper chamber. In short then, the executive was clearly committed and sympathetic to this issue—suggesting that rebels might be considered at least 'out of step' if not 'troublemakers' *per se*.

Votes that concerned the election of a new House of Commons Speaker occupy rows 5 and 6 of the table. Unlike, say, the United States House of Representatives, the Speaker of the House of Commons is a *non-partisan* referee of parliamentary debate. Selected from back-bench MPs, Speaker candidates do not usually receive official endorsements though MPs from the same party often vote similarly in the elections. Ultimately then, these are free votes, and the whip is not present. Moreover, these are votes for which voters are uninformed: the candidates are not typically *a priori* well known, nor do they have particular 'policies' beyond their general personal reputations. We would predict then, that rebeling—in the sense that MPs vote in a way that contrasts with the majority of their party—would be a boon to electoral performance. In fact, both coefficients—OLS and random forest—are in the correct direction (positive), and they are statistically significant.

The other roll calls in the table have a slightly more complicated status in terms of the party pressure applied to the vote. For example, consider the

June 2000 "Programming of Bills" motion that concerned the Modernisation Committee's recommendations for changes to the way that parliamentary business be conducted. Usually, 'House business' issues are seen as free votes, not least because they emanate from non-partisan Select Committees that are designed to hold the government to account. Interestingly, and as a break with standard practice, the Modernisation Committee was chaired by the Leader of the House, a cabinet member. Both in committee and during the ensuing Commons debate, MPs accused the Government—via the Leader of the House—of promoting its own agenda (see Kelso, 2007, 140–141). In particular, several MPs felt that the proposed reforms made it more difficult for parliament to oppose the executive and hold it to account. Thus, members backing an amendment to these plans that would have extended debating time were *de facto* frustrating Government ambitions. There are thus two consequences: first, the vote was 'free' for Labour backbenchers but pressure was surely present and second, the 'rebels' were actually the *majority* of the party on this vote (note that the Leader of the House opposed the amendment). With all this in mind, the sign is essentially as predicted: those showing independent-mindedness on a bill that was not an explicit part of the Government's agenda did better at the polls.

The City of London (Ward Election) Bill appears in row 4: the coefficient is significant. By this private bill, the City of London intended to reform its self-governing arrangements as a municipality. Unusually for an authority in a democratic state, the City gives businesses votes, and the proposed legislation would have *extended* this right to more businesses. For many left-leaning MPs in the Labour party, such a plan was anathematic. It is unclear exactly how or even if the Government asserted pressure on its MPs though. Certainly, the Leader of the House claimed it was an explicitly free vote, though Salter (2004) argues that the Government may have pressured MPs to at least pass *something* that could be considered 'reforming', just to be done with the issue. In July of 1999, the ideological components of the plans came under attack, with Labour backbenchers in the minority condemning them. In November of the same year, the bill was not yet passed and there was lively debate as to whether or not the proposals should be reconsidered in the next parliamentary session which, without explicit Government involvement they would not have been. To the extent that the rebels were now holding up procedural *Government* business, it is perhaps understandable that the sign is negative and they were punished for their intransigence.

The votes concerning "Weight and Measures" (row 2), "Human Fertilisation and Embryology" (row 9) have signs as predicted, though not at significant levels. The first vote dealt with a particular "statutory instrument", a method of delegated legislation that allows a minister to make changes to rules and laws without the requirement that the proposals be passed as a separate parliamentary bill. In this particular case, the vote dealt with how long Imperial measurement units would be allowed to continue as a 'supplement' to the metric units required by European Union legislation. It should be clear that such a vote should not fall into the category of essential governmental programmatic business as defined above—and we see that 'rebels' did better when they opposed this vote. The Human Fertilisation vote pertained to cloning as a medical research practice. Again, we see rebels garnering a (minor) uptick in support for their 'independent-mindedness' on a free vote for which the public is presumably fairly ill-informed.

The final case to consider concerns a vote on "Doctor-Assisted Dying" (row 8). Euthanasia is not generally a salient issue for most British voters at the ballot box (unlike, say, abortion for US voters), and the debate has been characterized by its cross-cutting dimensions. For example, conservative, religious groups, the British Medical Association (the main doctors' trade union) and several prominent left-wing MPs oppose doctor assisted dying. It is thus not precisely clear where the average Labour voter places herself on the issue, so it is hard to know what to make of the fact that 'rebels' (those who voted in favor of euthanasia) did worse at election time. In any case, the effect is apparently not significant.

4.5.2 Second term: 2001–2005

The 2005 election saw most Labour MPs do worse than in 2001, with a mean loss of 7 percent in our data set (variance was 12.32). Given that Labour entered the election with 30 seats where the majority was less than 7 percent, this was clearly a more competitive election. The *national* swing was a commensurate 6 percent away from the governing party, though their overall performance was sufficient to garner a workable majority of 66 seats. The biggest (Labour) loser of the election was Roger Godsiff who suffered a reduction in his vote total of some 21 percent (primarily to a candidate from the new anti-war Respect party) and the biggest gains were made by Shaun

Woodward, a Conservative defector who left his original party in 1999. In Figure 4.6 we graph the most important variables for this regression from the random forest technique. The (mean) pseudo-R^2 goodness of fit measure for the model was around 0.04.

