

Spatializing Partisan Gerrymandering Forensics

Local Measures & Spatial Specifications

by

Levi John Wolf

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2017 by the
Graduate Supervisory Committee:

Sergio J. Rey, Chair
Luc Anselin
A. Stewart Fotheringham
Wendy K. Tam Cho

ARIZONA STATE UNIVERSITY

December 2017

©2017 Levi John Wolf

All Rights Reserved

ABSTRACT

Gerrymandering is a central problem for many representative democracies. Formally, gerrymandering is the manipulation of spatial boundaries to provide political advantage to a particular group (Warf, 2006). The term often refers to political district design, where the boundaries of political districts are “unnaturally” manipulated by redistricting officials to generate durable advantages for one group or party. Since free and fair elections are possibly the critical part of representative democracy, it is important for this cresting tide to have scientifically validated tools. This dissertation supports a current wave of reform by developing a general inferential technique to “localize” inferential bias measures, generating a new type of district-level score. The new method relies on the statistical intuition behind jackknife methods to construct relative local indicators. I find that existing statewide indicators of partisan bias can be localized using this technique, providing an estimate of how strongly a district impacts statewide partisan bias over an entire decade. When compared to measures of shape compactness (a common gerrymandering detection statistic), I find that weirdly-shaped districts have no consistent relationship with impact in many states during the 2000 and 2010 redistricting plan. To ensure that this work is valid, I examine existing seats-votes modeling strategies and develop a novel method for constructing seats-votes curves. I find that, while the empirical structure of electoral swing shows significant spatial dependence (even in the face of spatial heterogeneity), existing seats-votes specifications are more robust than anticipated to spatial dependence. Centrally, this dissertation contributes to the much larger social aim to resist electoral manipulation: that individuals & organizations suffer no undue burden on political access from partisan gerrymandering.

DEDICATION

To all those who support movement in the spirit of reform, thank you. Further, to those who gave me strength & support on the long road to here, you will always be cherished.

ACKNOWLEDGMENTS

Thanks to my committee, chief among them Serge Rey, for the substantive and emotional support to finish the dissertation. Thank you, Alan Murray, for the great kickstart into my graduate school life and the countless little things you taught me in those early years. Thank you Julia Koschinsky, Daniel Arribas-Bel, and Taylor Oshan for the constant encouragement & willingness to listen. Thank you to the National Science Foundation for providing the means to enrich the computational and human aspects of this research through grant #1657689.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
1 ELECTORAL SYSTEMS ANALYSIS	1
2 MEASURING ADVANTAGE & BOUNDARY MANIPULATION	10
2.1 A Short Legal History of Partisan Gerrymandering	11
2.2 Disentangling Standards and Measures	14
2.2.1 A Vocabulary to Define Standards	17
2.2.2 Defining Standards	21
2.3 Measures of Partisan Advantage	23
2.3.1 Efficiency Gaps	25
2.3.2 Bonus Measures	27
2.3.3 Attainment Gap	29
2.3.4 Example: Washington 2016	31
2.4 Geometric Measures of Boundary Manipulation	34
2.4.1 Ideal Circle Measures	38
2.4.2 Convex Hull Measures	40
3 DATASET: SPATIOTEMPORAL DATABASE OF CONGRESSIONAL ELEC- TIONS, 1898-2016	41
3.1 Sources of Constituency-level Electoral Data	41
3.2 Methods	43
3.2.1 Code availability	47
3.3 Data Records	48
3.4 Technical Validation	48
3.5 Further Potential Uses	50
4 EMPIRICAL STRUCTURE OF ELECTORAL SWING	52

CHAPTER	Page
4.1 Political Advantage as a “Hypothetical” Edge	53
4.2 Disagreement About “Uniformity” in Electoral Swing	57
4.3 Analyzing Spatial Dependence in Electoral Swing Under Known Spatial Heterogeneity	61
4.3.1 Macro-geographical Structure of 2016 Presidential Swing	61
4.3.2 Exploring a Decomposition of Spatial Heterogeneity & Dependence in County-level Presidential Vote	66
4.3.3 Spatial Dependence & Heterogeneity in Congressional Swing	70
4.4 Conclusion	73
5 MODELING SEATS & VOTES: SPECIFICATION AND COMPARISON	76
5.1 Development of Seats-Votes Modeling Frameworks	77
5.1.1 Common Methods for Estimating Seats-Votes Curves	79
5.2 Uniform Swing & the Seats-Votes/Rank-Votes Curves	81
5.3 Generalized Uniform Partisan Swing Methods	85
5.3.1 Stochastic Methods for Estimating Seats-Votes Curves	86
5.3.1.1 The Gelman-King Model	87
5.3.1.2 Alternative Model-Driven Seats-Votes Constructions	94
5.3.1.3 Modeling Swing as Hierarchical Deviation	95
5.3.1.4 Bootstrapping	97
5.4 Comparing Seats-Votes Curves	98
5.4.1 Example: California Congressional Districts in 2014	100
5.4.2 Example: National Seats-Votes Curve in 2014	103
5.5 Concluding Remarks on Seats-Votes Specifications	106
6 ARE GENERALIZED UNIFORM PARTISAN SWINGS SPATIALLY REALISTIC? ..	108
6.1 Spatial Misspecification in Vote Share & Swing Models	113
6.1.1 Generating Spatially-Correlated Electoral Swing	115
6.1.2 Correlated \mathbf{h} or δ° in Seats-Votes Model Specifications	118

CHAPTER	Page
6.2 Validating the Imputation Model	125
6.2.1 Common Imputation Strategies	126
6.2.2 Presidential Imputation is Slightly Superior	128
6.3 Conclusion	133
7 LOCALIZING PARTISAN SYMMETRY MEASURES	136
7.1 Classical Leverage Measures for Seats-Votes Models	139
7.2 Local Measures of Partisan Impact: Jackknifing the Plans	142
7.3 Interpreting and Comparing Impact Distributions	146
7.3.1 Nonparametric Difference in Distribution Tests	146
7.3.2 Binomial Sign Testing for Deletion Distributions	147
7.3.3 Effect Size Estimation for Deletion Distributions	148
7.4 Advantage Impact, Model Leverage, and Model Influence	149
7.4.1 General Properties about Impact Measures	149
7.4.2 Case Study: California, 2000 & 2010	151
7.4.2.1 Leverage	152
7.4.2.2 Cook's Distance	152
7.4.2.3 Impact Measures	154
7.4.2.4 Are the Jackknife Distributions Meaningfully Distinct from the Statewide Distributions?	156
7.4.2.5 Examining Influence Direction	157
7.4.2.6 Digging Deeper Into Impact	159
7.4.2.7 Which Districts are Truly Beyond the Pale?	163
7.4.2.8 Yearly Impacts in California Since 2000	163
7.4.2.9 A Grand Decadal Estimate of District Partisan Impact	169
7.4.2.10 Relationship Between Classical Influence and Impact in Cali- fornia Since 2002	171
7.4.3 Case Study: Wisconsin, 2010	174

CHAPTER	Page
7.5 Conclusion	182
7.6 CHAPTER APPENDIX I: A FIRST BRUSH WITH IMPACT DISTRIBUTIONS	185
7.6.1 Distributions for California since 2002	185
7.6.2 Distributions for Wisconsin since 2012	190
7.7 CHAPTER APPENDIX II: INTER-IMPACT CORRELATIONS IN CALIFORNIA SINCE 2002	192
8 MANIPULATION AND LOCAL ADVANTAGE	195
8.1 Shape and Impact	196
8.2 Shape is Hardly Related to Impact	198
8.3 Compactness as a Decision Rule	205
8.4 Conclusion	212
9 SOCIAL, HUMAN, AND POLICY FACTORS	214
9.1 The General State of the Literature	214
9.2 Interview Results	216
9.2.1 Iowa	218
9.2.2 California	218
9.2.3 Arizona	219
9.2.4 Washington	224
9.3 Commonalities between California, Arizona, and Washington Participants	226
9.4 Practitioner Beliefs About Construct Validity	228
9.4.1 Premises of Symmetry Measures & the Attainment Gap	229
9.4.2 Premises of Wasted Votes	230
10 CONCLUSION	233
10.1 Empirics	233
10.2 Electoral Model Specification	234
10.3 Local Indicators of Partisan Impact	235
10.4 Shape and Advantage	238

CHAPTER	Page
10.5 Interviews	238
10.6 Avenues of Further Work	240
REFERENCES	243

LIST OF TABLES

Table	Page
1 Multiparty Electoral Notation	2
2 Two-Party Electoral Notation	4
3 Washington Congressional Election Results, 2016	31
4 Data Product Schema	46
5 Swing Specifications in Structural Election Models	55
6 Spatial Fixed Effects in Presidential Vote Share	67
7 Runs of Uncontestedness in Decades Since 1990.....	127
8 Imputation Accuracy and Precision	130
9 Partisan Impact Grand Effect Estimates, Wisconsin 2010 Plan	180
10 Estimate Shape Relationship to Partisan Impact	200
11 Shape and Impact Scores, Table A	210
12 Shape and Impact Scores, Table B	211
13 Gerrymandering in Popular Journals	215

LIST OF FIGURES

Figure	Page
1 Example Ideal Shape Measures around Maryland’s 5th District from 1903-1923	37
2 King (1994) and CLEA Democratic Vote Share Scatter	48
3 King (1994) and CLEA Democratic Vote Share over Time	49
4 Democratic Presidential Vote Share Distributions by County, 2008-2016	62
5 Two-Party Vote Swing by County, 2012 & 2016	63
6 Swing and Lag Swing, 2012 & 2016	63
7 Significant State Heterogeneity in Presidential Vote Share Models	68
8 Standard Moran Plot & Hierarchical Moran Plots by Region, Division, & State	68
9 Spatial Autocorrelation Function, Presidential Votes by County in 2016	69
10 Swing Distributions of Congressional Democratic Vote Share, 1994-2016	71
11 Moran Statistic by Year, Congressional Democratic Vote Share Since 1994	72
12 Seats-Votes Curve, US House 2012	82
13 Seats-Votes Curve, US House 2012 (Detail)	84
14 Congressional Seats-Votes Curve, California 2014	88
15 Congressional Seats-Votes Curve, Texas 2014	89
16 Pairwise Comparisons & Discrepancies in Three Seats-Votes Curves, California 2014 ..	101
17 District Vote Share Variance, California 2014	103
18 Pairwise Comparisons & Discrepancies in Three Seats-Votes Curves, National 2014 ..	105
19 District Vote Share Variance, National 2014	106
20 Shrinkage in Multilevel and Multilevel Spatially-Correlated Congressional Vote Models, National 2016 by State	110
21 Lagrange Multiplier Tests for National Two-Cycle Models, 1992-2016	115
22 Simulated Vote Shares under Dependence, National 2014	120
23 Seats-Votes Curve Discrepancies, National 2014	121
24 Variance Inflation in Spatially-Correlated Seats-Votes Models, National 2014	124
25 Congressional Democrat Vote Shares, 1992-2016	126

Figure	Page
26 Imputed Vote Share versus Observed Vote Share	130
27 Imputed Vote Share Precision by Year since 1994	131
28 Classification Error Rate by Congress, 104th-115th	132
29 Imputation RMSE Distributions over Districts in Congresses 104-115	132
30 California Leverage Plots, 2002-2016	152
31 California Cook's <i>D</i> , 2002-2016	153
32 Mann-Whitney <i>U</i> Choropleth, California 2002-2010.....	157
33 Mann-Whitney <i>U</i> Choropleth, California 2012-2016.....	158
34 Binomial Sign Test Choropleth, California 2002-2010	158
35 Binomial Sign Test Choropleth, California 2012-2016	162
36 Partisan Impact Estimates, California 2002-2010 (Observed Conditions)	165
37 Partisan Impact Estimates, California 2002-2010 (No Incumbents)	166
38 Partisan Impact Estimates, California 2012-2016 (Observed Conditions)	167
39 Partisan Impact Estimates, California 2012-2016 (No Incumbents)	168
40 Grand Estimates of Partisan Impact, California 2000 Plan.....	172
41 Grand Estimates of Partisan Impact, California 2010 Plan.....	173
42 Impact and Influence Scatter, California 2000 Plan	174
43 Impact and Influence Scatter, California 2010 Plan	175
44 Apportionment Histogram, 2010 Redistricting	175
45 Distributions of Efficiency Gap & Attainment Gap Impact Scores, Wisconsin 2016	177
46 Mann-Whitney <i>U</i> and Binomial Sign Tests' Choropleths, Wisconsin 2012-2016	178
47 Effect Estimate Choropleth, Wisconsin 2012-2016.....	179
48 Impact and Influence Scatter, Wisconsin 2012-2016.....	181
49 Impact Statistic Distributions, California 2002-2010	186
50 Impact Statistic Standard Deviations, California 2002-2010	187
51 Impact Statistic Distributions, California 2012-2016	189
52 Impact Statistic Standard Deviations, California 2012-2016	190

Figure	Page
53 Impact Statistic Distributions, Wisconsin 2012-2016	191
54 Intra-Impact Relationships by Scatter, California 2002-2010	193
55 Intra-Impact Relationships by Scatter, California 2012-2016	194
56 Shape Measure Scatter	196
57 Shape Measures over Time	197
58 Shape & Impact Correlations by State A	202
59 Shape & Impact Correlations by State B	203
60 Shape & Impact Correlation Matrices	204

Chapter 1

ELECTORAL SYSTEMS ANALYSIS

The analysis of partisan bias in electoral systems is one of the foundational issues in political science and electoral geography. Arising from both first-past-the-post and proportional representation systems, the analysis of electoral geography and electoral rules to determine whether they provide a *structural* advantage to one party over another is a longstanding interest of political systems analysis, and is fraught with real-world consequences.

One kind of structural advantage, that conferred by *gerrymandering*, refers to the manipulation of electoral boundaries in order to (dis)empower specific group or party. In the American context, many social groups may benefit from the manipulation of political boundaries. One contentious type of gerrymandering, racial gerrymandering, has strong legal remedies and effective quantitative methods that can be used to identify when racial gerrymandering occurs. Another type of gerrymandering benefits *incumbents*, those currently sitting, and can take the form of *anticompetitive* gerrymandering, by which a districts' boundaries are manipulated to prevent effective two-party competition, or *factional* gerrymandering, where party leaders force members of their own party to compete or to leave their district in order to increase the power of the party leaders. A final type of gerrymandering is *partisan* gerrymandering, which intentionally dilutes the power of one party. Insofar as the various types can be separated, this dissertation focuses explicitly on *partisan* gerrymandering, the intentional biasing of the political system towards or against one political party.

At its core, statistical analysis of election systems requires solving a critical challenge: elections provide deceptively little information. The number of political districts in any state is often quite small, as is the number of elections held under a single districting scheme (in the United States). In addition, many district seats are uncontested or only nominally contested in many elections. Thus, the critical issue for quantitative analysis of partisan gerrymandering is to make statistically reasonable conclusions about partisan advantage when the raw data is both sparse

Symbol	Name	Meaning	Relation
\mathbf{V}_{ij}	District Vote	# votes cast for party j in district i	$\mathbf{H}_{ij} * \mathbf{m}_i$
\mathbf{H}_{ij}	District Vote Share	% of votes cast for party j in district i	$\mathbf{V}_{ij} / \mathbf{m}_i$
$\bar{\mathbf{h}}_j$	Party Vote Share	% of all votes cast that were for party j	$\sum_i \mathbf{V}_{ij} / \sum_i \mathbf{m}_i$
\mathbf{s}_i	District Winner	index j of the party that won in district i	$\operatorname{argmax}_j(\mathbf{V}_{ij})$
$\bar{\mathbf{s}}_j$	Party Seat Share	share of all seats won by party j	$\sum_i \mathcal{I}(\mathbf{s}_i = j) / N$
\mathbf{m}_i	Turnout	# of votes cast in district i for any party	-

Table 1. Fully-general notation for electoral systems analysis. \mathcal{I} represents the indicator function, which is one when the predicate is true and zero otherwise.

and noisy. How can advantage be identified in only two elections or in only three districts? The methods used in this dissertation involve modeling the target data generating process and analyzing sets of simulated elections, but this only shifts the issue of data scarcity to the initial phase of analysis. While this is common method, it is important to acknowledge that this (and much of the formal notation and methods that follow) is essentially designed to maximize the information about the observed electoral process.

Before proceeding, I will define some terms. Each *congressional election* is a unique event, a contest between two (or more) discrete choices in each of the N constituencies in a state's congressional delegation. At the end of the election, an $N \times 1$ record is made of the total number of votes cast in each constituency, called \mathbf{m} , as well as an $N \times P$ matrix of the raw number of votes cast for each party $j = 1, 2, \dots, P$ over the constituencies, denoted \mathbf{V} . Sometimes, if full data is not available in historical settings, only the vote shares for party j , \mathbf{H}_j , are recorded as $\mathbf{V}_j / \mathbf{m}_j = \mathbf{H}_j$. In these cases, \mathbf{m}_j is often unavailable. Since this dissertation concerns biennial contemporary United States congressional elections from Chapter 3, it deals exclusively in *first-past-the-post* election rules, where the party with the most votes in a district wins the seat, and when \mathbf{V} is available, so too are \mathbf{H} and \mathbf{m} .

Under this win rule, let the vector \mathbf{s} be the *district seats* vector, an $N \times 1$ vector of indicator variables where $\mathbf{s}_i = j$ if party j wins more votes in district i than any other party. Another summary of the electoral performance of each party, the *party seat share* vector, $\bar{\mathbf{s}}$, is the P -length vector containing the overall percentage of seats won by each party, or the empirical

frequency vector of \mathbf{s} . Thus, \bar{s}_j is the percent of the congressional delegation controlled by party j . An analogue to the party seat share, called the *party vote share*, is also available. This vector, $\bar{\mathbf{h}}$, is a P -length vector containing the fractions of votes cast for each party $j = 1, 2, \dots, P$, out of all votes cast in the election. For clarity, these are summarized over districts i and parties j in Table 1.

Since these terms are discussed for a single election, multiple elections in time yield an additional index k in a total number of time periods T over which these observations can be summarized. In general, symbols with a bar over them are party-wise summaries, and are $P \times T$ in one dimension. Most other summaries are district-level, and so are $N \times T$.

Fortunately, in the United States, some of the party complexity can be reduced. Since there are only two major parties, J can be reduced to 1. The remaining party's raw vote, vote shares, or percentages are always recoverable from the grand totals, so recording only one party provides significant simplifications. The party that becomes the focal point of analysis is then called the *reference party*, and the party whose results are omitted the *complementary party*. Using this reference/complement split, matrices can be reduced to single vectors: \mathbf{V} becomes \mathbf{v} , an $N \times 1$ vector of the raw votes cast for the reference party over N districts, and \mathbf{H} becomes \mathbf{h} , the $N \times 1$ vector of the reference party vote shares. In addition, $\bar{\mathbf{h}}, \bar{\mathbf{s}}$ become scalars, with \bar{h} representing the share of the popular vote the reference party wins in the congressional election and \bar{s} the share of the seats in the congressional delegation the reference party wins. Further, \mathbf{s}_i becomes a binary indicator vector, reflecting whether the reference wins in the district. The designation of the reference party is arbitrary, and its reversal simply reverses the analysis.

Most questions about electoral fairness in the United States reduce to questions about the relationship between vote share and seat share, \bar{h} and \bar{s} . These party-level summaries are often implicated in claims about the electoral system:

Candidate A won the electoral college, even though he won fewer votes than his opponent. This is unfair!

Symbol	Name	Meaning	Relation
\mathbf{v}_i	District Vote	# of votes cast for RP in district i	$\mathbf{h}_i * \mathbf{m}_i$
\mathbf{h}_i	District Vote Share	% of votes cast for RP in district i	$\mathbf{v}_i / \mathbf{m}_i$
\bar{h}_j	Party Vote Share	% of all votes cast for RP	$\sum_i \mathbf{v}_i / \sum_i \mathbf{m}_i$
\bar{s}_i	District Winner	1 if the RP wins district i , 0 otherwise.	$\mathcal{I}(\mathbf{v}_i > .5)$
\bar{s}_j	Party Seat Share	% of all seats won by the RP	$\sum_i \bar{s}_i / N$
\mathbf{m}_i	Turnout	# of votes cast in district i for any party	-

Table 2. Two-Party notation for electoral systems analysis. \mathcal{I} represents the indicator function, which is one when the argument is true and zero otherwise. “RP” stands for reference party, which is always the Democrats throughout this dissertation. This choice is arbitrary, and only affects the orientation of the effects.

Party A wins two seats for every 10,000 votes they win, but party B only wins 1.2 seats on average. Clearly, votes for party B are being wasted, and the system is biased in favor of A .

Both of these claims, at a point, reduce to questions about the seats-votes relationship. But, each election only provides a small amount of information about the relationship between seats and votes for each party, a single observation (\bar{h}_t, \bar{s}_t) . While some focus on the historical relationship between (\bar{h}_t, \bar{s}_t) pools elections over t (Tufte, 1973), newer techniques use “extra” information: the information in the covariance of district-level vote counts. These techniques tend focus on the \mathbf{h}, \mathbf{s} vectors, summarizing them in novel ways (Brookes, 1960; Johnston, 2002; Hill, 2010), or extracting information using simulations (Gelman and King, 1994a; Gelman et al., 2010; Linzer, 2012). An entirely separate set of detection measures relies on the geometric properties of district shapes (Young, 1988). Districts whose shapes are irregular in some way are considered likely to be gerrymandered, since their boundaries are likely to be manipulated. These measures do not involve the relationship between \bar{h} and \bar{s} , and exist entirely independently of the political outcomes in the electoral system.

