Modeling Interdependence in Collective Political Decision Making

by
Bruce A. Desmarais

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Political Science.

Chapel Hill
2010

Approved by:

Thomas Carsey, Advisor

James Stimson, Reader

Skyler Cranmer, Reader

Isaac Unah, Reader

Kevin McGuire, Reader

Abstract

BRUCE A. DESMARAIS: Modeling Interdependence in Collective Political Decision Making.
(Under the direction of Thomas Carsey.)

Fundamental to many accounts of decision-making within political institutions is the interdependence between simultaneous choices. For instance, members in a legislature are hypothesized to take voting cues from party leaders, and the chief justice of the U.S. Supreme Court is thought to vote with the majority on the merits so as to assign opinion authorship. In this thesis I show that none of the conventional methods that have been used by political scientists for testing theories of simultaneous interdependence are statistically sound. I then propose a machine-learning algorithm that finds unmodeled interdependence in discrete-choice data. Next, I develop a novel statistical model that allows the researcher to explain – in a methodologically appropriate manner – the probability that an actor makes a particular choice as well as the probability that a collective-decision occurs in a particular form. In the last chapter of my dissertation, I demonstrate that U.S. Supreme Court case outcomes are interdependent and that the U.S. Supreme Court is best characterized as an institution striving to produce an ideologically optimal body of law rather than ideologically optimal independent case outcomes.

I dedicate this work to my wife, my lifesaver, Rebecca; and also to my parents, Bruce and Lisa.

Table of Contents

Abstract iii

List of Figures vii

List of Tables viii

1 Lessons in Disguise: Multivariate Predictive Mistakes in the study of Repeated Collective Choice 1

1.1 Repeated Collective Decisions 1

1.2 Information, Data and Repeated Collective Choice 5

1.3 Iterative Model Improvement Through Prediction Error Analysis 7

1.3.1 Evaluating the Updated Model 15

1.4 Replications with JPE-Suggested Extensions 17

1.4.1 The U.S. Supreme Court and Oral Argument Quality 17

1.4.2 The Reliability of Democratic Allies 26

1.5 Conclusion 33

2 The Exponential Random Configuration Model 36

2.1 Introduction 36

2.2 Theorizing About Discrete Configurations 39

2.3 A General Model for Discrete Configurations 42

2.3.1 Measurement and Configurational Hypotheses 43

2.3.2 A Distribution Over Discrete Configurations 46
2.3.3 Interpretation of the ERCM 48
2.3.4 Estimation of the ERCM 50
2.3.5 Predictive, Semi-Parametric Fit Comparison of ERCMs 54
2.4 Jurisprudential Regimes in the U.S. Supreme Court 57
2.4.1 Richards and Kritzer (2002) 58
2.4.2 Configurational Theories of Supreme Court Decision-Making . . 60
2.4.3 Results 64
2.5 Conclusion 67
2.6 Logit - A Special Case of ERCM 69

3 The U.S. Supreme Court as a Self-Correcting Institution 71

3.1 Introduction 71
3.2 Self-Correction on the U.S. Supreme Court 73
3.2.1 The Objective of Influence 73
3.2.2 Response to Supreme Court Decisions 75
3.3 The Process of Self-Correction 78
3.4 The Empirical Test 80
3.4.1 Dependent Variables 80
3.4.2 The Liberalism of the Body of Law 82
3.4.3 Control Variables and Estimation Strategy 84
3.4.4 Results 86
3.5 Conclusion 91

Bibliography 93

List of Figures

1.1 Posterior Predictive Distribution Sampling Algorithm 11

1.2 Joint Prediction Errors on the U.S. Supreme Court 19

1.3 U.S. Supreme Court Vote Predictions 25

1.4 Under-predictions from the model in Gartzke and Gleditsch (2004) . . . 29

2.1 MCMC-MLE Estimation Algorithm 52

2.2 Distributional Tendencies of Supreme Court Decisions 66

3.1 Descriptive Plots 85

List of Tables

1.1 U.S. Supreme Court Justices' Votes on the Merits 22

1.2 The Reliability of Allies and International Conflict 32

2.1 Alternative Models of U.S. Supreme Court Voting 63

3.1 Term-Level Liberalism: all Cases Included 87

3.2 Term-Level Liberalism in Criminal Procedure Cases 89

3.3 Term-Level Liberalism in Civil Rights and Liberties Cases 90

3.4 Term-Level Liberalism in Economic Activity Cases 91

Chapter 1

Lessons in Disguise: Multivariate Predictive Mistakes in the study of Repeated Collective Choice

"The applied statistician should avoid models that are contradicted by the data in relevant ways - frequency calculations for hypothetical replications can monitor a model's adequacy and help to suggest more appropriate models."

-Donald B. Rubin (1984)

1.1 Repeated Collective Decisions

Central processes studied in every field of empirical political science arise in the form of repeated collective decisions.. Roll-call votes in legislatures and decisions issued by multi-member courts of appeal are stable groups of political actors issuing individual decisions that are aggregated into salient collective outcomes. In the international arena, intervention into civil wars, the provision of relief for natural disasters, and the issuance of trade sanctions are interdependent decisions rendered repeatedly by a stable group of states. Due to the importance of the collective outcomes that result from individual decisions (e.g. laws written or the results of civil wars), and the fact that actors have multiple opportunities to learn optimal strategies for interaction, patterns of dependence or relationships are likely to emerge in repeated collective choice data.[1] However, in political science applications, we often pool the members of the stable group into a sample for regression modeling where relationships between the members are ignored to the extent that they are not captured by independent variables. If such relationships do exist (e.g. if members of the U.S. House of Representatives learn over time to take cues from certain policy specialists or party leaders, and issue roll-call votes in the same direction as these leaders), statistical inferences from pooled regression are subject to misspecification bias. Since the data contain repeated observations of collective behavior, it can be used to learn about interdependence among the actors. I propose an iterative method for learning about and modeling these dependencies. Similar in structure to the approach advocated by Achen (2005), rather than estimate an overly complicated model at the onset, I suggest specifying a simple model to start, and updating it to address predictive deficiencies, subjecting the updated model to rigorous conservative tests of the validity of those updates.

In many instances of repeated collective decision making, the salient collective outcome (e.g. the result of a court case or the passage of legislation) is a deterministic function of the individual decisions rendered by the members of the group (e.g. a lower court decision is reversed by the U.S. Supreme Court if five or more justices vote with the appellant). Because of this deterministic relationship between micro and macro-level outcomes, if a model is fit to the individual decisions in the collective, a model

[1] Many scholars have noted that patterns of sophisticated rational interaction are likely to emerge when collective choice situations are repeated many times, and actors can learn the rules and payoffs of the game (see e.g. Verba (1961) and Ostrom (1998)).

is automatically implied for the collective outcome.[2] For instance, if a model is fit to U.S. Supreme Court Justice's vote on the merits – estimating the probability that each individual justice will vote with the appellant in a case – the probability that the *Court* decides in favor of the appellant is given by the sum of the joint probabilities of all group configurations in which at least five of the nine justices vote with the appellant.

Due to the deterministic relationship between the micro and macro-level models of collective decisions, from the standpoint of probability theory, it is inconsistent to specify separate statistical models for individual decisions and higher-order configurations of those decisions – the former implies the latter. The critical implication of the micro-macro connection is that, in order to be correctly specified, the micro-level model must capture any tendency for individual decisions to produce sophisticated/intentional higher-order configurations. For instance, Hix, Noury and Roland (2005) find that there are varying levels of political party cohesion in the European Parliament. If the findings of Hix, Noury and Roland (2005) are valid, any micro-level analysis of roll-call voting in the European Parliament is misspecified if it does not account for a varying tendency towards intra-party cohesion in members' votes. Extending an individual-level decision model – often logistic regression where observations are assumed to be independent conditional on the covariates – to allow for flexible forms of interdependence commonly requires non-trivial and at times prohibitive computational effort to estimate the dependence parameters.[3] Rather than attempt to extend a micro-level model to accommodate every configurational tendency that has either found support in the

---

[2]It is important to note that in making this statement I am assuming that the analyst is working within either the likelihood or Bayesian estimation frameworks, or any other method used to fit a full parametric distribution to the data.

[3]See e.g. Alvarez and Nagler (1998) for an example where preferences for electoral candidates are posited to be correlated, Franzese and Hays (2007) for a discussion of the estimation challenges in accounting for spatial dependence in time-series cross-section data, and Ward, Siverson and Cao (2007) who find that latent reciprocal and transitive tendencies characterize international dyadic data

literature or can be reasonably conceived, I develop a method to identify configurations that are missed by a simple micro-level model – the analysis of which suggests the key configurational extensions necessary to make valid inference on the micro-level processes. The key mechanism underlying this procedure is a method of residual analysis that is particularly suited for repeated collective choice data.

The systematic analysis of residuals from regression models has long been used to monitor aspects of statistical fit with the goal of improving specification (see e.g. Cox and Snell (1971), Achen (1977), Beck (1982) and Achen (2005)). Generally speaking, residual analysis involves the comparison of observed data with predicted values from a statistical model. There are at least two major general challenges in residual analysis. First, with the generic goal of assessing the proximity of observed and predicted quantities, the particular function(s) of the observed and predicted data to be compared (e.g. expected value, variance, correlation, skewness etc.) must be insightfully chosen. Second, the analyst must specify the level of divergence between the true and predicted quantity that constitutes an interesting or potentially important deficiency. I introduce the concept of a *joint prediction error* (JPE), which is a collective outcome that is observed to occur with a much different frequency than predicted by a given statistical model, and provide benchmarks for deciding what constitutes a JPE. In doing so, I overcome a particular challenge that arises in the analysis of joint residuals. Given n members of the collective, and interest in finding JPEs composed of k members, $\binom{n}{k}$ groups need to be considered, which can become a giant number for realistic size collectives and even small k. For instance, if one is interested in monitoring predictive accuracy of a model predicting legislative activity on all possible groupings of five legislators in a 435 member chamber, there are $126,837,202,212$ groups to check, and if k is increased to ten the number of groupings is *multiplied* by 474,925,189. I show that algorithms common in the machine-learning literature, designed to find frequent joint occurrences in databases of millions of commercial transactions, can be used to efficiently search over all possible combinations of actors in the collective.

Finally, by replicating and extending two recently published studies, I demonstrate how improvements in models of repeated collective discrete choice processes can be discovered through the analysis of JPEs. I find that a logit model explaining Supreme Court votes on the merits published by Johnson, Wahlbeck and Spriggs (2006) critically understates the degree of case-level consensus on the Court. This observation leads to an improved model specification that accounts for correlation between the justices and includes additional important case-level covariates, and the updated model lends stronger support to one of the central theoretical claims in the original article. In Gartzke and Gleditsch (2004), a study on international defense alliance activation, the empirical model understates association among a state's allies. Additionally, in the defense alliance application, a pattern emerges in the JPEs which suggests that states with greater consultation obligations are less likely to enter a war in defense of their allies. Adding a measure of a state's consultation obligations to the model in Gartzke and Gleditsch (2004) (1) supports the insight that states with more consultation pacts are less likely to support their allies and (2) suggests that the original central empirical finding of the article – that democratic states are less likely to assist their allies – resulted from the omission of consultation obligation. In both replications, the published statistical analyses are improved by extensions suggested by the JPEs. Improvement is verified through multiple model fit metrics.

1.2 Information, Data and Repeated Collective Choice

There are two reasons in particular to expect theoretical innovations to arise through the inspection of joint prediction errors in the study of repeated collective choice. These are consequences of the fact that most collective choice modeling in political science involves an intense focus on the numerous interactions of a relatively small set of very well-known actors. First, simple labels – country names, legislative districts, justice names, etc.– on the actors in the dataset communicate information to the analyst above and beyond that which is contained in the rows and columns of the dataset. Second, there is likely to be an overwhelming amount of previous theoretical and empirical research that precedes any new study of historical political data. Both of these features present unique opportunities for improvement with joint prediction error analysis.

In their analysis of the representational efficacy of majority-minority Congressional districts, Cameron, Epstein and O'Halloran (1996, pp. 810) state, "In many southern state legislatures, [minority group leaders and Republicans] formed voting blocs when passing redistricting plans, and the [U.S.] Justice Department under Republican presidents was eager to create the maximum possible number of majority-minority districts." This represents rich information about the process under study – the motivations underlying the formation of majority-minority districts – yet no data or citation to outside work is provided. It is knowledge held by the authors, the validity of which was accepted at face-value by reviewers at the *American Political Science Review*. Anyone who has presented at a conference, and been confronted with the one case (e.g. legislator, country, year) that represents the perfect counter-factual to his or her theory, knows that political scientists have auxiliary expertise – constituting information about the observations above and beyond that which appears in the regression equation. If a scholar of civil war intervention runs a logistic regression model on the intervention decisions of states, he or she may recognize that the model poorly predicts outcomes in which developed states decide to intervene and others do not without collecting additional data about countries. Such a recognition would serve as motivation to collect and include in the model a measure of a state's development. This auxiliary information optimizes potential benefits from simply examining those combinations of cases that are poorly predicted by a given statistical model.

The second consequence of multiple studies of familiar observations is that the discipline accumulates a predictable set of control variables that are considered potentially serious omissions if left out of a model. For most salient topics in political science, dozens of studies precede any new research. Most of these studies propose partially unique explanations of a process and, thus, provide candidate control variables for anyone who endeavors to model the same or similar data in the future. It is uncommon and practically infeasible for one to include every variable that has ever been found to significantly influence a process in a new analysis. Indeed, such a model would likely lead to a convoluted interpretation, and be counter to the objective of data reduction (Achen, 2005). At the same time, previous findings cannot be ignored simply for the sake of time or parsimony. Examining joint prediction errors constitutes a reasonable compromise between ignoring past work outright and including the entire preceding empirical literature in an initial model. Knowledge about the approximate values of the omitted factors can be checked for consistency with patterns in the JPEs. For instance, judicial scholars are familiar with the seniority ranking of justices on the U.S. Supreme Court. Analysis of joint errors from a model of Supreme Court voting would reveal whether justices close in seniority were voting similarly, and, thus, whether seniority should be added to the model.

1.3 Iterative Model Improvement Through Prediction Error Analysis

The process I prescribe for developing the best statistical model of repeated collective choice data rests on the observation of Rubin (1984) that frequency calculations performed on the real data should not differ from model predictions in relevant ways. Once a model has been fit to data and is treated as the best possible model, it assumes the position of the analyst's null or assumed model. Quantities in the observed data that differ considerably from predictions drawn from the model serve as evidence against the null. When one fits a paremetric model to data, it is not only assumed that the regression function is properly specified, but also that the structure of the variance, association between the observations, and/or any other quantity that can be computed on the data is correctly captured by the model. The intuition provided by Rubin (1984) is that, if a model gives the correct distribution for the data, then it should not be possible to find distributional qualities of the data that contradict predictions derived from the model.

There are five stages in one iteration of the model-fitting procedure I advocate:

1. Fit the model (M) that represents the best specification the analyst can currently manage.

2. Draw many hypothetical datasets according to the probability distribution of the data implied by the model.[4]

3. Identify joint prediction errors by finding combinations of outcomes that occur with much greater or lesser frequency in the observed data than in the simulated data predicted from the model.

4. Update the model to accommodate deficiencies that are *hypothesized* to produce the prediction errors.

5. Assess, using model-fit metrics that favor a parsimonious specification, whether the updated specification provides a better fit to the data than the model estimated in step 1.

[4]It is possible that in simple cases the analytic distribution of the data will be available, but to assure the algorithm is applicable when it is not available, I advocate simulation.

This process can be repeated indefinitely, or until the analyst has no more intuition about the deficiencies creating the prediction errors. For those wary of purely data-driven procedures for model construction, it is critical to recognize the role of theory in the fourth step. Without a thorough theoretical understanding of the process under study, it will not be possible to recognize the significance of the JPE membership. For instance, a Congress scholar may recognize – through inspection of the JPE memberships – that a model of roll-call voting in the U.S. House poorly explains votes in which members on the Appropriations Committee disagree with those on the Budget Committee. Without at least a loose recollection of committee membership in the House, it would not be possible to even recognize, never-mind explain, such a pattern. Of course, any data-driven model-fitting procedure must guard against over-fitting the sample data. This is what step 5 adresses. After presenting the algorithm used to identify JPEs, I present a model fit metric that can be used to avoid over-fitting.

The specific metric used to determine whether a joint outcome constitutes a JPE is the *posterior predictive p-value* introduced by Meng (1994) for general use in a Bayesian context. This p-value can be used to asses the oddity of the frequency of a joint outcome given predictions from a model. It would allow one to state whether, for instance, the frequency of unanimous decisions on the U.S. Supreme Court is statistically significantly different from the frequency of unanimous decisions predicted by a statistical model. In the next few paragraphs, I review in detail, the construction of a predictive p-value.

In a Bayesian analysis, the prior distribution of the parameters ($\pi(\theta)$) represents the analyst's belief about the parameters *prior* to using the observed data (X). The posterior distribution of the parameters ($p(\theta|X)$) is the resulting belief regarding the distribution of the parameters after updating the prior distribution with the observed data. In a Bayesian analysis, point estimates are equal to the means of the posterior distribution, and credible intervals – the Bayesian analog to the frequentist confidence interval – are derived from the quantiles of the posterior distribution (Gill, 2002). The posterior distribution conditional on the observed data X is given by

$$p(\theta|X) = \frac{l(X|\theta)\pi(\theta)}{t_{\Theta} l(X|\theta)\pi(\theta)d\theta}, \tag{1.1}$$

where $l(X|\theta)$ is the likelihood function of the data given θ. If M is fit by maximum likelihood, the asymptotic sampling distribution of θ is used as an approximation of the posterior distribution of θ (King, Tomz and Wittenberg, 2000; Tomz, Wittenberg and King, 2003), which is multivariate normal with mean vector equal to the parameter estimates ($\hat{\theta}$) and covariance matrix equal to the variance-covariance matrix of $\hat{\theta}$ ($\hat{\Sigma}$). The posterior predictive distribution (PPD) of X is the expected distribution of future replicates of X. It represents the analyst's belief about the distribution of X after updating with the available data (e.g. the expected distribution of justice-votes given by the independent variables and regression coefficients in a model of voting on the U.S. Supreme Court). The posterior predictive distribution $f(X_{new})$ of the data is computed by averaging the likelihood function over $p(\cdot|\theta)$, and is given by

$$f(X_{new}) = \int_{\Theta} l(X_{new}|\theta)p(\theta|X)d\theta. \tag{1.2}$$

In practice, $p(\theta|X)$ and/or $f_X(\cdot)$ are often not available in closed form due to intractability of the integrals in equations 1.1 and 1.2. In the typical Bayesian analysis, using Markov Chain Monte Carlo (MCMC) methods, the researcher has a large sample from $p(\theta|X)$ rather than a formula for the posterior distribution. For example, in a regression model with five predictors, if the MCMC algorithm was run for 10,000 iterations after an initial burn-in period, instead of a formula for $p(\theta|X)$, the analyst would have a 10,000×5 matrix of regression coefficeints. In order to simulate from our model, with the objective of comparing simulated and observed data, we need to use

Figure 1.1: Posterior Predictive Distribution Sampling Algorithm

t = number of desired draws from $f_X(\cdot)$ (the PPD)
$\hat{\boldsymbol{\theta}}$ = D×P MCMC sample
X = N×M sample of observed data
$\hat{\boldsymbol{X}}$ = Sample from the PPD initialized to $\emptyset$

for(i in 1 to t) begin
1. Draw $\theta^{(i)}$ randomly from the rows of $\hat{\boldsymbol{\theta}}$
2. Draw $X_{new}^{(i)}$ (the same size as X) from $l(X_{new}|\theta^{(i)})$
3. Store $X_{new}^{(i)}$ in $\hat{\boldsymbol{X}}$
end

$\hat{\boldsymbol{X}}$ now contains t random draws from $f_X(\cdot)$

this simulated approximation to $p(\theta|X)$ to approximate $f_X(\cdot)$. The algorithm given in figure 1.3 can be used to draw from the posterior predictive distribution using the MCMC sample. When M is fit by maximum likelihood, the sample from the posterior distribution derived through MCMC is replaced with a random sample from the asymptotic sampling distribution.

P-values are commonly used in political science to measure the plausibility of some null parameter (e.g. population mean, difference in means of two populations, the regression coefficient in the population, the variance etc.) given an observed sample counterpart of that parameter (i.e. statistic) and additional assumptions about the data-generating process. Suppose it is of interest to assess the oddity or rarity of the observed value of some statistic computed on the data ($T(X)$) given an assumption about the distribution that generates X ($f(X)$). If it is possible to derive the distribution of $T(X)$ given $f(X)$ ($g(T(X))$) (i.e. the sampling distribution in a classical context), the placement of $T(X)$ on $g(T(X))$ can be used to estimate the area under $g(T(X))$ to the right (left) of a comparatively high (low) value of $T(X)$ to derive a p-value. In the familiar regression framework, with a large sample, the regression coefficient - $T(X)$ - has a normal sampling distribution - $g(T(X))$ - with standard deviation equal to the standard error of the regression coefficient. From this we know that the regression coefficient is quite unlikely to be zero if it is at least two standard errors away from zero. In the current application, $T(X)$ is a joint prediction error, and the process I describe below, through the approximation of $g(T(X))$, can be used to asses whether the observed frequency of a joint outcome is significantly different than that predicted by a model.

A considerable challenge in many settings is that, unlike the example of regression coefficients given above, the analytic sampling distribution (i.e. $g(T(X))$) is not available in closed form for many combinations of $T(X)$ and $f(X)$. For instance, the analytic sampling distribution of the sample median is rarely available in closed form (Greene, 2008, pp. 597). Originally suggested by Rubin (1984), and thoroughly explored by Meng (1994), the posterior predictive p-value provides a general solution for determining the rarity of an observed value of $T(X)$ given a fully parametric specification of $f(X)$. If $T(\cdot)$ is computed on many draws of hypothetical data from M using the posterior predictive distribution, the empirical distribution of $T(X)$ over the draws of X ($h(T(X))$) can be used as a substitute for $g(T(X))$. As the number of draws of X from M approaches infinity, the tail area outside of $T(X)$ on $h(T(X))$ approaches a p-value for $T(X)$ given M as the null model.

In the context of joint prediction error analysis, let $T(X)$ be the number of times a multivariate outcome (Γ) occurs in the data. As a hypothetical example, a possible $T(X)$ is the number of times Barack Obama and Hillary Clinton voted in the same direction on roll-calls in the U.S. Senate. Using M as the null model, if Γ has a posterior predictive p-value less than a tunable parameter α, it is classified as a joint prediction error. Suppose that Obama and Clinton both issued votes in 100 roll-calls.

To determine what the model predicts regarding their joint behavior, we can simulate these 100 roll-calls 1000 times. Suppose that in 95% of the simulated 100 roll-calls, they voted together in less than 75 of the roll-calls. If, in the 100 actual roll-calls, they voted together in more than 75 of the votes, we would conclude that the model under-predicts agreement between Obama and Clinton at the 0.05 level of statistical significance.