Once again, we note that the controls are certainly important, though their relative ranking has been permutated from that in Figure 4.5. And, again, we make a judgement call as to which of the bills are worthy of further inspection. Here, it is just 10, and we report their effects in Table 4.2. As in the 1997–2001 session, we see that the 'Programming of Bills' is an apparently important predictor, occupying rows 1 and 8 of the table. In both cases, 'rebeling'—in the sense of voting with the minority—was beneficial for MPs (and this effect is statistically significant). These votes were explicitly free though, once more, knowing precisely how much pressure was applied in practice is difficult to gauge. With this in mind, as with our previous findings, our theory predicts the positive sign we see in practice.

Unlike the programming matters of the previous paragraph, the "Children Bill" was part of the Government's legislative plans for the session and was therefore whipped appropriately. The vote in row 2 of the table actually concerned a backbench amendment (from David Hinchliffe, Chair of the Health Select Committee) that would have struck out part of the Government's bill that allowed parents to use "reasonable physical chastisement" in disciplining their children. The result would be a bill that essentially outlawed all such spanking. Somewhat unusually—given that it proffered a change

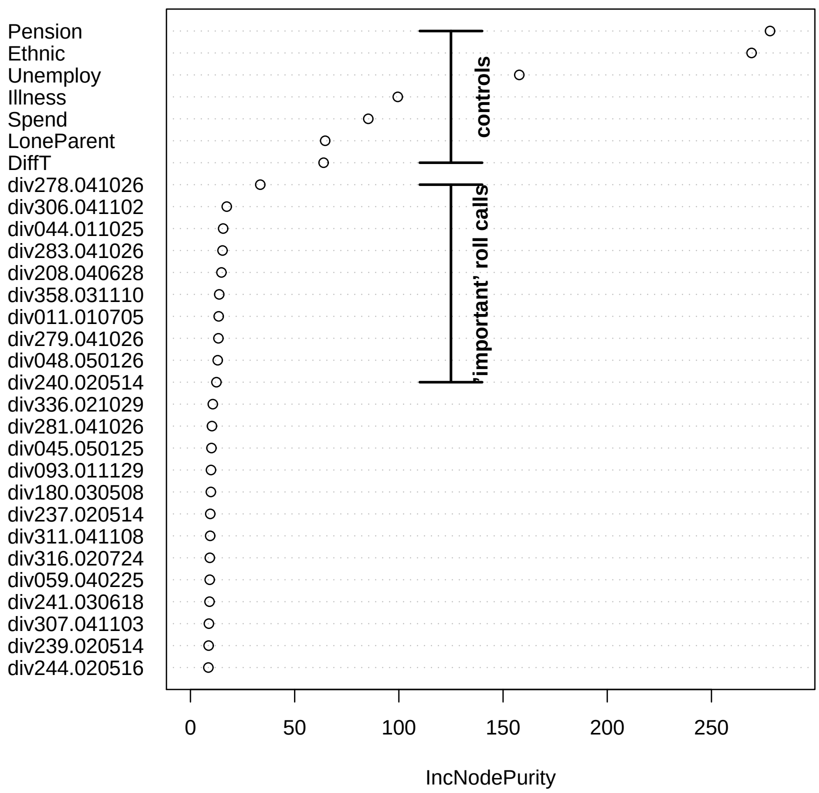

Figure 4.6: Variable importance plot for 2001–2005 data. Controls and 'important' roll calls used in subsequent analysis shown as subsets of all variables. x-axis gives the increase in the regression performance from including each variable.

Code	Name	Discipline	‘Rebel’ Effect (RF)	‘Rebel’ Effect (OLS)	p-value
div278.041026	Programming of Bills	free	+0.535	+1.660	0.005
div306.041102	Children Bill: Reasonable Punishment	moderate	-0.130	-1.113	0.039
div044.011025	Social Security	weak	+0.256	+0.931	0.130
div283.041026	Removal of References to Strangers	free	+0.288	+1.297	0.043
div208.040628	Human Tissue Bill: consent for removal	weak/mod	+0.240	+1.335	0.024
div358.031110	Fluoridation of Water Supplies	free	+0.141	+1.233	0.013
div011.010705	Members’ Allowances, Insurance	free	+0.100	+0.243	0.630
div279.041026	Programming of Bills	free	+0.250	+1.263	0.046
div048.050126	Modernization of the House	free	+0.047	+2.517	0.634
div240.020514	Liaison Committee: Power to Take Evidence	free	+0.082	+0.586	0.197

Table 4.2: ‘Most important’ parliamentary votes for predicting electoral performance by backbench Labour MPs in 2005 general election. First column is data code for the division with date after point. Name of bill in column 2; random forest ‘coefficient’ in column 3 for being a ‘rebel’(in the minority and/or frustrating the Government’s ambitions) ; OLS coefficient in column 6; p-value for that coefficient in column 7.

to government legislation—the vote was 'free' in theory. Nonetheless, the Government clearly had a preference on the vote, and we see a negative and significant coefficient implying that 'rebels' did worse after 'defying' the executive. The other piece of Government legislation in the table is the "Human Tissue Bill" (row 5) but, once again, the substantive circumstances of the vote are quite odd by Westminster standards. In particular, the roll call concerned a Liberal Democrat amendment to a government bill that would have made organ donation automatic unless the deceased had previously registered an objection. While this was a free vote for both the main Opposition parties, it was whipped by the government, on the philosophical basis that such an amendment would, by presuming consent, be denying individuals the right to exercise their conscience. Clearly then, this bill was not a cut and dried "government-versus-opposition" case and the low turnout adds fuel to this argument. With this in mind, the significant and positive coefficient for rebel behavior is perhaps less anomalous than it might first appear.