Crucially, the analysis of partisan advantage in political systems often comes into play around the time to redraw congressional district lines. This process, called redistricting, is a constitutionally-mandated state-by-state spatial reconfiguration of the American electorate. Redistricting occurs every ten years, at minimum, to equalize population between congressional districts in the US House of Representatives (*Wesberry v. Sanders*, 376 US 1 (1964)) as well as state legislatures (*Reynolds v. Sims*, 377 US 533 (1964)) and some local offices (*Avery*

v. Midland County, 390 US 474 (1968)). While equalizing populations between districts is a critical motivation for redrawing district lines, communities and constituencies also change in composition and spatial configuration. For a congressional geography to be “representative” of its underlying population distribution in addition to providing for equal “one person, one vote” representation, the lines must be redrawn to capture this structural and spatial shift.

Redrawn districts may result in geographies that cause one party to be more successful than it was in the previous plan. However, this change in fortunes is not necessarily indicative of partisan gerrymandering outright; the change may be driven entirely by demographic or ideological change in the state. Regardless, individuals may *believe* the system to advantage a particular party and use subjective interpretations of how congressional districts in the state *should* look to identify that partisan gerrymandering occurred. This perceptual standard of evidence leads to many “common sense” solutions to redistricting issues.

Unfortunately, no silver bullet has yet been loaded or fired. Determining whether a specific district or districting plan has been gerrymandered to advantage a given party *over and above* the advantages the party may enjoy due to social attitudes or incumbent candidates requires answers to a complex set of questions at the intersection of race, party, history, and community. This dissertation provides a new technique to identify partisan gerrymandering that can be used to answer those questions. This new technique may be applied to many different types of inferential analyses of partisan advantage, and thus sidesteps much of the debate in the literature attempting to identify the single most appropriate measure. In the tradition of model-based gerrymandering identification techniques, the new method is able to conditionally control for other potentially confounding sources of political advantage, in addition to being providing an indication of the impact each district has on the measure of partisan advantage in state congressional delegations.

To do this, I first examine the fundamental assumptions used in the counterfactual modeling process in Chapters 4, 5, and 6. The local measures of political advantage are *secondary* statistics about an electoral model, so their values may be affected significantly by model misspecification in either of two models. The main concern is with the seats-votes model that

drives inference. The seats-votes model assumes that the election results in each congressional district are independent from one another. The extent to which the vote shares are in fact independent will be assessed using spatial econometric techniques. Second, the inferential advantage measures also depend on a model of electoral swing used to generate electoral counterfactuals, elections that occur under conditions that are not observed. Here, electoral swing, δ , is the change in vote over N districts won by a party between two elections. Sometimes, δ is constant over all $i = 1, 2, \dots, N$, and sometimes it is modeled as a random effect with varying specifications. Given this, some model of swing is used to shift the average expected vote, $E[\bar{h}]$, to a known target value or to fix the simulated party vote share to the observed party vote share. Under these simulations in controlled conditions, the resulting bias statistics are analyzed. Thus, the swing model should represent plausible but unobserved elections. In many cases, modeling electoral swing as an independent and identically-distributed random innovation is *implausible* in light of its observed empirical structure. Thus, the extent to which the vote share model and the model of electoral swing reflect observed elections is assessed. In addition, potential spatial misspecification of both the vote share and swing models is examined.

In general, I find that electoral swing in the United States is significantly spatially correlated while accounting for various exogenous forms of heterogeneity. Thus, the use of independent, identically-distributed random swing effects generates empirically-unlikely maps of potential swings at either the county or the legislative level. However, the resulting maps of electoral *outcomes* do tend to be realistic, even though the maps of swing are unrealistic, since the magnitude of swing is often small with respect to the vote share to which it applies. Further, using a spatially-correlated swing model to provide more “realistic” counterfactual maps of electoral swing simply does not have a large impact on bias measures or the estimated seats-votes curve. Adding a small magnitude of white (spatial) noise to an electoral map generates nearly the same electoral results as adding a small magnitude of correlated noise. I also find that a common model of electoral outcomes used in gerrymandering analysis suffers from spatial misspecification. However, resolving this spatial misspecification does not significantly improve counterfactual simulations in any plausible scenario. I develop a new, retrospective method to

model the seats-votes curve using bootstrapping which does not depend on an explicit parametric model of swing. While this bootstrap is exceptionally simple (and highly-extensible), it also does not account for dependence or heterogeneity in electoral swing. This method is compared to two other types of seats-votes curve models, and tends to agree more with one model than another. Thus, it appears that corrections to the standard linear model to account for empirically-observed dependence in swing or vote share do not make a large difference in conclusions about the electoral system as a whole.

Given that these models are more robust than anticipated, I derive novel jackknife measures of district-level partisan impact. After deriving the impact measures, I examine their properties in a few case studies. I focus on California first, as its large number of congressional districts and adoption of vastly different electoral rules & districting schemes in the 2010 cycle yields an interesting significant breakpoint. I also examine the post-2010 redistricting in Wisconsin. I aim to determine whether impact behaves consistently over time and space. To do this, I develop a method of analysis for these jackknife impact measures. In addition, I examine whether the measure of impact is related to classical measures of observation influence in the underlying statistical model. If the districts that influence bias scores tend to be the districts that influence the underlying electoral model, then standard model influence measures might be more simple to use as local partisan advantage measures. Otherwise, it may be the case that districts that are influential on the underlying stochastic model *do not* significantly influence the bias of the statewide plan, or that influence of districts is inconsistent over time. If either is true, districts may be considered “not consistently impactful” on partisan advantage in a given state. Finally, I characterize two axes along which plans may vary in terms of their bias statistics. The first axis is *balance*. *Balanced* plans have impact measures that are symmetrically distributed around the statewide advantage measure, meaning that some districts may increase statewide advantage and some may decrease. *Unbalanced* plans have impact measures that fall primarily on one side of the statewide advantage, meaning most districts tend to move the advantage in a single direction. The second axis is *precarity*. Plans that are *precarious* have districts with consistently large impact scores; plans that are not tend to have low district impact scores.

I find that impact measures are significantly different from classical measures of influence in linear regressions. Second, I find that the classical measures of influence are not consistent over the model specification considered. Third, I find that a nominal control on incumbency filters out many districts from being “impactful.” Since there is no definitive answer as to whether bias under observed conditions or bias controlling for incumbency is more critical to examine, this difference is important to acknowledge. I also find that the impact measures follow the same general relationship to one another as the statewide measures which they decompose. Further, I note that impact measures sometimes disagree with one another, in that one suggests a district’s removal benefits Republicans and another suggests its removal benefits Democrats. In general, the measures are observed to work in two groups: the efficiency gap of McGhee (2014) and the bias-at-median defined in Gelman and King (1994a) tend to provide similar impact classifications, and the attainment gap from Linzer (2012) and bias at observed vote (discussed by Gelman et al. (2010)) tend to agree. Finally, I find that there are some precariously balanced plans, where each district has a strong impact on statewide advantage and, those districts tend to pull the bias in different directions. But, this bimodality tends to be stronger for individual impact measures than an inherent property of the plan itself.

Then, I compare the statistic to commonly-used measures of boundary manipulation in Chapter 8. This is done in a combination of exploratory regression and less-structured correlation analysis. Geometric measures are used to identify when an *individual district* might have had its boundaries manipulated during drafting. This manipulation is then assumed to be caused by an attempt to provide advantage to a given group. Thus, if the districts picked up for boundary manipulation *do* have large impact on statewide advantage, then this assumption holds. Otherwise, geometric measures may pick up irregular geometries, but not discover districts that provide advantage. By examining the relationship between the new impact measures and districts’ compactness scores, I aim to identify whether the boundary manipulation measures tend to pick up on districts that also have significant partisan impact. If this were the case, then shape measures would not necessarily discover *gerrymandering*, which is boundary manipulation *that generates* political advantage.

I find that geometric measures are unrelated to impact statistics for many different measures of partisan advantage. Thus, a causal relationship of the “boundary manipulation” detected by these measures generating partisan advantage *qua* impact is either unlikely, swamped by noise, or both. This lack of relationship occurs despite the presence of a separate aggregate relationship: compactness is negatively related to Democratic vote shares at the congressional district level. Further, I find that using geometric measures as a decision rule to identify bad districts would be unacceptable. Selecting “bad” districts based on their compactness scores would single out a large number of districts that have negligible impact on statewide bias. Further, I find that selecting on geometric measures would skip over some districts as *not* manipulated when they in fact have a significant impact on partisan advantage in the state. Drilling further down, I find that geometric measures are not useful predictors of partisan outcomes in the underlying statistical models themselves, and changes in district compactness tend to have no relationship to changes in partisan bias measures. Thus, geometric measures, insofar as they detect boundary manipulation *pursuant to* partisan advantage, should be retired.

Overall, the main realizations from this dissertation are that impact measures provide a powerful new tool to identify individual districts that influence statewide partisan advantage scores. In addition, these measures tell us something novel about the structure of the electoral models themselves and the conclusions we may draw about congressional districting plans. Critically, spatial dependence matters a lot less for modeling seats-votes curves and partisan bias than I thought it may from the outset. Thus, at least one commonly-used seats-votes modeling method is robust to spatial misspecification, even though it (in theory) may require a spatial correction. In finding this, I find that while some models are unrealistic and wrong, they are so useful as to be effectively indistinguishable from an empirically “correct” model. Finally, geometric measures are invalidated as an effective district-specific indicator of partisan gerrymandering, since they are unrelated to partisan advantage.

Chapter 2

MEASURING ADVANTAGE & BOUNDARY MANIPULATION

Many different measures of partisan advantage are available in its longstanding literature. While some suggest that measures of advantage *should* all standardize on specific indicators of advantage and argue court precedent demonstrates this need (Grofman and King, 2007), many core criticisms of these methods remain unanswered (Stephanopoulos and McGhee, 2015). As a recent comprehensive overview by Nagle (2015) demonstrates, consensus on appropriate measures has mainly fragmented since an early pre-*Davis v. Bandemer* review, Grofman (1983). Notably, many of the measures suggested by Grofman (1983) are dismissed by Nagle (2015) for concerns about construct validity: does the measure accurately & consistently indicate advantage when present? Critiques of measures tend to focus on construct validity more generally, such as with Altman (2002) on Johnston et al. (1999)'s application of Brookes (1960), critiques of the excess seats measures from Gelman and King (1994a) & Grofman and King (2007) made by McGhee (2014); McDonald and Best (2015), and Tam Cho (2017)'s review of McGhee (2014). These concerns about validity complement jurisprudence skeptical of these measures.

These advantage statistics abound because each strikes a novel compromise between descriptive and inferential purpose, normative grounding, and empirical applicability. Starting from a specific interpretation of what "advantage" means, they then provide a specific scalar measure of advantage that relies on a model of how advantage arises. Critically, though, these developments are often not unified with a consistent behavioral or process theory. Different *measures* can be constructed with reference to many different *standards* of electoral justice, and the process that generates the elections may be different from the process under which fairness can be measured. Measures often require a zero point, a hypothetically "fair" position against which some measure of distance is made. This zero point is frequently contentious,

implicit, and difficult to model because it is both unobserved *and* must be established through normative argument.

In what follows, I first briefly discuss the legal history of partisan gerrymandering in the United States. Then, I outline two common standards by which electoral systems are judged, both in academic and legal discourse. After this, I discuss how these standards are operationalized in a few common measures of partisan advantage in electoral systems. I pay specific attention to the construct validity of these measures, the implicit model of electoral fairness they operationalize, and how many fair positions are possible. In addition, I outline how to estimate these statistics in a two-party system given a generic model yet-unspecified stochastic model of elections in Section 2.3. After the discussion of political measures, I will discuss measures of boundary manipulation in Section 2.4, again paying close attention to construct validity concerns noted at least as early as Young (1988). After this introductory chapter on how to estimate system-wide partisan advantage measures and district-specific boundary manipulation measures, a novel technique for constructing local measures of a district's impact on partisan advantage will be presented in Chapter 7. Also in Chapter 7, these measures will be compared to classical measures of model influence for a specific model of electoral outcomes found in the literature. Finally, these measures will be compared to the geometric measures discussed here in Chapter 8, and the social and human contexts for this work discussed in Chapter 9. The work concludes in Chapter 10, where general understandings are stated and a workflow presented for using impact measures.

2.1 A Short Legal History of Partisan Gerrymandering

The history of legal review of partisan gerrymandering begins in earnest only in the latter half of the twentieth century. Although major cases (like *Baker v. Carr* & *Gomillion v. Lightfoot*) engage with districting questions, *Davis v. Bandemer* (478 US 109 (1986)) first established *partisan* gerrymandering as a justiciable subject. However, the decision contains an extreme reluctance to identify a single standard (or set of standards) that might be used to identify par-

tisan advantage. Thus, while partisan gerrymandering could be reviewed, the court did not set a standard on how to review it. Since courts in the United States have been reluctant to participate in reviewing “political questions,” which are legal issues driven primarily by direct political motives, judicial review of alleged partisan gerrymandering without empirical standards is a fraught endeavor. Cases after *Bandemer*, such as *Vieth v. Jubelirer* (541 US 267 (2004)) or *League of United Latin American Citizens (LULAC) v. Perry* (548 US 399 (2006)) significantly intensified the legal, social, and scholarly imperatives to provide convincing, valid measures of partisan advantage. Specifically, in an opinion authored by the Justice Scalia, the *Vieth* precedent hinges on the “lack of a judicially discoverable and manageable standard” for adjudicating partisan gerrymandering cases, which is one of many definitive properties of “political questions” defined in *Baker v. Carr* (369 US 186 (1962)). This skepticism echoes that found in *Bandemer*, that partisan gerrymandering cases are *theoretically* justiciable, but the lack of a clear, manageable standard interferes with judicial action. This call from the courts to provide a manageable judicial standard, a measure of partisan advantage grounded in the logic of the American single-member district (SMD), first-past-the-post (FPTP) electoral system, has not yet been definitively answered.¹

Thus, despite the intensification of need from *Bandemer v. Davis* through *Vieth v. Jubelirer* and *LULAC v. Perry* (Godfrey et al., 2005; Grofman and King, 2007), redistricting and partisan gerrymandering-adjacent court cases have proliferated. A string of recent cases on redistricting shows partisan gerrymandering and redistricting reform is an increasingly contentious legal issue. Especially as scholarship intensifies, the search for a manageable standard to detect partisan gerrymandering surfaces new debates and analyses in many court cases. In a pair of cases about the Arizona Independent Redistricting Commission (AIRC), the US Supreme Court validated the use of nonpartisan redistricting commissions² and reaffirmed the importance of

¹ First-past-the-post (FPTP) single-member district electoral systems are those in which the candidate with the largest share of the votes in a given district wins the election in that district.

² *Arizona State Legislature v. AIRC*, 576 US ____ (2015)

restraining partisan gerrymandering.³ Another case, *Evenwel v. Abbot* (578 US ____ (2016)), clarified the meaning of “person” in the “one person, one vote” doctrine established by *Baker v. Carr* (369 US 186 (1962)). The pace of litigation on redistricting is not set to slacken, either, North Carolina (*McRory v. Harris*, 15-1262, (M.D.N.C. 2017), pending) Wisconsin (*Whitford v. Gill*, 15 C.V. 421 B.B.C. (Wisc. 2016)), Maryland (*Shapiro v. McManus*, 1-13 C.V. 03233 J.K.B. (D.Md. 2015)), and Florida (*League of Women Voters of Florida v. Detzner*, 172 So. 3d 363, (Fla. 2015)) all have had redistricting plans overturned or challenged due in part to undue partisan advantage.⁴ Critically, *Cooper v. Harris* (58 US ____ (2017)) seems to have broken down the barriers between partisan and racial gerrymandering arguments, recognizing that in many places in the United States, race and party identification are nearly indistinguishable, and a racial gerrymander may look identically to a partisan gerrymander. Regardless, all of these court cases will likely change the rules significantly for the 2020 redistricting, especially as the cases percolate upwards through the system to the Supreme Court.

In each of these cases, many different measures of advantage and boundary manipulation were used across many different *amici curiae* briefs. No single best measure of advantage is championed in any of the eventual jurisprudence, however, so recent work on novel measures and methods abounds. Specifically, dedicated paper competitions on gerrymandering forensics have provided significant high-quality work in the area, focusing both on new measures (McGann et al., 2015; McDonald and Best, 2015; Wang, 2016; Arrington, 2016) and plan evaluation methods (Chen and Rodden, 2015; Cho and Liu, 2016b). Complicating matters, former President Barack Obama called for further political reforms to address partisan gerrymandering in redistricting in his final State of the Union address in 2016 and, upon leaving office, joined in the establishment of a specific, targeted redistricting reform organization and partisan action group on redistricting. Further reforms, such as independent redistricting commissions, which have an uncertain impact on redistricting outcomes (Miller and Grofman, 2013), are nonethe-

³ *Harris v. AIRC*, 578 US ____ (2016)

⁴ Regardless of the eventual success or defeat of *Whitford et al.* in *Gill v. Whitford* in the Supreme Court, explicit measures of partisan advantage are unlikely to fade with a single Supreme Court case

less important non-consequentialist methods of improving procedural fairness in redistricting (Webster, 2013; Stephanopoulos, 2013). Thus, new academic work on effective measures of partisan advantage will both drive and be driven by innovation in law, jurisprudence, and institutional reform, regardless of where the legal precedent moves the federal position on partisan gerrymandering.

2.2 Disentangling Standards and Measures

Thus, to provide a manageable judicial standard or *usable* measures,⁵ it is important to differentiate between the *standard of justice* and the appropriate *measure* of this standard. An electoral standard is a set of normative arguments that allow for the construction of a “fair” electoral result. A measure of advantage is a statistic that expresses the discrepancy between the state of an electoral system (observed or hypothetical) and fairness. Thus, measures exist *in reference to standards*, and standards do not implicate any specific measure in their construction of “fair” reference positions.

Standards of electoral fairness in redistricting have almost always focused on this deviation; the distance between the anticipated or observed impacts of a plan and what is “fair” constitutes a tangible (or impending) harm. However, “consequentialist” standards that focus on the results of elections & thus the consequences of redistricting are not the only standards of justice possible (Stephanopoulos (2013)). Alternative standards might focus instead on “procedural justice” which places constraints on the actual process of drawing lines. Methods to provide for procedural justice include the adoption of nonpartisan/independent redistricting commissions, explicit rules focused on empowering historically disadvantaged groups (Webster, 2013), or precluding the use of partisan information in the process of constructing districts.

However, procedural justice is quite difficult to ensure. First, it is unclear whether non- or bipartisan redistricting commissions generate significantly different plans from partisan ones

⁵A common theme about existing partisan advantage measures is that *no one* responsible for drawing boundaries uses them, as per the interviews in Chapter 9

(Abramowitz et al., 2006a; McDonald, 2006; Miller and Grofman, 2013; Hasen, 2013). This uncertainty about the impact of redistricting is in part due to the rarity of independent redistricting commissions and infrequent redistricting events. Regardless of the empirical stalemate, non- or bipartisan commissions are strongly encouraged by participants in those systems, as will be discussed in Chapter 9. Second, the type of redistricting for social justice suggested by Webster (2013) is unlikely to be constitutional under existing racial precedents. Further, it may be difficult to enact politically for redistricting commissions in states, which often are directed by objectives in state constitutions, rather than in statutes. Finally, precluding the explicit use of political information from redistricting hardly limits the unspoken political information in the discussions of “places that belong together” that occur between commissioners themselves, again discussed further in Chapter 9.

As such, the standards considered in this dissertation are exclusively consequentialist. Further, I suggest that empirically-validated partisan fairness is sufficient to constrain partisan gerrymandering, but other types of gerrymandering, such as racial or incumbent gerrymandering, may require much stricter attention to procedural fairness. As such, I reject out of hand the possibility of “expressive harms” (Pildes and Niemi, 1993) in partisan redistricting. These harms were first suggested as existing in the racial redistricting context. They are affective, may suppress individuals’ political efficacy or identity expression, and are exceedingly difficult to validate empirically in racial redistricting cases (Ansolabehere and Persily, 2015). They are hypothesized to be inflicted when individuals interact with a district map’s *geographical imaginary*, the social, racial, and political power relationships embodied by the boundary lines. These harms are sometimes used to justify developments of new redistricting tools focused on shape compactness (Chambers, 2010, e.g.), but reflect a difficult-to-measure undercurrent in the context of measures of political advantage. As such, affective factors like expressive harms are not considered here, and are often not ignored when considering partisan gerrymandering.

This common explicit focus, designing standards and measures of the *consequences* of redistricting, leads quite easily to the blending of standard and measure. The distinction between standard and measure is hinted at by Grofman and King (2007), but they (and others) imme-

diately proceed to conflate their chosen standard with their preferred measurement of it. They argue that partisan symmetry *as measured by* an excess seats measure operationalized in Gelman and King (1994a) is the single most appropriate method for detecting bias in the shadow of the *Vieth* decision. This conflation is then picked up and critiqued by McGhee (2014) when defining a novel measure of partisan advantage. However, McGhee (2014) defines the “efficiency gap” as if it were somehow *against or independent of* a symmetry. In fact, the efficiency gap in some formulations is a substantially more constrained seat symmetry measure placing a significantly stronger constraint on what a “fair” seat-vote relationship can look like (Jackman, 2017). However, since McGhee (2014)’s development occurs without a direct reference to a seats-votes model, the fact that the measure implicates a symmetry standard eludes the initial discussion and other establishing presentations (Stephanopoulos and McGhee, 2015).