To find joint prediction errors in a repeated collective choice dataset the p-value for every possible Γ must be computed. As noted earlier, the universe of possible Γs can be quite large. This poses a computational challenge in counting the frequency of Γ in both the real and simulated data for all Γs. Thankfully, this counting problem is very similar in structure to a challenge that has been considered in the machine learning literature for decades – counting, in databases of millions of commercial transactions for merchants offering thousands of products, the number of times product groupings occur in shopping baskets (e.g. the number of times a T.V. Guide, fishing pole and neck tie are all purchased together in transactions at Wal-Mart). *Frequent itemset mining* is the general term that encapsulates work on finding product groupings that meet certain criteria (Wen, 2004; Luo and Zhang, 2007). Treating the collective choice as the transaction, and the individual decisions made by the actors as the product occurrences, frequent itemset mining algorithms can be used to count the joint occurrence of individual decisions within collective choices. I take advantage of frequent itemset mining algorithms in the implementations of JPE analysis below. [5]

There are three parameters that must be set by the user of the algorithm outlined above: the size of the joint prediction errors (k), the number of draws from $f(X_{new})$ to

[5] Many of the algorithms available in the `R` package `arules` (Hahsler, Grn and Hornik, 2005) can be combined to efficiently implement JPE analysis in large datasets. I am developing an `R` package (`JPEMiner`), in which I wrap and structure a number of the algorithms in `arules` to efficiently perform JPE analysis after the estimation of many discrete choice models familiar to political scientists.

be used to compute the posterior predictive p-values (t), and the level of the p-value (α) at which to classify the joint outcome as a prediction error. As I will demonstrate through application later, a great deal of information is communicated in pairs. Pairs contain all of the available information about what outcomes occur together. For this reason, I suggest a default value of $k = 2$. It may be informative to move beyond k if particular higher order configurations are of interest. For instance, if one were interested in assessing whether a model accurately predicted intra-continental agreement in U.N. Security Council votes, it would be possible to look at the pairwise predicted versus observed agreements among all pairs within a continent, but it might be easier to compare the predicted and observed occurrences of continent-level consensus. The term α should be chosen to produce a manageable set of prediction errors – not so low that no prediction errors are discovered, and not so high that every joint outcome is considered a prediction error. Lastly, t should be set high (1,000–10,000) to start, and the analysis should be repeated two or three times to assure the results are not attributable to simulation error. If results differ across repetitions, t should be increase until variation across repetitions is negligible. These suggestions represent reasonable starting points for most applications, but should not be read as strict constraints on the values of the tuning parameters.

It is important to emphasize that discovering a pattern in the JPE analysis does not constitute rigorous statistical inference on the factors creating that pattern. The validation step comes *after* the model has been updated to account for patterns discovered in the JPE analysis. The objective in the JPE analysis stage is to tune the parameters (k, t, α) until either some intuition is reached regarding appropriate improvements to M or it is clear that no meaningful discrepancy between the data and the distribution implied by M can be found. The point is to push M to the breaking point in regards to its consistency with the data, with the intention of reconstructing a stronger model through a theoretical account of the prediction errors produced by M. The validation procedure presented next is used to judge the validity of the proposed updates to M.

1.3.1 Evaluating the Updated Model

As noted previously, observing a pattern in the joint residuals does not constitute statistical confirmation of that pattern as a component of the data generating process. Since the method of model improvement proposed here is fairly data-intensive, it is desirable to use a relatively conservative method of evaluating the fit improvement associated with the updates, so as to avoid over-fitting. The method I advocate is cross-validation.

Cross-validation avoids over fitting by evaluating the fit of a model on data that was not used to estimate the parameters of the model. The parameters of competing models are estimated on the training set, and the relative fit is judged using the validation set (data that was not used to estimate the model, but is considered to be drawn from the same population as the training set) (Jensen and Cohen, 2000). Leave-one-out cross-validation is a method of judging the predictive fit of a statistical model that provides predictions for the outcome under study. Similar in structure to the computation of Cook's D - the popular outlier identification statistic used in regression modeling (Cook, 1977) - in leave-one-out cross-validation every observation is iteratively used in the training and validation sets, and therefore does not require the analyst to arbitrarily exclude some of the data from estimation (Snee, 1977; Burman, 1989; Thall, Simon and Grier, 1992).

In order to implement cross-validation, a predictive measure of the fit of the model to the excluded observations must be identified. Many candidates have been considered including the cross-validated classification error for categorical outcomes (Leo et al., 1984) and the cross-validated squared error for continuous outcomes (Hjorth, 1993). A predictive measure that is particularly useful when the objective is to compare fully parametric models is the cross-validated log-likelihood. The cross-validated log-likelihood (CVLL) is computed by summing the log-likelihood of each observation given the parameters estimated on the rest of the data set (θ^{-i}) (Rust and Schmittlein, 1985; O'Sullivan, Yandell and Raynor, 1986; Verweij and Van Houwelingen, 1993; van Houwelingen et al., 2006). A very common metric of distance between two probability distributions is the Kullback-Leibler distance (Gelman, Meng and Stern, 1996; Clarke, 2003, 2007). In expectation, among a number of possible models, the model with the highest CVLL is that with the minimum Kullback-Leibler distance from the model that actually generated the data (the true model) (Cover and Thomas, 1991; Smyth, 2009). Thus, if the updates to M move the specification closer to the true model, then, on average, evaluation with the CVLL will indicate that the updates should be accepted. The formula for the CVLL is given by

$$CVLL = \sum_{i=1}^{N} \ln \left[l(x_i|\theta^{-i}) \right] . \tag{1.3}$$

The CVLL is extended to data that is organized hierarchically by clustering on a single level (e.g. court case) – a structure common in repeated collective choice data – by leaving out one cluster at a time and summing the log-likelihood of the left-out clusters rather than leaving out a single observation (Price, Nero and Gelman, 1996). To evaluate the fit of the various models specified in the current analysis, I compute the CVLL as well as the BIC, another conservative measure of model fit, in each of the applications below.

1.4 Replications with JPE-Suggested Extensions

1.4.1 The U.S. Supreme Court and Oral Argument Quality

Johnson, Wahlbeck and Spriggs (2006) test whether the quality of oral argument before the U.S. Supreme Court influences the votes of the justices. Justice Harry Blackmun graded the oral arguments of attorneys on an 8-point grading scale for cases argued before the Supreme Court from the 1970-1994 terms. Johnson, Wahlbeck and Spriggs (2006) specify a logistic regression model of votes (pooled over justices, cases and terms) where the dependent variable is coded 1 if the justice votes to reverse the lower court decision and 0 for affirm. The votes of Justice Blackmun are excluded due to concerns about endogeneity. A number of other control variables are included. See the original article for their justification.

Case-Level Prediction Errors

The collective choices made by the justices on the U.S. Supreme Court are case decisions. Each case is represented as a combination of justice-votes. On a typical case, there are eight justices (excluding Blackmun) who can each either vote to affirm or reverse, leading to $2^8 = 256$ possible eight-vote outcomes. The JPE analysis is performed on the full model specified in column 2 of table 3 in Johnson, Wahlbeck and Spriggs (2006). In the analysis I report I used $t = 5,000$ draws from the posterior predictive distribution of the data, a posterior-predictive p-value of $\alpha = 0.10$, and a prediction error size of $k = 2$ justice-votes.[6] Figure 1.4.1 gives the four most frequent over-predicted

[6] I repeated the analysis with three different simulated samples, and there was no variation in the set of prediction errors − leading me to conclude that the $t = 5,000$ is sufficiently large to avoid simulation error. Also, the substantive inferences I draw from the JPE analysis do not change for α as small as 0.05, and there is no utility in using a less restrictive p-value. Lastly, I looked at JPEs of size $k \in \{3, 4, 5\}$, but gathered no additional intuition regarding model improvement from the larger groups.

and under-predicted justice-vote pairs in the dataset. An under(over)-prediction is a pair that is predicted to occur less(more) frequently than it actually does. The left and right columns give under and over-predicted pairs respectively. Each panel is a histogram of the number of cases in which the justice-vote pair occurs in the 5,000 datasets drawn from the original model in Johnson, Wahlbeck and Spriggs (2006). The number of cases in which the pair occurs in the actual dataset is located at the solid vertical line in each panel. [7]

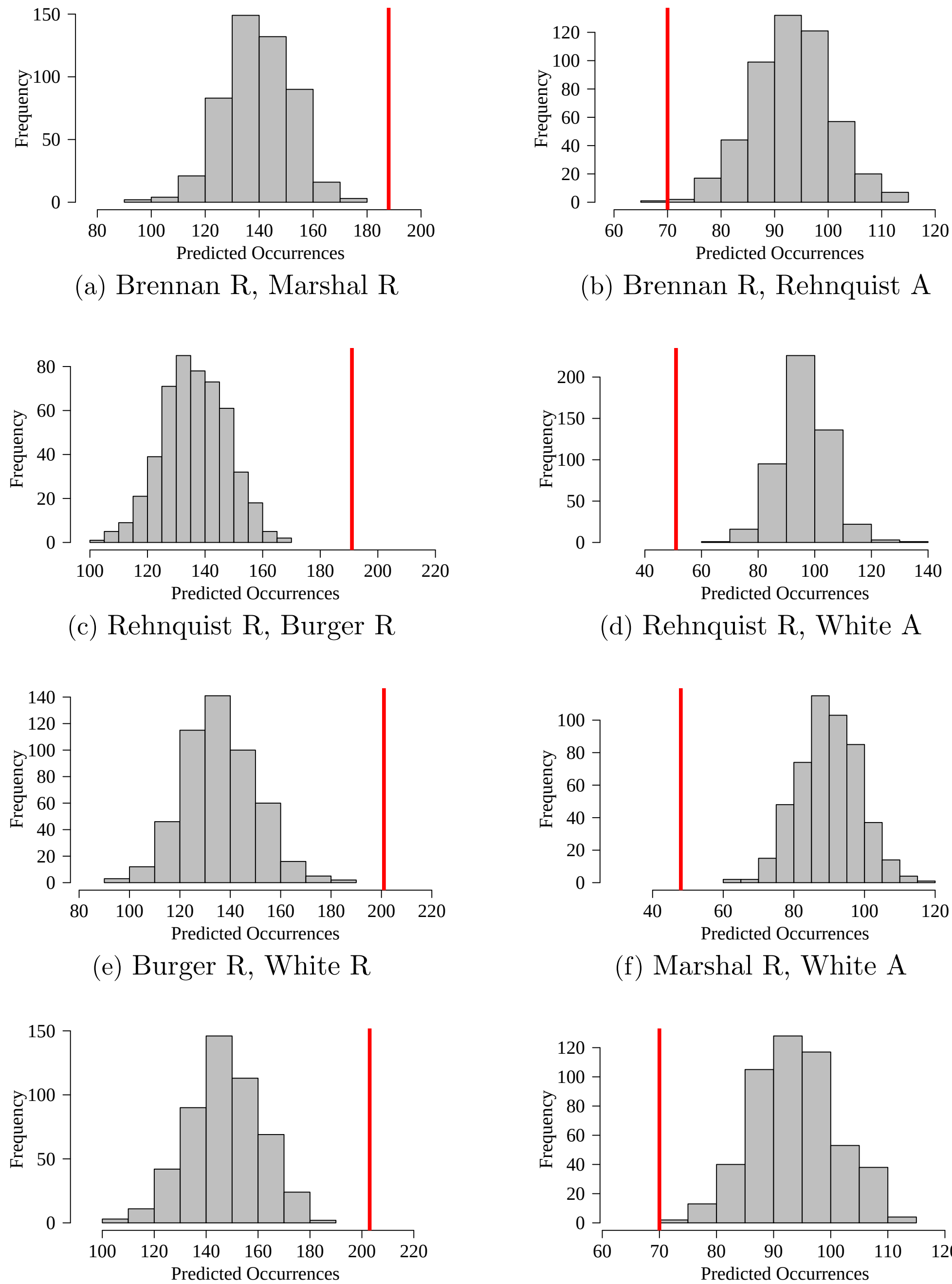
Examining figure 1.4.1 demonstrates a clear pattern in the prediction errors. All of the under-predicted pairs are justices in agreement. All of the over-predicted pairs are justices not in agreement. The results presented in the figure suggest that the original model heavily under-predicts agreement among justices in their votes on the merits. This pattern is confirmed in the larger set of JPEs. A total of 160 JPEs are identified. Among the 91 under-predicted pairs, 83 are pairs of justices voting in the same direction. The remaining 69 JPEs are over-predictions, and 68 of them are justices voting in opposite directions (i.e. one voting to reverse and one to affirm).

What these findings suggest is that the original model misses a strong degree of positive correlation between the votes of justices on any given case. This is an omitted feature of the data generating process that threatens the validity of inferences through misspecification bias (White, 1982). Two classes of underlying mechanisms could be contributing to the observed correlation. First, it is possible that overt influence or cooperation occurs on the Court. Previous studies have found that the Court tends towards consensus decision-making (Haynie, 1992; Epstein, Segal and Spaeth, 2001). It could also be that omitted legal factors are producing correlation. If there are legal facts that point every justice (or a large subset thereof) in a particular direction, the

---

[7] `R` package `Arules` Michael Hahsler and Hornik (2009) was used to perform the frequent itemset mining. I do not replicate the model in column 1 of table 3 in Johnson, Wahlbeck and Spriggs (2006) because an LR test strongly rejects the hypothesis that the restrictions in the reduced model are valid.

Figure 1.2: Joint Prediction Errors on the U.S. Supreme Court



(a) Brennan R, Marshal R

(b) Brennan R, Rehnquist A

(c) Rehnquist R, Burger R

(d) Rehnquist R, White A

(e) Burger R, White R

(f) Marshal R, White A

(g) Rehnquist R, White R

(h) Marshal R, Rehnquist A

Note: Histograms of the number of cases in which the justice-vote pair occurs over the 5,000 datasets drawn from the model. The solid line is the times that pair occurs in the actual data. The four most frequent under and over predictions are given in the left and right columns respectively. The title gives the last name of the justices and the direction of the vote (R–reverse, A–affirm).

omission of these factors from the model would cause the under-prediction of justices voting in a consensus manner. The dominance of the attitudinal model (Segal and Cover, 1989; Segal and Spaeth, 1996, 2002) over the last couple decades pulled political scientists' explanations of decision-making on the Court away from case-level legal factors. Yet, very recently, case-level apolitical factors have been regaining acceptance as important predictors of the votes of Supreme Court justices (Spriggs and Hansford, 2001, 2002; Collins, 2004; Johnson, Wahlbeck and Spriggs, 2006; Collins, 2007). The early dominance of the attitudinal model made light of case-level idiosyncrasy (Segal and Cover, 1989; Segal and Spaeth, 1996, 2002). Consensus prediction errors do not constitute a statistical test for the presence of unobserved association in justices' votes. In order to perform a principled test of the intuitions gathered from the JPE analysis, and asses the impact of these patterns on other inferences from the model, the model from Johnson, Wahlbeck and Spriggs (2006) must be improved to both test and account for positive case-level correlation among the justices.

Case-Level Determinants of Supreme Court Votes

I extend the model in Johnson, Wahlbeck and Spriggs (2006) in two ways to account for the pattern discovered in the JPE analysis. First, as mentioned previously, omitted case-level covariates could cause the observed association among the justices. Collins (2004, 2007) shows that the Court responds to *Amicus Curiae* briefs. Specifically, he shows that the probability that a particular side wins a case is directly proportional to the number of briefs filed on its behalf and inversely proportional to the number of briefs filed for the other side. Moreover, briefs filed by the U.S. Solicitor General have a larger effect on the Court's decisions than do those filed by others. I add a series of variables to the model to account for this. The variables *Appellee Amicus*, *Appellant Amicus*, *SG Appellee Amicus* and *Appellant Amicus* are the number of *Amicus Curiae* briefs filed on behalf of the appellee, appellant, appellee by the Solicitor General and appellant by the solicitor general respectively. Following Collins, I expect that briefs filed on behalf of the appellant (appellee) will have a positive (negative) effect on the likelihood a justice votes to reverse. I also add one more case-level control to the model; *Lower Court Conflict*, an indicator of whether the reason for granting *certiorari* is rooted in lower court conflict. Collins (2004) finds that the Court is less likely to reverse a decision that it hears due to lower court conflict.[8] I expect this variable to have a negative effect on the probability a justice votes for reversal.

The degree of consensus demonstrated in the JPE analysis is quite marked. It would be overly optimistic to assume that all of the case-level association would be explained by the covariates I add to the model. I therefore update the model to explicitly estimate the residual association among the justices' votes. A standard tool for modeling unobserved cluster-wise association in regression models is to include a hierarchical random-effect in the likelihood function (Gelman and Hill, 2007). It is assumed that there is a shared disturbance to the linear predictor to for every observation in a cluster. In the model implemented below, the random effect is assumed to be normally distributed with zero mean. It is integrated out of the likelihood function, leaving only a variance term of the random effect to be estimated. The higher the variance, the higher the correlation between the observations in the same cluster (Caffo, An and Rohde, 2007). Thus, the second update to the model presented in Johnson, Wahlbeck and Spriggs (2006) is to add a case-level random effect.

The results of the hierarchical logistic regression models are presented in table 1.4.1.[9] The model closest to the baseline specification that appeared in Johnson, Wahlbeck and

[8]The data for the added controls come from replication data for the analyses in Collins Jr (2008) made available on Paul Collins' website at `http://www.psci.unt.edu/~pmcollins/data.htm`

[9]R package *lme4* (Bates and Sarkar, 2006) was used to estimate the models in table 1.4.1

Table 1.1: U.S. Supreme Court Justices' Votes on the Merits

	Justice Level		Case Level		Case Level +	
	Estimate	SE	Estimate	SE	Estimate	SE
Constant	0.280	0.067	0.556	0.214	0.78	0.24
Ideological Compatibility with Appellant	0.310^{+}	0.017	0.599^{+}	0.027	0.599^{+}	0.0265
Oral Argument Grade	0.205^{+}	0.040	0.391^{+}	0.141	0.400^{+}	0.138
Case Complexity	0.075	0.101	0.169	0.366	0.137	0.359
Oral Argument Grade × Case Complexity	-0.089	0.091	-0.289	0.306	-0.252	0.301
Ideological Compatibility × Oral Argument Grade	0.020	0.016	0.026	0.025	0.026	0.025
U.S. Appellant	0.472^{+}	0.117	0.914^{+}	0.416	1.17^{+}	0.447
U.S. Appellee	-0.790^{+}	0.150	-1.633^{+}	0.544	-1.83^{+}	0.553
S.G. Appellant	0.325^{+}	0.127	0.544	0.447	0.096	0.485
S.G. Appellee	-0.208	0.167	-0.321	0.599	0.164	0.607
Washington Elite Appellant	0.406^{+}	0.136	0.765	0.483	0.499	0.478
Washington Elite Appellee	0.069	0.145	0.110	0.516	0.312	0.513
Law Professor Appellant	-0.757^{+}	0.269	-1.283	0.957	-1.53	0.940
Law Professor Appellee	-1.554^{+}	0.323	-3.007^{+}	1.135	-2.75^{+}	1.11
Clerk Appellant	-0.246	0.154	-0.571	0.541	-0.490	0.531
Clerk Appellee	-0.165	0.197	-0.145	0.690	-0.248	0.684
Elite Law School Appellant	0.025	0.088	0.090	0.316	0.014	0.310
Elite Law School Appellee	-0.127	0.089	-0.290	0.321	-0.342	0.315
Difference in Litigating Experience	-0.127^{+}	0.034	-0.234	0.122	-0.274^{+}	0.120
Appellee Amicus	–	–	–	–	-0.039	0.073
Appellant Amicus	–	–	–	–	-0.027	0.085
SG Appellee Amicus	–	–	–	–	-1.44^{+}	0.559
SG Appellant Amicus	–	–	–	–	1.05^{+}	0.522
Lower Court Conflict	–	–	–	–	-0.946^{+}	0.413
Justice-Level Variance	0.010	–	–	–	–	–
Case-Level Variance	–	–	6.88	–	6.52	–
CCVLL (No RE, RE)	-2,021	-2,019	-2,064	-1,557	-2,018	-1,555
BIC	4,153		3,253		3,274	
N	3,331		3,331		3,331	
Clusters	16		443		443	

Note: U.S. Supreme Court voting on the merits. Hierarchical logistic regression estimates are presented. $^{+}$ statistically significant at the 0.05 level (one-tailed). The CCVLL is the cluster cross-validated log-likelihood.

Spriggs (2006) is the Justice-Level specification. Johnson, Wahlbeck and Spriggs (2006) use cluster-robust standard errors (Williams, 2000) with the Justice as the clustering variable. In the case of logistic regression, this covariance estimator produces standard error estimates that are biased downward and the estimator itself is inconsistent in the face of unmodeled heterogeneity (Greene, 2008, p. 517; Harden, n.d.), so I use an alternative mechanism to account for within-justice correlation. I add a justice-level random effect to this model. This is compared to a model with a case-level random effect.[10]

The pattern discovered in the joint prediction error analysis led to a specification that greatly improves model fit, and alters many of the inferences derived from the original model. Adding the case-level random effect to the original model reduces both the CCVLL and BIC by almost 25%. Also there is much more unobserved heterogeneity and/or correlation at the case-level than the justice-level. The case-level random effect variance is estimated to be six hundred times greater than the justice-level random effect variance. A number of independent variables that are found in the justice-level model to be statistically significant at the 0.05 level are not significant in the case-level model. These are all case-level variables, and include *Solicitor General Appellant, Washington Elite Appellant, Law Professor Appellant, and the Difference in Litigating Experience.* It appears that these effects were concluded to be significantly different from zero due to specification bias. Also, three of the five variables added to the model – *SG Appellee Amicus, SG Appellant Amicus*, and *Lower Court Conflict* – are statistically significant in the expected direction. Evidence for the bloc of added variables is moderate in that the CCVLL is better in the full model, but the BIC is highest in the model that is only extended with a case-level random effect. Another important finding is that the

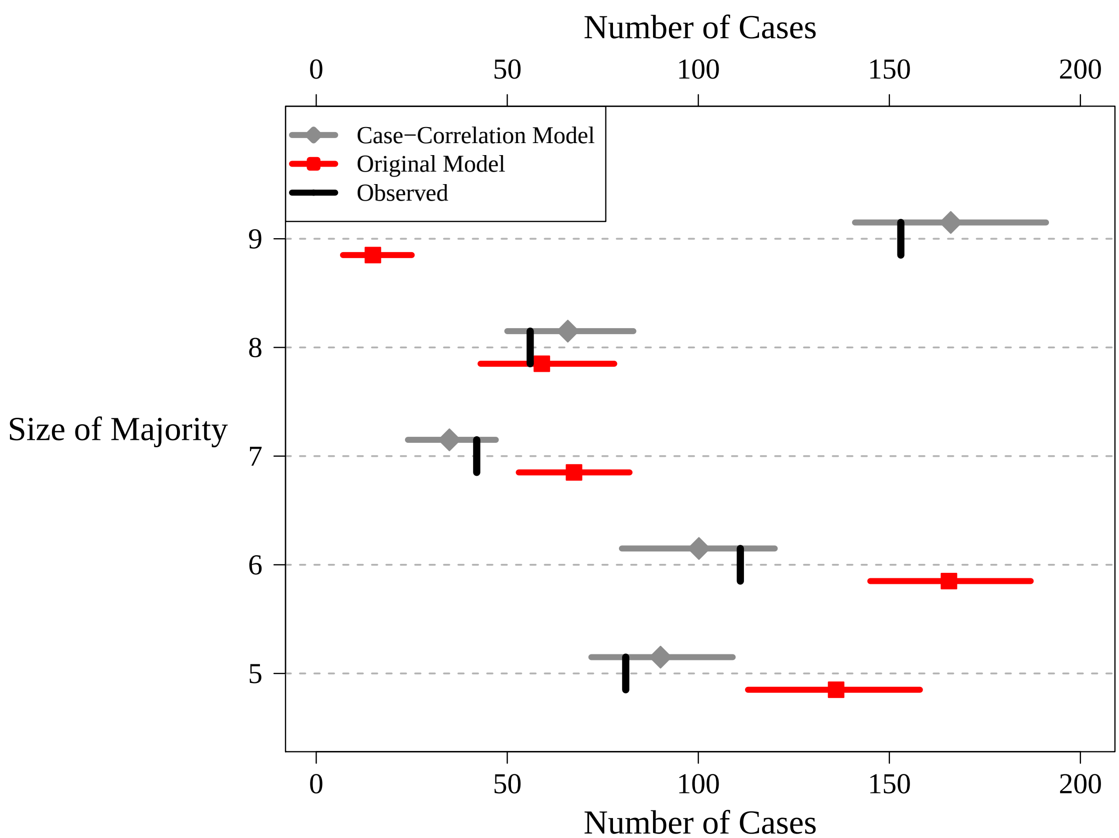
[10] I also considered a model with random effects at both the justice and case levels, but a likelihood ratio test indicates that the justice-level random effect does not improve the model.

coefficient on *Oral Argument Grade* - the variable used to test the primary theoretical proposition in the original article - nearly doubles in size in the updated models from 0.205 to 0.391 and 0.40.