Several of the remaining entries in Table 4.2 are free votes, primarily that dealt with internal House issues. For example, the "Removal of References to Strangers" (row 4) dealt with changes to parliamentary language, "Members' Allowances, Insurance" referred to various MP compensation concerns, "Modernization of the House" (row 9) referred to the sitting hours of the Commons and "Liaison Committee: Power to Take Evidence" (row 10) dealt with procedural issues for cross-party committees. In all cases, we see the sign is positive, as predicted for 'rebeling' on such unwhipped votes, though it is only significant in one case.

The last two votes to consider from Table 4.2 are quite obscure. The first, on “Social Security” (row 3) actually refers to a statutory instrument that regulates unemployment services in the UK. The Conservatives abstained on the issue and, though not technically a free vote, it was not a crucial part of the Government’s agenda. Thus the sign for ‘rebeling’ is in the correct direction (i.e. it is positive) but it is not significant. Finally, row 6 deals with “Fluoridation of Water Supplies”, a perennial debate in the Commons. The free vote concerned whether or not local authorities could force water providers to add fluoride to residential and other supplies. As predicted, the sign is positive—rewarding the minority on a free vote—and, in fact, significant.

4.5.3 Summary of Results

We found that controlling for various other factors,

- For both 2001 and 2005, MPs rebeling against important government legislative business (‘trouble-makers’) were punished by voters, and this effect was sometimes statistically significant.

- For both 2001 and 2005, MPs dividing with the minority on free votes (being ‘mavericks’) which involved issues on which voters were ill-informed tended to be rewarded by voters

- For 2001, Labour MPs voting against popular social justice legislation—in particular, gay rights—in free votes were generally punished by the electorate

- the OLS effects, when significant, were on the order of 1 to 2 percent of the vote *differential* from the previous election. The random forest (marginal) effects were generally of a much smaller magnitude, though the signs generally agree.

4.6 Discussion

It seems that old certainties about British politics are changing: government party MPs are now more rebellious than they have been in a generation (Cowley, 2002, 2005), and there is increasing evidence that local factors matter at election time (Denver and Hands, 1997; Pattie, Johnston and Fieldhouse, 1995). The current paper was concerned with connecting the two trends and systematically assessing the evidence for voter reaction to patterns of MP activity in Parliament. Using new non-parametric techniques and roll call records in their totality, we found that (Labour) voters do indeed seem to respond to overall patterns of activity: they punish 'trouble-making' rebellion on heavily whipped government business, while rewarding more maverick legislators when party unity is not *per se* at stake. Moreover, this finding makes sense of seemingly inconsistent public opinion preferences for both party loyalty and free thinking, and adds a layer of intuitive, though subtle, depth to our understanding of accountability in Westminster systems. This perhaps allows some 'catch up' with Americanist efforts, while producing new directions for future research.

As was explicitly noted above, the model of voter behavior was 'as if' in nature: we would not contend citizens have a sufficiently advanced knowledge of policy making or parliament to *directly* reward or punish legislators on certain bills. Rather, we think that voters receive signals from a possibly media-based elite, who digest and then describe overall patterns of MP behavior as 'rebellious', 'loyal', 'fractious', 'conforming' and so on. So, while our data and method tapped the generalities of this dynamic, it could not trace the causation path precisely. Further work in this area might thus profitably gather media or other data on the messages that voters receive in order to disentangle what exactly are the triggers that change the labels MPs receive.

Finally, this paper has potentially important ramifications for institutional reform in the United Kingdom. At the time of writing, the House of Commons Modernisation Committee is engaged in an attempt to strengthen the role of backbenchers in terms of their dealings with the executive, and is soliciting expert opinion on these matters (see, e.g., Cowley, 2006). It is hoped that this will resurrect public engagement in politics after two general election turnouts of around 60 percent (very low numbers by Western European standards). We have suggested that voters are *already* either heterogenous or discerning in their responses to backbench behavior, and further independence for MPs will presumably make them more so. Whether MPs supporting these reforms understand that their electoral fortunes will be increasingly tied to their legislative performance and choices—and not the party's overall popularity—is unclear. Given that institutional norms—whipping and unity—appear to both affect and be affected by 'what voter's want', changing those norms ought to have interesting and complex consequences.