This is not a new deficit: early discussions of advantage measures like Brookes (1960) provide hardly any normative grounding for what the author considers the “fair” electoral position. A similar level of implicitness affects the development of boundary manipulation measures as well, which are often constructed without specific reference to the type of manipulation the author hopes to identify. Later advantage measures, namely those counterfactual measures following from standards discussed in King and Browning (1987), are focused explicitly on the standards of “partisan symmetry.” Other measures, such as the efficiency gap of McGhee (2014), the attainment gap from Linzer (2012), or registration comparisons from Kousser (1996), derive from the same standards, but measure deviation from this standard in significantly different ways. The fact that two methods of analysis may share the same *standard* but may not be concordant on the *measure* is lost when these are conflated. Thus, confusion between the chosen measure and given standard has significantly affected both the discussion and development of novel techniques, and enforcing a clear distinction between standards and measures is critical.

2.2.1 A Vocabulary to Define Standards

Before defining precisely standards of partisan advantage, it helps to distinguish a few concepts in the theory of electoral system. The four concepts in Niemi and Deegan (1978) are helpful and still implicitly inform many current structural analyses of first-past-the-post electoral systems. However, Niemi and Deegan (1978) embed these concepts directly in measures of the system. This means that the four criteria are not necessarily orthogonal and do not, *in and of themselves* reflect distinct theoretical properties of the electoral system. These are properties of a seats-votes relationship, which is not a sufficient representation of an electoral system by itself, since it often ignores discrepancies in turnout or party registration (Kousser, 1996; McGhee, 2014), and is insufficient to distinguish the political system *as observed* from the theoretical representation. In addition, Niemi and Deegan (1978) present specific *constraints* of these properties that are necessary to be satisfied if the system is to be fair. They use these constraints to establish the “adequacy” of districting plans. This makes their development of a “theory” of districting systems to be rather empirical, and admit no operational distinctions between the concepts they suggest and the measures by which the system can be assessed. Thus, I suggest two new criteria to add to the four presented by Niemi and Deegan (1978) that can be used to ground or distinguish partisan advantage measures and then discuss and justify the two additional analytical criteria. I also back-out the underlying theoretical concepts of the various traits where necessary. I will avoid placing constraints on the structure of these properties until after discussing them in full and presenting the revised indicators.

Responsiveness

One property discussed by Niemi and Deegan (1978) is a longstanding critical interest of electoral systems analysis. Electoral *responsiveness* is the rate at which the number of seats a party wins changes with respect to a party’s popular vote (or average district vote) share. For a parametric model relating party seat share and popular/average vote share, this is simply slope

of the curve fit by that model. Empirically speaking, seat shares are a stepwise linear function of vote shares, since seats are either won or lost by a party. This means the number of seats awarded for each percentage change in popular vote may not be constant. Some regions of the system may be more “responsive” than others. Intuitively, this may seem to be the rule rather than the exception in first-past-the-post systems: the difference in the number of seats won when the vote share increases from 75% to 85% is likely much smaller than the change from 45% to 55% of the vote. This is shown to be the case empirically as well in Chapter 4. When a single scalar estimate of responsiveness is made, the estimate is often referred to as the “swing ratio,” reflecting the rate at which the legislature “swings” with respect to changes in vote. As a theoretical construct, responsiveness reflects the extent to which legislative composition changes with respect to changes in electorate preference, and seems well served by Niemi and Deegan (1978)’s choice of empirical measure in both multiparty and two-party systems.

Range

A closely related property to responsiveness is the *range* of the system. The range of an electoral system reflects the range of vote share over which seat share changes. Thus, this is the set of vote shares where responsiveness is nonzero. This property can be thought of as constituting the “barrier to entry” in a multiparty system, and extremely restrictive systems will be expected to have a narrow range. In addition, systems with smaller ranges must have higher responsiveness, since the entire range of seat shares must be traversed in a smaller vote share domain. In theory, this concept reflects the set of outcomes that the electoral system rewards, since movement in vote share below the minimum threshold or above the maximum threshold provides no change in seat share.

Competitiveness

Competitiveness, as defined by Niemi and Deegan (1978), refers to the fraction of districts whose “normal” vote, the expected vote share for a given party in the district, is within some fixed distance of 50%, the median in two-party systems. In their terms, competitive systems have more districts with vote shares closer to 50%, and less competitive systems have districts with party vote shares far away from 50%. In multiparty systems, identifying the correct location for a reference point may be difficult. While pinning to 50% may still appear reasonable (since that indicates outright control of the district), this is likely inapplicable. In addition, the best hinge point may vary over districts: since patterns of competition often vary in multiparty systems (Linzer, 2012), the relevant hinge point may as well. In this case, one might use the size of the gap between the winning and next-most-popular party as a measure of the competitiveness. More generally, the expression of competitiveness identified by this procedure attempts to express how closely to the electoral margin districts fall, and attempts to provide an indication of how marginal a district is. Incorporated in this measure are implicit proxy indicators of candidate recruitment in the district, since hotly contested districts may have more viable competitors, and districts with viable competitors tend to have tighter races.

Neutrality

Neutrality reflects the extent to which a political system does not favor one party over another when parties are similarly-situated. Much subsequent work on neutrality focuses on one necessary and sufficient condition Niemi and Deegan (1978) discuss, *symmetry*. While Niemi and Deegan (1978) suggest neutrality be understood intuitively as advantage, the necessary condition of symmetry becomes the criteria on which future research is established. Under symmetry, if one party wins \bar{s} seat share after having won \bar{h} fraction of the popular vote, it should win $1 - \bar{s}$ seat share after winning $1 - \bar{h}$ percent of the popular vote. Thus, this criteria is orthogonal to range, responsiveness, and competitiveness, since any system can have different

values for them while remaining symmetric or asymmetric. Symmetry is a difficult criteria to assess directly, since elections in their typical course provide no direct indication about neutrality. Further, it is rare for naturally-occurring elections under similar conditions to return \bar{h} one time and $1 - \bar{h}$ the next, let alone return (\bar{h}, \bar{s}) and $(1 - \bar{h}, 1 - \bar{s})$. Niemi and Deegan (1978) provide no direct empirical method to measure neutrality in their discussion, unlike most of the rest of the properties which can be measured directly from electoral results.

Fixity

While *competitiveness* measures of district-level vote shares are useful to characterize voting behavior, they do not provide an indication of how stable partisan control of the delegation or legislature is. Thus, one idea that captures *aggregate* competitiveness of an electoral system is the *fixity* of its majority. This embodies the extent to which an electoral system plan provides one party stable periods of control. One simple measure of fixity is the number of elections under which partisan control of the legislature or delegation flips. An alternative measure might be the distance between the “tipping” district vote share and 50%. In this context, the “tipping” vote share is the fraction of vote share required to be added or subtracted to all district vote shares in order to change the control of the state legislature from one party to another.⁶ Importantly, the empirical assessment of neutrality depends strongly on a system’s fixity: many statistical methods to assess neutrality assume that the case where control of the legislature or delegation may flip is at least a *plausible* situation. If one party’s majority is unwavering, this situation may reflect a significant extrapolation from the observed results. Thus, systems with strong fixity may pose significant challenges to the validity of symmetry measures. Further, fixity and competitiveness are not necessarily identical: an electoral system with small variance in vote shares may result in highly-competitive elections with generally fixed majorities. Thus,

⁶Clearly, this relies on an assumption that this change in vote share applies uniformly over districts, an assumption with deeply challenging geographic implications that are examined in Chapter 4.

competitiveness (in Niemi and Deegan (1978)'s sense) is a district-level measure and fixity is an aggregate system-level measure for a similar concept of party majority solidity.

Contest Size

One common complaint about seats-votes techniques is that they focus on parties and their aggregate performance over the behavior of individual voters or blocs of voters. This measurement is important, however, since the fraction of votes a party wins in a given district implicitly standardizes the measure of system properties over the universe of *voters*, not the universe of *people*. While legislative districts are required to be nearly exactly equivalent in terms of district residents, the number of voters in each district can vary widely. Thus, the extent to which *turnout* varies in a district may also reflect certain patterns of vote wasting and also create differences in the effect districts have on system responsiveness (Johnston, 1983). In light of this, a few measures of partisan advantage have focused on derived properties of observed or expected contest size (Brookes, 1960; Kousser, 1996; Johnston et al., 1999; McGhee, 2014). Thus, its explicit inclusion in a vocabulary for discussing standards of electoral fairness is important.

2.2.2 Defining Standards

With these six traits, a few common standards of fairness for representation systems are immediate. First, a standard of proportional representation requires partisan neutrality and linear responsiveness. Most common types of proportional representation *also* require that the slope of responsiveness be as close to one as possible, meaning a 1% change in vote share should be accompanied by as close to a 1% change in seat share as possible. However, in practice, many proportional representation systems implement a minimum threshold of representation, so many proportional representation systems have an effective responsiveness larger than 1 and that may fluctuate occasionally due to the finite number of seats available

to assign (Taagepera and Shugart, 1989; Gallagher, 1991; Curtice and Steed, 1986; Grofman, 1983). In addition, some forms of minimizing discrepancies in contest size, such as attaining a zero “simplified” efficiency gap (discussed by McGhee (2014)), would result in neutral, piecewise-linear response in elections (Jackman, 2017; Tam Cho, 2017).

In contrast, King and Browning (1987) and Grofman and King (2007) suggest neutrality alone is sufficient to provide for electoral fairness between parties in the United States. This is a significantly more flexible standard, since it allows for winners bonuses. In first-past-the-post single-member-district election systems, the responsiveness is nonlinear: increasing a party’s average vote share from 45% to 55% might result in a much smaller shift in the fraction of seats controlled in the legislature than an increase from 55% to 65%, as will be discussed in Chapter 5. If the distribution of individual district vote shares is bi-modal, with many safe districts for either party and a few marginal districts in the center, the shift from 45% to 55% may entail a party flipping only the few marginal districts while the winners of “safe” seats remain unchanged. Moving from 55% to 65% might result in “safe” districts from the opponent beginning to flip to the reference party *en masse*. This means that a nonlinear seats-votes response, which is unfair under most common types of generalized proportionality, is just fine under neutrality-as-fairness.

However, many critics of symmetry-only standards of fairness focus on the fact that fixity and seat costs are important to how fair a political system is *as actually experienced*, rather than in the hypothetical “tables-turned” electoral scenarios. Before the articulation of Niemi and Deegan (1978), advantage measures like those developed by Brookes (1960) focus on differences in the size of contests that parties tend to win or lose. If one party tends to lose many small contests but win a few big ones, then the number of seats that party wins may often be significantly smaller than their popular support may suggest. Alternatively, if a party wins many small contests and loses large ones, they may be more represented than their popular support may suggest. If districts that one party tends to win have much larger contest sizes, they may be more expensive in terms of total human and organizational costs. Thus, the measure focuses on consistent contest sizes for the losses and wins between parties and have been used in

the American context (Johnston, 2002; Johnston et al., 2005; Hill, 2010). Other methods to assess fairness like the efficiency gap based on turnout also focus on aspects of contest size, but incorporate information about the size of majorities and minorities, rather than the sizes of contests parties tend to win or lose. Finally, an early critique of symmetry-only analysis uses the anticipated differences in contest sizes using differences in voter registration (Kousser, 1996), arguing that registered voters tend to be effective predictive indicators of whether districts will have excessively-large majorities or large but consistent minorities.

2.3 Measures of Partisan Advantage

In this dissertation, a stand on which measure of partisan fairness is “superior” will not be made directly. Rather, I am concerned with developing a new technique that can be applied to decompose the impact each district has on these scores. Some measures may be more sensitive than others, have more stable distributional properties, or may have stronger face validity relative to specific arguments about how the political process works. It is true that the impact measures will inherit the same grounds as their plan-wide referents. But, the use of these impact measures will *also* allow for a better understanding of the plan-wide statistics themselves. By interpreting the types of districts that may (or may not) have strong influence on a measure, arguments about what types of manipulation or partisan impact a measure is *supposed to detect* can be directly assessed.

Five stochastic measures are considered in this dissertation. They all depend on the underlying data-generating process discussed in Chapter 5. In theory, two can be computed directly from the observed election returns, but are not done so here in order to keep the analysis on the same inferential footing. The stochastic measures must all be estimated using some simulation regime: an electoral model and simulation strategy are used together to construct hypothetical elections. These hypothetical elections occur under controlled conditions and are made to obey certain constraints. These simulations are then analyzed using a summary statistic about that condition. These “reference” conditions are electoral scenarios where bias can be measured

directly, and arguments about their plausibility (or validity) justify the logic of the each method.⁷ The distribution of summary statistics generated across all simulations provides an uncertainty bound on the statistic, and the distribution is either summarized directly in terms of quantiles or indirectly in terms of its mean and dispersion. Importantly, no *formal theory* provides structure to the measures' simulation distributions, so statements about confidence intervals on the “true value” of the simulation statistic are unavailable. Only statements about the likelihood of attaining a similar *result in simulation* are available. The simulation distribution is contingent on the entire chain of analysis, so misspecification in either the electoral model or the counterfactual model may affect the simulation distribution. This misspecification in models of the observed & counterfactual outcomes are discussed in Chapter 6.

Below, it is only necessary to understand that some stochastic process can generate simulated (possibly counterfactual) electoral outcomes for election cycle t , called \mathbf{h}_t° , given a set of observed vote shares \mathbf{h}_t , observed electoral conditions \mathbf{X}_t , and simulation conditions \mathbf{X}_t° . In the most general terms, the electoral model is some distributional statement about \mathbf{h}_t as a function of \mathbf{X}_t :

$$\mathbf{h}_t \sim \mathcal{P}(\mathbf{X}_t) \quad (2.1)$$

and the *counterfactual* model is a statement about \mathbf{h}_t° as a function of \mathbf{X}_t° , given that observed situation \mathbf{X}_t produced observed outcomes \mathbf{h}_t :

$$\mathbf{h}_t^\circ \sim \mathcal{P}(\mathbf{X}_t^\circ | \mathbf{X}_t, \mathbf{h}_t) \quad (2.2)$$

Some models have no analytical form for \mathcal{P} ; other models stipulate that the conditioning on \mathbf{X}_t and \mathbf{h}_t is unnecessary, so there is no difference between forward simulation from the electoral model (Eq. 2.1) and a counterfactual model (Eq. 2.2). Regardless, all of the simulated statistics are driven by many realizations of \mathbf{h}_t° and the empirical versions are driven solely by \mathbf{h}_t .

The five partisan advantage measures I compute can be reduced to three essential types. The first type are efficiency gap statistics. McGhee (2014) suggests two forms of the statistic;

⁷Practitioners and stakeholders were interviewed about their perceptions of the validity of these reference scenarios. These results are recorded in Section 9.4.

the “full” version, which measures discrepancies in “wasted” votes between parties, and the “simple” version, which assumes turnout is constant and reduces to a measure of the seats-votes relationship directly. These statistics can be assessed directly from observed \mathbf{h}_t and turnout vector \mathbf{m}_t in election t , and can *also* be computed for simulated elections \mathbf{h}_t° to provide a sense of the uncertainty around the observed value. If the model for \mathbf{h}_t° does not include information about \mathbf{m}_t , then simulated turnout \mathbf{m}_t° may be required.⁸ The efficiency gap statistics can be computed directly from observed election returns, but this provides no indication of their potential uncertainty, and so is not conducted in this dissertation.

The remaining types of advantage measures cannot be constructed without simulation. One of these types are symmetry measures from Gelman and King (1994a). These can be measured at the observed share of the vote, at the electoral median, or summarized as an average over an arbitrary range of vote shares. This dissertation computes the median & observed symmetry measures. Median bonus is the discrepancy between parties’ seat shares when they split the popular vote at 50%. The observed bonus requires a counterfactual “tables turned” scenario, summarizing a function of \bar{s} with one set of simulations at the observed \bar{h} and another set at $1 - \bar{h}$. Finally, the attainment gap, suggested by Linzer (2012) in a multiparty context, reflects the expected minimum share of votes required for a party to win a bare majority of seats.

2.3.1 Efficiency Gaps

The efficiency gap discussed by (Stephanopoulos and McGhee, 2015), derived from the relative wasted votes measure of McGhee (2014), and the partisan satisfaction measures suggested by Nagle (2015) use the differences in the size of parties’ majorities and minorities to measure define partisan advantage. McGhee (2014) suggests that most popular understandings of gerrymandering revolve around the concept of “wasted votes,” votes that are cast which

⁸For this perspective, note Linzer (2012)’s use of Gaussian mixture models for the joint distribution of \mathbf{m}_t and \mathbf{h}_t .

do not affect the outcome of an election. He argues that votes cast for winning candidates that they do not need in order to win and all votes cast for losing candidates are wasted, since they might, under other circumstances, be transferred to another district and change its winner without affecting the winner of the focal district. The votes wasted on winning candidates will be called here “surplus votes,” and votes wasted on losing candidates will be called “losing votes.” Added together, they provide the total number of votes wasted on a party. If the reference party wastes few votes and the opponent wastes many votes as a fraction of all votes cast, the districting plan makes the reference party more electorally efficient, and thus confers an advantage to that party.

In its original form from McGhee (2014), the efficiency gap is driven by the difference in parties’ total wasted votes. To define these let us first focus on a single election and let the losing votes for the reference party (W_l) be all votes cast for the reference party in districts i where the reference party loses:

$$W_l = \sum_i^N \mathcal{I}(\mathbf{h}_i < .5) * \mathbf{v}_i \quad (2.3)$$

where $\mathcal{I}(\cdot)$ is the indicator function returning 1 when the argument is true and zero otherwise, and \mathbf{v}_i is the raw number of votes cast for the reference party in district i . Further, let the surplus votes be the share of votes cast in districts a party wins, minus 50%:

$$W_s = \sum_i^N \mathcal{I}(\mathbf{h}_i > .5) * (\mathbf{h}_i - .5) * v_i \quad (2.4)$$

Total waste for the reference party (W_r) is then the sum of losing and surplus votes, $W_r = W_l + W_s$. Then, the same quantities are computed from $1 - h_i$, the opponent’s vote share vector, to get W_o , the opponent’s total wasted votes. The efficiency gap is then:

$$E_m = \frac{W_o - W_r}{\sum_i^N \mathbf{m}_i} \quad (2.5)$$

E_m reflects the difference in votes wasted between parties as a percentage of all votes cast. The system is biased against the reference party when E_m is negative, and biased towards the reference party when E_m is positive. To construct E_m from simulated elections, \mathbf{h}° is used in place of \mathbf{h} , and \mathbf{m}° may be used in place of \mathbf{m} if a model for \mathbf{m} is specified.

As discussed by McGhee (2014), the efficiency gap can be stated more simply in two-party systems when assuming that all districts have equal turnout:

$$E = (\bar{s} - .5) - 2(\bar{h} - .5) \quad (2.6)$$

This version of the efficiency gap is a straightforward constraint on the seats-votes curve, requiring neutrality and linear response of 2. This hyper-proportionality requires that a party's seat share increases by 2% for every increase in popular vote percentage. To assess this measure for simulated elections, \bar{s} and \bar{h} may be replaced by \bar{s}° and \bar{h}° .

2.3.2 Bonus Measures

Bonus measures derive directly from the neutrality property of Niemi and Deegan (1978). Reprising Niemi and Deegan (1978), King and Browning (1987) defines these measures as the “excess” seat share won (or lost) by the reference party *if its opponent were* to do as well as the reference party. For example, if the reference party wins $\bar{s} = .6$ of the legislature with an observed popular vote share $\bar{h}_1 = .52$, then it is necessary to estimate \bar{s}° , the fraction of seats the reference party wins when $\bar{h}^\circ = 1 - .52$. If $\bar{s}^\circ = .62$, then the system is biased against the reference party by 2%, since it wins 2% fewer seats than the opponent when tables are turned. Alternatively, if $\bar{s}^\circ = .58$, it is biased towards the reference party by 2% for the mirrored reason. This reflection strategy quantifies the “observed” excess bonus, since it characterizes the asymmetry of the seats-votes curve *at* the observed vote shares, \bar{h} for the reference party and $1 - \bar{h}$ for the opponent. Alternatively, if it were the case that $\bar{h} = 1 - \bar{h}$, then it would be sufficient to consider \bar{s} alone. If the reference party wins 50% of the popular vote but wins 53% of the legislature, the system is biased towards them by 6%. This is the “median” bonus measure, which demonstrates partisan symmetry at the electoral median, 50% vote share.

Both of these techniques require different sets of simulations. In addition, some view the “median” bonus as more realistic: since the median scenario simulates $\mathbf{h}_t^\circ | \bar{h}_t^\circ = .5$, it is always closer to the observed \bar{h}_t than $\bar{h}_t^\circ = 1 - \bar{h}_t$. So, \mathbf{h}_t° in simulations at the median is never a

more distant extrapolation than at $1 - \bar{h}_t$, and thus may reflect a less “extreme” counterfactual.⁹ Regardless of scenario, the counterfactual is core to the measure, resisted by academics, and might be found unrealistic by practitioners. A simulation regime to construct these symmetry measures are provided below.