The models in table 1 can be used to predict many features of the Court data. I examine the relative performance of the justice and case-level models through their prediction of the size of the voting majority in a case (e.g. 9-0, 8-1, 7-2 etc.). The ropeladder plot in figure 1.4.1 compares the predicted distribution of the size of the majority to the distribution of the majority sizes over the 443 cases in the the actual data. The y-axis gives the majority size, and the horizontal axes give the number of cases out of the 443. The points represented as squares give the predictions from the original model, and the diamonds give the predictions of the *Case +* model. It can be seen that overall the case-level model provides a much better fit than the original model, and where the improvement is most prevalent is in the tails of the distribution. Where the case-level model accurately predicts provides accurate predictions for all of the majority sizes, the original model does very poorly at predicting majorities of size 5, 6 and 9. Moreover, the modal case-level outcome in the data is a unanimous decision, which occurs in 153 out of the 443 cases. The original model predicts a frequency of unanimous decisions of fifteen. In short, there is a great deal of case-level consensus in voting on the Court, and failure to account for this results in a biased specification which leads to faulty inferences regarding the effect of independent variables as well as predictions regarding case-level outcomes.

The picture of the Court discovered here is much different than that painted by the dominant attitudinalist perspective on Court behavior, which contends that most of the Court's voting on the merits is driven by ideology (Segal and Spaeth, 2002). An enormous amount of variance exists at the case-level – so much that simply adding

Figure 1.3: U.S. Supreme Court Vote Predictions



Note: Ropeladder plot demonstrating the fit of the models to the size of the majority in Supreme Court cases. Points give predictions, and bars span 95% confidence intervals.

the case-level random effect increases the log-likelihood more than all of the covariates combined. The improvement suggested by the joint prediction error analysis (1) demonstrates that case-level factors, downplayed in the attitudinal model, are indeed important, (2) permits more reliable inferences on the effects of covariates than those published in the original article, and (3) directs attention in the way of past findings that have been inappropriately excluded from the model.

1.4.2 The Reliability of Democratic Allies

Examination of Original Findings

In the second replication, I examine international defense alliance fulfilment. Gartzke and Gleditsch (2004) test whether democratic allies are more or less likely than non-democratic allies to provide military aid to an ally that is attacked. Their hypothesis is that democratic states, due to the domestic audience costs of military intervention in a conflict involving an ally, are less likely to aid an ally than non-democratic states. To test their hypothesis Gartzke and Gleditsch (2004) study the participation of allies in wars from 1816 to the present. For each war considered, all of the allies of the participants are included in the dataset. The dependent variable is binary; coded 1 if the ally provided military aid and 0 otherwise. They specified a logistic regression model where the main independent variable of interest is an indicator of whether or not the ally has a democratic government (i.e. if the Polity II score is greater than 6). Other control variables include: whether the ally is contiguous to the attacked state, whether the ally is allied to the aggressor, and the COW composite combined capabilities score (CINC) of both the ally and the attacked state. They find that democratic states are less likely than non-democratic states to provide military aid to allies. It is relevant to note that Gartzke and Gleditsch (2004) assume, at least implicitly, that there is no interdependence among those allies considering intervention into the *same* conflict.

This proves to be an inappropriate assumption.

In the JPE analysis, the collective I consider is the group of states considering intervention on the same side of a conflict. This choice is far from arbitrary. First, there is very little in the way of conflict-specific information in the model, which would induce correlation through unobserved war-level covariates. Examples of potentially important factors that are omitted include whether the assets of third parties are endangered by the conflict (Butler, 2003), the history of interventions in the conflicts of the target state (Gleditsch and Beardsley, 2004), and the number of states involved in the conflict (Kim, 1991). Another possibility is that explicit coordination occurs among allies to states in a given conflict. Powerful international institutions such as the North Atlantic Treaty Organization (NATO) and the United Nations (UN) exist in part to coordinate the military intervention activities of member states (Hartley and Sandler, 1999; Solana, 1999; Lebovic, 2004; Lango, 2005). Lastly, intervention decisions by individual states are interdependent means to a common end – the result of the conflict. If the U.S. intervenes on behalf of one side of a conflict, Canada may no longer need to intervene to produce a victory for the side recieving help from the U.S.

The parameters of the JPE analysis are set at the same levels as in the Supreme Court example: the number of draws $t = 5,000$, the size of the JPE $k = 2$, and the p-value $\alpha = 0.10$.[11] A total of 1,071 JPEs are discovered. All of them are under-predictions, 807 of which are pairs of states making the same intervention decisions. Two interesting patterns emerge. First, since approximately 80% of the under-predictions are states in agreement, it appears the original model underestimates the degree of correlation between states considering assistance to one side of a conflict.

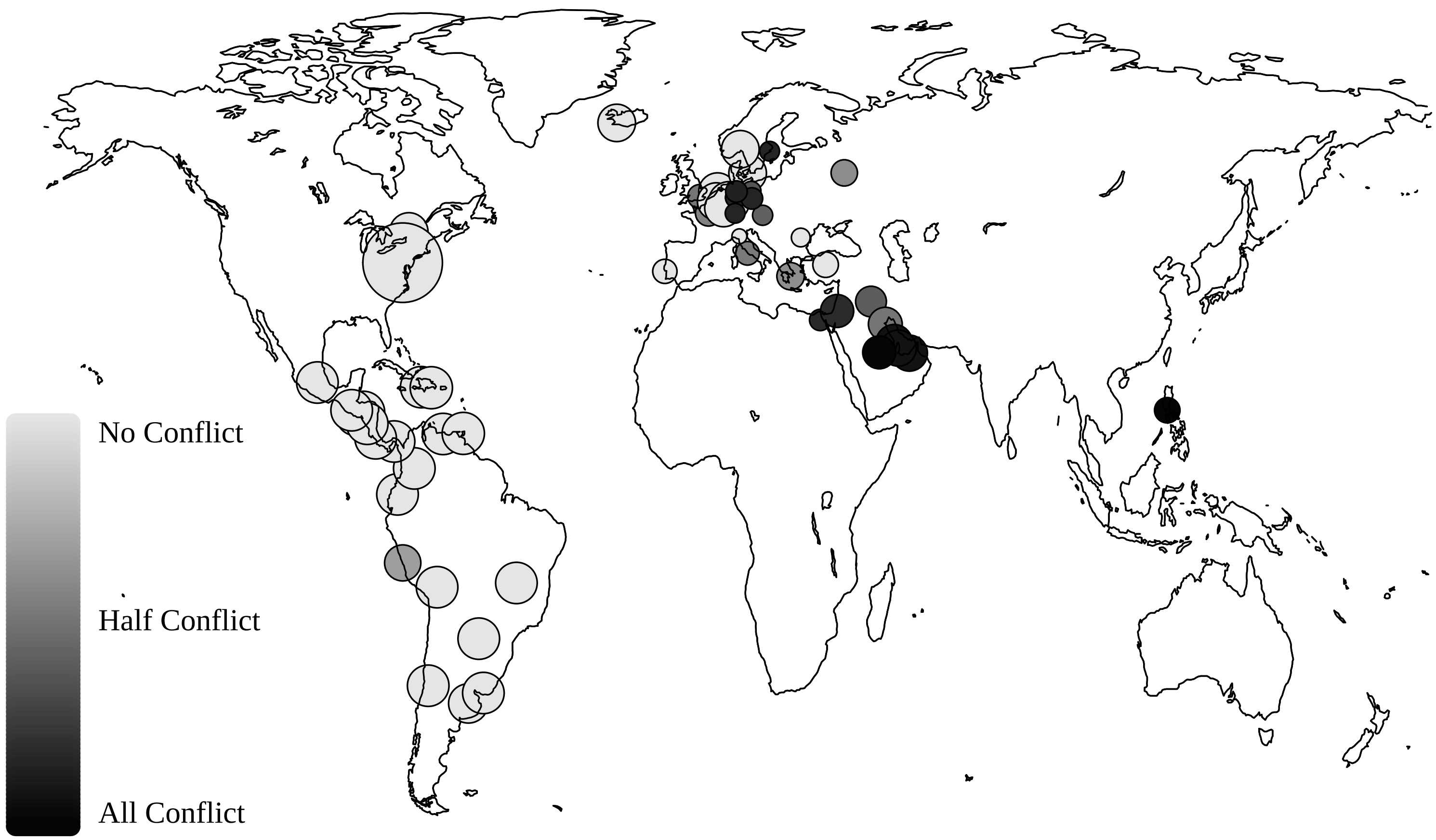
A second pattern in the JPEs regards the types of states that intervene more often

[11]As in the Court example, deviations from these parameter values do not produce different substantive inferences.

than are predicted and those that intervene less often. Examining the list of prediction errors, I noticed a stark difference between two areas of the globe that are less than completely democratic – Latin America and the Middle East. Latin American states intervene in conflicts much less often than predicted and Middle Eastern states intervene much more often than predicted. This is depicted in figure 1.4.2, where it is seen that the model in Gartzke and Gleditsch (2004) disproportionately underpredicts fulfilment decisions by Middle-Eastern States, and non-fulfilment by Latin American states. In the figure, a circle is placed at the capitol of every member in a prediction error. The darker the color of the circle, the greater the number of intervention prediction-errors in which that state is involved. Middle-Eastern states constitute the largest collection of conflict-prone dark states on the map, and Latin America is a collection of conflict-averse lighter states. This regional pattern leads to an additional hypothesis regarding the causes of defense alliance fulfilment.

As was briefly discussed above in reference to the role of international institutions, states often seek the approval and support of other nations when intervening in a conflict. There is debate regarding the ability of third party consultation to mitigate conflict in the international arena (Fisher and Keashly, 1991; Diehl, Druckman and Wall, 1998; Wilkenfeld et al., 2003), but the argument and findings presented by Ireland and Gartner (2001) support the hypothesis that the international consultation demands in alliance agreements are enough to discourage states from participating in conflicts. Ireland and Gartner (2001) argue that, in many instances, states will seek the approval of allies before entering into a conflict. In fact, many alliance agreements include pacts that require explicit prior consultation. In their empirical analysis of conflict initiation by European parliamentary governments from 1922 to 1996, Ireland and Gartner (2001) find that a consultation pact reduces the instantaneous hazard of conflict initiation by 85% – an effect that is statistically significant at the 0.05 level. States may be motivated

Figure 1.4: Under-predictions from the model in Gartzke and Gleditsch (2004)



Note:Under-predictions from the model in Gartzke and Gleditsch (2004). Each point is located at the capitol of the state involved in the JPE. The darker the point, the greater the number of intervention JPEs in which that state is involved. The larger the point, the greater the number of consultation pacts in which the state is involved.

to honor consultation agreements in order to create and maintain a reputation for reliable international commitments. A state's reputation affects inclusion in future international activities. As Crescenzi (2007, pp. 1) observes, "In international politics, states learn from the behavior of other nations, including the reputations states form through their actions in the international system." Gibler (2008) finds that states with a reputation for upholding defense alliances are more likely to be included in future alliances and that being allied with strong-reputation allies effectively deters military attacks from other states. Moreover, a state can damage its reputation for reliable international commitment by ignoring consultation obligations (Kagan, 2004; Tucker and Hendrickson, 2004; Sandler, 2005). Given that international consultation obligations can serve as an obstacle to states' entry into conflict, in the context of the current application, it would be expected that states with more consultation pacts would be less likely to fulfil defense alliances due to consultation's constraint on conflict initiation. A comparison of the regional patterns in the consultation alliance network with those in the prediction errors suggests that a state's consultation obligation is an important omitted variable.

Looking again at figure 1.4.2, the size of the point for each state is proportional to the average number of consultation pacts in which it is involved for the years that it appears in the data from Gartzke and Gleditsch (2004).[12] The ATOP codebook defines consultation pacts as agreements that, "obligate members to communicate with one another in the event of crises that have the potential to result in military conflict with the goal of creating a joint response." (Leeds (2005) p.10). The states with larger points also have lighter points, indicating that better connected states in the consultation network are less likely than predicted by the original model to intervene

[12]It may seem odd to see a number of small (i.e. poorly connected) states in the heart of Western Europe, but most of these are former German Kingdoms such as Bulgaria. These states appear in the data during conflicts in the 19^{th} century when consultation pacts were not common.

on behalf of an ally. This pattern is consistent with the hypothesis articulated above – that consultation pacts serve as a hindrance to conflict participation. Given theoretical reasons to expect consultation obligations to matter, the apparent association between conflict participation and consultation obligations in the JPEs suggests that the model of alliance fulfilment should account for the connectedness of a state – the expectation being that better-connected states will be less apt to fulfill alliance-based conflict obligations.

Improved Models of Defense Alliance Fulfilment

I have identified two interesting regularities in the joint prediction errors. First, it appears that the consultation obligations of an ally can inhibit the ally from entering into a conflict. Second, there seems to be unmodeled positive correlation between the decisions made by the allies of an attacked state. Again, we must statistically test whether the patterns discovered in the joint prediction error analysis truly exist in the data generally, and whether accounting for them improves the specification of Gartzke and Gleditsch (2004). To test whether consultation obligations reduce the likelihood of alliance fulfilment, I add a variable to the model (*Consultation Degree*) which is the number of states with which the ally has consultation pacts in year t. If state A must decide whether to intervene into a conflict in year t, *Consultation Degree* is the number of states with which state A has consultation pacts in year t. To account for correlation among states that are allied to the same state I add a target-conflict random effect to the model, where the target is the state being potentially assisted in the alliance and the conflict is a specific instance of war. Table 1.4.2 presents the results with various specifications that include the improvements identified in the JPE analysis.

The results support the inferences suggested in the JPE analysis. In terms of the

Table 1.2: The Reliability of Allies and International Conflict

	Original	RE	CD, ND	CD, ND, RE	CD	CD, RE
Constant	-2.16	-3.94	-1.53	-2.89	-1.43	-2.85
	(0.279)	(0.786)	(0.396)	(0.834)	(0.402)	(0.866)
Consultation Degree	–	–	-0.0523^{+}	-0.0701^{+}	-0.0483^{+}	-0.0698^{+}
			(0.0195)	(0.0438)	(0.201)	(0.0436)
A is Democracy	-1.02^{+}	-0.108			-0.737	-0.095
	(0.565)	(0.927)	–	–	(0.571)	(0.914)
A Allied to Other Side	-0.0536	-0.355	0.0549	-0.221	-0.0154	-0.24
	(0.315)	(0.905)	(0.314)	(0.828)	(0.318)	(0.832)
A and B Contiguous	0.911^{+}	1.09^{+}	0.882^{+}	1.05^{+}	0.776^{+}	1.03^{+}
	(0.31)	(0.507)	(0.307)	(0.48)	(0.314)	(0.49)
CINC A	-4.05^{+}	-7.15	-6.75^{+}	-11^{+}	-6.74^{+}	-11^{+}
	(2.45)	(7.11)	(2.8)	(6.72)	(2.82)	(6.7)
CINC B	7.43^{+}	11.4^{+}	5.76^{+}	11.3^{+}	6.31^{+}	11.3^{+}
	(2.76)	(5.06)	(2.81)	(4.92)	(2.86)	(4.9)
Coalition-Level Variance	–	7.15	–	5.69	–	5.60
CCVLL	-175.5	-141.7	-173.8	-139.1	-174.4	-140.6
BIC	351.9	306.4	347.3	303.1	345.4	309.2

Note: Results presented are logistic regression coefficients with standard errors in parentheses. $^{+}$ statistically significant at the 0.05 level (one-tailed). Model abbreviations are as follows; RE = Random Effect, CD = Consultation Degree, ND = No Democracy. A total of 451 observations with 91 target-conflict groups are used in each model. The CCVLL is the cluster cross-validated log-likelihood.

first pattern discovered in the JPE analysis, there is a high degree of association between the decisions rendered by states in the same target-conflict group. The addition of a target-conflict random effect improves model fit considerably. Over all three of the covariate specifications, the addition of the target-conflict random effect improves the BIC and CCVLL by 20-30 points. The suspicion that consultation obligation is an important omitted variable is also confirmed by the results. *Consultation Degree* is a statistically significant negative determinant of the probability of alliance fulfilment in all of the different specifications. Accounting for this relationship moves the specification closer to the true data generating process, as evidenced by the CCVLL. Overall, the contributions suggested by the JPE analysis improved the explanation of states' decisions to fulfil defense alliance obligations.

Another result from the improved specification is that the democracy indicator is no longer statistically significant. Simply adding the random effect to the model eliminates the statistical significance of the democracy indicator. In fact, the best fitting model, according to both the BIC and CCVLL, is the one where a random effect and *Consultation Degree* is included and the democracy indicator is constrained to have no effect. By improving the model specification, I have shown that the previous inference that democratic states are less likely to fulfil defense alliances is attributable to misspecification bias, and not an actual effect.

1.5 Conclusion

Political scientists have learned much from the study of repeated collective choice processes, where stable sets of well-known actors repeatedly issue individual decisions that have broad collective implications. A cornerstone of theory regarding repeated, salient interaction is that the actors involved equilibrate to sophisticated and highly interdependent choice strategies. Many contexts of repeated collective choice – roll-call voting in legislatures, decisions on the merits in multi-member courts of appeal, and intervention by nations into conflicts or other emergency events – are characterized by a body of individual actors making micro-level decisions that are aggregated into macro-level outcomes with far-reaching consequences (e.g. law, the interpretation of the law, and the results of conflicts). Since these collective interactions are repeated many times throughout history, an expectation of stable and sophisticated patterns of interdependence among the micro-level decisions is strongly justified. This poses a challenge to the statistical analysis of micro-level decisions in repeated collective choice data. Namely, if patterns of interdependence are a strong component of the data generating process, common parametric models that are used to analyze this sort of data, such as logistic regression, are misspecified and inferences regarding micro and macro-level factors that drive micro-level choices are suspect due to misspecification bias.

In order to make valid statistical inferences with repeated collective choice data, the model specification must account for the forms of interdependence that characterize the data generating process. It can be incredibly burdensome in terms of both computation and interpretation to specify and estimate a model that is robust to any conceivable form of multivariate dependence among discrete choices. I propose a solution to the problem of interdependence in repeated collective choice data that takes advantage of the wealth of knowledge political scientists hold regarding the observations (e.g. legislators, justices or countries) that is above and beyond that contained in the dataset. I propose that researchers estimate a simple model to start – that which represents the best theoretical specification that can be managed – then examine forms of multivariate deficiency, and iteratively improve the specification to include components hypothesized to account for the model's failures. Specifically, I introduce the *joint prediction error*, a collective outcome that is poorly predicted by a model, as a tool for discovering unmodeled forms of interdependence. Theoretical examination of commonalities among and individual characteristics of the membership and actions in these JPEs suggests substantive improvements to the specified model. Additionally, I suggest the use of the cross-validated loglikelihood, an unbiased metric of proximity to the true model, as a tool for judging the validity of the improvements derived from the analysis of JPEs. In two empirical applications I demonstrate the utility of the iterative model improvement procedure I propose. In the JPE analysis of the model from Johnson, Wahlbeck and Spriggs (2006) it is found that there is a very strong unmodeled tendency towards consensus on the U.S. Supreme Court. This is in contrast with a strictly political view of the Court. Updating the model to account for case-level positive correlation among the votes of the justices strongly improves the fit. Moreover, many inferences made in the original paper are shown to result from specification error. The second replication considers an analysis by Gartzke and Gleditsch (2004) of the likelihood that states fulfil defense alliances. The JPE analysis identifies correlation between states considering intervention on the same side of a conflict. Also, prediction errors are consistent with the importance of states' consultation obligations, which is omitted from the original specification. Both of the patterns in the JPEs lead to improvements in the empirical model.

Chapter 2

The Exponential Random Configuration Model

2.1 Introduction

Important processes in every field of empirical political science arise in the form of repeated collective decisions. Roll-call votes in legislatures and decisions issued by multi-member courts of appeal are comprised of individual decisions issued by a stable group of actors that are aggregated into politically relevant collective outcomes. In the international arena, intervention into civil wars, the provision of relief for natural disasters, and the issuance of trade sanctions are interdependent decisions rendered repeatedly by a stable group of states. The outcomes produced by many forms of collective political decision-making – laws written, laws interpreted, or the result of wars – are typically high-stakes. Additionally, the individual decision-makers have multiple opportunities to learn optimal strategies for interaction. This combination of important results and repeated interaction provides respectively the motive and opportunity for actors to develop complex and sophisticated interactive strategies. As a result, patterns of interdependence are likely to emerge in repeated collective choice data. Many scholars have noted that patterns of sophisticated rational interaction are likely to exist when the stakes of the interaction are high enough to induce careful cognitive effort, and the collective choice situation is repeated many times – providing actors with the motive and opportunity to learn the rules and payoffs of the game (see e.g. Verba (1961) and Ostrom (1998)).

The dependence among the individual decisions in repeated collective choices is just as critical to an accurate theoretical account of the process under study as the dependence of those decisions on exogenous covariates. Indeed, many well-known explanations of political phenomena directly imply interdependence in repeated collective choice. Theories of party leadership (Hix, 2002) imply that the roll call votes of legislators will depend upon those of their party leaders, accounts of acclamation on the U.S. Supreme Court (Dorff and Brenner, 1992) imply that freshman justices' votes will correlate strongly with those of their colleagues, and in the study of international conflict it is theorized that states are not likely to fight the enemies of their enemies (Maoz et al., 2006) (i.e. the choice of state A to fight state B, if state A is at war with C, depends on whether B is at war with C). However, in practice analysts often pool the members of the stable group into a sample for regression modeling where relationships between the members are ignored to the extent that they are not represented by independent variables. This causes two related problems. First, theories of interdependence are excluded from models of individual decision-making – leading to a substantively unrealistic account of the process. This, in turn, leads to the second problem – misspecification bias, where inference on the effects of exogenous covariates are possibly invalid due to incorrect model specification.

In many instances of repeated collective decision making, the politically relevant collective outcome (e.g. the result of a court case or the passage of legislation) is a deterministic function of the individual decisions rendered by the members of decision-making body (e.g. a lower court decision is reversed by the U.S. Supreme Court if five or more justices vote with the appellant). Because of this deterministic relationship between micro and macro-level outcomes, if a model is fit to the individual decisions that comprise the collective outcome, one is automatically implied for the collective outcome.[1] For instance, if a model is fit to U.S. Supreme Court Justice voting on the merits – estimating the probability that a justice will vote with the appellant in a case – the probability that the *Court* decides in favor of the appellant is given by the sum of the joint probabilities of all nine-justice configurations in which at least five vote with the appellant. It is inconsistent to specify separate statistical models for individual decisions and combinations of individual decisions(e.g. the proportion voting yea) – the the former implies the latter. The critical implication of the micro-macro connection is that, in order to be correctly specified, the micro-level model must capture any tendency for individual decisions to produce sophisticated/intentional combinations of decisions.