Appendix to Chapter 4

4.7 Regression Tree Split Criteria

Consider a 'parent' node t containing some number of cases $N(t)$ that is to be split. We measure the 'impurity' of t via the least squares deviation (also called the 'deviance'), written $R(t)$, which is

$$R(t) = \sum_{i=t} (y_i - \overline{y}(t))^2 \tag{4.3}$$

where y_i is the dependent variable value for the ith case at node t, and $\overline{y}(t)$ is the mean of all the cases in node t. Notice that Equation (4.3) is simply the residual sum of squares at t.

We will split the cases from t into two distinct subsets such that the subsequent 'child' nodes have summed deviations that are as small as possible. In particular, we require that our split s (on some value of one of the covariates) maximizes

$$R(t) - R(t_L) - R(t_R) \tag{4.4}$$

where $R(t_L)$ is the within node variance of the left child and $R(t_R)$ is the within node variance of the right child. This is equivalent to choosing a split maximizing the reduction in deviance.

As an example, suppose our parent node t has a variance of 20. Two possible splits in the covariates, s_1 and s_2 are proposed. These may be different points in the same variable, or different splits in different variables. Suppose s_1 yields two children: one with a deviance of 11, the other with a deviance of 19. For s_2, the left and right child both have a deviance of 15. These splits are considered equivalent since both result in an equal value of (4.4).

4.8 Model Fit, Variable Importance and Effect

Determining goodness of fit for random forests regression proceeds as follows. Notice from step 1 in Section 4.4 above that *in any given bootstrap sample* (i.e. in any given tree grown) around one third of the cases from the original training set will not appear there. These are called "out-of-bag" (oob) data. Every time a particular case i is in the oob data with respect to a tree, it is 'run down'—that is, regressed—according to that tree and an (oob) prediction produced. Take the average of all those oob predictions for i (that is, one for each tree where it is oob) and denote that value as $\hat{y}_i^{\text{oob}}$. Akin to linear regression, we would like to know the percentage of the variance in $\mathbf{y}$ that is explained by our model. For the current technique, this is computed as

$$1 - \frac{\text{MSE}_{\text{oob}}}{\hat{\sigma}_y^2}$$

where

$$\text{MSE}_{\text{oob}} = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{y}_i^{\text{oob}}\}^2$$

where MSE is the mean square error (Liaw and Wiener, 2002, 20).

The importance of variable v is estimated by considering the decrease in node 'impurities' when v is split (i.e. it is part of a tree), averaged out over all the trees in the forest. These impurities are measured via the residual sum of squares (RSS)—a formula we given in Appendix 4.7. We compare the (average) value of RSS when v is in $\mathbf{x}$ and when it is not. Intuitively, if the variable is 'important', it will be responsible for a better regression performance: it will generally occur more often in the 'best' split and/or the reduction in deviance when it occurs will be greater.

Finally, we will typically want to know the *marginal* effect that a variable v has on $\mathbf{y}$. Liaw and Wiener (2002) suggest estimating the function

$$\tilde{f(v)} = \frac{1}{n}\sum_{i=1}^{n} m(v, x_{iC})$$

where v is the variable of interest and x_{iC} are all the other variables. The element $m(v, x_{iC})$ is the (aggregated) random forest regression function (the predicted $\mathbf{y}$): the idea being that we vary the values of v ('aye' or 'no' in the case of a roll call) to see the (predicted) effect on constituency performance.

4.9 Imputing Missingness

Suppose that there is missingness in the underlying data (in the $\mathbf{x}$). Begin by filling missingness in each column with the median (if continuous) or modal (if discrete) category for that column. Then run a random forest regression procedure and compute the proximity matrix: that is the fraction of times (across all the trees) that each i, j pair of cases ends up in the same terminal node. For continuous missing data, the imputed value is the weighted average of the non-missing data for that column, with the proximities as the weights. For categorical missingness—the case for our work here—the category with the largest average proximity (which will be either 'aye' or 'no') is the imputed value. As a toy example consider a variable $x = [0, 0, \text{NA}, 1]$. Observation 3 is first filled with a 0 (since it is the most common category). Suppose that the resulting proximity matrix was

$$\left[\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1| & & 0.5 & 0.6 & 0.1 \\ 2| & 0.5 & & 0.4 & 0.1 \\ 3| & 0.6 & 0.4 & & 0.2 \\ 4| & 0.1 & 0.1 & 0.2 & \end{array}\right]$$

Observation 3 is with Observations 1 and 2 $\frac{1}{2}$ of the time, but with Observation 4 only $\frac{1}{5}$ of the time: the largest average proximity is thus zero and we impute this for the observation.

Chapter 5

Discussion

The preceding dissertation was concerned with three quite separate substantive areas, involving three somewhat different methods. What draws the chapters together is the focus on finding new solutions to 'old' problems: identifying 'turning points' in conflicts; measuring the 'power' of political actors; describing the 'effects' of roll call voting on electoral support. The essays all suggested thinking about these issues in new ways. In the first chapter, we employed a change point identification technique that was notable for its flexibility: the number of parameters was, itself, a parameter to be estimated. In the second chapter, we argued that 'power' makes little sense in the absence of 'contests' in which actors compete to garner finite resources. Our statistical model reflected our (more realistic) assumptions on this matter. In the third chapter, we argued that a very large number of variables (i.e. roll calls) might collectively demarcate members of parliament in terms of their appeal to voters. Rather than employing an *ad hoc* 'thinning out' of variables such that parametric models could be fit, we decided to take a more systematic non-parametric approach.