Algorithm 1 *To compute partisan bias in an election t at a given target vote share \bar{h}^* in a two-party system with a counterfactual generation method like that in Eq. 2.2:*

1. *Simulate K realizations of district vote share vectors \mathbf{h}° .*
2. *Add a perturbation (electoral swing) δ to each \mathbf{h}_k° so that $\bar{h}_k^\circ = \bar{h}^*$, $k = 1, 2, \dots, K$.*
3. *Simulate K new district vote share vectors (or translate previous K simulations using a new δ) so that $\bar{h}_k^\circ = 1 - \bar{h}^*$.*
4. *Compute the corresponding party seat shares, \bar{s}° , for all K realizations in both sets of simulations.*
5. *Over all simulations, compute the average of \bar{s}° in both sets of simulations, denoted \bar{s}_1° and \bar{s}_2° .*

The bonus towards the reference party at \bar{h}^ is:*

$$B_{\bar{h}^*} = \bar{s}_1^\circ - (1 - \bar{s}_2^\circ) \quad (2.7)$$

Breaking this down into its constituent terms, the second term, $(1 - \bar{s}_2^\circ)$, is an estimate of the share of seats the opponent wins when they get \bar{h}^* share of the vote. The first term, \bar{s}_1° is the share of seats won by the reference party when they get \bar{h}^* vote share. So, B is expressed in the percent of extra seat share the reference party wins. Positive values indicate that the reference party would expect to win more seats than the opponent if it wins $\bar{h}^\circ = \bar{h}^*$ share of the popular vote. In contrast, negative values of B mean the opponent party can expect to win more seats than the reference party would at $\bar{h}^\circ = \bar{h}^*$.

⁹In all interviews in Washington and Arizona, interviewees corroborated this statistical intuition: parties splitting the popular vote at 50% was more likely than parties swapping in their observed statewide vote share. This is discussed in Chapter 9.

This can also be simplified when $\bar{h}^\circ = 1 - \bar{h}^\circ$. Since $\bar{s}_1 = \bar{s}_2$ in that case, the equation can be restated:

$$\hat{B}_{.5} = 2 * \bar{s}^\circ - 1 \quad (2.8)$$

which only requires K simulated \mathbf{h}° vectors and a single shift term, δ . In addition, since $|\bar{h}^\circ - .5| < |\bar{h}^\circ - (1 - \bar{h}^\circ)|$ for any $\bar{h}^\circ \in (0,1)$, the required δ will always be smaller to produce simulations at the median election than simulations at $1 - \bar{h}^*$.

2.3.3 Attainment Gap

In theory, attainment gap is the inversion of the relationship measured by the median bonus measure in Section 2.3.2. Instead of measuring the excess share of seats the reference party wins when the vote is split evenly, the attainment gap estimates the difference in vote shares when the seat share is split as close to evenly as possible from above. If one party can win a majority of the seats with fewer votes than another party, it has an electoral advantage over the other party. This measure can also be used in multiparty systems, since it focuses solely on a single party and ignores the breakdown of other parties' seat and vote shares Linzer (2012).

However, this inversion is more difficult to estimate. It requires the generation of scenarios at fixed \bar{s}° rather than \bar{h}° . Typical simulation models allow for both \bar{h}° and \mathbf{h}° to be controlled, but \bar{s}° is often not controlled. Since the functional relationship between \mathbf{h} and \mathbf{s} is lossy, a specification in terms of \bar{s}° would not be complete for \bar{h}° ; a majority can be built many ways. In addition, this simulation strategy also requires counterfactual estimation, since parties are often not observed as winning the smallest possible majority. Further, this measure may be difficult to apply in states with small, even-numbered delegations; for a state with two districts, the barest majority is a single-party sweep of the delegation. The attainment gap in this case would estimate the smallest expected vote share at which this sweep occurs.

While Linzer (2012) suggests estimating the attainment gap by extrapolating linearly (according to an estimate of responsiveness) along the seats-votes curve to $(\bar{h}^\circ, .5)$, this estimating procedure admits no uncertainty about the value of the attainment gap. Instead, the problem

can also be stated as a direct optimization problem over a stochastic process. One algorithm to estimate the attainment gap would conduct a grid search for the minimum \bar{h}° with $\mathbf{E}[\bar{s}] > .5$:

Algorithm 2 *Given a counterfactual generation method like that in Eq. 2.2 and a convex loss function $\mathcal{S}(\hat{\theta}|\theta)$, such as a mean absolute deviation or squared error loss for the estimate $\hat{\theta}$ and target θ , the attainment gap in election t can be estimated by finding the \bar{h}° such that $\mathcal{S}(\bar{s}|\cdot)$ is minimized from above.*

1. *generate a batch of K realizations of \mathbf{h}° at a starting party vote share, \bar{h}° , and compute $\mathbf{E}[\bar{s}^\circ]$ for that batch.*
2. *generate two more batches of K realizations, \mathbf{h}_+° and \mathbf{h}_-° , at both $\bar{h}^\circ \pm \delta$, where δ is a small step size and compute the $\mathbf{E}[\bar{s}^\circ]$ for each batch.*
3. *score each batch using the loss function $\mathcal{S}(\bar{s}|\cdot)$. If $\bar{s} < .5$ for a batch, let the loss of that batch's \bar{h}° value be infinite.*
4. *If \bar{h}° has the lowest score, designate it as the optimal \bar{h}^* and save the value. Otherwise, shift \bar{h}° to the scenario with the lowest loss and return to 1.*

Repeating this procedure L times, the attainment gap is the expected value of \bar{h}^ over L replications:*

$$A = .5 - \mathbf{E}[\bar{h}^*] \quad (2.9)$$

Thus, the attainment gap is expressed as the extra vote share a party must win (or can afford not to win) in order to gain a majority. If the attainment gap is negative, the party expects to need *more than 50%* of the votes in order to win a bare majority of the seats. Alternatively, if the attainment gap is positive, the party can expect to need *fewer than 50%* of the votes to win a majority

While the form of the loss function \mathcal{S} may change slightly from batch to batch, using the expected values over K realizations in each batch significantly reduces its variability. Thus, standard bounded line search techniques, such as Brent's method, can also be used instead of grid search. In addition, estimates of A tend to be insensitive to whether absolute or squared

District No.	1	2	3	4	5
\mathbf{h}_{2016}	0.55	0.64	0.38	0.0	0.42
\mathbf{m}_{2016}	349,398	325,408	313,277	229,919	323,534
	6	7	8	9	10
	0.62	1.0	0.39	0.73	0.59
	327,834	378,754	320,865	281,482	290,564

Table 3. Washington congressional election results, 2016

error loss is used. However, each replication requires many batches of simulations to be generated as the loss function is minimized. One could also retain the initial \mathbf{h}° vector and simply optimize \bar{h}° as a function of δ directly. In practice, either of these methods yields consistently smooth loss functions, and the estimated attainment gaps are stable across replications for either method.

2.3.4 Example: Washington 2016

In the following discussion, I compute the measures for the observed congressional elections in Washington in 2016. In the analysis in the rest of the dissertation, these observed results stand in as one of the many realizations summarized using these advantage measures. Thus, while I present only one realization in this example, the analyses later in the dissertation summarize over many simulated elections. Table 3 contains the election returns for Washington in 2016. Notably, districts 4 and 7 are considered “uncontested.” These were, in fact, contests between the top-two primary finishers, which both happened to be of the same party. Due to Washington’s top-two primary, two Republicans ran against each other in the general election in district 4, two Democrats in district 7. I will leave them as uncontested in this example to keep the discussion simple, but in Chapters 4,5, and 6 I impute the values of uncontested elections, analyzing the expected vote share *if they were to have been contested*.

For the efficiency gap, it is necessary to total the vote waste. Districts 3,4,5, & 8 were lost

to Democrats, and so *all* votes cast for Democratic candidates in those districts are lost:

$$W_l^D = .38 * 313227 + 0 * 229919 + .42 * 323534 + .39 * 320865$$

and the waste in districts where Democrats won derives from the remaining districts:

$$W_s^D = (.55 - .5) * 349398 + (.64 - .5) * 325408 + (.62 - .5) * 327834 \\ + (1 - .5) * 378754 + (.73 - .5) * 281482 + (.59 - .5) * 290564$$

The same calculations for the complement sets of districts yields the two components of waste for Republicans:

$$W_l^R = (1 - .55) * 349398 + (1 - .64) * 325408 + (1 - .62) * 327834 \\ + (1 - 1) * 378754 + (1 - .73) * 281482 + (1 - .59) * 290564 \\ W_s^R = (.5 - .38) * 313227 + (.5 - 0) * 229919 \\ + (.5 - .42) * 323534 + (.5 - .39) * 320865$$

and the final efficiency gap is:

$$E_m = \frac{(W_l^D + W_s^D) - (W_l^R + W_s^R)}{\sum_i^N \mathbf{m}_i} = -.016$$

The simple gap can be obtained directly from the \bar{h} and \bar{s} values. In this case, we refer to the simple average (not turnout-weighted average) of \mathbf{h}_t , since this measure assumes all districts are equally-sized. Thus, the simple efficiency gap for 2016 in Washington is:

$$E = (.6 - .5) - 2 * (.53 - .5) = .04$$

Using the turnout-weighted average vote share,¹⁰ the measure is slightly different:

$$E = (.6 - .5) - 2 * (.55 - .5) = .00$$

Since there is no uncertainty information provided by these estimates, it is unknown whether an efficiency gap of .04 is egregiously large (or even significantly different from zero) without conducting simulation studies.

¹⁰which is also the share of popular vote for Democrats

Moving to the bonus measures, I first examine the median bonus. The worked example below might be thought of as *a single* realization of the bonus statistic, where many realizations are used with simulated \mathbf{h} and δ and summarized to provide an estimate. The empirical \mathbf{h}_{2016} will be used below, with two types of δ considered, a strict uniform and generalized uniform effect. Again, using the average (not turnout-weighted) vote share $\bar{h} = .53$, I add $\delta = -.03$ to \mathbf{h} to shift $\bar{h}^\circ = .5$.¹¹ In the case of districts where δ causes an invalid vote share to occur, these cases may be truncated to $(0, 1)$ (Gelman and King, 1994a). If imputation is used (as discussed in Chapter 6), this truncation is rarely required, since the shift to the median is often much smaller than the distance from any imputed district's vote share to 0 or 1. With this shifted vote share vector, no districts change hands. Thus, Democrats still win 6 seats, so \bar{s} still is .6, and the expected bonus at median is:

$$B_{.5} = 2 * .6 - 1 = .2$$

Note that if δ were a random effect, a shift of $-.03$ in mean *may* result in district 1 flipping from Democrats to Republicans, if $\delta_1 \leq -.05$. In that case, that realization's bias measure would be:

$$B_{.5}^\circ = 2 * .5 - 1 = 0$$

Since no other districts are likely to flip when $E[\delta] = -.03$, the fraction of times district 1 is won by Democrats versus the fraction of times it is won by Republicans in simulations is the effective determinant of the value of $B_{.5}$.

For the observed bonus, we must flip $\bar{h}^\circ = 1 - \bar{h} = 1 - .53 = .47$, which requires a δ of twice the magnitude.¹² If we simply add $\delta = -.06$ to \mathbf{h}_{2016} , we see district 1 flipping from Democrats to Republicans, but no other district flips. This would make $\bar{s}_2 = .5$, since Democrats win 5 seats in the “tables turned” counterfactual, when their vote share is 47%.

¹¹Accounting for varying turnout is not difficult here, too. A uniform swing of δ is converted using the turnout weights into a vector of adjustments. I deal in the average here to keep discussions simple.

¹²Thus, one sees how potential spatial correlation in δ may cause different simulation outcomes to be more likely, especially as swings become large.

Thus, the observed bias in this scenario would be:

$$B_{\bar{h}^*} = .6 - (1 - .5) = .1$$

However, if δ were a random effect, it may be the case that district 10 flips in addition to district 1, or that neither districts 10 or 1 flips. If both flip, the realizations' observed bonus would be

$$B_{\bar{h}^*} = .6 - (1 - .4) = 0$$

If neither flip, the realizations' observed bonus would be

$$B_{\bar{h}^*} = .6 - (1 - .6) = .2$$

On the whole, it is only likely that districts 10 and 1 are "in play" during this counterfactual; either both of them flips, only one flips, or none flip. The estimate of $B_{\bar{h}^*}$ thus is a summary of the frequency of those three outcomes.

For the attainment gap, the minimum value of \bar{h} such that $\bar{s} > .5$ is needed. Since there are an even number of districts (and 50% is not a majority), this means the smallest vote share where Democrats still win 6 seats is required. In this case, assuming a strict uniform swing, that would be $\bar{h} = .48$, the value at which district 1 flips to Republicans and Democrats & Republicans split the delegation at 50%. This would make the attainment gap estimate:

$$A = .5 - .48 = .02$$

indicating that Democrats win majorities in the Washington congressional delegation with around 2% fewer popular votes than Republicans. For a random swing with fixed expectation, the average of the minimal attainable \bar{h}° would be the estimate used for the attainment gap, given that either districts 10 or 1 might flip in simulations.

2.4 Geometric Measures of Boundary Manipulation

While partisan gerrymandering is embodied by the existence of partisan advantage, many previous studies and attempts at detecting partisan gerrymandering (and gerrymandering more

generally) have focused instead on electoral boundary manipulation. Akin to the early work on the measurement of partisan advantage, many measures that purport to identify when a congressional districts' boundary obscures or manipulates the underlying population distribution have been developed. Early measures, such as the moment of inertia measure suggested first by Weaver and Hess (1963), are essentially the same in spirit as more recent measures (Fan et al., 2015). In addition, concern about the sufficiency of geometric measures to identify legitimate boundary manipulation have also been ever-present (Young, 1988; Humphreys, 2011).

Regardless, many of the most commonly-used measures of boundary manipulation arise from simple geometric relationships between the observed shape of a congressional district and an ideal reference shape. Typically, one measure purports to identify a single dimension of shape regularity, such as shape elongation, boundary perforation, or winding (Niemi et al., 1990; Altman, 1998a; Wentz, 2000). Recent measures, such as that suggested by Chambers (2010) or Fryer and Holden (2011) attempt to make the scores relative to the context in which regularity is achieved. They aim to account for the fact that some states are less regularly-shaped than others. The way the frame is divided may affect the regularity of possible district shapes, so controlling for this frame dependence is necessary to make geometric measures that can be compared between states. For Chambers (2010), this is done by considering the set of shortest paths between voters; districts that are likely to fully-contain shortest paths are considered well-shaped. Likewise, Fryer and Holden (2011) suggest an index that describes plan compactness relative to the maximal compactness possible for a given frame. More recently, Fan et al. (2015)'s novel moment of area measure follows a similar logic to Weaver and Hess (1963) and Boyce and Clark (1964), but also construct a relative measure, standardized by the maximally-compact packing of population available in the district.

While these attempts to address shape measures' generalizability is admirable, critical evaluations of the effectiveness of these new measures in the vein of MacEachren (1985) is not common. Indeed, since allegations of boundary manipulation are *based* on these measures, new ones are often not "ground-truthed" by identifying commonly-agreed-upon districts that

were manipulated. While Ansolabehere and Palmer (2015) does this in a sense, comparing recent districts to the original “Gerry mander” of 1812 to identify whether compactness has increased or decreased as a whole, the use of a pre-*Baker v. Carr* district is not helpful; the legal and social regime for redistricting was massively different after the 1960s cases discussed in Section 2.1. Thus, while work strives to make these boundary manipulation measures comparable between states or over time, fundamental issues with validity, uncertainty, and accuracy have not been addressed.

At its core, many reservations with shape measures revolve around their overly-simplistic view of human spatial social structure. Part of the issue with identifying boundary manipulation is that people do not tend to live in regular polygons, tiled spatially with similar social, ideological, racial, and ethnic characteristics (Archer, 1988; Gimpel and Schuknecht, 2009; Walker, 2013). Indeed, the spatial and ethnic configuration of people in cities is not exogenous; social and spatial aspects of ethnic and racial divisions can reinforce or arrest political attitudes, sometimes changing their expressions (Giles and Hertz, 1994; Sastry et al., 2002; Rocha and Espino, 2009). Some cities (or subregions within cities) have significant disparities in terms of their local population density, and the structure of these differences manifests in different styles of population grouping (Reardon et al., 2006). In addition, the link between the observed district boundary and the intuitive idea about how a given numerical summary presents the manipulation of “natural” or “latent” social or physical boundaries is tenuous. Compactness (and shape regularity more generally) has many dimensions (Angel et al., 2010), and many practitioners interviewed in Chapter 9 could not articulate what they viewed as important when thinking about district shape regularity.

Regardless, geometric measures provide the only commonly-used *local* measures of districts. The partisan scoring methods discussed in Section 2.3 generate a single, plan-wide score. In contrast, many of the geometric scores are expressed district-by-district. With the exception of Fryer and Holden (2011), this means that individual “bad” districts can be identified. Thus, these measures are commonly used in both popular discussion and litigation about redistricting concerns. In fact, occasionally these measures of shape regularity are balanced

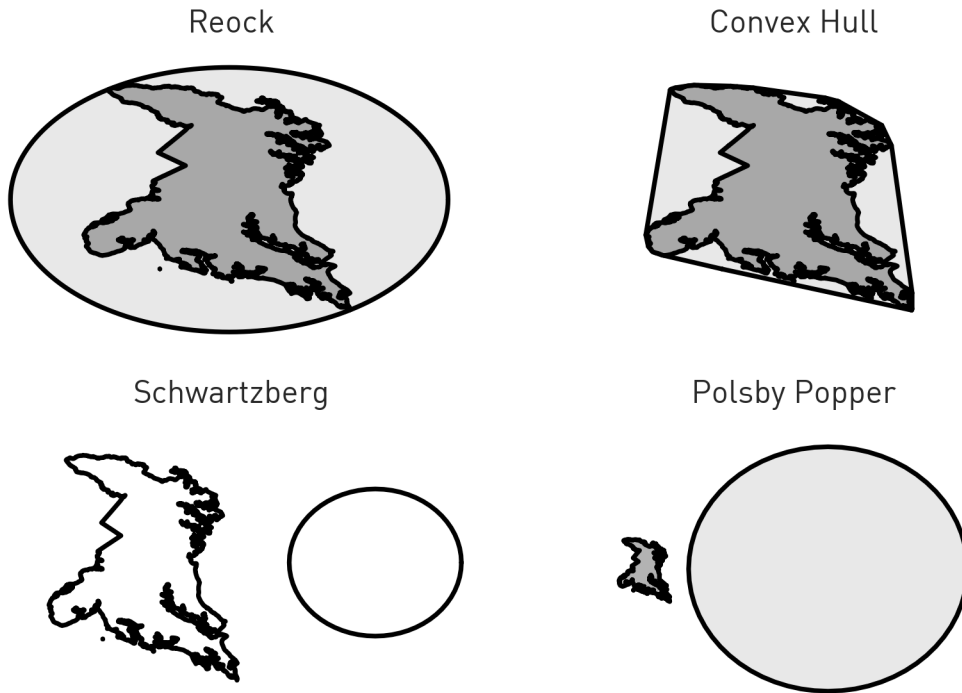


Figure 1. Ideal shapes for a congressional district, Maryland's 5th District from 1903 to 1923.

against measures of political advantage; when political advantage is detected over the entire plan, spatially-irregular districts are often singled out for legal reproach. Since the partisan measures used up to this point have no capability to identify which districts strongly affect the statewide bias measure, this is a reasonable mode of analysis.

So, in order to compare the localizations of the bias measures in Section 2.3 to geometric measures, I first discuss which geometric measures are used and how they will be computed. The five measures used in my comparison derive from four distinct ideal shapes. These ideal shapes are shown below in Figure 1. Typically, ideal shapes refer explicitly to an optimization of the given metric used to identify irregularity. When using a population measure, the ideal should have a uniform and efficient distribution of population; when using a perforation measure, the ideal should have as nearly smooth boundaries as possible. In practice, many ideal shapes are circles or simplexes.

Most of the common geometric forensics range between zero and one, where one indicates perfect similarity to the ideal shape, and zero indicates perfect dissimilarity. As a result, districts that score well are close to unity, and are considered “not gerrymandered” because their shapes are close to the ideal comparison shape. Measures that focus on boundary perforation are often highly-sensitive to the scale at which the boundary is measured (Mandelbrot, 1967), and so are often avoided in cases of complex coastline districts. While these statistics have a restricted domain, no further distributional theory is available for their values. They admit no uncertainty, being simple summaries of the geometric or population information about a district, and provide no indication of how unusual a given shape is in a relative or absolute sense.