My contributions in this work are three-fold. First, I propose a statistical model, the exponential random configuration model (ERCM), that can be used to account for and test hypotheses regarding virtually any form of interdependence in collective discrete choice data. The model also permits individual decisions to depend upon exogenous covariates in the exact same way as in a logistic regression model. Indeed, in the case where interdependence does not exist, the effects of *configurational terms* on the probability of observing any particular collective decision go to zero and the model reduces to logistic regression. Second, I describe a useful *configurational* approach to theorizing about the data generating process in collective choice that facilitates the integration of individual covariate and collective interdependence theories into a single

[1]In making this statement I assume that the model is fit within either the likelihood or Bayesian estimation frameworks, or any other method used to fit a full parametric distribution to the data.

framework. Third, I present a semi-parametric, predictive measure of model fit that can be used to select between any two statistical models of dichotomous collective choice. To illustrate the applicability of these contributions, I replicate and extend a model of U.S. Supreme Court justices' votes on the merits (Richards and Kritzer, 2002) – demonstrating how the ERCM can be used to account for well-known case-level dynamics while modeling the individual choices of the justices.

2.2 Theorizing About Discrete Configurations

An approach to theorizing about collective choice outcomes that unifies theories of individual behavioral motivations (i.e. covariate effects) and those that posit the relevance of particular combinations of decisions will facilitate the construction of comprehensive empirical models using the ERCM. When modeling the individual decisions that constitute an instance of collective choice, two ways to state the scientific task at hand are (1) to explain each of the individual decisions within the instance of the collective choice and (2) to explain why the collective decision unfolded in the way it did without aggregating the votes in any way (i.e. to explain why we observed the particular *configuration* of votes on hand). Task (2) nests task (1) as a necessary component. In order to explain why a vote unfolded in a particular manner it is critical to account for any individual (i.e. exogenously determined) motivations of the decision-makers. Task (2) orients the theorist towards higher-level explanations of the collective outcome – focusing on relevant combinations of decisions. In an instance of roll-call voting, such higher-level or *configurational* explanations might include a recognition that a high degree of party discipline characterized the vote, legislators took cues from, and thus voted with, noted policy experts on the subject matter of the bill under consideration, and legislators X, Y and Z voted with other legislators who typically vote for legislation sponsored by X, Y and Z (i.e. vote-trading).

When theories involve joint outcomes, where the decisions of one actor depend upon those of others within the same instance of collective choice, hypotheses derived from these theories cannot be accommodated in the conventional discrete choice regression modeling framework. Probabilistic inferences using logit or probit require the assumption that the individual outcomes which constitute the dependent variable are independent and identically distributed conditional upon the covariates (i.i.d.). Entering one or more of the choices as an independent variable for another choice renders the outcomes dependent upon each other by construction - violating the i.i.d. assumption.[2]

It is true that hierarchical/multilevel models, as described by Gelman and Hill (2007), allow for dependence between the individual decisions, but the form of the dependence is quite limited. In a multilevel model, individual observations within groups are interdependent insofar as they share the value of some unobserved contextual independent variable that adds a shared within-group disturbance to one of the model parameters (e.g. intercept, regression coefficient, variance etc.). A hierarchical model, for instance, could not accommodate the sort of interdependence that might arise due to cue-taking in a legislature – where the interdependence cannot be represented by cutting the observations into separate groups.

Configurational theory is not new to political science, but testing and accounting for configurational hypotheses in micro-level decision models is. When exploring configurational hypotheses it is common for political scientists to identify aggregations and/or transformations of the micro-level data that render it suitable for testing their particular configurational theories. There are a number of examples of this in studies of the U.S. Supreme Court. Johnson, Spriggs and Wahlbeck (2005) provide support

[2]It is important to note that this is not *always* the case. If the sequence of decision-making is known, later decisions only depend upon those made earlier in the sequence (e.g. there is no reaction to anticipated decisions), then a Markov assumption is justified (Ware, Lipsitz and Speizer, 1988) and later decisions can be made dependent upon earlier decisions in a logit or probit.

for their hypothesis that the chief justice will vote with the majority by demonstrating that the chief chooses to pass when it becomes his or her turn to vote in conference – observing which side will win, then voting with the majority. In a study of conformity voting on the Court, Epstein, Segal and Spaeth (2001) show that justices tend to change their votes in the direction of the original conference majority when issuing their final vote on the merits in cases. An area in political science that relies heavily upon aggregation of collective decision-making data is that on parties in legislatures. Many configurational theories regarding party cohesion and the effect of party leaders in legislatures have been tested by aggregating outcomes in ways so as to directly measure the form of association under study (e.g. party cohesion, polarization, and discipline scores). Specifically, past studies have examined the determinants of variation in partisan polarization and party discipline in legislatures (Rohde, 1991; Aldrich and Battista, 2002; Hix, 2002). In every field of political science that regularly studies collective choice, configurational theory has been developed, and tests of these theories are regularly defined on careful manipulations or aggregations of micro-level data.

Though one might identify a satisfactory manipulation of the data that permits an effective test of a configurational theory, this does not help the analyst that needs to account for configurational tendencies in the study of micro-level decisions. Virtually any support for configurational theories found by manipulating micro-level data constitutes evidence that these theories need to be accommodated in models of micro-level decisions. In the next section I propose the Exponential Random Configuration Model (ERCM), a statistical model that can be used to seamlessly integrate configurational theory into models of micro-level behavior.

2.3 A General Model for Discrete Configurations

The exponential random configuration model (ERCM), which I define in the current section, can be used to model collective choice, accounting for and testing the effect of any theoretically plausible form of interdependence among the decisions. This model is derived from an incredibly flexible exponential family discrete multivariate probability distribution that has been used in areas as diverse as social network analysis in the form of the exponential random graph model (ERGM) (Wasserman and Pattison, 1996), natural language processing in the form of conditional random fields (Altun, Smola and Hofmann, 2004), the energy in microstates of discrete systems in physics (Shirts and Shirts, 2002), and modeling the spatial distribution of plant infections in the form of lattice systems (Besag, 1974).[3] The exponential family of multivariate distributions used to derive the ERCM has seen such wide application due to its ability to accommodate any form of measurable interdependence among the individual components in the grouping under study – whether that grouping is a set of relationships in a social group, words on a document, or particles in a physical system. In the current section I make the case for adding to this list collective decision outcomes in politics.

The key step in using the ERCM is to approach each collective decision as a single observation with multiple components – rather than a number of choices that are independent conditional on the covariates. The ERCM takes as inputs measures defined on the collective decision that the researcher believes effect the likelihood of a particular instance of collective choice being observed. These measures are identified by the analyst to capture combinations of decisions that are implied to be more or less likely by their configurational theories. For instance, if a theory implies that a decision-making body

[3] In terms of other treatments of this distributional form, the literature on the ERGM is likely the most familiar and/or accessible to political scientists. See Holland and Leinhardt (1981); Wasserman and Pattison (1996); Goodreau (2007) and Robins (2009) for additional reading on the ERGM.

tends toward consensus, the measure could be the size of the majority coalition, hypothesizing that within a consensus-building institution the size of the majority coalition will have a positive effect on the likelihood of observing an instance of collective choice. These measures are operationalizations of the configurational hypotheses reviewed in the previous section.

In the remainder of this section I derive the ERCM in full. First, I describe in detail the process of operationalizing configurational hypotheses, including a characterization of conventional additive covariate effects as configurational hypotheses. Next I describe how configurational measures defined on the collective choice outcome can be used to form a distribution over all possible collective outcomes. Then I show how the parameters in the ERCM can be used to interpret the effect of configurational measures on the likelihood of observing any particular instance of collective choice. After deriving the model and discussing interpretation, I describe alternative methods for estimating the parameters of the ERCM. Last, I provide a model-fit measure that can be used to compare differently parametrized ERCMs.

2.3.1 Measurement and Configurational Hypotheses

Configurational hypotheses imply the prevalence of certain combinations of choices in instances of collective decision-making. An alternative way to state this is that configurational hypotheses state the effect of observing a particular combination of decisions within a collective choice on the likelihood of observing that instance of collective choice. For example, if the configurational theory claims that party leadership is important in the U.S. House, a hypothesis derived from this theory could be that we are more likely to observe roll-call outcomes where more than 50% of majority party members vote with the Speaker of the House than we are to observe roll-calls in which more than 50% vote with the minority leader. The first step in testing a configurational hypothesis is to define the measurement(s) that can be computed on a collective choice outcome that best capture the process implied by the configurational theory. The inputs to the ERCM are finite, scalar-valued statistics defined on the binary decisions that constitute the collective choice. In the same way that the researcher defines the appropriate operationalizations of the independent variables to be included in the model, these statistics are also designed by the researcher to measure features of the collective choice that vary in a systematic way according to the configurational hypotheses. To continue with the example of party leadership, this statistic may be the proportion of legislators that vote with their party leaders. The ERCM is designed to assess the effect of these statistics on the likelihood of observing any particular instance of the collective decision. The fully operationalized configurational hypothesis regarding party leadership would be that the likelihood of observing a particular instance of roll-call voting increases as the proportion of members voting with their party leaders in that particular instance increases. Equivalently, it could be stated that, given a vote is taken, we are more likely to observe an outcome where the proportion of members voting with their party leaders is high than one in which the proportion of members voting with their party leaders is low.

Let $\boldsymbol{y}$ be a binary collective decision issued by n actors (i.e. a vector of n zeros and ones), and $\Gamma(\boldsymbol{y})$ be a function/statistic defined on $\boldsymbol{y}$ that evaluates to a real number that is finite, and there is varies in value over the possible configurations of $\boldsymbol{y}$. Suppose there are k configurational hypotheses implied by an analyst's theory of collective decision-making. The statistic $\Gamma_i(\boldsymbol{y}),\ i \in \{1, 2, \dots, k\}$ should be specified such that if the i^{th} hypothesis is accurate, the likelihood of observing a particular instance of $\boldsymbol{y}$ varies predictably with $\Gamma_i(\boldsymbol{y})$. Using again the example of the count of legislators voting with their party leaders, a hypothesis of strong party leadership would imply that the likelihood of observing an instance of $\boldsymbol{y}$ increases as the proportion of legislators voting with their party leaders in that instance of $\boldsymbol{y}$ increases (i.e. we are more likely to observe a roll-call in which 50% of legislators vote with their party leaders than one in which 30% of legislators vote with party leaders). An alternative way to state this expectation is that we are more likely to observe instances of $\boldsymbol{y}$ where many legislators vote with their party leaders. Particular specifications of $\Gamma(\boldsymbol{y})$ derive from theories of collective choice in the same way as the operationalization of covariates in conventional regression modeling. In the same way a researcher might search for the best possible indicator of the ideology of legislators, in order to operationalize a hypothesis regarding cue-taking, the researcher should identify what measure can be computed on roll-call votes that would provide the clearest evidence of cue-taking – this measure should become $\Gamma_{cue-taking}(\boldsymbol{y})$. The innovation available in ERCM modeling is to adjust inferences regarding the presence of cue-taking for other factors that might produce a particular value of $\Gamma_{cue-taking}(\boldsymbol{y})$.

The regression framework represents the dominant approach to the statistical analysis of discrete choice in political science. It is therefore important to show how $\Gamma(\boldsymbol{y})$ can be specified to capture the same relationships in the data that are tested in a regression model. Let $\boldsymbol{x}$ be an n element exogenous covariate where the j^{th}, $\{j \in 1, 2, \ldots, n\}$ element of $\boldsymbol{x}$ corresponds to the j^{th} individual choice in $\boldsymbol{y}$, then accounting for the additive effect of $\boldsymbol{x}$ on the individual components of $\boldsymbol{y}$ in constructing the likelihood of observing a particular instance of $\boldsymbol{y}$ is accomplished with

$$\Gamma_{\boldsymbol{x}}(\boldsymbol{y}) = \left[\sum_{i=1}^{n} y_i x_i\right]. \tag{2.1}$$

A hypothesis that $\boldsymbol{x}$ will have a positive effect on $\boldsymbol{y}$, in a discrete-choice regression framework, is stated as an expectation that the probability that $y = 1$ will increase as x increases. In configurational form, this is equivalent to the expectation that the ones in $\boldsymbol{y}$ will align with higher values in $\boldsymbol{x}$. $\Gamma_{\boldsymbol{x}}(\boldsymbol{y})$ will be higher if the ones in $\boldsymbol{y}$

correspond to the higher values in $\boldsymbol{x}$. If $\boldsymbol{x}$ is expected to have a positive effect on $\boldsymbol{y}$, this, again, can be stated in configurational form as an expectation that the likelihood of observing a particular instance of $\boldsymbol{y}$ increases with the degree to which the ones in $\boldsymbol{y}$ align with the higher values in $\boldsymbol{x}$. By using statistics in the form given in equation 2.1, the analyst can test the usual covariate-effect hypotheses alongside theories of complex interdependence in discrete choice.

I claim that, in order to test configurational hypotheses, the analyst must define measurements on the collective outcome that help explain the way in which the collective decision unfolded. It may appear, at first glance, like these could constitute second-level covariates in a multilevel model where individual choices are nested in collective decisions. There is a fundamental problem with this approach. Since $\Gamma(\boldsymbol{y})$ is itself a measurement/function defined on the collective outcome, it is, by construction, endogenous to the collective outcome that would constitute the dependent variable in such a multilevel model. Putting $\Gamma(\boldsymbol{y})$ on the right-hand-side of a hierarchical model of $\boldsymbol{y}$ would amount to explaining $\boldsymbol{y}$ with a function of $\boldsymbol{y}$. Having identified the $\Gamma(\boldsymbol{y})$s that are believed to be relevant to the probability of observing a particular instance of $\boldsymbol{y}$, I have yet to define how the relevance and overall effects of the $\Gamma(\boldsymbol{y})$s can be estimated within a principled inferential framework.

2.3.2 A Distribution Over Discrete Configurations

In this section I introduce a multivariate distribution that is permits the estimation of the effect of $\Gamma(\boldsymbol{y})$ on the likelihood of observing $\boldsymbol{y}$ to be computed without making the blatantly incorrect assumption that $\Gamma(\boldsymbol{y})$ is exogenous to $\boldsymbol{y}$ – thus completing the derivation of the ERCM. Let $\boldsymbol{\Gamma}(\boldsymbol{y})$ be the vector of k statistics hypothesized to effect the likelihood of observing $\boldsymbol{y}$. In order to make inference on the effects of the statistics of interest, $\boldsymbol{\Gamma}(\boldsymbol{y})$ must be integrated into a proper probability distribution defined on possible configurations of $\boldsymbol{y}$ along with parameters that can be estimated and provide an interpretation the effect of $\boldsymbol{\Gamma}(\boldsymbol{y})$. Denote the vector of effect parameters $\boldsymbol{\theta}$, and the linear configuration predictor $\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})$. The $\boldsymbol{\theta}$ give the magnitude and direction of the effect of $\boldsymbol{\Gamma}(\boldsymbol{y})$ on the likelihood of observing $\boldsymbol{y}$. If θ_j is positive (negative), then as $\Gamma_j(\boldsymbol{y})$ increases, the likelihood of observing $\boldsymbol{y}$ increases (decreases). A *relative* probability weight for any particular configuration is given as

$$\omega(\boldsymbol{y}) = \exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}. \tag{2.2}$$

This defines a non-negative weight for any observed collective choice. Thus, the probability of observing a particular configuration $\boldsymbol{y}$ is proportional to $\omega(\boldsymbol{y})$. A proper probability distribution for $\boldsymbol{y}$ is then defined by dividing $\omega(\boldsymbol{y})$ by the sum of all weights for all possible configurations of the collective decision –

$$P(\boldsymbol{y}) = \frac{\omega(\boldsymbol{y})}{\sum_{\forall \boldsymbol{y} \in \boldsymbol{Y}} \omega(\boldsymbol{y})} = \frac{\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}}{\sum_{\forall \boldsymbol{y} \in \boldsymbol{Y}} \exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}}. \tag{2.3}$$

The probability distribution given in equation 2.3 permits statistical inference within the Bayesian or likelihood frameworks. Suppose the data under study contains T collective choices (e.g. roll-calls or court cases), then the likelihood function defined on the entire set of collective decisions is given as

$$l(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{\boldsymbol{T}} | \boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{\omega(\boldsymbol{y}_t)}{\sum_{\forall \boldsymbol{y} \in \boldsymbol{Y}_t} \omega(\boldsymbol{y})}, \tag{2.4}$$

where $\boldsymbol{Y}_t$ is the set of all possible configurations of binary decisions that could have been observed in the t^{th} collective decision.[4]

At first glance the ERCM might seem quite different from logistic regression, the

[4]For an alternative derivation of this distribution – one oriented towards the modeling of a single social network – see Park, Gelman and Bafumi (2004).

modal method for the study of discrete choice in political science and elsewhere, but there is a subtle commonality. If none of the higher-order (i.e. endogenous) $\Gamma(\boldsymbol{y})$s have an effect on $P(\boldsymbol{y})$, and only those statistics that account for exogenous covariates entered in the form of equation 2.1 matter, then the ERCM reduces to logistic regression on the discrete choices that compose the collective decisions, and the $\boldsymbol{\theta}$ are exactly equivalent to the coefficients from logistic regression.[5] I prove this in the final section of this chapter. This property establishes a seamless relationship between the ERCM and discrete choice modeling with logistic regression. Since logit is nested in the ERCM, a likelihood ratio test can be used to assess whether a significant improvement in model fit is achieved by moving from the familiar logit framework to the more complicated ERCM.

2.3.3 Interpretation of the ERCM

Unlike the regression approach to modeling, where the outcome of interest is the mean or expected value of the dependent variable, the quantity that is directly modeled in the ERCM is the probability of observing any particular configuration of the collective decision. This probability depends flexibly upon characteristics of the collective outcome that can include indicators or counts of theoretically relevant combinations of choices, the alignment of choices with exogenous covariates, or interactions between the two. The parameter θ_j can be used to derive the proportional change in the probability of observing $\boldsymbol{y}$ given a change in $\Gamma_j(\boldsymbol{y})$. Consider two potential configurations $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ where $\Gamma_j(\boldsymbol{y}_1) - \Gamma_j(\boldsymbol{y}_2) = \delta,\ \delta \neq 0$ and $\Gamma_i(\boldsymbol{y}_1) = \Gamma_i(\boldsymbol{y}_2)\ \forall i \neq j$, meaning $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ differ by some non-zero quantity δ on the j^{th} statistic and are equal on all other

[5] I am not the first to note this property of models parametrized in the form of the ERCM. In the context of the ERGM for social networks, see Faust and Skvoretz (2002) and Goodreau, Kitts and Morris (2009) for other works that discuss the logistic regression as a special case.

statistics. The ratio of the probability of observing $\boldsymbol{y}_1$ to that of observing $\boldsymbol{y}_2$ is

$$\frac{\frac{\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y_1})\}}{\sum_{\forall \boldsymbol{y}\in \boldsymbol{Y}}\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}}}{\frac{\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}}{\sum_{\forall \boldsymbol{y_2}\in \boldsymbol{Y}}\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y})\}}} = \frac{\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y_1})\}}{\exp\{\boldsymbol{\theta}'\boldsymbol{\Gamma}(\boldsymbol{y_2})\}}$$

$$= \frac{\exp\{\theta_1\Gamma_1(\boldsymbol{y_1})\},\dots,\exp\{\theta_j\Gamma_j(\boldsymbol{y_1})\},\dots,\exp\{\theta_k\Gamma_k(\boldsymbol{y_1})\}}{\exp\{\theta_1\Gamma_1(\boldsymbol{y_2})\},\dots,\exp\{\theta_j\Gamma_j(\boldsymbol{y_2})\},\dots,\exp\{\theta_k\Gamma_k(\boldsymbol{y_2})\}} = \frac{\exp\{\theta_j\Gamma_j(\boldsymbol{y_1})\}}{\exp\{\theta_j\Gamma_j(\boldsymbol{y_2})\}}$$

$$= \frac{\exp\{\theta_j\left[\Gamma_j(\boldsymbol{y_2})+\delta\right]\}}{\exp\{\theta_j\Gamma_j(\boldsymbol{y_2})\}} = \exp\{\theta_j\delta\}.$$

Thus, the multiplicative effect of a change in $\Gamma_j(\boldsymbol{y})$ (i.e. δ) on $P(\boldsymbol{y})$ is $\exp\{\theta_j\delta\}$. To give some substance to this interpretation, I revert to the hypothetical example of party leadership in a legislature. If $\Gamma_j(\boldsymbol{y})$ is the number of partisans voting with their party leaders, consider again two possible configurations of the vote $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, and assume that $\Gamma_j(\boldsymbol{y}_1)$ is 51, $\Gamma_j(\boldsymbol{y}_2)$ is 50 (i.e. $\delta = 1$), and $\theta_j = 0.0953$. Then $\frac{P(\boldsymbol{y}_1)}{P(\boldsymbol{y}_2)}$ is $\exp\{0.0953 \times 1\} = 1.1$ – the probability of observing $\boldsymbol{y}_1$ is approximately 1.1 times the probability of observing $\boldsymbol{y}_2$ – a difference of 10% attributable to the greater degree of agreement with party leaders in $\boldsymbol{y}_1$.

Though the ERCM is parametrized to directly model the *probability* of a particular instance of collective choice, researchers may be interested in other quantities such as the expected value of one individual's decision, the expected correlation between any two individuals, or the uncertainty (i.e. variance) surrounding the prediction of an individual's decision. The estimates of the ERCM can be used to predict any of these quantities. Since a full parametric distribution is defined over the collective decision, the parameters can be used to study and predict any quantity of interest that can be computed on a collective decision. A Monte Carlo method that has been developed for the interpretation of ERGMs applied to social networks will prove useful in the interpretation of ERCMs (see e.g. Hunter et al. (2008) and Goodreau, Kitts and Morris (2009)). First, a large number of collective outcomes are simulated from the distribution implied by the model estimates. Then quantities of interest are computed on the simulated outcomes – constituting an empirical approximation to the theoretical distribution of the quantity of interest. This theoretical distribution is not generally available due to its dependence on the form of the $\Gamma(\cdot)$s, which will vary based on the substantive application, and other computational challenges discussed below. An example of using this approach to interpretation would be to compute the model-based distribution of the number of consensus votes in a legislative session. [6]

2.3.4 Estimation of the ERCM

The theoretical complexity accounted for in the ERCM comes at the cost of occasional computational challenges in estimation.. In theory, since the ERCM is a fully parametric exponential family distribution (van Duijn, Gile and Handcock, 2009), Newton-Raphson type hill-climbing algorithms can be used for parameter estimation by maximum likelihood – the same method typically used to estimate parameters in the generalized linear model (i.e. logit, probit, Poisson regression etc.) (Haberman and Renshaw, 1996). However, in practice, the analytic value of the log-likelihood function often cannot be computed within a reasonable time-frame. Recall from equation 2.3 that calculating the likelihood for any one collective decision requires the summation of $\omega(\boldsymbol{y})$ over all possible realizations of $\boldsymbol{y}$ to compute the denominator. The number of realizations is 2^n, where n is the number of actors in the decision-making body. When n is of moderate size, say a nine-member court, meaning there are 512 possible realizations of a collective decision (i.e. the vote on a case) then the likelihood function can be

[6]Note that this approach to model/statistical interpretation is equivalent to the method advocated by King, Tomz and Wittenberg (2000) and implemented in `Clarify` (Tomz, Wittenberg and King, 2003), the `STATA` add-on that is very popular in political science.

computed exactly within a reasonable time frame.[7] In larger scale applications, exact computation of the likelihood function would take years or lifetimes. For example, a roll-call vote among the 100 U.S. Senators can arise in 1.27 nonillion (i.e. 10^{30}) unique configurations. Since the distributional forms and estimation challenges between the ERCM and the ERGM are equivalent, I review methods of approximate inference that have been applied to the ERGM, which can also be used to estimate the ERCM. The current review draws heavily from Desmarais and Cranmer (N.d.).