In the individual chapters we made some comments about future uses of these techniques within their substantive areas, but it may be helpful to think more generally about applications here. Reversible jump Markov chain monte carlo, which we introduced in the first chapter, seems an intriguing tool for 'cluster analysis' in parliaments. Historically, ideal point estimation (frequentist or Bayesian) has been a popular choice for analyzing legislatures (Clinton, Jackman and Rivers, 2004; Poole and Rosenthal, 1997). A model of legislator behavior is posited in which elected members compare the utility of the status quo relative to that the policy changed potentially introduced by passing a new bill. A distribution for the error terms is assumed—typically Type-1 extreme value or Guassian—to produce a (parametric, random-utility) statistical model. On occasions, such models produce misleading or unhelpful results. As noted in Chapter 4, Spirling and McLean (2007) show that strong whipping in Britain's House of Commons produces rank ordering on an 'agreement with the government' dimension, rather than an ideological spectrum. Clustering algorithms, which search for groupings, are an alternative way to proceed. Some of these techniques—optimal partitioning methods being among them (Venables and Ripley, 2002, 316)— require an *a priori* commitment to a fixed number of clusters K to which cases are then assigned. By contrast, other (typically 'hierarchical') methods fit numerous models (one for each value of k total clusters) which must then be compared. Since Bayesian procedures are philosophically appealing in terms of model comparison (see Gill, 2002, 199–224), they seem a natural choice here. Moreover, because we would like to be flexible about the number of clusters, RJMCMC would seem a particular good option.

The second essay (Chapter 3) made the case that we can consider political actors as players competing in contests. This is an idea with mileage outside the current context. For example, at least since the time of the ancient Greeks, scholars have studied the power of argument to convince citizens. The earliest accounts were primarily interested in rhetoric as an art (e.g. Aristotle, 322BC/1991), though in more recent times, psychologists and economists have studied the consequences of 'framing' effects, especially with respect to risk-taking behavior (e.g. Tversky and Kahneman, 1981). In these cases, 'arguments' essentially 'compete' with respondents either suggesting which they prefer directly, or voting commensurate with their views. Using the Bradley-Terry model in the second chapter would be one way forward here. Indeed, if one could properly separate competing arguments in, say, an election campaign, analysts could begin to model why certain *candidates* win or lose. As constructed, the standard Bradley-Terry model assumes that contests involve two players, with one losing and one winning. Over a series of (independent) contests, we make the reasonable assumption that success is binomially distributed. Of course, contests in politics are often more complicated. For example, US presidential primaries and caucuses are typically three- or four-way competitions, at least in the early stages. The same is true of British electoral contests, where Labour, Conservative and Liberal Democrat candidates vie for success in parliamentary constituencies. Moreover, there are not always clear 'winners' and 'losers': outcomes are often 'tied.' Paradoxically, this idea often holds in majoritarian electoral systems where there is *one* winner and all others, in effect, lose 'equally': for example, when candidates run for President of the France. An economic version of this holds when firms bid for public works contracts. With these cases in mind, it would be helpful to think more generally about the Bradley-Terry model, perhaps facilitating the estimation of abilities when contests have several players and when ties are possible.

The third essay (Chapter 4) claimed that, when the number of variables to consider was greater than the number of parameters to be estimated, one helpful way to proceed was the employment of a 'random forests' classification procedure. King (1989, 3) makes the case that political science "outdistances most other social sciences" in terms of data collection. For data *exploration* purposes then, the model matrices are potentially very large. Moreover, translation from 'theory' to statistical model is typically not direct in political science: for example, a scholar may believe that 'corruption' has some systematic effect on citizen-to-citizen trust in a city , or that 'more proportional' electoral systems give rise to fiercer *intra*-party competition. In such cases, there are presumably several indicators and ways of measuring the attributes of the independent variables and possibly no way to *a priori* decide between them. Since non-parametric techniques allow one to include literally hundreds more predictors than observations, these operationalization decisions can be refined within the (first) estimation 'round' rather than on some *ad hoc* basis with knock-on effects for the study as a whole. As data becomes increasingly available in massive amounts—

especially via the internet—such work would seem important and valuable. Of course, non-parametric fitting should not be considered the *final* stage in any investigation of political actor behavior: nonetheless, it can help us think more clearly about possible micro-mechanisms.