2.4.1 Ideal Circle Measures

Three measures used in this dissertation fall into this category. Many compactness properties are satisfied by circles (Angel et al., 2010), but the reference circles constructed for each of the three measures used in this dissertation are distinct. The first reference circle measure, suggested famously by Polsby and Popper (1991) is the isoperimetric quotient, sometimes called the Polsby-Popper metric in the districting literature. The isoperimetric quotient is a well-known property of shapes that describes how “efficiently” an area is enclosed by a perimeter. Polsby and Popper (1991) suggest that this measure should filter out districts that meander around the map, attempting to avoid or pick-up target areas. For a district D , let its area be denoted A_D and its perimeter be denoted P_D . Then, the isoperimetric quotient is the ratio between the area of the district and the area of a circle having the same perimeter as the district:

$$IPQ = \frac{4\pi A_D}{P_D^2} \quad (2.10)$$

The IPQ is always less than one, and is exactly one when the shape is circular. Since a circle encloses an area with the minimal perimeter, this measure can be thought of as the “shrinkage” in the size of the district due to its kinked and winding perimeter. Schwartzberg (1965) suggests a similar metric, instead using the isoareal quotient. A less well-known quantity, the isoareal

quotient relates the perimeter of the district to the perimeter of a circle enclosing the same area:

$$IAQ = \frac{\sqrt{\frac{A_D}{\pi}}}{P_D} \quad (2.11)$$

One novel measure for this purpose might acknowledge that a perfectly-spherical district would not tile alongside other districts. Thus, an isoareal/isoperimetric quotient might also be derived using the regular hexagon with the same area/same perimeter, respectively. While the originators suggest that these two measures ostensibly single out different districts and use different arguments to justify the measures, the two are perfectly rank-correlated, since they are one-to-one nonlinear transformations of one another. Thus, if examining the *value* of these scores or their raw correlations with other measures, we might expect them to be different. However, examining quantiles or rank distributions of these statistics should yield nearly identical selected districts, depending on the precision of the computation, and identify identically-manipulated districts, despite the fact that both authors provide distinct arguments about what the indices measure.

The final ideal circle measure used here is the Reock measure (Reock, 1961). This measure is argued to identify elongated shapes. It relates the area of the minimum bounding circle to the area of the target district:

$$R = \frac{A_D}{A_{MBC}} \quad (2.12)$$

Since the bounding circle is guaranteed to contain the district, it must have an area at least as large as the district. Therefore, R varies between zero and one, with values approaching one indicating that a district is very nearly shaped like its bounding circle. Computing the minimum bounding circle is a linear-time optimization problem isomorphic to a facility location problem. However, a large constant factor (and high-resolution boundaries) make it difficult to compute this value for many shapes. At most, the minimum bounding circle intersects three points of the input shape (Skyum, 1990), and at worst has the diameter of the shape (the furthest pairwise distance between boundary points) as its own diameter. There is no intrinsic meaning to the minimum bounding circle in this context, in the same sense as the circles in the isocircle measures.

2.4.2 Convex Hull Measures

Two other measures used to compare shapes involve the convex hull of the district. A convex hull is the simplex containing all pairwise pairwise connections between points on the shape (De Berg et al., 2008). Stated in terms of dissimilarity metrics, the following two measures are studied by Brinkhoff et al. (1995) for examining geographic shape regularity. One was also used by Ansolabehere and Palmer (2015): the convex hull areal ratio. This relates the area of a shape to the area of its convex hull:

$$CA = \frac{A_D}{A_{CH}} \quad (2.13)$$

Like the Reock measure, this lies between zero and one, with values close to one indicating that the district is almost coincident with its convex hull. Since the convex hull is guaranteed to contain the district, its area is guaranteed to be at least as large as the district. An alternative measure involving convex hulls relates the length of the district perimeter to the perimeter of the convex hull:

$$BA = \frac{P_{CH}}{P_D} \quad (2.14)$$

This specification is a direct similarity measure. By pivoting the dissimilarity measures from Brinkhoff et al. (1995) into similarity metrics, I preserve the same interpretation as the previous ideal shape measures: values approaching one indicate close-to-ideal shapes, and values near zero indicate potentially-manipulated boundaries. Brinkhoff et al. (1995) call their dissimilarity specification of this measure the “boundary amplitude”, since it provides a rough indication of how twisted the boundary of the polygon must be to fit within the convex hull.

These five measures will be used to attempt to characterize how “regular” district shapes are when boundary regularity is compared directly to political impact. By relating them to the measures of district impact developed in the next section, I will demonstrate the link between suggested measures of boundary manipulation and the actual impact districts whose boundaries may be manipulated have on the measures of partisan advantage.

Chapter 3

DATASET: SPATIOTEMPORAL DATABASE OF CONGRESSIONAL ELECTIONS, 1898-2016

3.1 Sources of Constituency-level Electoral Data

Longitudinal study of Congressional elections in the United States focusing on the estimation of bias and responsiveness is not new. With the publication of King (1994), high-quality data on US Congressional elections at constituency level was made available for various studies of redistricting, voter behavior, and electoral system analysis. Many of the influential post-*Bandemer* studies on the impact of redistricting on Congress uses this data (Gelman and King, 1994b,a). In the decade after it was published, many other studies of American elections also used this data.

However, later studies of elections have not provided data to extend King (1994) directly. In most cases, these studies both extend and enrich the original data set, providing a superset of the original Congressional elections data. Often, these analyses focused on sociodemographic study of redistricting's impact on various aspects of the electoral system (Abramowitz et al., 2006b; McDonald, 2006; McKee et al., 2006) or are general studies of the social and demographic structure of American Congressional geography (Gimpel and Schuknecht, 2009; Crespin et al., 2011). While privately-owned data exists for this purpose, the price of obtaining coverage comparable to (King, 1994) is high. Thus, the Constituency-Level Electoral Archive (CLEA) was developed in part to provide an extended, more detailed, and open data set on legislative electoral geography (Kollman et al., 2016). For US Congressional elections, this data set also provides much more data about minor parties and candidates themselves, and has been used in a variety of contemporary electoral studies (Linzer, 2012; Gerring et al., 2015; Kayser and Lindstädt, 2015; Bochsler, 2016). Since the CLEA is a multi-country data set, it is used often for comparative studies that examine the generalizability or comparative validity of particular theories about campaigns or elections, as well as generic polimetric or psephological

studies. For US elections, the CLEA has been used for longitudinal analysis of electoral structures, examining how specific electoral properties, like effective number of political parties or competitiveness, change over time.

For the geospatial research on US Congressional elections, work has focused on the development and propagation of macro-scale sectionalism and the construction of ideological and geographic voting blocs over time (Archer and Taylor, 1981; Bensele, 1987; Shelley and Archer, 1995), as well as the analysis of scale-sensitive political identities, both in redistricting and in identifying “normal” vote (Openshaw and Taylor, 1979; Archer, 1988). Recently, calls for a revitalized quantitative electoral geography, focusing on electoral systems analysis and voter behavior, have been made (Warf and Leib, 2011; Cho and Gimpel, 2012), and many foundational problems in the spatial analysis of electoral systems, such as those articulated by Gudgin and Taylor (1979), have been explicitly reclaimed by contemporary authors (Rodden, 2010; Calvo and Rodden, 2015). This has seen an explosion of spatio-temporal analysis of the electoral geography of the U.S. Congress (Calvo and Escolar, 2003; Coleman, 2014), as well as analyses at non-Congressional spatial scales (Gelman, 2007; Bishop, 2009; Hawley and Sagarzazu, 2012), and non-American electoral systems (Shin and Agnew, 2007; Harbers, 2016). Altogether, the literature on electoral analysis has become both robust and wide ranging.

Complementing the revitalization of electoral systems analysis, most of the data generated in this literature has been openly shared under permissive licenses. However, the construction and maintenance of spatially-referenced data sets for Congressional analysis can be a more difficult process than analysis of elections at a state or county level. Indeed, King (1994) and the CLEA only provide spatial information in terms of the states in which districts are found. They do not provide information about the shape or extent of the districts, nor the neighborhood & topological relations between districts. Geography & spatial effects may be richer than nesting relationships alone, however, so longitudinal study of spatial effects in congressional districts is quite restricted (Owen et al., 2015). While the CLEA provides a selection of “georeferenced elections data” (the GRED), this data set is not comprehensive, with limited temporal scope when compared to US elections coverage in the CLEA. To remedy this, I have constructed a

general-purpose spatial database for this work that extends King (1994) forward in time using data from the CLEA and novel data on incumbency. With this extended dataset, I connect the individual district geometries compiled by Lewis et al. (2013) to yield a single spatio-temporal database of US Congressional elections since 1898 using a single continuous encoding and indexing scheme. In what follows, I discuss the process for constructing this data set, compare the relative values of the source data sets, and briefly discuss potential use cases or novel analyses that this new data may provide.

3.2 Methods

To extend King (1994) using the CLEA and bind both to Lewis et al. (2013), a common key across all data sets was needed. This required a coherent data modeling strategy that could encompass the abstractions in each of the data sources. Thus, the data set I construct is a collection of the results of general elections to the US Congress. Elections to a Congress are composed of some number of *contests*, which are electoral challenges in which some number of winning candidates are declared. Each contest occurs within some territorially-bound constituency, or *district*, which may or may not be unique in each Congress. Relating the three data sets required disambiguating the relationship between contests, districts, and the Congresses. Occasionally, a district may have more than one contest within it, like the Alabama congressional for the 88th congress, where all eight seats were elected at large.

To ensure unique indexing of contests and disambiguation from the districts themselves, the Inter-University Consortium for Political and Social Research (ICPSR) numerical codes for states used in King (1994) were first converted to US Census Bureau Federal Information Processing Standard Codes (FIPS codes). These codes are contained in column `state_fips` of the example table segment shown in Table 4. Then, a composite database key was constructed to refer uniquely to a contest's Congress, state, and district. The first three characters of the composite index reflect the Congress number of the record, with zero-padding on the left if the Congress number is less than three digits. The second three digits of the composite index are

the zero-padded state FIPS code. The third three digits of the composite index are the zero-padded district number. This provides a unique index for each territory while allowing more than one seat to be in each of these constituencies.

Finding a common district numbering scheme was the first design challenge for this data set. In all cases, the sources used different conventions to refer to “at-large” districts, districts that are the entire state. In the CLEA, at-large districts were variously referred to as the first district (`vermont 01`), as the “zeroth” district (`vermont 00`), or having no number (`vermont`). In all cases, Lewis et al. (2013) referred to at-large districts as the “zeroth” district. In both data sets, district numbers refer to the spatial constituency in which the contest occurs, so multiple contests may have the same district number. In contrast, King (1994) constructs multiple district indexes in the case of a multi-member district. In at-large contests, a “district” index is labeled decreasing from 98, meaning that in these cases, the index uniquely identifies a contest, not a district. If two candidates run at-large in a state, their “district” numbers are 98 and 97 even though the spatial territory of their electorate is the same.

The Lewis et al. (2013) convention is the most simple and robust for this application since it treats the district consistently as a spatial object, rather than as a hybrid of contest and district concepts. In addition, the Lewis et al. (2013) indexing strategy retains an advantage of King (1994)’s index, since at-large contests can be separated efficiently. At-large contests can be structurally different from typical Congressional elections that occur at the sub-state level, and at-large occasionally merit separate consideration. So, the CLEA records were made consistent with the Lewis et al. (2013) convention, and the King (1994) records were converted to this convention as well. The final index is contained in the `geom_id` column of Table 4, and the original indices retained in `king_dist` and `lewis_dist`, which reflect the two consistent styles of district numbers.

In addition, a unique index for the geometries themselves, the index from Lewis et al. (2013), is retained in each record. This is composed of four three-digit codes. The first component is the zero-padded state FIPS code. The second component is the Congress in which the district shape first appeared. The third component is the last Congress in which the district shape

was used. The final component reflects the district number assigned to the district during its lifetime. This index uniquely identifies the geometries of constituencies in US Congressional elections, whereas the other index provides a unique identifier for the contests, referring to their Congress, state, and constituency. In addition, this index allows for the construction of a high-quality redistricting indicator variable since a contest in a “new” district is one whose congress variable matches the second triplet in the component index.

For elections before 1992, the `vote_share` and `turnout`, `delsouth`, and `inc` covariates are taken directly from King (1994). For later elections, `vote_share`, `delsouth`, and `turnout` is constructed from the CLEA and `inc` is coded by hand. All variables aim to replicate the method used to generate King (1994). The simplest to replicate is `delsouth`, a binary variable indicating that a record is in a southern state. The vote shares, in this case, are the share of the two-party vote cast for Democratic candidates. To construct this from the CLEA, the total number of votes cast for the candidate endorsed by the major party candidates is recorded. Thus, in cases of “fusion voting,” where the same candidate appears on multiple party tickets, these counts are added to the major party’s total. The sum of these votes is `turnout`. Then, the share of `turnout` that the Democratic candidate receives is the `vote_share`. To match the structure of King (1994), the detail in the party identification in the CLEA was reduced to three parties: Democrat, Republican, and Other. In most cases, the reduction in parties was not significant. However, one case should be mentioned: Farmer-Labor candidates before the Democrat Farmer-Labor merger in 1944 were considered to be Democrats. This decision does not affect the resulting data product, since only King (1994) was used during this period, but will be apparent in the validation plots shown below. The CLEA does not contain incumbency information, so the `inc` variable was derived by hand from Congressional rosters. The post-1992 `inc` variate was coded to match King (1994): a Republican incumbent who runs for reelection is coded as -1, a Democrat who runs for reelection is coded as 1, and a zero is recorded when there is no single incumbent. Together, this comprises the dataset produced from the CLEA. It provides similar data to King (1994) in addition to extending past 1992 and enriching the data with spatial information. The period of overlap in King (1994) and the data derived from the

congress	delsouth	fips	contest_uid	inc	king_dist	lewis_dist
114	1	22	114022006	0	NaN	6
114	0	23	114023001	1	NaN	1
96	0	6	96006027	-1	27	27
96	1	13	96013002	1	2	2

state_name	turnout	vote_share	year	geom_uid	wkb
louisiana	329327	0.288853	2014	022113114006	01060000...
maine	281425	0.663317	2014	022113114001	01060000...
california	175272	0.489981	1978	006094097027	01030000...
georgia	42234	1	1978	013093097002	01030000...

Table 4. The example schema of the final data product, broken into two lines. The column containing the shapes encoded in well-known binary is truncated for brevity.

CLEA, all US elections from 1896 to 1992, will be analyzed in the validation section to ensure that the CLEA data after 1992 comports with the data sourced from King (1994).

After this, all data was inserted into a SQLite database and a outer join conducted, retaining all district shapes. The join resulted in over 98% of matches on keys, so only a tiny fraction of the districts in Lewis et al. (2013) did not find matches in the extended King (1994). Most of the remaining missing entities reflected duplications, malformed original entries that slipped through the data cleaning process, or non-voting constituencies that were not recorded in the CLEA or King (1994). The geometry information was stored in a text format in a column, `wkb`, of the resulting comma-separated table. This column is shown truncated in Table 4, and contains Polygons or MultiPolygons (as defined by Open Geospatial Consortium (2010)) encoded in well-known binary (WKB), stated in hexadecimal. This more concise statement of WKB-encoded geometry is common in database software (such as PostGIS), but still results in a column with long elements. The coordinates are stored without a coordinate system using the NAD83 datum, inherited from Lewis et al. (2013).

3.2.1 Code availability

The methods used to generate the data set will be made available through the Open Science framework. All scripts were implemented in Python, and requires a few Python data analysis libraries: `pandas`, a tabular data processing library, `geopandas`, a geospatial tabular processing library, `numpy`, a numerical computation library, and `SQLite`, used for the final out-of-core database join. In addition, a `makefile` is provided for convenience, to ensure the build process executes in the correct order.

When run in the correct order, the scripts generate intermediate data and final data products from a collection of sources. First, Kollman et al. (2016), King (1994), and the manually-constructed collection of incumbency information for elections beyond 1992 are contained in a `sources` directory. The first script, `00_get_all_shapes.py`, collects all district shapes from the repository maintained by Lewis et al. (2013), placing them in the `sources` directory as well. Two intermediate data products are constructed. First, after running `01_data_munge_clea.py`, a cleaned and party-reduced version of the CLEA data is stored in `intermediates/clean_clea.csv`. Second, the next script in the sequence, `02_rebuild_database.py`, combines all of the district shapes together in a single table in a `SQLite` database. Then, the fourth script, `03_extended6311.py`, concatenates the original King (1994) with the cleaned CLEA data. This first product, the `extended_6311.csv` data set, has the same schema as the final data set constructed by `04_final_merge.py`, which accesses the `SQLite` database and merges the extended King (1994) with the collection of district shapes. In this merge, two final outputs are generated, `products/pre1948.csv` and `products/post1948.csv`, which split the results of the merge in two parts. The split divides the series roughly in half, and corresponds to the division between full Congresses organized by the Legislative Reorganization Act of 1946. Finally, if more columns from the Lewis dataset are required by analysts, the retained columns of the merge in `04_final_merge.py` can be changed without affecting the merge process.

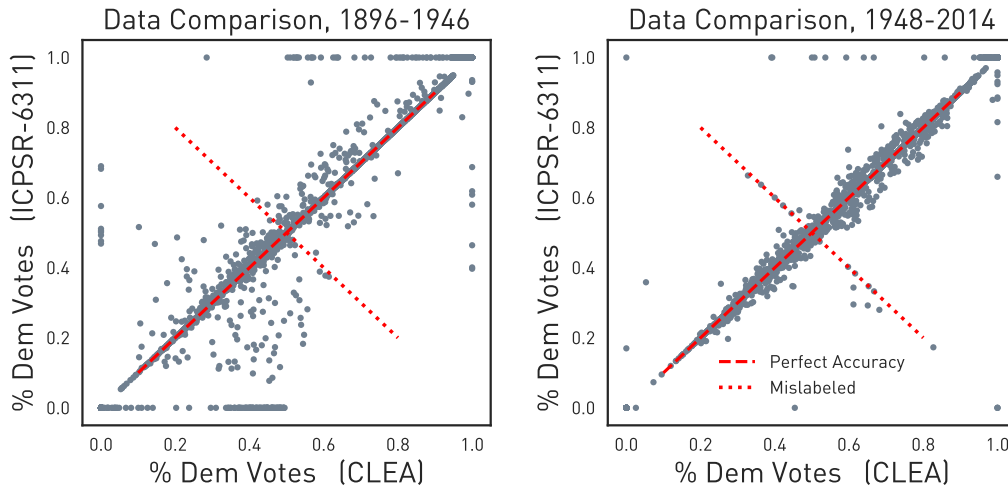


Figure 2. Relationship between Democrat share of the two-party vote in King (1994) (ICPSR-6311) and that constructed from Kollman et al. (2016) (CLEA).

3.3 Data Records

As discussed in the Section 3.2.1, three data products are combined within the dataset. The spatial dataset due to Lewis et al. (2013) is split into pre-1948 and post-1948 components to reduce the size of the resulting product. To join these pieces together, the latter table's header must be removed and the tables concatenated. Both tables have the schema discussed above in Table 4 with columns in the same order. To assist those who have no need for the spatial referencing, `products/extended_6311.csv`, named for the original ICPSR numerical designation of King (1994), contains the complete elections data with the geometric column omitted.

3.4 Technical Validation

After merging and validation, the resulting two-party vote shares constructed from the CLEA were compared to the original source King (1994) during the period of overlap in the dataset, from 1896 to 1992. In this period, the CLEA data is not retained in the final product. But,

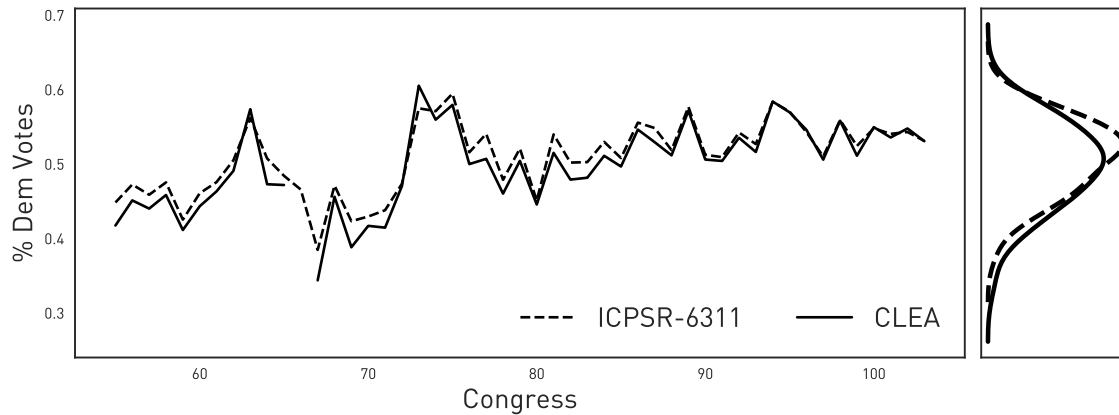


Figure 3. Comparisons of the national average Democratic vote share between King (1994) (ICPSR-6311) and Kollman et al. (2016) (CLEA)

comparisons over this period of overlap illustrates how closely the method constructs a dataset from the CLEA with the same semantics as King (1994). First, the comparison shown in Figure 2 presents the scatterplot of the Democratic party vote shares in Congressional elections constructed using the CLEA and that from King (1994) (ICPSR-6311). The plot for the early period of overlap is shown on the left, and the comparison over the latter period is shown on the right. The early period of overlap contains all Congresses conducted before the Legislative Reorganization Act of 1946, and the latter period contains all full Congresses that take place after the passage of the Act. The correspondence in the two data sets is high, but is much better in the second half of the data than in the first. This is likely both due to the way the Democrat-Farmer-Labor faction was processed and the relative disappearance of minor factional party classifications in the CLEA in the period after 1948. In addition, the prevalence of fusion voting declines in this period, which makes tabulation of the two-party vote much simpler. Thus, this comparison indicates that using the data derived from the CLEA should provide an accurate post-1992 extension of King (1994), since accuracy is better for the contemporary Congresses

However, this plot does clearly show cases where the CLEA and King (1994) are almost perfectly negatively correlated. When isolated, these cases occurred when the two source data sets disagreed about the party identification of the legislators in a contest. Since this is two-

party vote data, a disagreement about party would lead h_i to be $1 - h_i$ in the other dataset over a universe $h_i \in [0, 1]$. Upon further examination, these cases were consistently determined to be errors in the CLEA and were reported. However, since these coding errors are detected only in the pre-1992 portion of the CLEA-derived data, these few coding errors do not propagate into the derived spatiotemporal database.