If the size of the decision-making body is too large, rendering maximum likelihood estimation (MLE) infeasible, there are two popular methods of approximation available – maximum pseudolikelihood (MPLE) (Frank and Strauss, 1986) and Markov Chain Monte Carlo maximum likelihood (MCMC-MLE) (Geyer and Thompson, 1992). The MCMC-MLE algorithm uses a large random sample of collective outcomes from the distribution of possible outcomes to approximate the large sum in the denominator of the likelihood function. This approximate likelihood function is maximized to find parameter estimates. The covariance matrix of the estimates computed as the inverse of the negative Hessian of the final approximate log-likelihood function. Pseudocode for the MCMC-MLE algorithm for the ERCM is given in figure 2.3.4.

Instead of using random sampling, the maximum pseodolikelihood estimator uses an analytic approximation to the likelihood function. In the MPLE, the joint likelihood of the individual choices that compose the collective decision is replaced with the product over the conditional probability of each decision given the other decisions in the instance of collective choice. That is, for a given actor in collective t, the conditional probability of a one, a moderate adaptation of the conditional probability of an edge in an ERGM

[7] It is impossible to arrive at exact figures for what does or does not constitute "moderate" since such a distinction depends completely on the computer technology being used, but anything over 15 (32,768 possible collective outcomes) would pose a formidable challenge for any high-end contemporary desktop.

Figure 2.1: MCMC-MLE Estimation Algorithm

m = number of simulated collective decisions used to approximate likelihood
α = threshold for stopping iterative optimization
$\boldsymbol{\theta}$ = parameter vector
Δ_{ll} = change in log-likelihood
o = indicator for the observed network
LL = log-likelihood

Initialize Δ_{ll} to ∞
Initialize LL to $-\infty$
Initialize $\boldsymbol{\theta}$ to starting values

while($\Delta_{ll} > \alpha$){
1.
    for($\forall\ i \in 1, 2, \ldots, T$){
        Draw and store m collective choices from the i^{th} distribution
        parametrized with $\boldsymbol{\theta}$
    }
2. Using a hill-climbing algorithm, find $\boldsymbol{\theta}^*$ to maximize

$$LL^* = \sum_{t=1}^{T} \log \left[\frac{\omega(\boldsymbol{y}_t)}{\sum_{j=1}^{m} \omega(\boldsymbol{y}_j)} \right]$$

3. Store $\boldsymbol{\theta} = \boldsymbol{\theta}^*$
4. Store $\Delta_{ll} = LL^* - LL$
5. Store $LL = LL^*$
}
$\boldsymbol{\theta}$ is now the MCMC-MLE

provided in (Goodreau, Kitts and Morris, 2009), is given by

$$Pr(y_i^t = 1|\boldsymbol{y}_{-i}^t, \boldsymbol{\theta}) = \text{logit}^{-1}\left[\boldsymbol{\theta}'\Delta_i(\boldsymbol{\Gamma}(\boldsymbol{y}))\right]. \tag{2.5}$$

The expression $\Delta_i(\boldsymbol{\Gamma}(\boldsymbol{y}))$ is the vector of change statistics – the change in the vector of $\Gamma(\boldsymbol{y})$s when y_i is toggled from 0 to 1, holding the rest of $\boldsymbol{y}$ constant at the observed values. It turns out that the conditional probabilities can be computed very quickly since the summation in the denominator does not factor into the conditional likelihoods. The maximum pseudolikelihood is computed by using a hill-climbing algorithm to find the vector of parameters that maximize

$$\log\left(\prod_{t=1}^{T}\prod_{i=1}^{n^t} Pr(y_i^t = 1|\boldsymbol{y}_{-i}^t, \boldsymbol{\theta})^{y_i^t}(1 - Pr(y_i^t = 1|\boldsymbol{y}_{-i}^t, \boldsymbol{\theta}))^{1-y_i^t}\right)$$

This is very convenient in that it can be computed using standard logistic regression software, with the change statistics composing the matrix of independent variables. Conventionally, as with the MCMC-MLE method, an approximate covariance matrix for the maximum pseudolikelihood is formed by inverting the observed information matrix (i.e. the standard covariance matrix derived from the logistic regression software used to compute the MPLE). Desmarais and Cranmer (N.d.) propose an alternative method for computing the covariance matrix with the pseudolikelihood that does not rely on the use of the parametric covariance matrix from logistic regression. Similar to the hierarchical bootstrap methods presented by Field and Welsh (2007) and Harden (Forthcoming), Desmarais and Cranmer (N.d.) suggest the bootstrap resampling of conditionally/Markov independent units.[8] In their application the units are networks; in the current application they are collective choices. Though individual decisions

[8]Markov independence implies that collective outcomes can depend upon previous outcomes (e.g. Court decisions could depend upon precedent).

within instances are not independent and thus cannot be bootstrapped, the T collective decisions are independent and can be resampled.

The choice of which estimator to use is straightforward only if the decision-making body is small. In this instance, analytic maximum likelihood estimation is computationally feasible and should be used. If MLE is infeasible, either MCMC-MLE or MPLE can be used. MPLE represents a more conservative choice in that it is less efficient than MCMC-MLE (van Duijn, Gile and Handcock, 2009), but the confidence intervals derived from the clustered bootstrap exhibit good coverage properties. Little is known regarding the scalability of MCMC-MLE (e.g. should the number of draws used to approximate the denominator increase with the size of the decision-making body?). If MCMC-MLE is computationally feasible, it is advisable to try a few order-of-magnitude different number of draws in the approximation (e.g. 1,000 - 10,000 - 100,000), and rerun the model a few times using different seeds to the random number generator. If there is no substantively meaningful simulation variance in the estimates, the MCMC-MLE is probably the best choice. Studies have found the MPLE to be unbiased (van Duijn, Gile and Handcock, 2009; Desmarais and Cranmer, N.d.), and it is computationally stable in that it does not rely upon simulation. The MPLE combined with bootstrapped covariance estimates is a robust and fast alternative in demanding settings where large decision-making bodies are under study and/or the MCMC-MLE shows signs of instability.

2.3.5 Predictive, Semi-Parametric Fit Comparison of ERCMs

The ERCM is a flexible tool that permits complicated interactive theories to be seamlessly accommodated in models of collective choice. Inevitably, the analyst needs to decide what level of complexity best characterizes the data on-hand. It is critical to assure that the overlapping complexities in the ERCM specification are not combining to explain and falsely attribute structure to pure and possibly outlying noise in the data. This is not a straightforward task for two reasons. First, whereas many methods for detecting outliers in conventional regression analyses have been developed, it is not apparent that any techniques have been designed to test for outlying configurations. Second, since many practical applications will require the use of either MCMC-MLE or pseudolikelihood approximations to the MLE, it is unclear what conventional measures of model fit such as AIC or BIC would mean since they would themselves be approximations and would not be based on the MLE. In this section I describe an approach to model fit assessment that penalizes over-fit models and is agnostic regarding the criterion (e.g. MLE, MCMC-MLE, pseudolikelihood etc.) used to estimate the parameters.

As I alluded to above, a measure of model fit designed for comparison of ERCMs should satisfy three criteria. First, since using ERCMs involves the addition of model complexity to the standard regression framework, it is desirable that the fit metric consciously avoid over-fitting the data – favoring a parsimonious specification. Second, since it will often be impossible to be confident that the ERCM estimates are equivalent to the MLE, the fit measure should be able to determine the relative performance of two models independent of the estimation procedure. Third, the fit metric should be appropriate for the comparison of non-nested models. Such a measure could be used to compare among arbitrarily different ERCMs fit to the same data or between alternative models and ERCMs, and the validity of said comparison would not depend on the estimation procedure.

The cross-validated F-measure (CVFM) of classification performance proposed by Forman and Scholz (2009) meets the three criteria presented above. Leave-k out Cross-validation is a general procedure for judging the performance of an empirical model through out-of-sample predictive fit that efficiently uses the data (Kohavi, 1995; Ward, Siverson and Cao, 2007). It is similar in structure to the computation of Cook's D – the outlier identification statistic commonly used in regression modeling (Cook, 1977). Given some measure of the model's fit to the observations in $\boldsymbol{y}$ (e.g. R^2, proportional reduction in error, BIC etc.) denoted $\Lambda(\boldsymbol{y})$, cross-validation works by computing $\Lambda(\boldsymbol{y}_i)$ on each observation using the parameters estimated with the data excluding the i^{th} observation. If the observations are clustered into groups (i.e. individual choices in collective decisions), the i indexes groups instead of observations (Price, Nero and Gelman, 1996). Thus, every observation is used to estimate (i.e. influence) the parameters and to evaluate the fit of the model, but never at the same time. If a procedure that over-fits the data can be characterized as favoring a more complex model than the true one that generated the data, Jensen and Cohen (2000) prove that cross-validation effectively avoids over-fitting in the context of comparing multiple models.

The selection of $\Lambda(\boldsymbol{y})$ is critical to the performance of cross-validation. Forman and Scholz (2009) compare many candidate measures in the context of leave-k-out cross-validation with binary outcomes (e.g. binary votes in collective decisions). They find that the F-measure, computed precisely in the manner reviewed below, is unbiased and performs well at selecting the right model using cross-validation. The F-Measure is designed to balance the concepts of *precision* and *recall* in the classification of binary outcomes. Precision is defined as the number of true positives divided by the total number of positive predictions – a measure of a model's ability to discriminate between zeros and ones. Recall is defined as the number of true positives over the total number of positive outcomes – penalizing classifiers that simply choose the modal outcome in the context of a rare event. The CVFM is agnostic regarding the method used to estimate the parameters of the predictive model. The model only needs to provide a prediction regarding the modal state of the binary outcome. Let $TP^{(t)}$ be the number of ones in the t^{th} collective decision that are predicted to be ones (i.e. true positives),

$FP^{(t)}$ be the number of zeros in the t^{th} collective decision that are predicted to be ones (i.e. false positives), and $FN^{(t)}$ be the number of ones in the t^{th} collective decision that are predicted to be zeros (i.e. false negatives). The CVFM is then given by

$$CVFM = \frac{2\sum_{t=1}^{T} TP^{(t)}}{2\sum_{t=1}^{T} TP^{(t)} + \sum_{t=1}^{T} FP^{(t)} + 2\sum_{t=1}^{T} FN^{(t)}}. \tag{2.6}$$

The higher the CVFM, the better the fit of the model.[9] The CVFM provides a comprehensive summary of the predictive performance of a model, is semi-parametric in that it only relies on a distribution insofar as one is used to provide the predictions, and can be used to compare multiple explanations of collective decision-making.

2.4 Jurisprudential Regimes in the U.S. Supreme Court

In the current section I demonstrate the use of the ERCM by replicating and extending a model from Richards and Kritzer (2002) of U.S. Supreme Court votes on the merits in all first amendment (i.e. free press, free expression and free speech) cases from the 1953 – 1998. It is a replication in that (1) I use the exact same data from the original article, (2) I include all of the covariates that appeared in the original analysis, and (3) the ERCM can reduce to the original method – logistic regression – if the addition of endogenous terms in the ERCM do not significantly improve the fit of the model. The extension I offer is to include a number of new terms in the model that represent well known configurational theories of decision-making on the Court, but have yet to be appropriately included in models of justice-votes on the merits – an incredibly popular

[9]Unfortunately, like AIC and BIC, the sampling distribution of the CVFM is not available and it is therefore not possible to test hypotheses regarding differences in the CVFM. A bootstrap approach might be possible, but inquiry into this possibility is beyond the scope of the current analysis.

method for testing theory about decision-making on the Supreme Court (see e.g. Segal and Spaeth (1993, 2002) and Johnson, Wahlbeck and Spriggs (2006) for some widely cited recent examples). A conceptualization of the unit analysis that applies to both the original logistic regression model and the ERCM is the collection of justice-votes on a case. There are 570 cases in the data, and therefore 570 collective decisions under study. As is proven in the final section of this chapter, the logistic regression model where the unit of analysis is the individual justice-vote can be considered an ERCM in which the parameters on terms capturing case-level interdependence among the justices' decisions are assumed to be zero.

2.4.1 Richards and Kritzer (2002)

Richards and Kritzer (2002) offer a novel conceptualization of the influence of law or precedent on the decision-making of the U.S. Supreme Court. Instead of influencing decision-making directly (e.g. decision X dictates the Court should rule in a certain direction in today's case) – the conception typically used by Court scholars (Knight and Epstein, 1996; Segal and Spaeth, 1996; Spriggs and Hansford, 2002) – they argue that important cases influence the factors considered by the Court in subsequent cases – a framing that can skew the ideological balance of future decisions. Richards and Kritzer (2002) term the collection of factors considered in cases in a particular area a *jurisprudential regime*. A jurisprudential regime dictates what facts of the case are relevant to the decision, and does not establish what decision the Court should make.

To test their theory of the operation of jurisprudential regimes, Richards and Kritzer (2002) identify the 1972 cases *Chicago Police Department v. Mosley* and *Grayned v. Rockford* as establishing a new jurisprudential regime in first amendment cases. These cases establish that content-based limitations on speech (e.g. passengers cannot to speak of "bombing" in airports) should be subjected to greater scrutiny than content-neutral limitations (e.g. citizens cannot speak at a volume that disturbs the peace). They argue that this regime encourages a liberal shift in first amendment decisions – increasing the likelihood that limitations of speech will be overturned by the Court.

Richards and Kritzer (2002) use logistic regression to study whether a Supreme Court justice is more likely to vote to overturn limitations of speech post-*Grayned* than they were prior to the claimed establishment of the content-neutrality regime. They consider all 4,986 justice-votes in first amendment cases decided in the 1953 – 1998 terms. The dependent variable is binary and coded as (1 = anti-expression rights (conservative), 0 = pro-expression rights (liberal)). They use a dummy variable to indicate whether the justice-vote being considered was cast after the decision in *Grayned* was announced. Their expectation is that the *post-Grayned* indicator variable will have a negative effect on the likelihood a justice will vote against expression rights.

Richards and Kritzer (2002) include a number of controls in their model to protect against faulty inference through omitted variable bias. The controls include a measure of the liberal-conservative ideology of the justices *Justice Ideology* – the Segal-Cover score of the justice – as well as sets of indicator variables that capture characteristics of the Jurisprudence in the case, the action taken to limit expression, characteristics of the governing body limiting expression and indicators for categorizations of the identity of the expression-limited party in the case. [10]

I was unable to replicate the original model exactly. Specifically, there is some inconsistency in the results for the categorical variables. In table 3, I report the original results along with (1) my best attempt at replicating the logistic regression results from the original article, and (2) the results from the extended ERCM.

[10] See the original article for an in-depth justification for the inclusion of these variables as well as specific hypotheses.

2.4.2 Configurational Theories of Supreme Court Decision-Making

As noted above, it is common in many areas of political inquiry to transform or aggregate data in order to optimize its use in testing a particular theory, and scholarship on U.S. Supreme Court decision-making is no exception. The dilemma posed by this practice is that, though these theories may be directly testable on some transformation of justice-vote data, a positive finding in favor of a configurational theory supports the contention that conventional micro-level regression models are misspecified. If the configurational theory operates as a component of the data generating process for justice-votes and, as a result, can be seen in a transformation of justice-vote data, the assumption of correct specification that is common to virtually all data analysis methods used by scholars of the Court is violated – leading to misspecification bias. In this section I present the configurational theories of Supreme Court voting on the merits that I extend the model in Richards and Kritzer (2002) to accommodate. Note that these are *not* covariates that are added to the model. These are characteristics of the case-level outcome that effect the probability of observing particular combinations of votes on cases. They are statistics defined on case-level outcomes that are intended to capture common configurational theories of decision-making on the Court. Though the estimates of their corresponding parameters can be listed in a table in the same way that covariate effects can, there is no way to include them in the columns of a conventional rectangular dataset.

Perhaps the most significant task undertaken by the chief justice (CJ) on the U.S. Supreme Court is the assignment of authorship for the majority opinion (Maltzman, Spriggs and Wahlbeck, 2000). If the majority in the final vote on the merits includes the CJ, the CJ assigns authorship of the written opinion. This creates an incentive, independent of the CJ's personal preference regarding the two sides of a case, for the CJ to vote with the majority. Even if the decision of the Court is opposite the preference of the CJ, he or she may be able to steer the legal influence of the decision by assigning authorship to a more moderate justice. Johnson, Spriggs and Wahlbeck (2005) find that the CJ passes on the conference vote in order to observe the votes of the other justices and eventually vote with the majority. When it comes to the final vote on the merits, if the CJ votes with the majority to assign opinion authorship, the final majority coalition in the vote should include the CJ more often than would be expected based on his or her ideological preferences or the facts of the case. This is a configurational proposition that cannot be tested or accommodated in a logit model of the liberal-conservative direction of justice-votes, but can be in an ERCM. I add to the model of Richards and Kritzer (2002) an indicator that assumes a value of one if the CJ is in the majority coalition and zero otherwise (*CJ in Majority*). It is expected that this measure will have a positive effect on the probability of a case outcome.

Much scholarship on the Court has noted a "freshman effect" for justices in their first term or two after appointment (Dorff and Brenner, 1992; Hagle, 1993; Hurwitz and Stefko, 2004). In this period justices are said to be more moderate in their ideological behavior. More specifically, they demonstrate a greater tendency towards conformity – voting with the majority – than the other justices on the Court at the time (Dorff and Brenner, 1992) or themselves later in their careers (Hurwitz and Stefko, 2004). Dorff and Brenner (1992) demonstrates this by showing that freshman justices are more likely to change their vote from the original conference to the final vote on the merits if they are in the minority in the initial conference vote. Hurwitz and Stefko (2004) show that the ideological extremity of justices is much greater later in their career than earlier. If it is true that freshman justices are more likely to vote with the majority than would be expected based on ideological or case factors, then this should be accounted for in the model of justice-votes on the merits. I add to the model an indicator that assumes a value of one if a freshman justice, defined as a justice serving in his or her first term, is in the majority coalition and zero otherwise (*Freshman Effect*). It is expected that this measure will have a positive effect on the probability of a case outcome.

The final configurational theory I add to the model regards the size of the voting majority in a case. All decisions are not treated equally by lower courts and other subsequent users of Supreme Court doctrine. Johnson (1979) demonstrates that lower courts (U.S. Circuit Courts of Appeals) are more likely to follow Supreme Court decisions that involve a larger voting majority. This creates a clear incentive for justices to work towards consensus in conference (Dorff and Brenner, 1992) – at times going as far as to offer opinion authorship to the opposite side in order to increase the size of the majority coalition (Maltzman, Spriggs and Wahlbeck, 2000). Dorff and Brenner (1992) find that it is particularly difficult for one or two member minorities to hold their ground – leading to a tendency toward either consensus or contentious decisions (i.e. 9-0 or $\{6-3, 5-4\}$ respectively). To accommodate this possibility I add to the model a quadratic in the proportion of justices in the majority *% in Majority* and *[% in Majority]*^{2}. First, if there is truly a tendency toward consensus, then this polynomial should be at its highest value when *% in Majority* is one. Second, if group-level dynamics play out such that small minorities are more easily convinced into the majority, then this polynomial should assume a U-shape – indicating that both consensus and contentious votes are more likely than small minorities. The estimates of the coalitional effects constitute, to my knowledge, the first estimation of general distributional tendencies of the Court that *controls* for the effects of ideological disagreements and other important case-level factors.

Table 2.1: Alternative Models of U.S. Supreme Court Voting

	Original		Logit Replication		ERCM	
	Coefficient	Std. Error	Coefficient	Std. Error	Coeficient	Std. Error
Intercept/Sum	0.420*	0.160	0.893*	0.153	0.794*	0.171
Post-Grayned	-0.350*	0.090	-0.345*	0.088	-0.873*	0.101
Justice Ideology	-1.070*	0.060	-1.072*	0.056	-1.421*	0.065
Jurisprudence						
(Baseline = Less Protected)						
Content-Based	-0.740*	0.090	-0.735*	0.086	-0.303*	0.096
Content-Neutral	0.440*	0.140	0.442*	0.139	0.037	0.152
Below Threshold	1.300*	0.230	1.265*	0.223	0.705	0.239
Identity						
(Baseline = Other)						
Politician	0.100	0.300	-0.600*	0.150	-0.178	0.165
Racial Minority	-0.600*	0.150	-0.120	0.125	0.035	0.139
Alleged Communist	-0.120	0.130	0.534*	0.210	0.048	0.243
Military Protester	0.540*	0.210	-0.331*	0.101	-0.203	0.112
Business	-0.330*	0.100	-0.690*	0.189	-0.278	0.208
Religious	-0.690*	0.190	-0.365*	0.123	-0.109	0.142
Media	-0.030*	0.140	-0.028	0.139	0.025	0.155
Government						
(Baseline = State)						
Other	0.170	0.480	-0.164	0.158	0.038	0.179
Private	0.310*	0.160	-0.577*	0.188	-0.386	0.211
Education	-0.110	0.190	-0.485*	0.092	-0.098	0.103
Local	-0.010	0.100	-0.471*	0.084	-0.190*	0.094
Federal	0.470*	0.080	-0.380*	0.114	-0.151	0.128
Action						
(Baseline = Civil)						
Criminal	-0.380*	0.110	-0.392*	0.122	-0.218	0.137
Deny Expression	-0.390*	0.120	0.610*	0.146	0.243	0.161
Deny Benefit	0.610*	0.150	-0.818*	0.234	-0.223	0.259
Disciplinary	-0.820*	0.240	0.336*	0.167	0.195	0.185
Regulation	0.050*	0.190	0.049	0.187	0.050	0.208
Configurational Statistics						
% Majority	–	–	–	–	-3.934*	0.688
[% Majority]^2	–	–	–	–	9.116*	0.582
CJ in Majority	–	–	–	–	0.276*	0.075
Freshman Effect	–	–	–	–	0.271*	0.110
AIC	–		5,957		5,055	
BIC	–		6,107		5,231	
CVFM	–		0.459		0.488	

Note: Coefficients with standard errors reported. All models estimated on 4,986 justice-votes in 570 cases. * – Statistically significant at the 0.05 level (two-tailed). Dependent variable is coded (0 = Liberal (pro-expression), 1 = Conservative (anti-expression)).

2.4.3 Results

The results from the original logistic regression and extended ERCM are presented in table 3.[11] First, in terms of overall fit, it is apparent that the ERCM out-performs logistic regression. The parametric AIC and BIC are lower in the ERCM and the cross-validated F-measure (CVFM) is higher in the ERCM – all pointing to the ERCM as the better-fitting model. Additionally, with a likelihood ratio statistic of 910 and four degrees of freedom, the validity of the restriction to the original logit model is rejected any any conventional level of statistical significance. This is strong evidence that co-variates do not sufficiently explain commonly observed forms of interdependence on the Court. The configurational terms cannot be ignored without inducing misspecification bias. Case-level interdependence among the decisions made by the justices is as critical to the explanation of decision-making on the U.S. Supreme Court as are the effects of exogenous covariates.

In terms of the effects that appear in both models, the results in the ERCM differ considerably from those in the original logistic regression. First, the main independent variable of interest *Post-Grayned* has an effect nearly twice the magnitude of that in the original model – a difference that is statistically significant at the 0.05 level. In the logit model, the effect of the addition of a conservative vote to a case on the likelihood of observing that case post-*Grayned* is 30% lower that the effect pre-*Grayned*.[12] This percentage decrease is estimated to be 58% in the ERCM. Thus, the original jurisprudential regimes hypothesis garners stronger support in the ERCM than in the logit. However, fourteen of the sixteen categorical effects that differed significantly from their

---

[11] This model, since the maximum number of actors in a Supreme Court case is 9, was estimated by maximum likelihood. The function `optim()` in the `R` statistical software (R Development Core Team, 2009) was used to maximize the likelihood and estimate the Hessian. The standard errors come from the asymptotic covariance matrix derived by inverting the negative Hessian.

[12] This is computed as $100 \times (\exp\{-0.345\} - 1) = -30$.