All told then, this dissertation has argued that our statistical solutions should fit our problems, and has suggested several realms in which either current practices might be improved, or where new profitable avenues of investigation exist. It would remiss of the author, however, to fail to address a criticism of quantitative techniques that arises particularly in political methodology. Phrased in various ways, typically implicit and casually suggested, the claim of some is that methodologists fetishize complexity in their models; that is, much of political methodology is 'complicated for the sake of being complicated.' This argument is imprecise, incorrect and unfair. It is *imprecise* insofar as it can be applied—in its most basic form—to *any* data analysis technique: logistic regression is 'more complicated' than OLS, which is 'more complicated' than cross-tabs, which are 'more complicated' than frequency summaries. Otherwise put, it is unclear as to where the analyst should draw the line: what extra utility must be garnered in order to approve of a more complex technique? Moreover, such standards clearly differ from scholar to scholar: reversible-jump Markov chain monte carlo approaches might seem unaccessible to someone who know little of Bayesian methods, but are a helpful and important extension for an analyst already familiar with MCMC techniques. The claim is *incorrect* because it fundamentally misunderstands the goal of political methodology. The aim is to provide tools that help us understand political behavior; to the extent that progress towards this goal remains incomplete (few if any would disagree with this contention), 'more' or 'better' theory, methods and data are the way forward. It can surely never *hurt* to refine our techniques and our understanding of their capabilities. Note also that, in computational terms, the costs involved with 'complicated' techniques have fallen precipitously over time and will continue to do so. Finally, the claim is *unfair* because it cynically assumes that methodologists either deliberately dissemble (and are thus dishonest in describing) their work, else they are simply incompetent in solving the problems that the discipline presents. Yet, one of the strengths of work in methodology is that assumptions of models and techniques are both clearly laid out, and written in a language common to its practitioners. It is certainly in this spirit that this dissertation is intended as a contribution to the discipline.

Bibliography

Achen, Christopher H. 2002. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Sience* 5:423–50.

Agresti, Alan. 2002. *Categorical Data Analysis*. 2 ed. New York: John Wiley and Sons.

Alesina, Alberto and Allan Drazen. 1991. "Why are Stabilizations Delayed?" *American Economic Review* 81(5):1170–1188.

Amadae, S. M. and Bruce Bueno de Mesquita. 1999. "The Rochester School: The Origins of Positive Political Theory." *Annual Review of Political Science* 2:269–295.

Aristotle. 322BC/1991. *Rhetoric*. New York: Oxford University Press (translation by George A. Kennedy).

Bachrach, Peter and Morton Baratz. 1962. "Two Faces of Power." *American Political Science Review* 56:947–952.

Banzhaf, John F. 1965. "Weighted Voting Doesn't Work: A Mathematical Analysis." *Rutgers Law Review* 19:317–343.

Barry, Brian. 1991. Is it Better to be Powerful or Lucky? In *Democracy and Power*. Oxford: Clarendon.

Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–36.

Benedetto, Giacomo and Simon Hix. 2007. "The Rejected, the Ejected, and the Dejected: Explaining Government Rebels in the 2001-2005 British House of Commons." *Comparative Political Studies* 40:755–781.

Black, Duncan. 1948. "On the Rationale of Group Decision-making." *Journal of Political Economy* 56:23–34.

Brace, Paul. 1984. "Progressive Ambition in the House: A Probabilistic Approach." *The Journal of Politics* 46(2):556–571.

Bradley, R A and M E Terry. 1952. "Rank Analysis of Incomplete Block Designs I: The Method of Paired Comparisons." *Biometrika* 39:324–345.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks/Cole.

Breiman, Leo. 1996. "Bagging predictors." *Machine Learning* 24(2):123–140.

Breiman, Leo. 2001. "Random forests." *Machine Learning* 45:5–32.

Brooks, Stephen P. 1997. "Markov chain Monte Carlo Method and its applications." *The Statistician* 47(1):67–100.

Budge, Ian. 1999. *Critical Elections: British Parties and Voters in Long-term Perspective*. London: Sage Publications.

Burke, Edmumd. 1774/1975. Speech to the Electors of Bristol. In *Edmund Burke: On Government, Politics and Society*, ed. B.W. Hill. London: Fontana.

Burnham, Gilbert, Riyadh Lafta, Shannon Doocy and Les Roberts. 2006. "Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey." *The Lancet (online)* 10.1016/(S0140-6736(06)69491-9).

Butler, David and Donald Stokes. 1974. *Political Change in Britain*. 2 ed. Macmillan.

Cain, Bruce E., John A. Ferejohn and Morris P. Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Cambridge, MA: Harvard University Press.

Canes-Wrone, Brandice, David W. Brady and John F. Cogan. 2002. "Out of Step, Out of Offce: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1).

Carlin, Bradley, Alan Gelfand and Adrian Smith. 1992. "Hierarchical Bayesian Analysis of Changepoint Problems." *Applied Statistics* 41(2):389–405.

Chib, Siddhartha. 1998. "Estimation and comparison of multiple change point models." *Journal of Econometrics* 86:221–241.

Clinton, Joshua, Simon Jackman and Doug Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98.

Cowan, Katherine. 2007. "Living Together: British Attitudes to Lesbian and Gay People." *Stonewall* Poll by YouGov.

Cowley, Philip. 2002. *Revolts and Rebellions: Parliamentary Voting Under Blair*. London: Politico's Press.

Cowley, Philip. 2005. *The Rebels: How Blair Mislaid His Majority*. London: Politico's Publishing.

Cowley, Philip. 2006. "Strengthening the role of the backbencher." *Written submission to the Select Committee on Modernisation of the House of Commons* Session 2006–07, no. 1.