A second verification step, comparing the share of votes Democrats receive and the share of seats Democrats win in the US Congress between the two data sets was conducted. This comparison is shown in Figure 3. Both sources generate similar estimates in these two cases, and again tend to track better in later Congresses. Notably, however, the two most recent CLEA releases (versions 8 and 9) omit elections in the United States for 1918, which this graph makes clear. Since the derived dataset uses King (1994) for all years before 1992, this missing data does not affect the final data product. Thus, with these two comparisons, it seems the two-party vote data generated from the CLEA is sufficient to extend King (1994) past 1992, and the final spatio-temporal database of US Congressional elections since 1896 is coherent.

3.5 Further Potential Uses

To use this enhanced version of King (1994), the data must first be loaded correctly into an efficient spatial format. The format chosen here is standards-compliant and can be read by any tabular data reader with access to GDAL, the Geographic Data Abstraction Library. In addition, the table can be read directly into various SQL engines (such as PostgreSQL or SQLite), and the well-known binary column converted directly to geometries using appropriate PostGIS or Spatialite functionality. Then, spatial analysis can be conducted using standard statistical packages (Thomas et al., 2004; Rey and Anselin, 2007; Bivand and Piras, 2015). This may include spatial econometric analysis of electoral models (Anselin and Rey, 2014), exploratory local spatial modeling (Calvo and Escolar, 2003), or cluster analysis and voter diffusion detection (Coleman, 2014). In this dissertation, the dataset allows for a novel study of the relationship between measures of partisan advantage and measures of electoral boundary manipulation. I

also am interested in linking the dataset with *-NOMINATE scores Poole and Rosenthal (1987) to investigate the ways boundary change and political ideology may shift together over time.

Chapter 4

EMPIRICAL STRUCTURE OF ELECTORAL SWING

All of the partisan advantage measures considered in this dissertation (starting from Chapter 2 onwards) and nearly all measures of partisan advantage discussed in the literature (e.g. Nagle, 2015) stand upon a model of the relationship between seat shares and vote shares in the electoral system. They either do so explicitly, by specifying and estimating a stochastic model of elections against which fair reference scenarios are evaluated, or implicitly, by constructing a measure of advantage that prizes certain functional relationships between seats and votes over others. Abstract standards of electoral fairness often place constraints on the structure of this relationship as well (Grofman, 1983; Stephanopoulos, 2013). In some studies, such as Tuftes (1973), the model of \bar{h}, \bar{s} is stated, and statistics used to characterize the potential fairness of an electoral system pertain to summaries of the estimated relationship. In others, an implicit standard of justice belies an “ideal” theoretical seats-votes relationship, but a specific parametric form is not provided and the “fair” curve not constructed directly. Regardless, misspecification in the model of elections may provide erroneous estimates of the seats-votes curve, and thus incorrect or unrealistic values for bias and responsiveness estimates from these curves.

Recent work using implicit seats-votes models to diagnose advantage focus on comparing the observed slope and location of an assumed-ideal seats-votes curve (McGhee, 2014; Stephanopoulos and McGhee, 2015) or require a given skewness for the seats-votes relationship (McDonald and Best, 2015; Wang, 2016). Some methods of measuring partisan electoral advantage do consciously attempt to avoid implicating a model of the seats-votes relationship (Brookes, 1960; Johnston et al., 1999; Hill, 2010), but these methods do not provide a clear alternative theory for what they quantify (Altman, 2002). Other work simulates many district plans and compare the outcome of the observed plan to the set of outcomes expected under the simulated plans (Chen, 2013; Cho and Liu, 2016b). Alternative methods also focus on comparing the enacted plan to alternative candidate plans known to policymakers, attempting

to demonstrate specific directional tradeoffs between objectives rather than characterizing the abstract fairness of plans given a standard (Altman et al., 2015). Many analyses that attempt to avoid explicit seats-votes arguments end up invoking an implied seats-votes structure regardless (Kousser, 1996; McGhee, 2014), so identifying a sufficiently-robust seats-votes model will benefit many different types of gerrymandering analyses.

Thus, before proceeding to develop & examine local estimates of partisan advantage generated by one of these modeling strategies, I will spend time considering deeply the structure of seats-votes modeling strategies that are used in subsequent chapters to estimate local impact scores. I develop a new seats-votes modeling strategy based around bootstrap inference. I also aim to characterize the empirical structure of electoral swing. The difference between each years' election results, electoral swing holds a significant place in modeling elections, and models of swing are required to estimate seats votes curves and measures of partisan advantage. Thus, I will examine whether common assumptions about swing are empirically verified and whether common electoral models are spatially misspecified. In addition, I will consider whether corrections to account for spatial misspecification have any effect on the resulting curve.

4.1 Political Advantage as a “Hypothetical” Edge

Historically, work on modeling the relationship between seats and votes focused on reliable and robust estimation of system responsiveness. Namely, in an attempt to validate a “natural law” of democratic societies, the “Cube Rule”(Kendall and Stuart, 1950) motivated many foundational analyses of bias and responsiveness in democratic systems (Tufte, 1973). The movement away from full-system analysis to the current focus on district- or precinct-level models cemented in the early 1990's with a sequence of influential papers, and is accelerating as data availability becomes better.

In tandem with the development of new responsiveness and bias estimation methods, the development of better seats-votes modeling methods surged around the *Davis v. Bandemer*

(1986) case. As discussed in Chapter 2, theoretical and empirical arguments about partisan advantage and boundary manipulation abound in this period. The literature engaging with the case (Grofman, 1985; Niemi and Deegan, 1978), centered primarily on discussions of appropriate measures of boundary manipulation and political advantage and the theory of standards in districting. Focusing on methods to reliably estimate these quantities, Niemi and Fett (1986) critiques “historical” analyses, where a seats-votes relationship is estimated directly from (\bar{h}_t, \bar{s}_t) across many previous elections. These methods tend to provide sensitive estimates that change dramatically from year to year and are highly contingent on the few data points available.

In its place, Niemi and Fett (1986) suggests “hypothetical” analysis, where seats-votes curves are constructed directly from district level information in a single year or pair of years using an explicit model of *electoral swing*, or the model of how changes in party average vote share apply to each district. Absent any other model for how a change in system-wide average vote share is reflected in each district, Niemi and Fett (1986) assume that an increase in a party’s system-wide vote share is well-modeled by a proportional increase in all districts. However, the assumption that changes in party average vote share should apply uniformly to all districts in an electoral model is an assumption with a significant history of debate in electoral geography. The model of *strict uniform swing*, where each district increases *exactly* by the average increase, is still used thoroughly today as a first approximation of electoral dynamics.

Preliminary interest in hypothetical modeling using strict uniform swings is followed by a pair of influential papers, Browning and King (1987) and King (1989). In this, a structural theory of elections based entirely on the “hypotheticals” of Niemi and Fett (1986) is used to justify a system to measure partisan advantage. These hypothetical methods rely entirely on an implicit model partisan swing that is used to shift observed outcomes into desired counterfactual scenarios (Gelman and King, 1994a). However, to provide for more realistic counterfactuals, Gelman and King (1994a) suggest that uncertainty about the electoral process should be partitioned into inherent error and a separate component for uncertainty within the electoral system. Inherent error, they suggest, is present in all attempts to model the electoral system. Electoral

Kind	Mean	Variance	Covariance
Strict	constant	-	-
Generalized	constant	constant	-
Hierarchical	regional	regional	within-region

Table 5. Swing specifications common in political science literature. Note that none assume correlation between observations in different regions.

uncertainty, though, is unique to each redistricting period and allows counterfactual elections to be simulated with less uncertainty: since the outcome for election t is observed, any counterfactual for t should be some combination of the observed outcome \mathbf{h}_t and the predicted $\hat{\mathbf{h}}_t$ given the counterfactual electoral conditions, \mathbf{X}° . Thus, Gelman and King (1994a) suggest a method to conduct counterfactual analysis in a way that shrinks counterfactual \mathbf{h}_t° towards the observed \mathbf{h}_t , meaning the realization-specific uncertainty is removed.¹³

In this model, a *generalized* uniform partisan swing is used to construct counterfactual elections, where shifts in \bar{h} are generated by a random effect with a fixed expectation. Contemporaneously, Jackman (1994) also models seats-votes curves using multiple stochastic components, but treats swing instead as a spatial hierarchical random effect, suggesting that swings for a given constituency result from nested local, state, regional, and national processes. This also results in a *generalized* uniform partisan swing, where any single district experiences a swing correlated with its state and regional context.

Together, these illustrate that the sense of the term “uniform” swing is ambiguous at best. Random effects are not uniform in value, though they may be in expectation. So, they are hardly *strictly* uniform in value, the sense used by Niemi and Fett (1986). A spatial hierarchical effect is not spatially uniform (even in expectation). Yet both authors discuss their methods as a generalization or extension of “uniform” swing. This ambiguity is actually older than the reaction to the models & work flowing from *Bandemer*, and can be divided into three distinct conceptual

¹³This will be detailed further in Section 5.3.1.1.

models for how shifts in aggregate party vote share (\bar{h}) is (or should be modeled as) related to the vector of party vote shares (\mathbf{h}):

Strict Uniform Swing: Empirically, constituency-level swing is so tightly clustered around the aggregate swing that it is effectively constant. As a modeler, this means a single scalar shift, δ , may be used to construct counterfactual elections.

Generalized Spatially Uniform Swing: Constituency-level swing has variability, but this variability has a common distributional structure over space. Thus, swing can be modeled as identically distributed random shocks in each constituency, exogenous to electoral conditions or district context.

Spatially Independent Swing: Constituency-level swing may be spatially heterogeneous, but conditional on the heterogeneity, swing is independent. Thus, swing may be modeled as an exchangeable random effect.

In many ways, these three claims drive a significant amount of discussion about effective ways to model electoral volatility. In general, the *spatially-independent* argument is still quite popular, with some suggesting that it holds so strongly that electoral modeling is a “solved” (Gelman, 2014) problem.

Since the model for electoral swing is implicit in the construction of electoral counterfactuals, the simulation distributions for the measures discussed in Chapter 2 are contingent on these specifications. Thus, misspecification in either the electoral model or the simulation method for counterfactual elections may result in an incorrect seats-votes curve and invalid measures of advantage or responsiveness. It is necessary to clarify and examine the structure of electoral swing in modeling seats-votes curves before developing the local measure of partisan impact that rely on these measures.

4.2 Disagreement About “Uniformity” in Electoral Swing

The longevity of argument between these perspectives on what uniformity means still reflects current divides; a clear lack of consensus about the spatial structure of electoral swing has been manifest throughout the 20th century. Early use of uniform swing arguments include Brookes (1960)’s method to decompose electoral advantage using strict uniform swing or the model of electoral volatility provided by Hawkes (1969). However, Rasmussen (1964) opposes arguments using strict uniform swing in British multiparty electoral politics, suggesting that it oversimplifies electoral dynamics and presents a misleading picture of how elections are won and lost.

In contrast to the strict uniform strategies, early foundational work from Stokes (1965) presages Jackman (1994), modeling electoral swing with local- and state-level components. Thus, “uniform” for Stokes (1965) is in fact a generalized swing: a shift in support for one party manifests as a shift in *all places* towards that party, but the shifts are essentially random. Going further, (Katz, 1973b,a) provides a revision of Stokes (1965) allowing for further spatial heterogeneity. Levels of the hierarchy may vary in their response to a given swing. But, given the swing in a district’s state, region, or nation, the distribution is linked together by a common hierarchy, and swings are independent conditional on this hierarchy (Wilson, 1978). While early formal treatment of explicit models for spatial dependence existed at this time (Whittle, 1954), dependence between districts was not a foundational concern like nested multilevel structures were.

In another context, arguments about the meaning of “uniformity” surface in a robust debate about the usefulness of swing in analyzing Australian elections. These elections pose a a distinct set of challenges to the estimation of responsiveness and advantage, since instant-runoff voting in a multi-party system means that swing is not zero-sum (Mackerras, 1973). In this context, Mackerras (1976) suggests that changes in a constructed two-party vote share for dominant coalition parties in Australia in 1975 tended to be strictly uniform, arguing that each sub-national constituency experienced nearly the same swing as the national average

swing. Sharp critiques of this view echo Rasmussen (1964), and focus on the fact that a uniform swing in a multi-party democracy is an unnecessarily strong simplification (Sharman, 1978; Mackerras, 1978). One relevant critique, that of Austen (1978), focuses instead on how Mackerras (1976) is inconsistent: if swing were strictly uniform, then the significant fraction of text Mackerras (1976) uses to analyze how some areas do not swing like the nation would be moot.

Their discussion of just how “uniform” is uniform enough, generated no conclusive answers. This tension between “strict” and “generalized” senses of uniform swing has also re-appeared since then in different venues, such as the exchange initiated by Butler and Van Beek (1990). They argue that swing should be used in the analysis of American elections, citing its use in Australian and British analyses. Rose (1991) disagrees, again suggesting that swing is an unnecessary simplification and is generally unhelpful because volatility is rarely constant over space. Countering, Gibson (1992) argues that swing is no worse than Rose (1991)’s own favored electoral simplifications. More recently, these arguments resurface in discussions of uniform swing as a “first approximation” for presidential electoral models (Jackman, 2014; Ghitzza and Gelman, 2013). In addition, recent work discussing relative swings by social group in an ecological analysis of voting behavior in the Weimar Republic vote for the NDSAP (Nazi party) King et al. (2008) reiterate a view of generalized uniform swing as essentially correct.

Foundational work by quantitative electoral geographers is also critical of the empirical basis for strict uniform swing (Johnston, 1982, 1983). Two important monographs, Johnston (1979) and Gudgin and Taylor (1979), make explicit theories about the structure of electoral politics, swing, and partisan efficiency. While Gudgin and Taylor (1979) suggest uniform swing may be a helpful analytical technique to reason about the distribution of district vote shares, they do not suggest that swing is strictly uniform over space. This avoidance of strict uniformity is all the more interesting due to the contemporaneous work detailing how the spatial structure of macro-political alignments is relatively consistent over time (Archer and Taylor, 1981; Bensel, 1987).

Johnston (1983) also provides a novel theoretical argument illustrating the tension between

two types of strict uniformity. Strict uniform swing might be understood in two ways: all places shift the same *absolute* percentage points, or they shift the same *relative* fraction of support from their current level. A *relative* swing of +5% would move a constituency where Republicans get 51% of the vote by 2.55 percentage points, but would move the district where they get 42% by only 2.21 percentage points. Thus, a strict relative swing must result in a *generalized* swing, depending on the distribution of district vote shares. Further, a strict absolute swing, where all districts rise or fall by the same number of percentage points, results in a distribution of relative swings. Districts usually have unequal turnout, so this distinction affects the party's system-wide popular vote, and thus makes the relationship between the party's system-wide vote share and the district vote share less certain. More generally, Johnston (1983) argues that the distribution of vote shares and the distribution of volatility are necessarily linked, and constant shifts are empirically unlikely when districts are different, too.

These discussions and ideas from electoral geographers only weakly percolate into the later discussions of electoral swing in political science (Johnston, 2005). Although King et al. (2008) cites O'Loughlin et al. (1994) and states that "spatial research" analyzing electoral swing in the Nazi vote is "informative" (p. 971), they then suggest in a footnote:

The idea that **partisan swing is approximately uniform across geographic units** dates to Butler [(1951)] . . . [and] has been generalized to a stochastic model that fits electoral data in Gelman and King [(1994)]. For an example of the notion that citizen support for political candidates shifts uniformly across most social groups in the same direction and extent as the national swing, see Gelman and King [(1993)]. (*emphasis added*)

Thus, again, the sense of "uniform" is confused. The author suggests that prior work has demonstrated swing is spatially uniform, so neither heterogeneous or dependent. However, consulting Gelman and King (1993), where discussion focuses on swing in support among social groups, they suggest that electoral swing is strongly *geographically correlated* and commonly distributed, so that "dependence among states" (p. 416) must be modeled within years. While states might shift together uniformly in mean in presidential elections, the swings are not correlated when conditioning on their group nested structure (Bernardo, 1996).

For the purposes of seats-votes modeling, Gelman and King (1994a) hardly engages with

spatial structure for the electoral swing term. A preceding paper, Gelman and King (1990), suggests instead that:

We assume, therefore, that vote swings about the statewide mean are spatially independent across districts . . . Modeling districts with additional information such as spatial correlation or covariates, if they were available, would probably yield more accurate estimates of the seats-votes curve. Omitting this unavailable information is unlikely to systematically bias our results.

Later, Thomas et al. (2013) also explicitly suggests “generalized” swing is sufficient, claiming that swings tend to be independent conditional on state or regional groups.

A thread pulls consistently in objections to the spatially “uniform” swing asserted by King et al. (2008). Claims that swing is distinctly spatially heterogeneous have been present ever since (Katz, 1973b,a)’s critique of Stokes (1965). The line of reasoning remains in Austen (1978)’s response to Mackerras (1976) and surfaces again with Rose (1991)’s critique of swing in general as being unhelpful because it varies spatially. However, systematic spatial analysis of electoral swing is not common in political science outside of hierarchical arguments (Katz, 1973b; Gelman and King, 1993).

Thus, an analysis of the empirical structure of electoral swing will prove rewarding in both its own right. Targeting the structure of swing in seats-votes models may also improve the generation of realistic counterfactuals and improve the validity of those approaches. A “generalized” uniform swing with no heterogeneity or dependence may only match the *distribution* of swing in a given year and appear entirely unlike a map of swing that has ever been observed. Hypothetical election maps would then be somewhat unrealistic, in that a heterogeneous white noise is applied to all districts when, in fact, neighboring districts might tend to shift together. Theoretically, if nearby districts swing together, then there are significant implications for questions of polarization and sorting in redistricting (Carson et al., 2003; Bishop, 2009; McCarty et al., 2009; Johnston et al., 2016).

Thus, in what follows, an empirical examination of the spatial structure of electoral swing will be conducted. First, recent presidential election swings at the county level will reveal spatially-

correlated electoral swings, even when controlling for heterogeneity. Then, the empirical structure of swing at the congressional level since 1992 will be analyzed.

4.3 Analyzing Spatial Dependence in Electoral Swing Under Known Spatial Heterogeneity

Voting in American elections has distinct structure at various levels of spatial hierarchies (Archer and Taylor, 1981; Bense, 1987; Archer, 1988; Gimpel and Schuknecht, 2009). While significant attention in American electoral geography has focused on identifying the spatial structure of the “normal vote,” the expected behavior of places in regions or in political eras (Converse, 1966; Miller, 1979), analysis of electoral volatility directly is less common. When done, the work focuses on *sectional* realignment (i.e. macro-scale heterogeneity), on the long-wave processes of political realignment and coalition building. Instead, Johnston (1983) suggests the study of volatility not in terms of strategic coalitions and realignment but in terms of direct analysis of electoral swing, which is not as common in the analysis of American electoral geography. Since models of electoral swing are critical to the generation of the plausible counterfactuals required by partisan advantage measures, the lack of direct spatial analysis of electoral swing is somewhat surprising. Below, I will discuss the structure of the 2016 election in terms of the spatial structure of electoral swing at county level. Then, I will focus explicitly on swing in legislative elections since 1992. Doing this analysis at two scales, I aim to identify that swing in congressional elections is substantively different than swing in the county level, and may require much more subtle treatment for a few reasons that will be identified after the discussion of empirical results.

4.3.1 Macro-geographical Structure of 2016 Presidential Swing

At the county level, contemporary presidential voting patterns tend to be fairly stable. In this vein, the three county-level maps of presidential elections since 2008 shown in Figure 4 demonstrate a trend to red in some marginally-aligned areas of the Midwest and Ohio River

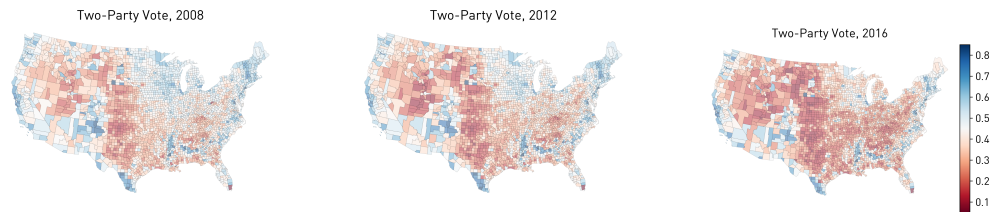


Figure 4. Democratic share of the Two-Party Vote in presidential elections since 2008 at the county level

Valley states, but the overall pattern of strongly-aligned areas does not change significantly.¹⁴ A reasonable visual interpretation of these spatial trends is that everything appears to get more red, moving from the 2008 to the 2016 contest. While 2008 and 2016 are memorable electoral contests, neither fit the pattern of the “watershed” election, which would present strong spatial realignments in partisan coalitions.¹⁵

However, examining the maps of *swing* in the presidential elections since 2008 (shown in Figure 5, the two maps appear quite different. Notably, these maps show that change in the electorate from election to election is strongly patterned, but was inconsistent between the 08-12 swing and the 12-16 swing. Some areas, such as Wisconsin and Michigan, appear to have moved steadily red over the past two elections; in contrast, the Minnesota break to the Republicans is a 2016 innovation. Regardless, the swing is strongly spatially patterned, with both 2016 and 2012 exhibiting similar levels of spatial dependence. The Moran’s *I* for electoral swing in county-level presidential vote in 2012 is .66 ($p < .0001$) and 2016 is .61 ($p < .0001$). Clearly, swing is not spatially uniform; some places move towards Democrats and some away from Democrats, and the spatial structure of this deviation is likely not random. In addition, the distribution of swings at the county level shown in Figure 6 appears to be well-behaved and accord with the generalized uniform partisan swing arguments. The distribution is unimodal and nearly symmetrical in both years, hanging close to the slightly-negative national median

¹⁴Robert Watrel & J. Clark Archer’s recent work on spatial and temporal principal components of normal voting in the US Presidential is further illustrative here, but remains unpublished.