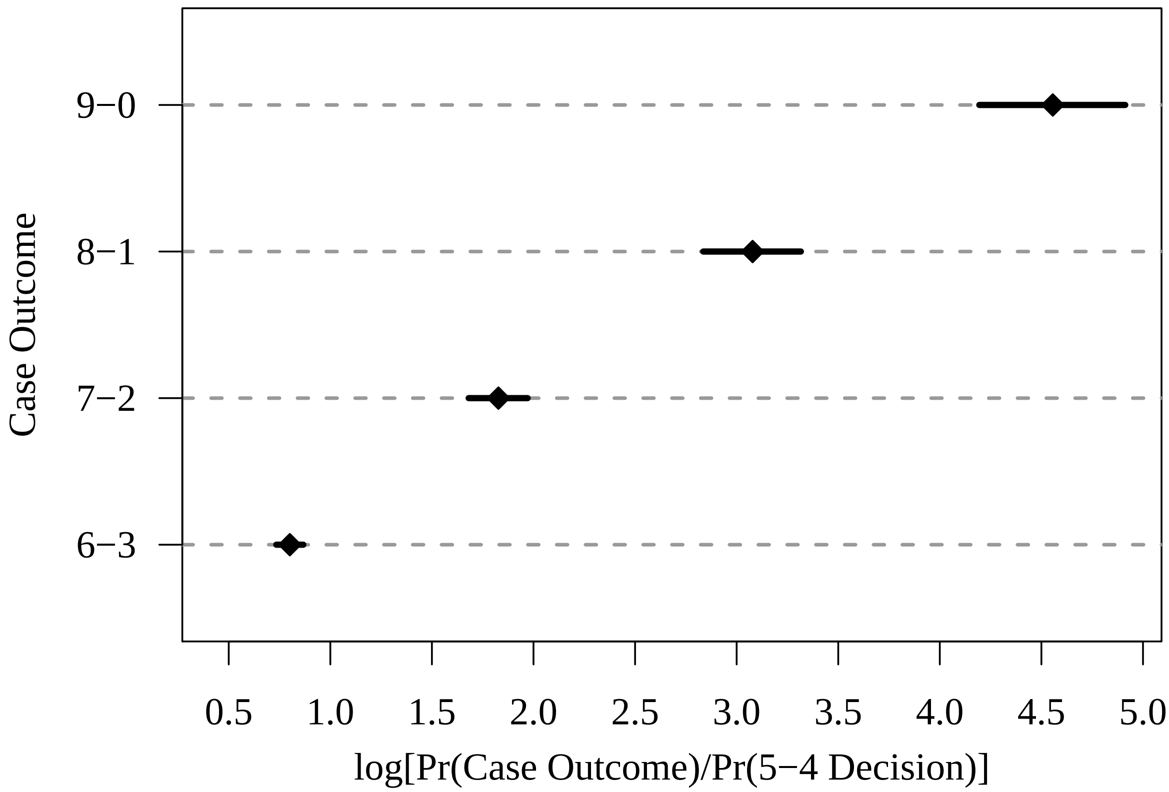
respective baselines in the logit model were not different from the baselines in the ERCM – indicating that many of the inferences on the effects of these variables were incorrect and attributable to misspecification bias. Indeed, if the categorical variables Identity, Government and Action are dropped from the ERCM, a likelihood ratio test indicates with a p-value of 0.257 that the restriction is valid.

Overall the configurational effects work as expected. All else equal, we are more likely to observe cases in which the chief justice votes with the majority than cases in which the chief votes with the minority. Specifically, the probability of observing a particular configuration in a case is increased by 32% if that configuration includes the chief justice in the majority coalition. This supports the contention that the CJ often joins the majority in spite of his or her ideological predisposition. Similarly, all else equal, when a freshman justice is on the Court, we are more likely to observe cases in which the freshman justice votes with the majority than cases in which the he or she votes with the minority. The probability of observing a configuration is 30% higher if the configuration includes a freshman justice in the majority. This is consistent with past notions of the "freshman effect". These findings represent two well-understood and important components of the data generating process for justice-votes on the U.S. Supreme Court, but cannot be accommodated in the logistic regression framework. The ERCM, for the first time, allows scholars of the Court to account for these factors in a model of justice-votes.

The finding regarding the likely majority coalition size is one of a tendency toward unanimity. All else equal, the larger the majority coalition in a particular collective outcome, the more likely we are to observe that particular outcome. The distributional tendencies of the majority coalition size are depicted in figure 2.4.3.[13] The x-axis of this

[13]Using the method presented in King, Tomz and Wittenberg (2000), 10,000 draws from the asymptotic sampling distribution were used to construct the confidence intervals in this plot.

Figure 2.2: Distributional Tendencies of Supreme Court Decisions



Note: The points give the expected logarithm of the ratio of the probability of observing the case outcome on the y-axis to observing a 5-4 decision. The black bars span 95% simulated confidence intervals.

figure gives the natural logarithm of the ratio of the probability of observing the case outcome on the y-axis to observing a 5-4 decision. Values greater than zero indicate that that particular majority-minority combination is more likely to occur than is a 5-4 decision. A value below zero would correspond to a ratio less than one and indicate that a 5-4 decision is more likely than the decision given on the y-axis. None of the 95% confidence intervals in the figure contain a value below zero, indicating that a 5-4 outcome is the least likely outcome. Moreover, the ratio is increasing in the size of the majority coalition, indicating that the likelihood of observing a case outcome is strictly increasing in the size of the majority coalition. The likelihood of observing a 9-0 decision is $\exp\{4.7\} = 109.9$ times the likelihood of observing a 5-4 decision if there is no ideological disagreement or other divisive case factor involved. There is no evidence here for a U-shaped distributional tendency. The *modus operandi* on the Court is the unanimous decision. This tendency towards consensus is interrupted by ideological or legal disagreements to produce non-unanimous decisions.

2.5 Conclusion

In virtually every area of collective choice studied by political scientists there are theories that imply forms of interdependence between the individual decisions that form the collective outcome. This interdependence is as important to the data generating process as the dependence of decisions on exogenous covariates. Yet the literature on collective decision-making has either (1) ignored interdependence in the interest of estimating the effect of exogenous covariates on micro-level choices or (2) transformed individual decision data in order to study the interdependence extant in collective choice. I propose the exponential random configuration model can be used to study determinants of individual decisions as well as any form of dependence between them simultaneously.

In order to accommodate a theory of dependence, researchers need to identify configurations within the collective decision that theory predicts should be more or less likely to occur. The theory is then operationalized as arithmetic functions that measure these configurations. This process of configurational model-building provides an accessible, iterative approach to the construction and testing of interdependent theories of collective choice. An added benefit of using the ERCM to study interdependent choice is that the model reduces to logistic regression on the individual decisions if exogenous covariates constitute the only systematic determinants of the collective outcomes.

To demonstrate the utility of the ERCM in the study of repeated collective choice I replicate and extend a model of the ideological direction of justice-votes on the U.S. Supreme Court, first appearing in Richards and Kritzer (2002). I show how three important configurational theories of decision-making on the Court – the majoritarianism of the chief justice, conformity voting on the part of freshman justices, and the tendency towards consensus – can be accommodated along with covariates to explain justices' votes on the merits in Supreme Court cases. In doing so I show that the effect of jurisprudential regimes on justices' choices was notably underestimated in the original model, and I found evidence that many of the original findings regarding case factors that influence justices' choices are attributable to misspecification bias. There is much opportunity across the subfields in political science for future work in the spirit of this replication. For instance, legislature-specific interaction effects in the ERCM applied to roll-call voting would permit a comparative study of the ways in which institutional and/or electoral rules structure the interdependence in legislative decision-making. In sum, the ERCM represents an empirical approach that does not require researchers to limit theoretical claims to the effects of covariates or to manipulate micro-level data to study coordination and interdependence between political choices.

2.6 Logit - A Special Case of ERCM

I show that ERCM with no endogenous terms reduces to logistic regression. I demonstrate that the conditional probability of an individual actor in a particular instance of collective choice choosing a one, which has a logistic form, is equivalent to the marginal probability if no endogenous terms are included in the $\mathbf{\Gamma}(\boldsymbol{y})$. Equivalence of the conditional and marginal probabilities implies independence of the individual choices – concluding the proof that ERCM reduces to logit if only covariate terms matter among the $\mathbf{\Gamma}(\boldsymbol{y})$.

$$Pr(y_i^t = 1|\boldsymbol{y}_{-i}^t, \boldsymbol{\theta}) = \text{logit}^{-1}\left[\boldsymbol{\theta}'\Delta_i(\mathbf{\Gamma}(\boldsymbol{y}))\right] = \frac{1}{1+\exp\{-\left[\boldsymbol{\theta}'\Delta_i(\mathbf{\Gamma}(\boldsymbol{y}))\right]\}}. \tag{2.7}$$

If $\Gamma_j(\boldsymbol{y})$ is in the form of the statistic given in equation 2.1, then $\Delta_i(\Gamma_j(\boldsymbol{y}))$ – the change in $\Gamma_j(\boldsymbol{y})$ when y_i changes from zero to 1, is x_{ij}^t, so equation 2.7 can be rewritten as

$$Pr(y_i^t = 1|\boldsymbol{y}_{-i}^t, \boldsymbol{\theta}) = \frac{1}{1+\exp\{-\boldsymbol{\theta}'\boldsymbol{x}_i^t\}}, \tag{2.8}$$

where $\boldsymbol{x}_i^t$ is equal to the vector of covariates for the i^{th} observation in the t^{th} collective decision. Note this is exactly equal to the probability of a one in a logistic regression where the $\boldsymbol{\theta}$ represent the regression coefficients. Also, since the conditional probability that $y_i = 1$ does not depend on any other value in $\boldsymbol{y}$, the conditional probability is equal to the marginal probability that $y_i = 1$, which demonstrates that the individual components of $\boldsymbol{y}$ are independent conditional on the covariates. Since the values in $\boldsymbol{y}$ are independent conditional on the covariates, the log-likelihood is constructed as

$$l(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T|\boldsymbol{\theta}) = \log\left(\prod_{t=1}^{t}\prod_{i=1}^{n^t}\left[\frac{1}{1+\exp\{-\boldsymbol{\theta}'\boldsymbol{x}_i\}}\right]^{y_i^t}\left[1-\frac{1}{1+\exp\{-\boldsymbol{\theta}'\boldsymbol{x}_i^t\}}\right]^{1-y_i^t}\right), \tag{2.9}$$

Which is exactly equivalent to the logistic regression likelihood function with the pooled individual decisions as the dependent variable and $\boldsymbol{\theta}$ representing the regression coefficients.

Chapter 3

The U.S. Supreme Court as a Self-Correcting Institution

"*And if every action has a reaction, today's Supreme Court, with two new Bush appointees and a distinctly conservative cast, is something of a reaction to the Warren Court. The modern conservative legal movement was born on campus as a reaction to the Warren era. Several of the court's current members were bred in the cauldron of that movement.*"

- Nina Totenberg (June 30, 2008)

3.1 Introduction

What influence do past decisions rendered by the U.S. Supreme Court exert on its current behavior? Political scientists have provided answers to this question that fall into two general classes. First, many scholars advocate purely ideological explanations of the Court's behavior, contend that existing law places little or no constraint on the current decisions of the Court (Segal and Spaeth, 1996, 2002). Others have found evidence for a controlling influence of *stare decisis* – the tendency of the Court to follow precedent in its current decisions (Brenner and Stier, 1996; Knight and Epstein, 1996; Richards and Kritzer, 2002). I consider a third possibility – that the current Court must over-compensate in its ideological decision-making in the direction and magnitude opposed to the balance of past doctrine in order to desirably locate the relevant line of legal reasoning.

The opening quote exemplifies a common explanation among legal scholars for the recent conservatism of the Court. Many ideologically conservative decisions issued by the U.S. Supreme Court over the last three decades are understood as *negative* reactions to the liberal doctrine of the Warren Court (Fallon, 2002; Feld, 2002; Keck, 2002; Lain, 2004; Kennedy, 2006; Cross, Smith and Tomarchio, 2008; Devins, 2008). Contrary to past conceptions which posit a zero or positive relationship between past decisions and the current behavior of the Court, I theorize that the ideological content of the Court's past work exerts a negative influence on the liberalism of the contemporary Court.[1] Current decisions compete with those of the past to inform the regular activities of legal entities such as administrative agencies, private corporations and lower courts. This creates an incentive for the Court to counter the ideological bias of past decisions in order to move the relevant line of legal reasoning to its ideal point. The oft-cited conservative reaction to the Warren Court is simply an extreme example of the regular tendency of self-correction exhibited by the modern Supreme Court.

The tendency of the Court to move in opposition to the decisions of the past is demonstrated through empirical analyses of the term-level liberalism of the Court from the 1953-2008 terms. This insight has broad implications for the study of the Court. An implicit or explicit assumption underlying existing theories of decision-making on the

[1]At first glance, this may appear to be a restatement of the finding of Segal and Spaeth (1990) that the Court at times practices a reversal strategy. The current theory diverges from this work in two ways. First, the findings of Segal and Spaeth (1990) apply to the Court's relationship with lower courts, and not with past Supreme Court. Second, the explanation for the tendency toward reversal of the lower court is completely different from my explanation for the Court's negative relationship with its past decisions.

Court is that cases are safely considered in isolation. This assumption is questionable. Future theoretical accounts of decision-making on the Court should focus on the ways in which the Court might use cases to achieve the end of an optimal body of legal policy.

3.2 Self-Correction on the U.S. Supreme Court

There are two previous findings on the creation and use of legal policy by the Supreme Court that underlie my argument that the Supreme Court is a self-correcting institution. First, the Supreme Court renders decisions with the objective of moving the implementation of legal policy toward its ideologically preferred position. Given that the Court strives to achieve optimal use of its work, I turn to the nature of said usage by other legal entities. A well-established finding in the literature on compliance with the Court is that subsequent users of law set by the Supreme Court are influenced by an accumulation, or body, of law, rather than critical individual cases.[2] At the end of this section I demonstrate that (1) if the Court is concerned with influencing a legal entity and (2) that entity follows something closer to the average of recent decisions than the most recent decision in isolation, then the Court has a clear incentive to cancel out the ideological bias in past decisions.

3.2.1 The Objective of Influence

Countless studies have documented a strong positive association between objective measures of the liberal-conservative ideology of Court membership and the direction of decision-making (Segal and Cover, 1989; Hagle and Spaeth, 1991; Flemming and Wood,

[2] I am not arguing that landmark cases, which reconstitute legal rules, do not occur; they are simply the exception and not the norm.

1997; Erikson, MacKuen and Stimson, 2002; Grofman and Brazill, 2002; Martin and Quinn, 2002; Lens, 2003; McGuire and Stimson, 2004; Epstein et al., 2007). Though some scholars hold that justices simply want to vote their preferences without regard to compliance with Court decisions (see e.g. Schubert (1962, 1965) and Segal and Spaeth (1993)), the results of many studies align with the observation of Wahlbeck (1998) that, "Court decisions derive significance from the impact of their rules on expected patterns of behavior and their sanctions for violations of those patterns."(pp. 614).

A growing body of work has demonstrated the Court's concern with Congressional compliance with its rulings. Eskridge (1991) provides empirical evidence that the Court modifies its behavior to optimize its influence on subsequent actions of Congress. Spiller and Gely (1992) find evidence of an ideological Congressional constraint on Supreme Court decision-making in labor policy cases. Hansford and Damore (2000) find that when Congress has overridden a statutory decision by the Court, the Court is sensitive to the ideological preferences of Congress in subsequent treatments of the relevant legal questions. And Bergara, Richman and Spiller (2003) show that the probability of a liberal decision increases with the liberalism of members of Congress. These studies illustrate the Courts' attentiveness to the likelihood of favorable implementation of its decisions.

Enjoying stronger empirical support than the Court's responsiveness to Congress is the influence of public opinion on the Court. Attentiveness to public opinion demonstrates the Court's concern with favorable treatment of its rulings. The logic is as follows: the Court's political influence is directly proportional to the degree to which its decisions are heeded by the elected branches. If its decisions are at odds with the preferences of the public, and hence lack public support, elected officials will have an incentive to disregard the decisions of the Court. Supreme Court Justices do not want to be ignored, so they take public opinion into account when issuing decisions (McGuire and Stimson, 2004). A number of studies have reported that public liberalism influences the liberalism of the Court. For example, Mishler and Sheehan (1993) find evidence that the Supreme Court responded positively to public opinion between 1956 and 1981, at which point a negative relationship between public opinion and Supreme Court ideology emerged. Flemming and Wood (1997) find that public opinion is a positive determinant of the ideological behavior of individual justices. Erikson, MacKuen and Stimson (2002); McGuire and Stimson (2004) show that public mood - an annual measure of the aggregate liberalism of the U.S. public - is a positive determinant of aggregate annual Supreme Court liberalism. If the Court were not concerned with the favorable implementation of its rulings, it would have no incentive to bend with public opinion.

It has been firmly established that the U.S. Supreme Court acts to influence relevant legal entities. The Court prefers that subsequent users of its rulings behave in a manner consistent with an ideologically ideal legal policy. Before I can render a theoretical account of the Court's management of legal development over time, it will be helpful to consider the manner in which common legal entities respond to the Court.

3.2.2 Response to Supreme Court Decisions

When the Court strikes down a law, at least when it does so in a high-profile case, it does much more than merely invalidate a particular statute. It sends a pulse into the lawmaking process that can have pervasive effects on a wide range of legislation, and it creates a rhetorical tool that can be used to great effect by ideologically motivated politicians and legislators.

- Larry Kramer (2000, p. 290)

Countless familiar legal entities make active use of the decisions rendered by the

U.S. Supreme Court. These include lower courts (Johnson, 1987; Songer, 1987; Songer and Sheehan, 1990; Songer, Segal and Cameron, 1994; Benesh and Reddick, 2002; Haire, Songer and Lindquist, 2003; Hoekstra, 2005), administrative agencies (Spriggs, 1996, 1997), the U.S. Congress (Hausegger and Baum, 1999), and private corporations (Caldeira and Wright, 1990). Subsequent use of a particular case tends to be by and with rulings that concern a variety of legal issues that do not necessarily address the primary legal rule of concern in the case. It is relatively uncommon that any legal entity takes direction from a single Supreme Court decision in isolation. Thus, when deciding any given case, the Court may be speaking to any number of legal policy areas. By-and-large, through individual decisions, the Court contributes to a large body of relevant legal policy, rather than inform precedent on a single narrow area. Moreover, at the time a case is decided, it is not possible to perfectly predict the range of legal policies that will be informed by that decision. This impossibility is critical to the strategy the Supreme Court must take in ascertaining adherence to its ideal legal policy. The Court will achieve optimal influence by assuring that the aggregate body of law points in its preferred direction.

Political Scientists have demonstrated that it is incomplete to characterize the lower courts as responding to single decisions of the Court. Instead, lower courts tend to follow significant trends in the law. Songer (1987) shows that a statistically significant change in the *aggregate* liberalism of the Supreme Court's antitrust policy is a direct cause of a corresponding change in the antitrust liberalism of the appellate courts, but that the relationship between the legal policy set by the Supreme Court and the appellate courts is missed if the focus is on the impact of individual decisions. Canon (1973) describes the impact of the Supreme Court on the legal policy of the lower courts as being through the consistent build-up of a large body of precedents. In another explication of the gravitational force of prior decisions rendered by the Court,

H. (1974) observes that, "The majority of lower court disavowals [of established lines of precedent] have followed an implicit but substantial modification of a precedent by the Supreme Court."(pp. 516). In other words, the Court must deviate considerably from an established legal rule in order to move lower courts away from the use of that precedent. The implication of this work is that the Supreme Court must build a robust body of supporting law if the objective is to motivate lower courts to follow suit.

Government administrators also recognize and respond to the accumulation of relevant Court decisions rather than individual cases independently. Wahlbeck (1998) explains that individual cases do not make or break legal rules regarding environmental regulation – rather legal rules adapt incrementally through the contributions of new cases. Cooper (1986) explains how government information policy is formed by the body of Freedom of Information Act cases, First Amendment rulings, and the interrelatedness between those two bodies of law. Jaegal and Cayer (1991) review the marked effects that five Supreme Court cases in the latter half of the 20^{th} century have had on the practice of federal public personnel administration. Lens (2001) explains how, through a series of decisions, the Court is entering a "New Era" in its treatment of civil rights and social welfare laws, thus limiting the scope of progressive legislation. These are all clear examples of the influence of the accumulation of legal policy.

Two of the most prevalent targets of Supreme Court rulings – lower courts and administrative agencies – respond not to individual cases but to the summary rule embedded in a line of relevant decisions. A Court that wishes to affect the behavior of legal entities such as these must do so by constructing an aggregate of legal policy located at its ideological ideal. In order to derive a precise empirical prediction regarding the behavior of a self-correcting Court, in the next section I use a simple formal account of the process by which the Court would target an optimal body of law through time.

3.3 The Process of Self-Correction

In this section I posit a simple formal model that captures the process of self-correction undertaken by a Supreme Court concerned with influencing subsequent actors that respond to the average legal rule in a relevant body of law. The Court sets the liberalism of the current term at a point on a unidimensional conservative-liberal continuum, and the subsequent legal entity implements a policy along the same dimension - taking into consideration the liberalism of existing law, including that of the current term. The Court prefers that the subsequent legal entity set policy as close as possible to the Court's ideal point. I will arrive at the expected behavior of the Court through the method of backwards induction. First I describe the policy choice (P) of the subsequent legal entity given the average liberalism of previous decisions (L_0) and the liberalism of the current term (L_t). Next I give the utility the Court derives from setting the current liberalism at L_t given L_0, the Court's ideal policy (C) and the resulting behavior of the subsequent actor. Once the Court's utility is specified, I derive its optimal choice for current liberalism – that which sets $P = C$ (i.e. that which motivates the relevant legal entity to set policy at the Court's ideal point).

Following the findings described in section 3.2.2, it is assumed that the subsequent actor implements a legal policy that represents a combination of past and current decisions. As an abstraction of this process I specify the implemented legal policy P as a weighted average of the past and current liberalism of the Supreme Court. Let $\alpha \in [0, 1]$ be the weight applied to past liberalism, then the expected legal policy is

$$P = \underbrace{\alpha L_0}_{Past\ Terms} + \underbrace{(1-\alpha)L_t}_{Current\ Term}. \tag{3.1}$$

The utility of the Court is given by the negative of the absolute difference between its ideal point and the policy implemented by the legal entity

$$U_C = -|C - P|.$$

Substituting the right-hand-side of equation 3.1 into the Court's utility function gives

$$U_C = -|C - [\alpha L_0 + (1-\alpha)L_t]|.$$

The Court chooses the current liberalism that maximizes its utility. Since the Court's utility is the negative of an absolute value, it is maximized when the term in the absolute value is equal to zero such that

$$C - [\alpha L_0 + (1-\alpha)L_t] = 0.$$

Solving for L_t shows that the optimal level of liberalism in the current time point is

$$L_t^* = \frac{1}{1-\alpha}\left[\underbrace{C}_{\textit{Court Ideology}} - \underbrace{\alpha L_0}_{\textit{Self-Correction}}\right]. \tag{3.2}$$

The under-bracketed terms in equation 3.2 represent the two essential processes undertaken by a Court that targets an optimal body of law through time. The first term (*Court Ideology*) represents the adjustment of current liberalism for current factors determining the Court's preferred policy. This is the Court's ideal equilibrium (i.e. time-invariant) level of liberalism for the body of law. The second term (*Self-Correction*) is the correction for any ideological bias from the past. The expected change in current liberalism due to a one unit increase in past liberalism is given by the first partial derivative of L_t^* with respect to L_0 and is equal to

$$\frac{dL_t^*}{dL_0} = \frac{-\alpha}{1-\alpha}. \tag{3.3}$$

At first glance, it is counter-intuitive that the expected change in current liberalism due to an increase in past liberalism does not depend on the Court's ideology. Conceptualizing adjustment of the body of law as a two-step process motivates the intuition behind this result. The first adjustment step is to negate any past bias in the liberalism of the law – effectively zeroing it out. The next step is to move the liberalism of the law from zero to the Court's preferred position of C. This is not actually a sequential process – it occurs with a single decision, but the two-step metaphor helps to decompose the current behavior of the Court into (1) its reaction to the past and (2) its pursuit of current objectives.