Cox, Gary W. and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.

Crewe, Ivor. 1983. The Electorate: Partisan Dealignment 10 Years On. In *Change in British Politics*, ed. Hugh Berrington. London: Frank Cass pp. 183–215.

Dardagan, Hamit and John Sloboda. 2006. "Iraq Body Count.".
URL: *http://www.iraqbodycount.net/*

de Marchi, Scott, Christopher Gelpi and Jeffrey D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(2):371–378.

Denver, David and Gordon Hands. 1997. *Modern Constituency Electioneering*. Portland, Oregon: Frank Cass.

Dowding, Keith. 1996. *Power.* Minneapolis: University of Minnesota Press.

Downs, Anthony. 1987. *An Economic Theory of Democracy.* New York: Harper.

Efron, Bradley and Robert Tibshirani. 1993. *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Felsenthal, Dan S and Moshé Machover. 1998. *The Measurement of Voting Power: Theory and Practice, Problems and Paradoxes.* Cheltenham: Edward Elgar.

Felsenthal, Dan S. and Moshé Machover. 2004. "A Priori Voting Power: What Is It All About?" *Political Studies Review* 2:1–23.

Fenno, Richard F. 1973. *Congressmen in Committees.* Boston, MA: Little, Brown and Company.

Filson, Darren and Suzanne Werner. 2004. "Bargaining and Fighting: The Impact of Regime Type on War Onset, Duration, and Outcomes." *American Journal of Political Science* 48(2):296–313.

Fiorina, Morris. 1974. *Representatives, Roll Calls and Constituencies.* Lexington, MA: Lexington Books.

Firth, David. 2005. "Bradley-Terry Models in `R`." *Journal of Statistical Software* 12.

Firth, David and Arthur Spirling. 2006. "tapiR and the Public Whip: Resources for Westminster Voting." *The Political Methodologist* 14(2).

Fowler, James H. 2006. "Connecting the Congress: A Study of Cosponsorship Networks." *Political Analysis* 14(4):456–487.

Garrett, Geoffrey and George Tsebelis. 1999. "Why Resist the Temptation to Apply Power Indices to the EU." *Journal of Theoretical Politics* 11:291–308.

Gill, Jeff. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach*. New York: Chapman & Hall/CRC.

Green, Peter J. 1995. "Reversible Jump MCMC computation and Bayesian model determination." *Biometrika* 82(3):711–732.

Groseclose, Tim and Charles Stewart. 1998. "The Value of Committee Seats in the House 1947–1991." *American Journal of Political Science* 42(2):453–474.

Hastie, David. 2005. Towards Automatic Reversible Jump Markov Chain Monte Carlo PhD thesis University of Bristol.

Holler, Manfred and Mika Widgren. 1999. "Why Power Indices for Assessing European Union Decision-Making?" *Journal of Theoretical Politics* 11:321–330.

Huitt, Ralph K. 1957. "The Morse Committee Assignment Controversy: A Study in Senate Norms." *American Political Science Review* 51(2):313–329.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7:483–505.

Jacobson, Gary. 1989. "Strategic Politicians and the Dynamics of House Elections, 1946-86." *American Political Science Review* 83:773–93.

Janis, Irving L. 1972. *Victims of Groupthink*. Boston: Houghton Mifflin Company.

Johnson, Catherine and Gemma Rosenblatt. 2007. "Do MPs have the 'Right Stuff'?" *Parliamentary Affairs* 60:164–169.

Kam, Christopher. 2006. "Demotion and Dissent in the Canadian Liberal Party." *British Journal of Political Science* 36:561–74.

Kam, Christopher. 2007. *On Division: Cohesion, Discipline, and Intra-party Politics*. University of British Columbia: Book manuscript under review.

Kass, Robert E. and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90(430):773–795.

Kelso, Alexandra. 2007. "The House of Commons Modernisation Committee: Who Needs It?" *British Journal of Politics and International Relations* 9:138–157.

Key, V. O. 1949. *Southern Politics in State and Nation*. New York: Knopf.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press.

King, Gary. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization* 55(2):497–507.

Krehbiel, Keith. 1991. *Information and Legislative Organization.* Ann Arbor: University of Michigan Press.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking.* Chicago: University of Chicago Press.

Lane, Jan-Erik and Sven Berg. 1999. "Relevance of Voting Power." *Journal of Theoretical Politics* 11:309–320.

Lasswell, Harold. 1936. *Who Gets What, When and How.* New York: Whittlesey House.

Leech, Dennis. 2002*a*. "Computation of Power Indices." *Warwick Economic Research Papers* (644).

Leech, Dennis. 2002*b*. "An Empirical Comparison of the Performance of Classical Power Indices." *Political Studies* 50(1):1–22.

Leonte, Daniela, David J. Nott and William T. M. Dunsmuir. 2003. "Smoothing and Change Point Detection for Gamma Ray Count Data." *Mathematical Geology* 35(2).

Liaw, Andy and Matthew Wiener. 2002. "Classification and Regression by randomForest." *R News* 2(3).

Marx, Karl. 1967. *Das Kapital.* New York: International Publishers. English Translation.