¹⁵Again here, citing Watrel & Clark would be illustrative.

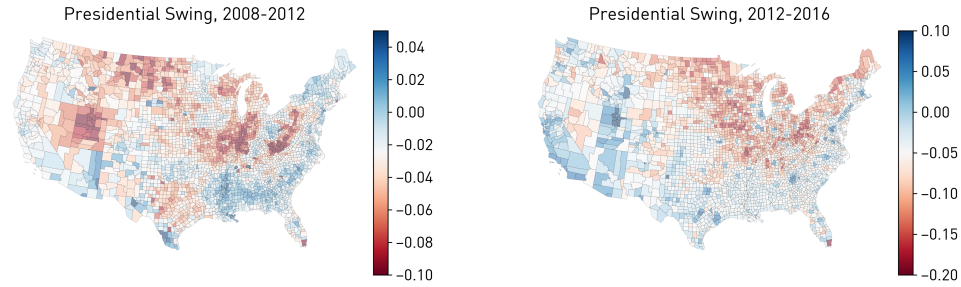


Figure 5. Swing in Democrat two-party vote share from 2008 to 2012 & from 2012 to 2016. Here, negative values indicate states that became much more strongly republican, and positive values indicate counties that became more Democrat, measured in percentage points of popular vote for president. The swing in the 2016 election was around twice as dispersed as the swing for 2012, so the color-bars are nearly double in range between the two maps. In addition, the “midpoint” of the color ramp is the median national swing; it is slightly negative in both elections.

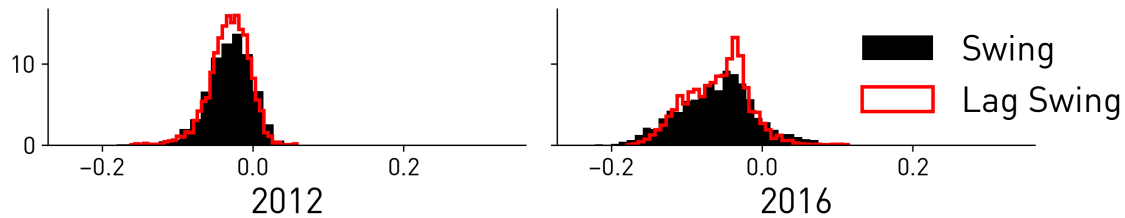


Figure 6. Distributions of county presidential swing and its spatial lag using Queen contiguity weights from 2008 to 2012 and from 2012 to 2016.

swing. In addition to the distribution of swing being well-behaved, the distribution of the *spatial lag* of swing is also well-distributed, with almost identical apparent structure to the distribution of swing itself.

However, determining whether or not this positive dependence value is robust to spatial heterogeneity is somewhat complicated. While spatial dependence is distinct from spatial heterogeneity (Anselin and Arribas-Bel, 2013), the extent to which exploratory statistics like Moran’s *I* are robust to heterogeneity can be examined. One thing that is clear in the maps of swing is strong state and region heterogeneity. For example, Utah becomes significantly more Republican during the 2012 election. Mitt Romney, a candidate with significant connection to the state, attained an usually-strong level of support there, which reverted in the next election. To

examine the extent to which the Moran statistic is robust to exogenously-identifiable spatial heterogeneity, I develop two methods. One, a “hierarchical” Moran technique, re-specifies the Moran regression as a hierarchical mixed-effect model. The second involves a conceptually-simpler de-meaning of the response vector by group means before conducting analysis.

To define the hierarchical model, first consider the typical Moran-form regression:

$$\mathbf{W}Y = \alpha + \rho Y + \epsilon \quad (4.1)$$

where α is the conditional mean, \mathbf{W} is a row-standardized spatial weighting matrix that records the $N \times N$ spatial relationships between observations, Y is the $N \times 1$ vector of the variate under study, and ϵ is an independent and identically-distributed Gaussian error term. I call this specification a “Moran-form” regression to distinguish from the so-called “spatial lag”-form of mixed regressive, spatial autoregressive model considered by Cliff and Ord (1973); Anselin (1988) which is more standard in multivariate spatial regression work. The mixed-regressive, spatial autoregressive specification is more well-used because it is well-defined when $\rho = 0$, whereas Eq. 4.1 is often undefined when $\rho = 0$ in reduced form:

$$\mathbf{W}Y = \alpha + \rho Y + \epsilon \quad (4.2)$$

$$\mathbf{W}Y - \rho Y = \alpha + \epsilon \quad (4.3)$$

$$(\mathbf{W} - \rho I)Y = \alpha + \epsilon \quad (4.4)$$

$$Y = (\mathbf{W} - \rho I)^{-1}(\alpha + \epsilon) \quad (4.5)$$

When $\rho = 0$, this leaves \mathbf{W}^{-1} alone. Many common specifications of \mathbf{W} are singular, and thus the specification becomes undefined. Regardless, I proceed from the Moran-form regression as the underlying specification of Moran’s I , one exceedingly-common diagnostic statistic for spatial autocorrelation in univariate data (Anselin, 1996).

Thus, a hierarchical Moran statistic involves the same type of specification as in Eq. 4.1, but instead of fitting a single global intercept α , the intercepts are modeled hierarchically. Thus, with data Y , the same \mathbf{W} mapping the $N \times N$ spatial proximity relationships between observations, and a new $N \times J$ matrix, Δ , relating each of the N observations to their group $j = 1, 2, \dots, J$.

This lets us express the model as a function of both the global mean μ and group-mean effects γ :

$$\begin{aligned}\mathbf{W}Y &= \Delta\alpha + \rho Y + \epsilon \\ \alpha &= 1_J\mu + \gamma + u\end{aligned}\tag{4.6}$$

where 1_J is a J -length vector of ones. In this model, u is the group-wise variance component, ϵ is the unit-specific error term, and ρ corresponds to a Moran's I -style dependence coefficient *conditional* on the hierarchical model for α , containing global effects μ and region-specific effects, γ . This results in the following single-equation Moran-form regression in reduced form:

$$Y = (\mathbf{W} - \rho I)^{-1}(1_N\mu + \Delta\gamma + \Delta u + \epsilon)\tag{4.7}$$

However, due to the introduction of the joint random effect $\Delta u + \epsilon$, custom estimators are required to fit this model (Wolf, 2016). For a simplification that can be implemented without specialized estimators, the group-level variance component can be removed. This results in the following model, a spatial fixed-effect Moran model:

$$\mathbf{W}Y = \Delta\gamma + \rho Y + \epsilon\tag{4.8}$$

where γ is still the J -length vector of conditional group means. Another extension may privilege variation due to heterogeneity over variation in dependence by constructing the group-means unconditionally. This “group de-measured” functions like a standard Moran specification where the empirical (unconditional) group means are subtracted off of Y before estimation:

$$\mathbf{W}(Y - \Delta\hat{\gamma}) = \mu + \rho(Y - \Delta\hat{\gamma}) + \epsilon\tag{4.9}$$

where $\hat{\gamma}$ is the naive group mean of Y grouped by J groups. This provides a measure of the relationship between the N lower-level units after having removed the potential effects of exogenous spatial heterogeneity in $\Delta\hat{\gamma}$. While this model is not as rich as that in Eq. 4.6, this initial group de-meaning is incredibly simple to implement, can be estimated without concern for the dual error terms, and has a “direct” empirical interpretation as removing the spatial heterogeneity expressed in $\Delta\hat{\gamma}$. This method allows for the analyst to control for the presence of

potential known heterogeneity while still exploring the structure of dependence in a dataset. By examining the group de-meaned data directly, the structure of dependence *after* heterogeneity can be examined after filtering out the potentially different means.¹⁶

4.3.2 Exploring a Decomposition of Spatial Heterogeneity & Dependence in County-level Presidential Vote

First, simply examining the following Moran-form regression using OLS provides a comparable value of ρ to the univariate Moran's I , conditioning on heterogeneity across states:

$$\mathbf{W}(\mathbf{h}_{2016} - \mathbf{h}_{2012}) = \mathbf{W}\delta_{2016} = \delta_{2016} + \Delta_S\gamma + \epsilon \quad (4.10)$$

where Δ_S is the grouping matrix for counties-in-states, meaning γ models the state-specific conditional means. In addition, δ_t is the swing vector from the previous election to the t th election. The parameter estimates from this regression are contained in Table 4.3.2. Using this model, we obtain an autoregressive coefficient of .3817, which is about half the Moran coefficient estimated without the fixed effects, .6171. Twenty-seven statistically-significant state fixed effects surface in this model, and these are shown graphically in Figure 7.

So, conditional on the heterogeneity at the state level, the county-level spatial dependence (as measured by the ρ effect in regression 4.10) is still around half of the Moran's I for the 2016 swing without attempts to control for spatial heterogeneity. But, what if this heterogeneity is diagnosed at the wrong scale? Referring again to figure 5, some larger-than-state regions appear to swing together frequently. This might be addressed using endogenous cluster detection in the future (Duque et al., 2012), but here I suggest a repeated application of the group de-meaning strategy from Equation 4.9 using states, census divisions, or census regions. This

¹⁶This style of dependence-after-control strategy is similar to that advocated in the remarkable study conducted by Hodges and Reich (2010), which suggests partitioning the sums of squares in spatial error specifications to avoid “clobbering” your fixed effects with correlated errors.

	Coefficient	Std. Err.	<i>p</i> -value		Coefficient	Std. Err.	<i>p</i> -value
Swing	0.3817	0.0091	0.0000**				
AL	-0.0262	0.0025	0.0000**	AZ	0.0216	0.0059	0.0003**
AR	-0.0033	0.0034	0.3343	CA	0.0279	0.0037	0.0000**
CO	0.0029	0.0035	0.4058	CT	0.0022	0.0076	0.7713
DE	-0.0010	0.0119	0.9308	FL	0.0006	0.0035	0.8603
GA	0.0076	0.0029	0.0096**	ID	0.0088	0.0039	0.0254*
IL	-0.0269	0.0032	0.0000**	IN	-0.0270	0.0033	0.0000**
IA	-0.0480	0.0033	0.0000**	KS	0.0003	0.0032	0.9358
KY	-0.0233	0.0031	0.0000**	LA	0.0070	0.0035	0.0471*
ME	-0.0248	0.0056	0.0000**	MD	0.0140	0.0048	0.0037**
MA	0.0231	0.0063	0.0003**	MI	-0.0288	0.0033	0.0000**
MN	-0.0404	0.0033	0.0000**	MS	0.0042	0.0033	0.2038
MO	-0.0345	0.0031	0.0000**	MT	-0.0122	0.0037	0.0009**
NE	-0.0175	0.0032	0.0000**	NV	0.0050	0.0055	0.3596
NH	0.0011	0.0068	0.8673	NJ	0.0143	0.0051	0.0047**
NM	0.0022	0.0043	0.6106	NY	-0.0219	0.0036	0.0000**
NC	0.0029	0.0032	0.3550	ND	-0.0410	0.0038	0.0000**
OH	-0.0371	0.0033	0.0000**	OK	-0.0084	0.0034	0.0131**
OR	0.0023	0.0042	0.5793	PA	-0.0207	0.0035	0.0000**
RI	-0.0178	0.0094	0.0574*	SC	0.0046	0.0039	0.2327
SD	-0.0314	0.0035	0.0000**	TN	-0.0180	0.0032	0.0000**
TX	0.0115	0.0028	0.0000**	UT	0.0396	0.0046	0.0000**
VT	-0.0105	0.0059	0.0782*	VA	-0.0013	0.0031	0.6753
WA	0.0067	0.0041	0.1008	WV	-0.0353	0.0037	0.0000**
WI	-0.0270	0.0035	0.0000**	WY	0.0020	0.0049	0.6763

Table 6. Regression results for a spatial fixed effects Moran model detailed in Eq. 4.6. The model achieved an adjusted R^2 of .75 and a significant F statistic. Two asterisks indicates the p -value is less than .01, and one asterisk indicates $p < .1$.

might also be implemented using a hierarchical Moran regression specification that nests these levels (like Stokes (1965)) to identify the empirical structure of this variation through an exploratory model.¹⁷

¹⁷ In theory, since the group de-meaning occurs before applying the Moran statistic, local indicators could be naively constructed for the de-meaned sequences simply by treating them as input data (Anselin, 1995). However, it is both unclear whether this is statistically appropriate, given the multiple layers of distributional assumptions implicit in repeated demeaning, and if the identified clusters have any useful interpretable meaning without conducting further basic research into the use of hierarchical Moran techniques. Further, with the nested hierarchical implementation, distributional assumptions are made about the structure of the model which likely push the technique out of the exploratory spatial data analysis context supplied by common deployments of the Moran & Local Moran

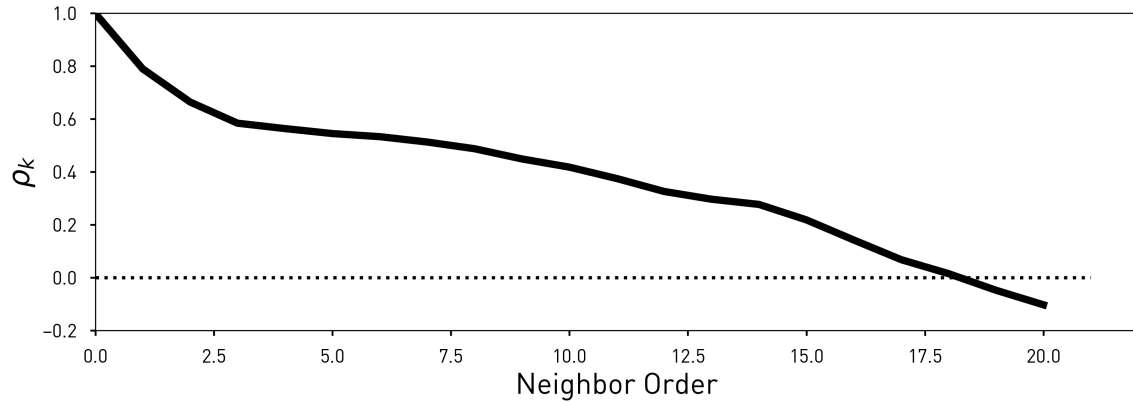


Figure 9. The full spatial autocorrelation function, truncated at the first k th order at which the correlation becomes negligible. This is computed on the 2016 presidential swing at county level.

The successively de-measured Moran scatterplots are shown in Figure 8. In this case, each plot subtracts its groupwise mean from its constituent observations and treats the data as if it were organic input data to a typical Moran statistic. All of the p -values for these repeated analyses are statistically significant. Thus, the x and y axes plot the de-measured swing and lagged de-measured swing. The line of best fit is plotted through each point cloud, whose slope represents the Moran coefficient for that regression. Thus, as you increase the scale of controls for the spatial heterogeneity, the dependence measure declines in this dataset. Conceptually, this means that Moran statistics may not be robust to spatial heterogeneity, especially when that heterogeneity is strong.

Another method of examining this similar facet is to compute the (partial) spatial autocorrelation function. Akin to the temporal (p)ACF, the spatial p(ACF) measures the (partial) dependence between successive orders of spatial lag in a lattice dataset. Here, the full spatial ACF is most illustrative, since it characterizes the speed of the decay in correlation between counties as the order of neighbors increases. In the spatial context, the k th *exact* order neighbors of observation y_i is the set of observations y_j that are first reached in k steps. This means that statistics. Thus, while I have implemented “local” hierarchical Moran statistics for this analysis, I do not present their results on this data in this dissertation.

the graph distance between observation y_j and y_i is exactly k :

$$\{y_{ik} : \min(\|y_j - y_i\|) = k \quad \forall j = 1, 2, \dots, n\}$$

Thus, the k th order spatial autocorrelation function is:

$$\rho_k = \text{cor}(y, \mathbf{W}^k y) \quad (4.11)$$

where \mathbf{W}^k is the adjacency matrix for k -minimal neighbors. The k th-order partial spatial autocorrelation function is:

$$\hat{\rho}_k = \text{cor}(y, \mathbf{W}^k y \mid \mathbf{W}^{k-1} y, \mathbf{W}^{k-2} y, \dots, \mathbf{W}^1 y) \quad (4.12)$$

This structure, focusing on the counties *first* reached in k steps, prevents the repeated inclusion of low-order neighbors in high-order lag evaluations. As shown in Figure 9, the correlation between adjacent counties declines as a smooth function of proximity. At around the 20th-order neighbors, observations begin to become uncorrelated. Conceptually, this means that dependence between counties likely operates at a scale somewhere between the state and the Census division, since Census divisions are populated by around 200 to around 500 counties, whereas states typically have well under 100 counties. In contrast, the partial spatial autocorrelation function, which conditions on each subsequent set of k th order neighbors as discussed in Eq. 4.12, suggests that correlation goes to zero almost immediately. Once the first-order neighbors are conditioned on, the second-order neighbors provide nearly no information about the source observation.

4.3.3 Spatial Dependence & Heterogeneity in Congressional Swing

When conducting a similar analysis for the swing in Congressional election returns in recent elections, it becomes necessary to handle uncontested elections in some way. This concern of how to address uncontested elections has long affected seats-votes modeling work. Since many elections are uncontested in Congress, the “swing” from year to year in legislative vote

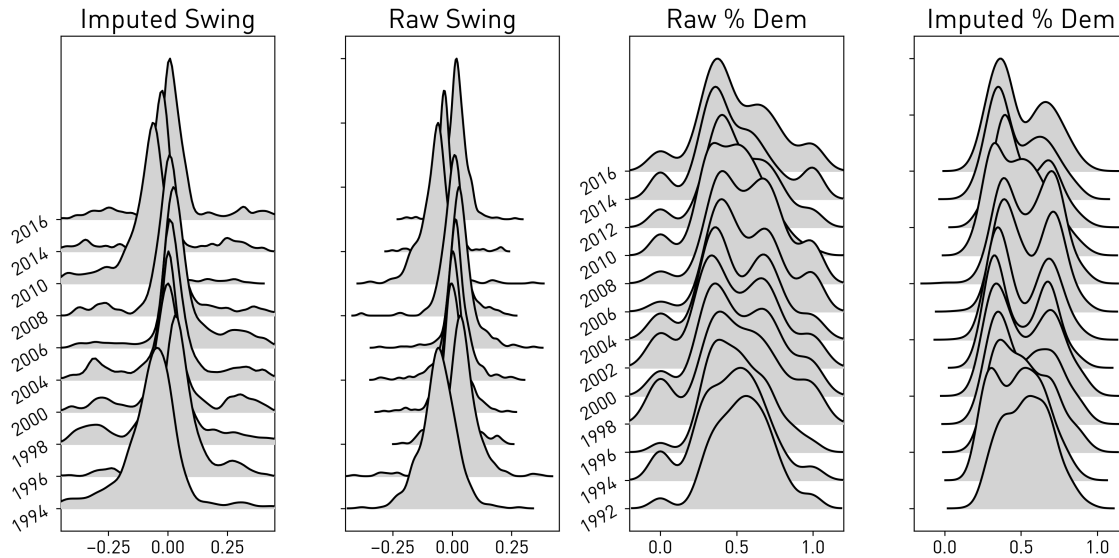


Figure 10. Distributions of congressional-level Democratic vote share and swing each year. For swing, matching of congressional districts is done when no redistricting occurs, so only years with known successor-predecessor pairs are shown. This means swing is unavailable for all years ending in 2, immediately following redistricting.

shares may be much more volatile than for county-level presidential returns. In the case where an election is uncontested one year and then contested the next, the swing might be on the order of ± 0.4 . This does communicate some information about the change in vote shares, but it may provide more noise than valuable information. In addition, the move from uncontested to weakly-contested is more common than the move from uncontested to strongly-challenged, so most swings for uncontested districts that are contested in the next period are smaller than the worst case. Finally, the distribution of swings without imputing the uncontested elections is more similar to than without the correction than that for the vote shares directly.