The sign of the reaction to an increase in past liberalism – given in equation 3.3 – is always negative, because $\alpha \in [0, 1]$ and as a result $\frac{\alpha}{1-\alpha}$ is strictly positive. Current Courts will always react in the opposite direction of ideological biases of the past. The conservatism of recent Supreme Courts can be understood through the conservative values of the justices sitting on the Court (i.e. C) and a negative reaction to the extreme liberalism of the Warren Court. The supposed reaction to Warren is sufficient as a motivating example, but does not constitute robust empirical evidence of a self-correcting Court. In the next section I show that the Court has behaved as if it were targeting an optimal body of law from the 1953 to 2008 terms.

3.4 The Empirical Test

3.4.1 Dependent Variables

The objective in this empirical analysis is to test whether the liberalism of past decisions rendered by the U.S. Supreme Court has a negative effect on the current liberalism of the Court, which is the expected relationship under a Court that targets an optimal body of law rather than independent cases. I analyze the term-level liberalism of the

Supreme Court from the 1953-2008 terms. The dependent variable in each analysis is the within-term proportion of cases decided in favor of the the more liberal side. The ideological coding of decisions comes from *The Supreme Court Database* (Spaeth et al., 2010).

There are two common approaches to the analysis of Supreme Court decisions – consideration of term-aggregated liberalism as a time-series (Baum, 1988, 1992; Mishler and Sheehan, 1993; Grofman and Brazill, 2002; McGuire and Stimson, 2004) and pooling decisions over terms, treating them as independent conditional on the covariates (Richards and Kritzer, 2002; Johnson, Wahlbeck and Spriggs, 2006; Wohlfarth, 2009). There are two main reasons why I analyze term-level liberalism. First and more importantly, the test for self-correction requires knowledge of the sequence of decisions. The hypothesis is that the liberalism of past cases has a negative impact on that of current decisions. Within a term, two critical dates are the oral argument date and the date the decision was announced. There is often a non-trivial interval of time between argument and decision, and it is common that cases will be argued and decided between the argument and decision dates of other cases (Hoekstra and Johnson, 2003). These unequal intervals eschew the sequence of events. Averaging over a term eliminates uncertainty about the sequence of events. All cases from one term are both argued and decided before those of the next term. The second justification lies in the difficulty of identifying the relevant legal policy neighborhood of an individual case. The theory states that current liberalism will be negatively related to the liberalism of the existing body of law. It is reasonable to posit that the average case within a term - in which numerous topics are considered - will respond to the average cases from previous terms, but it is less reasonable to consider each individual case responsive to all other cases. Averaging over terms washes out legal-policy idiosyncrasies - an example of aggregation gain (MacKuen, Erikson and Stimson, 1989).

I analyze four different variants of term-level liberalism. The choice of dependent variables is intended to balance aggregation gain and the intra-series relevance of cases. First, I examine the time series including all cases, which offers the greatest gain from aggregation – washing out noise from idiosyncratic legal and political factors that are not included in the model. This constitutes the most general test of self-correcting dynamics on the Court. The one drawback of considering all cases is that the series is derived from combining possibly unrelated cases from different legal areas. Following McGuire and Stimson (2004), I also consider separately term-level liberalism in the legal issue areas Criminal Procedure, Civil Rights and Liberties, and Economic Activity – delineating according to the *Issue Area* coding in the *Supreme Court Database*. There are obviously fewer cases within each issue area than all the cases in a term, but cases in these series are more likely to be topically related – and we would thus be more likely to observe active targeting of an optimal body of relevant law. Next I describe the measurement of the key independent variable - the liberalism of the body of law - and the other control variables.[3]

3.4.2 The Liberalism of the Body of Law

In order to test whether the U.S. Supreme Court self-corrects and thus reacts in opposition to the ideological balance of the existing law, I must measure the liberalism of the law. I use two alternative measures. Since, in deriving the dependent variable for the analysis, I have already measured the term-level liberalism of the Court's output, I use term liberalism as the basic building block of the first measure of the overall liberalism of the law. The liberal balance of the law (*Cumulative Liberalism*) is measured as the

[3]McGuire and Stimson (2004) suggest using only reversals of lower court decisions to compute aggregate measures of Court liberalism. They argue that reversals provide a more sincere ideological signal. In the spirit of robustness, I estimated all of the models with reversals only. The key result of this study – the negative relationship between the liberal balance of the body of law and current liberalism – holds up if only reversals are included.

sum of the proportion of liberal decisions in past terms less 0.5, which is given as

$$CL_t = \sum_{i=1952}^{t-1} \pi_i^{(L)} - 0.5,$$

where $\pi_i^{(L)}$ is the proportion of liberal decisions in the i^{th} term. The constant 0.5 is subtracted for each term so that a term that is more conservative than liberal decreases *Cumulative Liberalism* and vice versa. This variable is expected to have a negative effect on term-level liberalism.

One possible critique of the *Cumulative Liberalism* measure as an operationalization for the current liberalism in the law is that it treats the contribution of the last term and a term from thirty years back as contributing equally to the ideological balance of existing law. Fowler and Jeon (2008), through an analysis of citation patterns between U.S. Supreme Court cases, find that the importance of a case, which increases with the number of cases that cite it and, recursively, the importance of cases that cite it, is not constant over time.[4] Cases generally see an initial upward trend, then decline in importance through time. Fowler and Jeon (2008) derive a measure of the authority of every case ever decided by the Court, which varies from term to term. The measure of authority for case a is proportional to both the number of cases that cite it and the *hub* score of those cases, where the *hub* of case b is a measure of how many important cases are cited by b. The authority scores are estimated by solving a set of equations defined on the entire citation network that express the authority of case a as a sum of the *hub* scores of the cases that cite it.[5] Because the authority of every case is

[4]This measure of importance is based on a recursive algorithm for finding influential works in citation networks developed by Kleinberg (1999). Fowler and Jeon (2008) validate their measure by showing that the list of important cases generated by their method corresponds closely with expert rankings appearing in *The Oxford Guide to United States Supreme Court Decisions*, *Congressional Quarterly's Guide to the United States Supreme Court*, and a list generated by the Legal Information Institute.

[5]See (Fowler and Jeon, 2008, pp. 20) for more details on the authority measure.

measured repeatedly through time, this measure can be used to adjust the measure of the liberalism of the law for the importance life-cycle of a decision. The second measure of the liberalism of the law (*Liberal Authority*) is the proportion of total authority of past cases – the sum of authority scores at time t over past liberal and conservative decisions – that belong to liberal decisions giving

$$LA_t = \frac{\sum_{\forall i \in \boldsymbol{L}_t} \alpha_{it}}{\sum_{\forall i \in \boldsymbol{L}_t} \alpha_{it} + \sum_{\forall i \in \boldsymbol{C}_t} \alpha_{it}},$$

where α_{it} is the authority of the i^{th} case at time t, $\boldsymbol{L}_t$ and $\boldsymbol{C}_t$ are the sets of past liberal and conservative decisions at time t respectively. *Liberal Authority* is expected to have a negative effect on term liberalism. A time-series presentation of the measures of term and body-of-law liberalism variables is rendered in figure 3.4.2.[6]

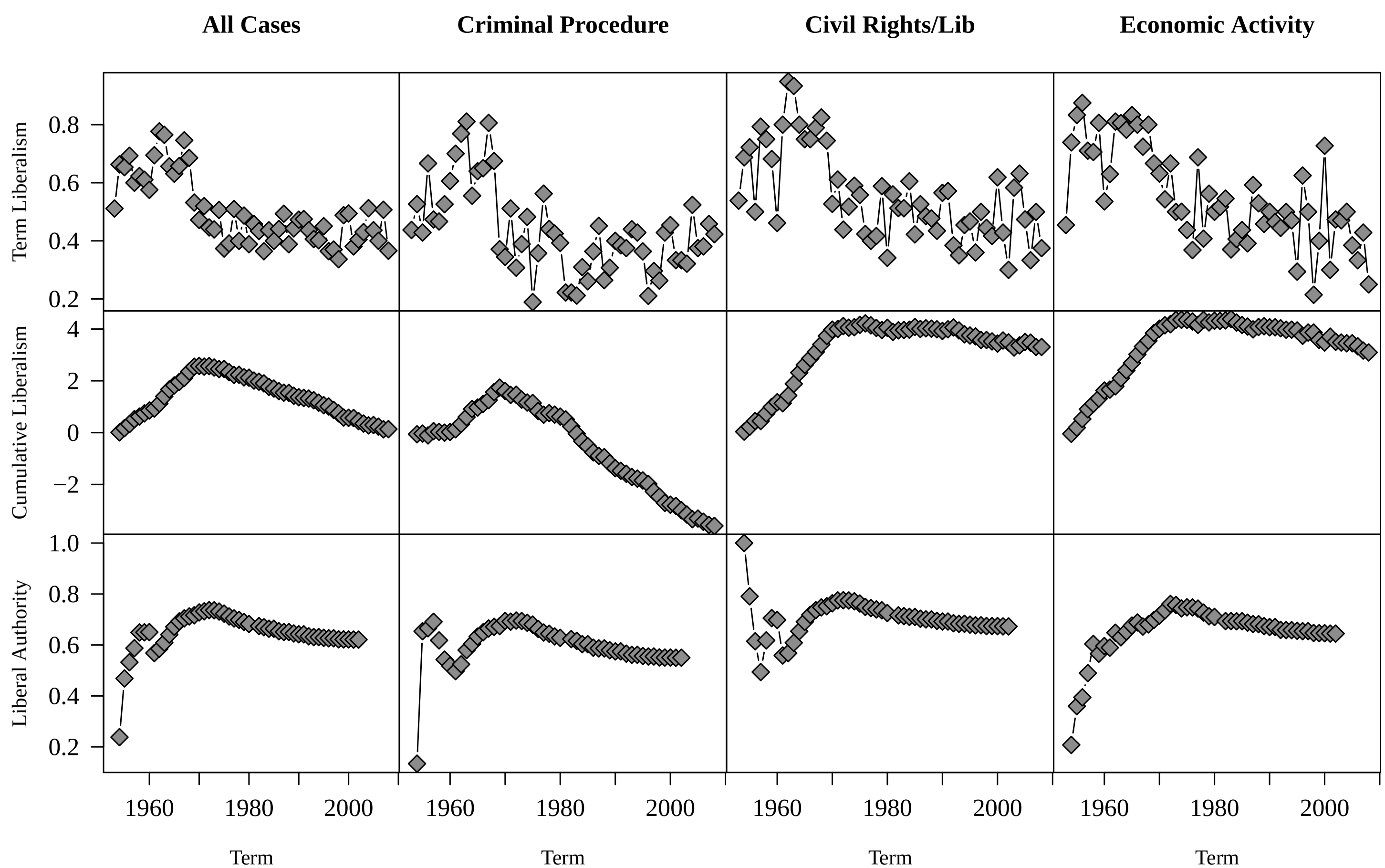
3.4.3 Control Variables and Estimation Strategy

Three control variables are considered. First, a lagged dependent variable is included in most models to account for any autocorrelation in *Term Liberalism*.[7] The second control variable (*Public Opinion*) is an annual measure of the liberalism of the American Public, the measure of public policy mood used by Erikson, MacKuen and Stimson (2002), which McGuire and Stimson (2004) find has a positive effect on the term-level liberalism of the Court. I include *Public Opinion* in all analyses. The last control (*Court Ideology*) is a measure of the ideological preference of the Court and is given as the average Segal-Cover score (Segal and Cover, 1989) of the justices on the Court. *Court Ideology* is included in every model.

[6]Since the authority scores are only available through 2002, models including *Liberal Authority* only include the 1953-2002 terms.

[7]The dynamic specification in these models is notably simple – a lagged dependent variable at most. Though simple, it seems quite adequate. The Ljung-Box test (Ljung and Box, 1978) does not reject the null hypothesis that the residuals are white noise for any of the models estimated.

Figure 3.1: Descriptive Plots



Note: Term-Level Measures of the Liberalism of the U.S. Supreme Court's Past and Present Output.

Ordinary least squares (OLS) is used to estimate the model with all cases included. If the error term in the models contain factors that determine the general liberalism of the Court, there may be correlation between the residuals in the models separated by policy area. More efficient estimates, relative to those produced with OLS, are provided if I account for this correlation using seemingly unrelated regression (SUR) (Conniffe, 1982). The parameters of the Criminal Procedure, Civil Rights and Liberties, and Economic Activity models are estimated as a three-equation system with SUR.[8]

3.4.4 Results

The regression results are presented in tables 3.4.4-4. For each dependent variable, I present five models. The first model (I) serves as a baseline for comparison and does not include either of the measures of the liberalism of the body of law. The second (II) and fourth (IV) – representing purely theoretical specifications – omit the lagged-dependent variable and include *Cumulative Liberalism* and *Liberal Authority* respectively. The third (III) and fifth (V) models – referred to as the *full* models – include lagged liberalism and that of the body of law.[9]

The results for the models including all cases in the dependent variable, presented in table 3.4.4, strongly support my claim that the Court is a self-correcting institution. The estimates of the coefficients on *Cumulative Liberalism* and *Liberal Authority* are negative and statistically significant at the 0.05 level (one-tailed). A term in which

---

[8]The authority scores are available at James Fowler's website `http://jhfowler.ucsd.edu/`. The updated time series of *Public Opinion* is available at James Stimson's website `http://www.unc.edu/~jstimson/`. Segal-Cover scores are available at Jeffrey Segal's website `http://www.stonybrook.edu/polsci/jsegal/`. The statistical software R is used for all computations (R Development Core Team, 2009). The add-on package `systemfit` is used for the SUR (Henningsen and Hamann, 2007). An intercept is estimated in each model, but not reported.

[9]*Cumulative Liberalism* and *Liberal Authority* are not included in the same model due to the fact that they are alternative measures of the same concept – rendering the *ceteris-paribus* interpretation of the coefficients non-sensical – and they are highly correlated.

Table 3.1: Term-Level Liberalism: all Cases Included

	I	II	III	IV	V
Liberalism$_{t-1}$	0.174	–	0.065	–	0.159
	(0.141)	–	(0.132)	–	(0.128)
Cumulative Liberalism	–	-0.036*	-0.0348*	–	–
	–	(0.00969)	(0.0101)	–	–
Liberal Authority	–	–	–	-0.423*	-0.41*
	–	–	–	(0.0978)	(0.0978)
Public Opinion	0.0041*	0.0035	0.00321	0.00723*	0.00657*
	(0.00245)	(0.00214)	(0.00224)	(0.00223)	(0.00228)
Court Ideology	0.367*	0.491*	0.461*	0.411*	0.337*
	(0.079)	(0.0467)	(0.0767)	(0.0489)	(0.0765)
Adjusted-R^2	0.733	0.784	0.781	0.819	0.822
N	54	54	54	47	47

Note: Analysis of term level liberalism in all cases. OLS coefficients are reported with standard errors in parentheses. * = Significant at the 0.05 level (one-tailed).

60% of the cases are decided in a liberal direction contributes 0.10 to *Cumulative Liberalism*, causing an approximate 0.0035 decrease in the following term's liberalism, a (0.10-0.0035)0.35 = 0.0338 decrease in the liberalism of the term after the next and so forth. A liberal term sets the Court on a conservative trajectory, and a conservative term a liberal trajectory. In terms of the control variables, the effect of *Public Opinion* is positive in all of the models and statistically significantly so in three of the models – indicating that the Court follows the ideology of the public. Lastly, the effect of *Court Ideology* is as expected and strongly significant in every equation. It is seen that the typical control for dynamics, the lagged dependent variable, exerts little effect. It is not statistically significant in any of the models, and reduces the Adjusted R^2 from model II to III.

When it comes to the individual legal policy areas, the results – presented in tables 2-5 – are largely indicative of a Court that targets an optimal body of law. Every coefficient estimate is negative for both *Cumulative Liberalism* and *Liberal Authority* across policy areas. Moreover, the effects are statistically significant in the areas of criminal procedure and economic activity, but not in civil rights and liberties cases. The lack of significance in the area of civil rights and liberties may be due to the non-technical nature of the issues considered in that area. In the formal model presented in the previous section, the parameter α determined the relative influence of current decisions on the behavior of subsequent actors. In effect, this parameter determines the degree to which the Court can overwrite past legal policy with a single decision. If this parameter is set at its maximum of 1, legal policy can be rewritten with a single decision, the only important decision in the eyes of legal actors is the Court's current policy, and there is no self-correction on the Court. Since issues considered in the area of civil rights and liberties are not very technical relative to other areas (Tate, 1981) and thus less likely to lead to ambiguity as to when and where decisions apply, it is easier to redirect the body of law with a single decision. It is likely that α is much higher in civil rights and liberties than in economic activity and criminal procedure, leading to a more moderate effect that requires a longer time series to uncover with precision. In terms of the control variables, the results herein diverge from those of McGuire and Stimson (2004). I find that *Public Opinion* has a statistically significant positive effect in economic activity but not in civil rights and liberties cases, whereas McGuire and Stimson (2004) find the opposite.

Table 3.2: Term-Level Liberalism in Criminal Procedure Cases

	I	II	III	IV	V
Liberalism$_{t-1}$	0.227*	–	0.194*	–	0.236*
	(0.127)	–	(0.126)	–	(0.132)
Cumulative Liberalism	–	-0.0298*	-0.0263*	–	–
	–	(0.013)	(0.0129)	–	–
Liberal Authority	–	–	–	-0.301*	-0.309*
	–	–	–	(0.171)	(0.167)
Public Opinion	0.0101*	0.00768*	0.00684	0.00953*	0.00831*
	(0.00398)	(0.00424)	(0.00419)	(0.00437)	(0.00427)
Court Ideology	0.289*	0.612*	0.498*	0.459*	0.338*
	(0.101)	(0.127)	(0.144)	(0.0956)	(0.114)
Adjusted R^2	0.558	0.559	0.578	0.56	0.59
N	54	54	54	47	47

Note: Analysis of term-level liberalism in criminal procedure cases. OLS coefficients are reported with standard errors in parentheses. * = Significant at the 0.05 level (one-tailed).

Table 3.3: Term-Level Liberalism in Civil Rights and Liberties Cases

	I	II	III	IV	V
Liberalism$_{t-1}$	0.0121	–	0.0221	–	0.0303
	(0.133)	–	(0.134)	–	(0.147)
Cumulative Liberalism	–	-0.0152	-0.0152	–	–
	–	(0.0148)	(0.0151)	–	–
Liberal Authority	–	–	–	-0.273	-0.288
	–	–	–	(0.22)	(0.223)
Public Opinion	0.00581	0.00591	0.00576	0.00414	0.00383
	(0.004)	(0.00386)	(0.00401)	(0.00464)	(0.00475)
Court Ideology	0.54*	0.486*	0.475*	0.59*	0.576*
	(0.107)	(0.101)	(0.126)	(0.0933)	(0.123)
Adjusted R^2	0.591	0.597	0.59	0.644	0.637
N	54	54	54	47	47

Note: Analysis of term liberalism in civil rights and liberties cases. OLS coefficients are reported with standard errors in parentheses. * = Significant at the 0.05 level (one-tailed).

The empirical results are consistent with a Court that targets an ideologically optimal body of law by canceling out ideological imbalances from the past. The statistical power gained from combining the policy areas appears to outweigh any loss from combining unrelated cases into a single measure. In the models including all cases, the effects of *Cumulative Liberalism* and *Liberal Authority* are statistically significantly negative and larger in magnitude than in any of the individual legal domains. This is further evidence of a strong tendency to correct general ideological imbalances of its past.

Table 3.4: Term-Level Liberalism in Economic Activity Cases

	I	II	III	IV	V
Liberalism$_{t-1}$	-0.0841	–	-0.076	–	-0.1
	(0.134)	–	(0.132)	–	(0.139)
Cumulative Liberalism	–	-0.0282*	-0.0276*	–	–
	–	(0.0148)	(0.015)	–	–
Liberal Authority	–	–	–	-0.315*	-0.303*
	–	–	–	(0.161)	(0.164)
Public Opinion	-0.00102	-0.00132	-0.00119	0.00321	0.00335
	(0.00406)	(0.00396)	(0.00398)	(0.00456)	(0.00458)
Court Ideology	0.734*	0.58*	0.629*	0.55*	0.614*
	(0.12)	(0.0998)	(0.131)	(0.106)	(0.137)
Adjusted R^2	0.607	0.626	0.623	0.605	0.604
N	54	54	54	47	47

Note: Analysis of term level liberalism in economic activity cases. OLS coefficients are reported with standard errors in parentheses. * = Significant at the 0.05 level (one-tailed).

3.5 Conclusion

Many scholars claim that the conservative jurisprudence of recent Supreme Courts can be understood as a reaction to the liberal doctrine of the Warren Court. In other words, the body of law has taken a rightward turn *because* of a prolonged leftward trajectory. This phenomenon, that the Court moves in the opposing direction of the ideological balance of its past doctrine, has never been considered in the context of a general theoretical or empirical model. It is well-established that (1) the Court is concerned with subsequent preferable treatment of its decisions and (2) future users of its decisions tend to respond to a rich build-up of law rather than individual cases in isolation. I show that these two facts combine to create an incentive for the Court to target an ideologically optimal body of law, correcting (i.e. negatively reacting to) ideological imbalances of its past. Empirical analysis of the ideological content of the

Supreme Court's decisions on the merits from the 1953-2008 terms provides strong support for the theoretical model. The modern Court's reaction to the Warren years is not an isolated instance. In general, the Court strives for an ideologically optimal body of law, and in so doing acts to cancel out ideological biases of the past.

The primary implication of the main finding in the current analysis is that scholarship on the U.S. Supreme Court should proceed with the assumption that the Court is striving to produce an ideologically optimal body of law rather than individual cases. This finding represents a foundational change to the common theoretical approach of treating the individual case as the independent unit of analysis. Reconsideration of many relationships – from the impact of public opinion to the strategic considerations in opinion assignment – in the context of a Court striving for an ideologically ideal line of decisions may lead to theoretical innovations in the explanation of Court decision-making. For instance, a theoretical perspective – that the Court responds to precedent – has long eluded strong empirical support. A simple reading of the theory of *stare decisis* would predict that liberal precedents would lead to liberal decisions. A self-correcting Court would react in the opposite way, turning in a conservative direction in response to liberally biased preceding decisions. The negative reaction to previous law is subtle, a few percentage points at most. It is possible that (1) the tendency for self-correction has muddled the effect of precedent in past studies and/or (2) a constraining impact of precedent tempers the tendency for self correction in the current study. Future studies should consider the ways in which precedent would effect a self-correcting Court.

Bibliography

Achen, Christopher H. 1977. "Measuring Representation: Perils of the Correlation Coefficient." *American Journal of Political Science* 21(4):805–815.

Achen, Christopher H. 2005. "Let's put garbage-can regressions and garbage-can probits where they belong." *Conflict Management and Peace Science* 22(4):327.

Aldrich, John H. and James S. Coleman Battista. 2002. "Conditional Party Government in the States." *American Journal of Political Science* 46(1).

Altun, Yasemin, Alex J. Smola and Thomas Hofmann. 2004. Exponential families for conditional random fields. In *UAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*. Arlington, Virginia, United States: AUAI Press pp. 2–9.

Alvarez, R. Michael and Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42(1):55–96.

Bates, Douglas and Deepayan Sarkar. 2006. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.9975-10.