Mayhew, David R. 1974. *Congress : the Electoral Connection.* New Haven: Yale University Press.

McCarty, Nolan, Keith Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, MA: MIT Press.

McLean, Iain, Arthur Spirling and Meg Russell. 2003. "None of the Above: The UK House of Commons Votes on Reforming the House of Lords, February 2003." *The Political Quarterly* 74:298–310.

Norris, Pippa. 2001. *New Labour and Public Opinion: The Third Way as Centrism?* London: Macmillan pp. 32–44.

Norton, Philip and David M. Wood. 1993. *Back from Westminster: British Members of Parliament and their Constituents.* Lexington: University of Kentucky Press.

Paltzelt, W. 2000. What Can an Individual MP Do in German Parliamentary Politics. In *The Uneasy Relationship Between Parliamentary Members and Leaders*. London: Frank Cass.

Park, Jong Hee. 2006. "Modeling Structural Changes: Bayesian Estimation of Multiple Changepoint Models and State Space Models." *Prepared for American Political Science Association Meeting* August, 2006, Philadelphia.

Pattie, Charles, Edward Fieldhouse and R. J. Johnson. 1994. "The Price of Conscience: The Electoral Correlates and Consequences of Free Votes and Rebellions in the British House of Commons, 19897–92." *British Journal of Political Science* 24:359–80.

Pattie, Charles J., Ronald J. Johnston and Edward A. Fieldhouse. 1995. "Winning the Local Vote: The Effectiveness of Constituency Campaign Spending in Great Britain, 1983-1992." *American Political Science Review* 89(4):969–983.

Poole, Keith and Howard Rosenthal. 1997. *Congress: A Political Economic History of Roll Call Voting.* New York: Oxford University Press.

R Development Core Team. 2006. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Riker, William. 1962. *The Theory of Political Coalitions.* New Haven: Yale University Press.

Riker, William. 1964. "Some Ambiguities in the Notion of Power." *American Political Science Review* 58(2):341–349.

Riker, William. 1982. *Liberalism Against Populism.* San Francisco: W. H. Freeman.

Riker, William. 1986. *The Art of Political Manipulation.* New Haven: Yale University Press.

Ripley, Brian D. 1996. *Pattern Recognition and Neural Networks.* Cambridge: Cambridge University Press.

Rohde, David W. 1979. "Risk-Bearing and Progressive Ambition: The Case of Members of the United States House of Representatives." *American Journal of Political Science* 23(1):1–26.

Rohde, David W. 1991. *Parties and Leaders in the Postreform House.* Chicago: University of Chicago Press.

Ruger, Theodore W., Pauline T. Kim, Andrew D. Martin and Kevin M. Quinn. 2004. "Competing Approaches to Predicting Supreme Court Decision Making." *Perspectives on Politics Symposium* 2.

Salter, Lee. 2004. "Parliament and The Parliamentarians: The Worrying Case of the City of London (Ward Elections) Bill." *Political Quarterly* 75:155–167.

Searing, Donald. 1994. *Westminster's World: Understanding Political Roles.* Cambridge, MA: Harvard University Press.

Shapley, L.S. and Martin Shubik. 1954. "A Method for Evaluating the Distribution of Power in a Committee System." *American Political Science Review* 48:787–92.

Shepsle, Kenneth. 1979. "Institutional arrangements and equilibrium in multidimensional voting models." *American Journal of Political Science* 23:27–60.

Shepsle, Kenneth and Barry Weingast. 1987. "The institutional foundations of committee power." *American Political Science Review* 81:85–104.

Spirling, Arthur and Iain McLean. 2007. "UK OC OK? Interpreting Optimal Classification Scores for the United Kingdom House of Commons." *Political Analysis* 15(1).

Springall, A. 1973. "Response surface fitting using a generalization of the Bradley-Terry paired comparisons model." *Applied Statistics* 22:59–68.

Stam, Allan C. and D. Scott Bennett. 2006. "Predicting the Length of the 2003 U.S.-Iraq War." *Foreign Policy Analysis* 2(2):101–116.

Stigler, S. 1994. "Citation Patterns in the Journals of Statistics and Probability." *Statistical Science* 9:94–108.

Stuart-Fox, Devi M., David Firth, Adnan Moussalli and Martin J. Whiting. 2006. "Multiple Signals in Chameleon Contests: Designing and Analysing Animal Contests as a Tournament." *Animal Behaviour* 71:1263–1271.

The Economist. 2006. "Indoor rainforests: Rejecting Eden." *Print Edition* Jul 13.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton, N.J.: Princeton University Press.

Tversky, Amos and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211:453–458.

Venables, William and Brian Ripley. 2002. *Modern Applied Statistics with S*. New York, New York: Springer.

Wawro, Gregory. 2000. *Legislative Entrepreneurship in the US House of Representatives*. Ann Arbor: University of Michigan Press.

Western, Bruce and Meredith Kleykamp. 2004. “A Bayesian Change Point Model for Historical Time Series Analysis.” *Political Analysis* 12:354–374.

Winston Churchill, Sir. 1954. “Duties of a Member of Parliament.” *Parliamentary Affairs* 8.