Baum, Lawrence. 1988. "Measuring Policy Change in the U.S. Supreme Court." *American Political Science Review* 82:905–912.

Baum, Lawrence. 1992. "Membership Change and Collective Voting Change in the United States Supreme Court." *The Journal of Politics* 54(1):3–24.

Beck, Nathaniel. 1982. "Parties, Administrations, and American MacroeconomicOutcomes." *The American Political Science Review* 76:83–93.

Benesh, Sara C. and Malia Reddick. 2002. "Overruled: An Event History Analysis of Lower Court Reaction to Supreme Court Alteration of Precedent." *The Journal of Politics* 64(2):534–550.

Bergara, Mario, Barak Richman and Pablo T. Spiller. 2003. "Modeling Supreme Court Strategic Decision Making: The Congressional Constraint." *Legislative Studies Quarterly* 28(2):247–280.

Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2):192–236.

Brenner, Saul and Marc Stier. 1996. "Retesting Segal and Spaeth's Stare Decisis Model." *American Journal of Political Science* 40(4):1036–1048.

Burman, Prabir. 1989. "A Comparative Study of Ordinary Cross-Validation, v-Fold Cross-Validation and the Repeated Learning-Testing Methods." *Biometrika* 76(3):503–514.

Butler, Michael J. 2003. "U.S. Military Intervention in Crisis, 1945-1994: An Empirical Inquiry of Just War Theory." *The Journal of Conflict Resolution* 47(2):226–248.

Caffo, Brian, Ming-Wen An and Charles Rohde. 2007. "Flexible random intercept models for binary outcomes using mixtures of normals." *Computational Statistics and Data Analysis* 51(11).

Caldeira, Gregory A. and John R. Wright. 1990. "Amici Curiae before the Supreme Court: Who Participates, When, and How Much?" *The Journal of Politics* 52(3):782–806.

Cameron, Charles, David Epstein and Sharyn O'Halloran. 1996. "Do Majority-Minority Districts Maximize Substantive Black Representation in Congress?" *The American Political Science Review* 90(4):794–812.

Canon, B.C. 1973. "Reactions of state supreme courts to a US Supreme Court civil liberties decision." *Law and Society Review* pp. 109–134.

Clarke, Kevin A. 2003. "Nonparametric Model Discrimination in International Relations." *Journal of Conflict Resolution* 47(1):72–93.

Clarke, Kevin A. 2007. "A Simple Distribution-Free Test for Nonnested Hypotheses." *Political Analysis* 15(3):347–363.

Collins Jr, Paul M. 2008. *Friends of the Supreme Court: interest groups and judicial decision making.* Oxford University Press, USA.

Collins, Paul M. Jr. 2004. "Friends of the Court: Examining the Influence of Amicus Curiae Participation in U.S. Supreme Court Litigation." *Law & Society Review* 38(4):807–832.

Collins, Paul M. Jr. 2007. "Lobbyists before the U.S. Supreme Court: Investigating the Influence of Amicus Curiae Briefs." *Political Research Quarterly* 60(1):55–70.

Conniffe, D. 1982. "Covariance Analysis and Seemingly Unrelated Regressions." *The American Statistician* 36(3):169–171.

Cook, R. Dennis. 1977. "Detection of Influential Observation in Linear Regression." *Technometrics* 19(1):15–18.

Cooper, Phillip J. 1986. "The Supreme Court, the First Amendment, and Freedom of Information." *Public Administration Review* 46(6):622–628.

Cover, Thomas. M. and Joy A. Thomas. 1991. *Elements of Information Theory*. Hoboken, NJ: Wiley-Interscience.

Cox, D. R. and E. J. Snell. 1971. "On Test Statistics Calculated from Residuals." *Biometrika* 58(3):589–594.

Crescenzi, Mark J. C. 2007. "Reputation and Interstate Conflict." *American Journal of Political Science* 51(2):382–396.

Cross, Frank B., Thomas A. Smith and Antonio Tomarchio. 2008. "The Reagan Revolution in the Network of Law." *Emory Law Journal* 57(1227).

Desmarais, Bruce and Skyler Cranmer. N.d. "The Temporal Exponential Random Graph Model for the Longitudinal Analysis of Social Networks." *Presented at the Midwest Political Science Association Annual Meeting, Chicago* .

Devins, Neal. 2008. "Symposium: Precedent and the Roberts Court." *North Carolina Law Review* 86:1399.

Diehl, Paul F., Daniel Druckman and James Wall. 1998. "International Peacekeeping and Conflict Resolution: A Taxonomic Analysis with Implications." *The Journal of Conflict Resolution* 42(1):33–55.

Dorff, Robert H. and Saul Brenner. 1992. "Conformity Voting on the United States Supreme Court." *The Journal of Politics* 54(3):762–775.

Epstein, L., J.A. Segal and H.J. Spaeth. 2001. "The norm of consensus on the US Supreme Court." *American Journal of Political Science* pp. 362–377.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal and Chad Westerland. 2007. "The Judicial Common Space." *Journal of Law, Economics and Organization* 23(2):303–325.

Erikson, Robert S., Michael B. MacKuen and James A. Stimson. 2002. *The Macro Polity*. New York: Cambridge University Press.

Eskridge, William N., Jr. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *The Yale Law Journal* 101(2):331–455.

Fallon, Richard H., Jr. 2002. "The "Conservative" Paths of the Rehnquist Court's Federalism Decisions." *The University of Chicago Law Review* 69(2):429–494.

Faust, Katherine and John Skvoretz. 2002. "Comparing Networks across Space and Time, Size and Species." *Sociological Methodology* 32:267–299.

Feld, Barry C. 2002. "Race, Politics, and Juvenile Justice: The Warren Court and the Conservative Backlash." *Minnesota Law Review* 87:1447.

Field, Christopher A. and A. H. Welsh. 2007. "Bootstrapping clustered data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69(3):369 – 390.

Fisher, Ronald J. and Loraleigh Keashly. 1991. "The Potential Complementarity of Mediation and Consultation within a Contingency Model of Third Party Intervention." *Journal of Peace Research* 28(1):29–42.

Flemming, R.B. and B.D. Wood. 1997. "The public and the Supreme Court: Individual justice responsiveness to American policy moods." *American Journal of Political Science* pp. 468–498.

Forman, George and Martin Scholz. 2009. Apples-to-Apples in Cross-Validation Studies: Pitfalls in Classifier Performance Measurement. Technical Report HPL-2009-359 Hewlett-Packard Laboratories.
URL: *http://www.hpl.hp.com/techreports/2009/HPL-2009-359.html*

Fowler, J.H. and S. Jeon. 2008. "The authority of Supreme Court precedent." *Social Networks* 30(1):16–30.

Frank, Ove and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81(395):832–842.

Franzese, Robert J., Jr and Jude C. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15(2):140–164.

Gartzke, Erik and Kristian Skrede Gleditsch. 2004. "Why Democracies May Actually Be Less Reliable Allies." *American Journal of Political Science* 48(4):775–795.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, MA: Cambridge University Press.

Gelman, Andrew, Xiao L. Meng and Hal Stern. 1996. "Posterior predictive assessment of model fitness via realized discrepancies." *Statistica Sinica* 6:733–759.

Geyer, Charles J. and Elizabeth A. Thompson. 1992. "Constrained Monte Carlo Maximum Likelihood for Dependent Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 54(3):657–699.

Gibler, Douglas M. 2008. "The Costs of Reneging: Reputation and Alliance Formation." *Journal of Conflict Resolution* 52(3):426–454.

Gill, Jeffrey. 2002. *Bayesian methods: A social and behavioral sciences approach*. CRC press.

Gleditsch, Kristian Skrede and Kyle Beardsley. 2004. "Nosy Neighbors: Third-Party Actors in Central American Conflicts." *The Journal of Conflict Resolution* 48(3):379–402.

Goodreau, Steven M. 2007. "Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network." *Social Networks* 29(2):231–248.

Goodreau, Steven M., J. Kitts and M. Morris. 2009. "Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks." *Demography* 46(1):103–25.

Greene, William H. 2008. *Econometric Analysis*. Prentice Hall Upper Saddle River, NJ.

Grofman, Bernard and Timothy J. Brazill. 2002. "Identifying the Median Justice on the Supreme Court through Multidimensional Scaling: Analysis of "Natural Courts" 1953-1991." *Public Choice* 112(1/2):55–79.

H., P. K. 1974. "Lower Court Disavowal of Supreme Court Precedent." *Virginia Law Review* 60(3):494–539.

Haberman, Steven and Arthur E. Renshaw. 1996. "Generalized Linear Models and Actuarial Science." *Journal of the Royal Statistical Society. Series D (The Statistician)* 45(4):407–436.

Hagle, Timothy M. 1993. ""Freshman Effects" for Supreme Court Justices." *American Journal of Political Science* 37(4):1142–1157.

Hagle, Timothy M. and Harold J. Spaeth. 1991. "Voting Fluidity and the Attitudinal Model of Supreme Court Decision Making." *The Western Political Quarterly* 44(1):119–128.

Hahsler, Michael, Bettina Grn and Kurt Hornik. 2005. "arules - A Computational Environment for Mining Association Rules and Frequent Item Sets." *Journal of Statistical Software* 14(15):1–25.

Haire, Susan B., Donald R. Songer and Stefanie A. Lindquist. 2003. "Appellate Court Supervision in the Federal Judiciary: A Hierarchical Perspective." *Law & Society Review* 37(1):143–168.

Hansford, T.G. and D.F. Damore. 2000. "Congressional Preferences, Perceptions of Threat, and Supreme Court Decision Making." *American Politics Research* 28(4):490.

Harden, Jeffrey J. Forthcoming. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics and Policy Quarterly* .

Harden, Jeffrey J. n.d. "A Comparison of Methods for Modeling Multilevel Data." Unpublished manuscript.

Hartley, Keith and Todd Sandler. 1999. "NATO Burden-Sharing: Past and Future." *Journal of Peace Research* 36(6):665–680.

Hausegger, Lori and Lawrence Baum. 1999. "Inviting Congressional Action: A Study of Supreme Court Motivations in Statutory Interpretation." *American Journal of Political Science* 43(1):162–185.

Haynie, S.L. 1992. "Leadership and consensus on the US Supreme Court." *The Journal of Politics* pp. 1158–1169.

Henningsen, Arne and Jeff D. Hamann. 2007. "systemfit: A Package for Estimating Systems of Simultaneous Equations in R." *Journal of Statistical Software* 23(4):1–40.

Hix, Simon. 2002. "Parliamentary Behavior with Two Principals: Preferences, Parties, and Voting in the European Parliament." *American Journal of Political Science* 46(3):688–698.

Hix, Simon, Abdul Noury and Gerard Roland. 2005. "Power to the Parties: Cohesion and Competition in the European Parliament, 1979–2001." *British Journal of Political Science* 35(02):209–234.

Hjorth, J.S. Urban. 1993. *Computer Intensive Statistical Methods*. Boca Raton, FL: Chapman and Hall/CRC.

Hoekstra, Valerie. 2005. "Competing Constraints: State Court Responses to Supreme Court Decisions and Legislation on Wages and Hours." *Political Research Quarterly* 58(2):317–328.

Hoekstra, Valerie and Timothy Johnson. 2003. "Delaying Justice: The Supreme Court's Decision to Hear Rearguments." *Political Research Quarterly* 56(3):351–360.

Holland, Paul W. and Samuel Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76(373):33–50.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau and Martina Morris. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24(3):1–29.

Hurwitz, Mark S. and Joseph V. Stefko. 2004. "Acclimation and Attitudes: "Newcomer" Justices and Precedent Conformance on the Supreme Court." *Political Research Quarterly* 57(1):121–129.

Ireland, Michael J. and Scott Sigmund Gartner. 2001. "Time to Fight: Government Type and Conflict Initiation in Parliamentary Systems." *The Journal of Conflict Resolution* 45(5):547–568.

Jaegal, Don and N. Joseph Cayer. 1991. "Public Personnel Administration by Lawsuit: The Impact of Supreme Court Decisions on Public Employee Litigiousness." *Public Administration Review* 51(3):211–221.

Jensen, David D. and Peter R. Cohen. 2000. "Multiple comparisons in induction algorithms." *Machine Learning* 38(3):309–338.

Johnson, Charles A. 1979. "Lower Court Reactions to Supreme Court Decisions: A Quantitative Examination." *American Journal of Political Science* 23(4):792–804.

Johnson, Charles A. 1987. "Law, Politics, and Judicial Decision Making: Lower Federal Court Uses of Supreme Court Decisions." *Law & Society Review* 21(2):325–340.

Johnson, Timothy R., James F. Spriggs and Paul J. Wahlbeck. 2005. "Passing and Strategic Voting on the U.S. Supreme Court." *Law & Society Review* 39(2):349–377.

Johnson, T.R., P.J. Wahlbeck and J.F. Spriggs. 2006. "The Influence of Oral Arguments on the US Supreme Court." *American Political Science Review* 100(01):99–113.

Kagan, Robert. 2004. "America's Crisis of Legitimacy." *Foreign Affairs* 83(2):65–87.

Keck, Thomas M. 2002. "Activism and Restraint on the Rehnquist Court: Timing, Sequence, and Conjuncture in Constitutional Development." *Polity* 35(1):121–152.

Kennedy, Joseph E. 2006. "Just Right?: Assessing the Rehnquist Court's Parting Words on Criminal Justice: Cautious Liberalism." *Georgetown Law Journal* 94:1537–1993.

Kim, Chae-Han. 1991. "Third-Party Participation in Wars." *The Journal of Conflict Resolution* 35(4):659–677.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):347–361.

Kleinberg, J.O.N.M. 1999. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM* 46(5):604–632.

Knight, J. and L. Epstein. 1996. "The Norm of Stare Decisis." *AMERICAN JOURNAL OF POLITICAL SCIENCE* 40:1018–1035.

Kohavi, Ron. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. pp. 1137–1143.

Kramer, Larry D. 2000. "Putting the Politics Back into the Political Safeguards of Federalism." *Columbia Law Review* 100(1):215–293.

Lain, Corinna Barrett. 2004. "Countermajoritarian Hero or Zero? Rethinking the Warren Court's Role in the Criminal Procedure Revolution." *University of Pennsylvania Law Review* 152(4):1361–1452.

Lango, John W. 2005. "Preventive Wars, Just War Principles, and the United Nations." *The Journal of Ethics* 9(1/2):247–268.

Lebovic, James H. 2004. "Uniting for Peace? Democracies and United Nations Peace Operations after the Cold War." *The Journal of Conflict Resolution* 48(6):910–936.

Leeds, Brett Ashley. 2005. "Alliance Treaty Obligations and Provisions (ATOP) Codebook." *Houston: Rice University, Department of Political Science* .

Lens, Vicki. 2001. "The Supreme Court, Federalism, and Social Policy: The New Judicial Activism." *The Social Service Review* 75(2):318–336.

Lens, Vicki. 2003. "Reading between the Lines: Analyzing the Supreme Court's Views on Gender Discrimination in Employment, 1971-1982." *The Social Service Review* 77(1):25–50.

Leo, Breiman, Jerome Friedman, Charles Stone and R.A. Olshen. 1984. *Classification and Regression Trees*. Boca Raton, FL: Chapman and Hall/CRC.

Ljung, G. M. and G. E. P. Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65(2):297–303.

Luo, Ke and Xue-Mao Zhang. 2007. An Efficient Frequent Itemset Mining Algorithm. Vol. 2 pp. 756 –761.

MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1989. "Macropartisanship." *American Political Science Review* 83:1125–1142.

Maltzman, F., J.F. Spriggs and P.J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. Cambridge University Press.

Maoz, Zeev, Ranan D. Kuperman, Lesley Terris and Ilan Talmud. 2006. "Structural Equivalence and International Conflict: A Social Networks Analysis." *The Journal of Conflict Resolution* 50(5):664–689.

Martin, A.D. and K.M. Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.

McGuire, Kevin T. and James A. Stimson. 2004. "The Least Dangerous Branch Revisited: New Evidence on Supreme Court Responsiveness to Public Preferences." *The Journal of Politics* 66(4):1018–1035.

Meng, Xiao-Li. 1994. "Posterior Predictive p-Values." *The Annals of Statistics* 22(3):1142–1160.

Michael Hahsler, Bettina Gruen and Kurt Hornik. 2009. *arules: Mining Association Rules and Frequent Itemsetse*. R package version 0.9-6.

Mishler, William and Reginald S. Sheehan. 1993. "The Supreme Court as aCountermajoritarian Institution? The Impact of Public Opinion on SupremeCourt Decisions." *The American Political Science Review* 999:999.

Ostrom, Elinor. 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action: Presidential Address, American Political Science Association, 1997." *The American Political Science Review* 92(1):1–22.

O'Sullivan, Finbarr, Brian S. Yandell and Jr. Raynor, William J. 1986. "Automatic Smoothing of Regression Functions in Generalized Linear Models." *Journal of the American Statistical Association* 81(393):96–103.

Park, David K., Andrew Gelman and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12:375–385.

Price, Phillip N., Anthony V. Nero and Andrew Gelman. 1996. "Bayesian prediction of mean indoor radon concentrations for Minnesota counties." *Health Physics* 71(6):922–936.

R Development Core Team. 2009. *R: A language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Richards, M.J. and H.M. Kritzer. 2002. "Jurisprudential regimes in Supreme Court decision making." *American Political Science Review* 96(02):305–320.

Robins, Garry. 2009. Exponential Random Graph (p*) Models for Social Networks. In *Encyclopedia of Complexity and Systems Science*. Springer New York pp. 8319–8333.

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

Rubin, Donald B. 1984. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applies Statistician." *The Annals of Statistics* 12(4):1151–1172.

Rust, Roland T. and David C. Schmittlein. 1985. "A Bayesian Cross-Validated Likelihood Method for Comparing Alternative Specifications of Quantitative Models." *Marketing Science* 4(1):20–40.

Sandler, Todd. 2005. "Collective versus Unilateral Responses to Terrorism." *Public Choice* 124(1/2):75–93.

Schubert, Glendon. 1962. "The 1960 Term of the Supreme Court: A Psychological Analysis." *The American Political Science Review* 56(1):90–107.

Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Northwestern University Press.

Segal, Jeffrey A. and Albert D. Cover. 1989. "Ideological Values and the Votes of Supreme Court Justices." *American Political Science Review* 83:557–565.

Segal, Jeffrey A. and Harold J. Spaeth. 1990. "Rehnquist Court Dispostion of Lower Court Decisions: Affirmation Not Reversal." *Judicature* 74(2):74–86.

Segal, Jeffrey A. and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge University Press.

Segal, Jeffrey A. and Harold J. Spaeth. 1996. "The Influence of Stare Decisis on the Votes of United States Supreme Court Justices." *American Journal of Political Science* 40(4):971–1003.

Segal, Jeffrey A. and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge University Press.

Shirts, Randall B. and Michael R. Shirts. 2002. "Deviations from the Boltzmann distribution in small microcanonical quantum systems: Two approximate one-particle energy distributions." *Journal of Chemical Physics* 117(12):5564.

Smyth, Padhraic. 2009. "Model Selection for Probabilistic Clustering Using Cross-Validated Likelihood." *Statistics and Computing* 9:63–72.

Snee, Ronald D. 1977. "Validation of Regression Models: Methods and Examples." *Technometrics* 19(4):415–428.

Solana, Javier. 1999. "NATO's Success in Kosovo." *Foreign Affairs* 78(6):114–120.

Songer, Donald R. 1987. "The Impact of the Supreme Court on Trends in Economic Policy Making in the United States Courts of Appeals." *The Journal of Politics* 49:830–841.

Songer, Donald R. and Reginald S. Sheehan. 1990. "Supreme Court Impact on Compliance and Outcomes: Miranda and New York Times in the United States Courts of Appeals." *The Western Political Quarterly* 43(2):297–316.

Songer, D.R., J.A. Segal and C.M. Cameron. 1994. "The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science* 38(3):673–696.

Spaeth, Harold, Lee Epstein, Ted Ruger, Keith Whittington, Jeffrey Segal and Andrew D. Martin. 2010. *The Supreme Court Database*.
URL: *http://scdb.wustl.edu*

Spiller, Pablo T. and Rafael Gely. 1992. "Congressional Control or Judicial Independence: The Determinants of U.S. Supreme Court Labor-Relations Decisions, 1949-1988." *The RAND Journal of Economics* 23(4):463–492.

Spriggs, James F., II. 1996. "The Supreme Court and Federal Administrative Agencies: A Resource-Based Theory and Analysis of Judicial Impact." *American Journal of Political Science* 40(4):1122–1151.

Spriggs, James F., II. 1997. "Explaining Federal Bureaucratic Compliance with Supreme Court Opinions." *Political Research Quarterly* 50(3):567–593.

Spriggs, James F., II and Thomas G. Hansford. 2001. "Explaining the Overruling of U.S. Supreme Court Precedent." *The Journal of Politics* 63(4):1091–1111.

Spriggs, James F., II and Thomas G. Hansford. 2002. "The U.S. Supreme Court's Incorporation and Interpretation of Precedent." *Law & Society Review* 36(1):139–160.

Tate, C. Neal. 1981. "Personal Attribute Models of the Voting Behavior of U.S. Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions, 1946-1978." *The American Political Science Review* 75(2):355–367.

Thall, Peter F., Richard Simon and David A. Grier. 1992. "Test-Based Variable Selection via Cross-Validation." *Journal of Computational and Graphical Statistics* 1(1):41–61.

Tomz, Michael, Jason Wittenberg and Gary King. 2003. "CLARIFY: Software for interpreting and presenting statistical results." *Journal of Statistical Software* 8(1):1–30.

Totenberg, Nina. June 30, 2008. "Earl Warren's Legacy." National Public Radio Interview.
URL: *http://www.npr.org/templates/story/story.php?storyId=92043809*

Tucker, Robert W. and David C. Hendrickson. 2004. "The Sources of American Legitimacy." *Foreign Affairs* 83(6):18–32.

van Duijn, Marijtje A.J., Krista J. Gile and Mark S. Handcock. 2009. "A Framework for the Comparison of Maximum Pseudo-Likelihood and mMximum Likelihood Estimation of Exponential Family Random Graph Models." *Social Networks* 31(1):52 – 62.

van Houwelingen, Hans C., Tako Bruinsma, Augusinas A.M. Hart, Laura J. van't Veer and Lodewyk F.A. Wessels. 2006. "Cross-validated Cox regression on microarray gene expression data." *Statistics in medicine* 25(18):3201.

Verba, Sidney. 1961. "Assumptions of Rationality and Non-Rationality in Models of the International System." *World Politics* 14(1):93–117.

Verweij, Pierre J.M. and Hans C. Van Houwelingen. 1993. "Cross-validation in survival analysis." *Statistics in Medicine* 12:2305–2305.

Wahlbeck, Paul J. 1998. "The Development of a Legal Rule: The Federal Common Law of Public Nuisance." *Law & Society Review* 32(3):613–638.

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.

Ware, James H., Stuart Lipsitz and Frank E. Speizer. 1988. "Issues in the analysis of repeated categorical outcomes." *Statistics in Medicine* 7(1):95–107.

Wasserman, Stanley and Philippa Pattison. 1996. "Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p*." *Psychometrika* 61(3):401–425.

Wen, Lei. 2004. An efficient algorithm for mining frequent closed itemset. Vol. 5 pp. 4296 – 4299.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica* 50(1):1–25.

Wilkenfeld, Jonathan, Kathleen Young, Victor Asal and David Quinn. 2003. "Mediating International Crises: Cross-National and Experimental Perspectives." *The Journal of Conflict Resolution* 47(3):279–301.

Williams, Rick L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56(2):645–646.

Wohlfarth, Patrick C. 2009. "The Tenth Justice? Consequences of Politicization in the Solicitor General's Office." *The Journal of Politics* 71(1):224–237.