For example, consider the distribution of swings and vote shares on the left side and right side of Figure 10, respectively. Regardless of whether swing or vote shares are being used, the imputation removes the extreme secondary peaks of the distributions. These secondary peaks surface due to uncontested districts becoming contested in the subsequent election. Regardless, though, the magnitude of these peaks is much smaller in the swing distributions

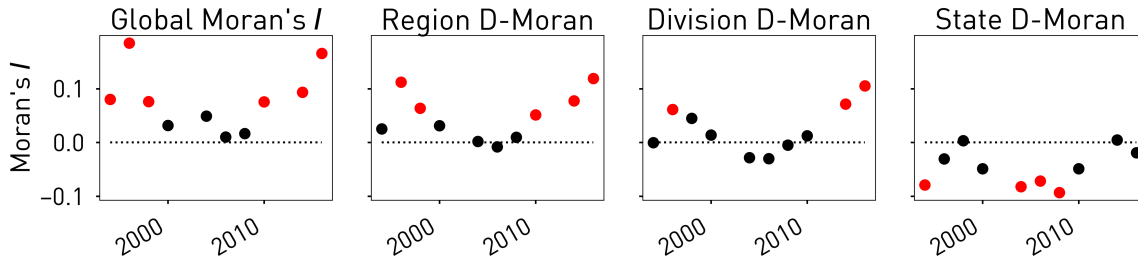


Figure 11. Global and group-wise de-meanded Moran statistics for Congressional elections since 1992. Points in red are “statistically significant” with no multiple comparison correction. The magnitude of the statistic is provided on the y -axis. Uncontested elections have been imputed using an autoregressive strategy detailed in Section 6.2.

than in the vote share distributions. Examining the structure of the imputed distributions results in a more well-conditioned analysis. In addition, the spatial dependence in the swing distribution increases when uncontested districts are dropped. This is in part due to the fact that dropping uncontested districts affects both the connectivity structure *and* distribution of values. Thus, it seems most appropriate to analyze the distributions after imputing the uncontested election vote shares, rather than only the contested elections.¹⁸

Given that we are using the autoregressive imputation strategy, the results from the year-by-year Moran statistic plot is shown in Figure 11. Without any multiple comparison correction, Moran statistics with pseudo- p -values under .05 are colored red. With a Bonferroni correction, only the second and last Regional D -Moran statistic are significant; only the second Division D -Moran, and the D -Moran for 2008 are significant under a Bonferroni correction for the Division and State versions. Regardless, the spatial patterning in Congressional vote shares after accounting for uncontested districts is inconsistent over time; swing is not *always* significantly correlated, regardless of whether a correction for exogenous spatial heterogeneity along census region, division, or state lines is used.

In addition, the fact that the I statistic becomes *negative* after the state-level heterogeneity correction is remarkable. If, as some suggest (see Haining (2003, pg. 87), also Griffith and

¹⁸For more on the actual imputation model, refer to Section 6.2.

Arbia (2010)), negative spatial autocorrelation arises from competitive processes in ecology or human geography, then an interesting theory might be explored given the shift towards negative spatial dependence in the congressional swing after accounting for spatial heterogeneity at the state level. Noting that districts are drawn to be “safe” more often than not,¹⁹ swings tend to act on these safe districts in opposing directions, conditional on the statewide swing. Thus, since safe Democrat and safe Republican districts likely share borders, the negative spatial dependence after accounting for statewide swing would imply that neighboring congressional races tend to “tighten” or “loosen” together, rather than shift linearly together. Since the autocorrelation is negative, a “tightening” would involve safe Democrat districts swinging down and safe Republican districts swinging up, with respect to the Democrat vote share in the district. A “loosening” would involve safe Democrat districts swinging towards a Democrat vote share of 1 and safe Republican districts swinging towards 0. While this exploratory analysis indicates there may be negative dependence in the movement of Congressional vote shares after accounting for spatial heterogeneity at the state level, this result should not be over-interpreted; the p -values on these statistics are marginal (at most two in each series are significant after a Bonferroni correction) and the implied model is likely underspecified.²⁰

4.4 Conclusion

Regardless, electoral swing is clearly not strictly uniform. Indeed, electoral swing does not behave empirically like any of the common models of swing suggested above. Despite controlling for differences in the mean swing over states or regions, swing still exhibits significant spatial dependence in presidential elections. Thus, it seems that a single theory for how electoral swing is best specified will likely *not* hold across time, elections, or geographies. Indeed,

¹⁹For example, see the analysis of interviews in Chapter 9.

²⁰ To truly *demonstrate* whether this holds, that congressional races tend to either tighten or loosen together rather than drift together linearly after accounting for spatial heterogeneity would require a follow-up study that attempts to control for incumbency and midterm structures, likely through some form of spatial panel model with state fixed effects.

claiming that electoral swing can be modeled sufficiently as a uniform random (or hierarchical random) effect should be justified by empirical illustration. In this instance, electoral swing in presidential vote by county *is* strongly spatially-dependent even when accounting for potential spatial heterogeneity. Thus, modeling it as a uniform random effect is inapt. However, spatial dependence in electoral swing is much weaker in congressional geographies.

After this thorough exploration of introducing spatial effects into seats-votes counterfactual generating processes, a few takeaway points are clear. First, electoral swing (like raw vote shares) are strongly correlated over space, in many elections, and at many scales. In presidential election returns, electoral swing is quite strong, positive, and is robust to corrections for spatial heterogeneity at the state, Census division, or Census region level. As far as the “correct” scale of spatial heterogeneity present in the macrogeography of presidential vote, it seems that swing is best modeled as heterogeneous somewhere between the state- and division level. This was detected using two novel analyses; one, a family of hierarchical/group de-measured Moran statistics were specified. Two, a spatial autocorrelation function akin to the temporal autocorrelation function was developed for discontinuous lattice data and estimated. This provided an estimate of scale in county-level presidential vote that suggested that counties are unrelated after moving around 20 counties out. Exploratory regionalization or scale-discover regression techniques (such as GWR or multi-scale GWR) might corroborate this analysis. In congressional elections, the spatial dependence is much weaker and is less-consistent over time as well. Indeed, after a correction for spatial heterogeneity, swing becomes *negatively* spatially correlated, although this correlation is again marginal.

Thus, when the models discussed in Chapter 5 are examined for potential spatial misspecification in chapter 6, I anticipate there being misspecification. However, since the strength of dependence in the congressional models is so weak, I also anticipate the *correction* for this misspecification (either in vote share models directly or in the counterfactual generating model) having little-to-no effect on the resulting advantage measures or seats-votes curves. That is, the models will likely be misspecified in the sense that there exists a spatial pattern to swing that is not modeled sufficiently by treating swing as a uniform random effect. But, using an empirically-

realistic model of electoral swing might only induce *slight* patterning, since the dependence is small.

MODELING SEATS & VOTES: SPECIFICATION AND COMPARISON

Given that electoral swing is empirically nonuniform, it is important to explore whether more realistic models of partisan swing may affect estimates seats-votes curves. That is, does it make a difference to the shape of seats-votes curves if vote models or swing structures exhibit spatial dependence? Even ignoring Johnston (1983)'s argument from first principles, positive spatial correlation in swing might indicate that, in general, nearby constituencies swing in a similar manner, and that volatility in electoral outcomes may cluster. But, without perfect spatial correlation, there *must* be fluctuation in the structure of the swing over the electoral map. A model that assumes swings are independent between observations when electoral conditions are held constant (c.f. Gelman and King, 1994a; Jackman, 1994; Linzer, 2012; McGann et al., 2016) would provide for generalized spatially-independent swings.

Thus, in the following discussion, I review seats-votes modeling techniques and make their models of electoral swing explicit. First, the basic theoretical model is presented. Originally justified as a model under strict uniform swing (Jackman, 1994), the empirical seats-votes curve estimate is shown to be a translation of the rank distribution of vote shares. This will allow for a thorough explanation of the way the seats-votes curve represents an electoral system, and illustrate one critical step of all more complicated seats-votes curve estimation strategies. After this, two stochastic methods to generate seats-votes curves will be discussed, and a novel bootstrapping method will be developed.

The estimated seats-votes curves from the three approaches are compared, and the simulations assessed along multiple dimensions of quality. Then, one of these methods is selected for an intensive study of potential spatial misspecification. In addition to straightforward testing of model adequacy, I also conduct extensive simulations under four distinct spatially-explicit data generating processes to examine how this misspecification may result in different estimates than the original specification in Chapter 6. If there is no substantial difference when explicitly-

spatial models are used— if correlated swings tend to result in similar seats-votes models than uncorrelated swings— then the introduction of complex models and estimators that account for spatial dependence may not be worth the additional effort.

5.1 Development of Seats-Votes Modeling Frameworks

In general, seats-votes models, are composed of two distinct models; one of legislative vote shares and one of electoral swing, the change in vote shares from year to year examined in Chapter 4. The model of electoral swing is used to generate “counterfactual” or “hypothetical” elections, which then in turn are analyzed to identify partisan advantage. The analysis of “hypothetical” elections (suggested by Niemi and Fett (1986)) has a few advantages over the straightforward empirical analysis of observed elections. Constructing a seats-votes curve as the functional relationship between the observed system-wide average party vote, \bar{h} , and the share of the delegation or legislature that the party wins, \bar{s} , ignores essentially all information about the district-level dispersion, correlation, and electoral conditions that give rise to \bar{s} and \bar{h} . For a given districting plan in the US, there might be five observations of (\bar{h}_t, \bar{s}_t) under the typical decadal redistricting and reapportionment regime in the US. This is simply not enough information to reliably estimate a relationship if only these five observations are used.

More information *is* available in each election, however: the district-level vote share vector, h_t . This is useful since it implicitly contains information about how probable values of \bar{s}_t may be under minor changes in electoral conditions or outcomes. By treating \bar{s}_t as the *only* useful realization of the response in election t , a potentially rich source of information about the seats-votes relationship is ignored.

Thus, district-level analysis models the seats-votes relationship using *plausible* hypothetical district vote share vectors, h_t^o that attain known hypothetical average vote shares, \bar{h}_t^o but occur under changed electoral conditions or small disturbances. These simulation ensembles constitute the “zone of chance” surrounding an observed (\bar{h}, \bar{s}) (Wang, 2016). Elections in this “zone” are just as useful as the observed election in computing summaries of advantage, efficiency,

or responsiveness. Thus, for district-level seats-votes analysis techniques, *two* models are required: a distributional model for h and a *generative* model of h° . While some methods use the same construct for both (McGann et al., 2016), many use separate process and generative models (Jackman, 1994; Gelman and King, 1994a; Kastellec et al., 2008; Thomas et al., 2013).

Further still, some analyses compare the results of *plans* directly, either to candidate plans (Kousser, 1996; Altman et al., 2015) or simulated plans (Chen and Rodden, 2015; Cho and Liu, 2016b). These strategies aim to reveal the impacts of the enacted plan as contrasted to known or attainable alternative plans. This contrasts from seats-votes perspectives that estimate advantage conditional on the given boundaries. Instead, these plan comparison strategies identify when an enacted plan deviates significantly from the simulated or observed comparison plans. Thus, these strategies are still essentially comparative, but the “zone of chance” surrounds the *boundaries* of an electoral system, rather than the observed results within fixed boundaries. In theory, these methods are not mutually exclusive, since seats-votes models may be estimated for simulated district plans, too. But, the goal of using district map comparisons is often to avoid seats-votes constructs entirely (Kousser, 1996), since advantage in these approaches derives directly from boundary differences rather than the structure of simulated h° for the observed plan.

Focusing on seats-votes approaches, many recent analyses of districting plans used in legal proceedings or academic literature focus on critiquing the accuracy and validity of the model of h .²¹ Typically, models of h need only be accurate *distributional* models, in that the typical parameters in the linear model are not substantively interesting. Regardless, thorough specification searches for models of h are uncommon. Although simple forms of misspecification have received attention for seats-votes curve estimation (King and Roberts, 2015), empirical comparisons of new seats-votes methods to one another is relatively uncommon (McGann et al., 2016,

²¹For this, many defendant amici curiae in both *Florida League of Women Voters v. Detzner* (2015) and *Whitford v. Gill* (2016) are illustrative, simple critiques of the inaccuracy and uncertainty in the predictions of these models. Neither swayed the court, but this line of argument is expected.

c.f.), although the comparison of the summary measures from seats-votes curves is common (Grofman et al., 1997a; Linzer, 2012).

This is unfortunate, since many different measures of partisan advantage can be computed from a given seats-votes curve (Nagle, 2015). They focus on different sections of the curve or compute deviation from fairness differently, so it may affect these measures if a given seats-votes curve estimate is significantly different from another. Most of the time, due to structural stability in US legislative elections, reasonably accurate post-hoc analysis of \mathbf{h} using demographic, political, social, or candidate factors (contained in a design matrix, \mathbf{X}), is not too difficult, and predicted \mathbf{h} vectors often agree between many modeling methods. However, the structure of the seats-votes curve may be different depending on the generative model of \mathbf{h}° . By comparing only the estimates of a given quantity of interest, there is no systematic understanding of how the differences in the *generative* model differs.

5.1.1 Common Methods for Estimating Seats-Votes Curves

Typically, seats-votes curve estimation methods divide into two groups. One group, suggested by Tufte (1973), estimates the curve as a linear function of the observed relationship between system-wide seat shares and vote shares pooled over a number of elections. In each election, one observation of a party's average national/statewide vote share \bar{h} and the fraction of the legislature/congressional delegation it controls (\bar{s}) are observed. Pooling over many elections, a linear regression relating \bar{s} and \bar{h} provides the expected share of seats a party wins given that it wins some level of the popular vote.

This technique discards all information about the dispersion and modes of the underlying district vote shares, \mathbf{h} , and has largely been superseded by alternative methods (Jackman, 1994). On this, Jackman (1994) and Gelman and King (1994a) suggest that district-level information should be used to construct a seats-votes curve. They argue that these single-election (or paired-election) strategies avoid pooling between dissimilar elections. These techniques typically combine *two* conceptual models of the electoral process: a substantive model of elec-

toral outcomes under uncertainty, a model of the inherent randomness in an electoral system even under perfect estimation. The first model is justified as a method of constructing accurate estimates of the n -length vector of district-level vote shares, \mathbf{h} , and the second model is justified as a method of constructing plausible *hypothetical* outcomes, \mathbf{h}° , from \mathbf{h} .²²

Various techniques of increasing complexity are used in the first vote shares model, depending on data availability and the demands of the model. These techniques vary from focusing exclusively on *observed* election results (i.e. empirical model-free analysis) to large modeling frameworks designed to control for multiple structural factors under counterfactual simulation. Typically, the second component is called the model of “electoral swing,” a term used to imply randomness in the deviations between the last election and the current election. While many different model specifications exist for the first case, most methods of elections assume that swing is an independent, identically-distributed random effect, while prediction uncertainty can be complex, correlated, and heteroskedastic.

In addition, Nagle (2015) suggests and I show below that previous intuitions about the seats-votes curve are equivalent to statistics about the rank-vote distribution of a given electoral system. Summaries of expected seat shares that a party wins at a given expected statewide/national vote share map directly to questions about the expected rank of the district won with that vote share. This both simplifies the conceptual model of the seats-votes curve, the presentation of its results, and the computation of statistics about the seats-votes curve. While developed historically with reference to a swing model, the seats-votes curve shape is, in fact, available without one. Most critically, this means that the full length of the seats-votes curve is informative, since the vote shares attained by the most Democratic or most Republican contest may be wider than the vote shares attained at the median district, but the curve through this domain is less “valid” from the perspective of the model. This contrasts with the focus in Tufte (1973) inherited by Gelman and King (1994a); Gelman et al. (2010), which suggests that

²²In general, k° will denote a “hypothetical” k , for any symbol or property k . Any hypothetical k° may be *simulated*, in that it obtains under identical conditions to that of k , or it may be *counterfactual*, in that it obtains under alternative electoral conditions. For this discussion, this distinction is not important. But, in later discussions of the validity of these approaches, this distinction matters greatly.

seats-votes curves are only “valid” within a close band of the observed (\bar{h}, \bar{s}) . Extreme ranks are not intrinsically “less certain” than middle ranks, nor are ranks around the pivotal district more certain than ranks near the edges. As will be shown below, the certainty associated with a rank (shown by range of vote shares that can be expected for a district at that partisan rank) may vary in a nonstandard way.

In the following discussion, I will detail a few methods of estimating seats-votes curves, from the simplest empirical uniform-swing strategy to arguably the most complex model types found in the literature deriving from that suggested by Gelman and King (1994a). Then, I examine three stochastic methods to model the seats-votes curve, comparing their estimates and model specifications. Then, I explore the extent to which these model specifications are valid and whether the elections generated are plausible electoral events. This line of inquiry suggests that, in general, better vote share models will lead to more valid seats-votes curves, but the impact of various vote share model misspecifications on the structure of the seats-votes curve estimates may not be straightforward to determine.

5.2 Uniform Swing & the Seats-Votes/Rank-Votes Curves

The uniform swing method forms the basis for all other types of seats-votes curve estimation strategies. This technique examines the hypothetical relationship between party average vote share and the fraction of seats it wins in a Congressional delegation or state legislature by using the size of the average change required to flip each district from one party to another. In the end, it is an analysis of the rank distribution for \mathbf{h} (Nagle, 2015), or a scaled-shifted version of it. Thus, the techniques for constructing this seats-votes curve and its corresponding rank-votes curve are used in *every* other seats-votes estimating technique considered in Section 5.3, and are critical to the construction of advantage measures discussed in Chapter 2. Since each of the realizations from a model of \mathbf{h}_t corresponds to a single rank-vote distribution, the set of simulated rank curves is a direct summary of the relationship between expected seat shares given a party’s level of popular support.

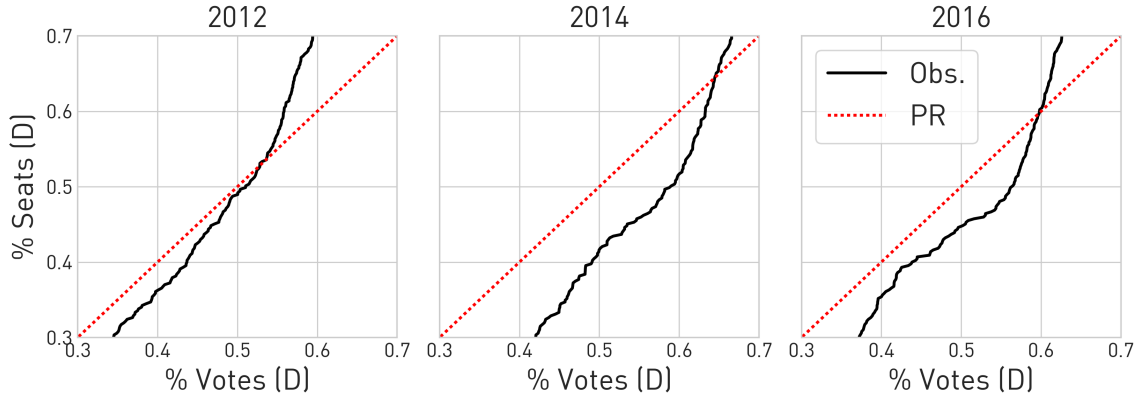


Figure 12. Seats Votes Curves for the US House in the 2012 elections under uniform partisan swing.

Proceeding, the uniform swing presentation of the empirical seats-votes curve requires the vector of vote shares, \mathbf{h} , the average statewide or national vote share for that party, \bar{h} , and a shift term, δ . Let the fraction of the delegation won by the reference party (\bar{s}°) when the average vote share increases by some amount, δ , be the point $(\bar{h} + \delta, \bar{s}^\circ)$ that lies on the seats-votes curve (Jackman, 1994). This point reflects a strict uniform swing in \mathbf{h} (in the sense from Chapter 4, since a shift in average vote share is achieved by adding δ directly to each \mathbf{h}_i , $i = 1, 2, \dots, N$). Using a uniform swing, the fraction of seats the Democrats win at a given hypothetical shift δ° is:

$$\bar{s}^\circ = \frac{\sum_i^N \mathcal{I}(h_i + \delta^\circ > .5)}{N} \quad (5.1)$$

where \mathcal{I} is the indicator function. This means the seats-votes curve under uniform swing is a monotonically-increasing step function related to the cumulative distribution function for the *opponent's* vote shares. The districts most strongly aligned with the reference party sit at the bottom left of the seats-votes curve, since those districts are the most strongly aligned with Democrats. Since they sit at very low values of \bar{h}° , they must have large *district-level* Democrat vote shares, h_i , since they are still won by Democrats when their average support is weak. As δ° increases (Democrat aggregate support increases), districts can only flip from the Republicans to Democrats, until the districts in the top right of the seats-votes curve — the ones most strongly aligned with the Republicans — flip to the Democrats as they win overwhelming levels

of aggregate support. An example of this method, three empirical seats-votes curves for each election since 2012, are shown in Figure 12.²³

While this provides one way that the relation between seats and votes can be measured, this construction requires a theory of strict uniform electoral swing, which is a contentious assertion (as discussed in Chapter 4). However, the assumption of uniform partisan swing is *not necessary* to develop the curves shown in Figure 12. In fact, these curves are a scaled and shifted rank distribution of the district vote shares (Nagle, 2015). The seats-votes curve constructed by uniform swing is always a non-decreasing step function; this is apparent in Figure 13, a detail of the national 2012 seats-votes curve shown in Figure 12 that focuses on the median “tipping point” district, the district closest to $\mathbf{h}_i = .5$ from above. For these curves, any step point in a uniform swing seats-votes curve occurs at some hypothetical party vote share, when some district’s vote share crosses .5 for a given value of δ° . To the left of the step, the reference party wins district i and all other districts k that require $\delta_k \leq \delta_i$ to cross .5. To the right, the party loses i and all k where $\delta_k > \delta_i$.

Let the rank of a district in the distribution of vote shares be denoted r_i , and let ranks be assigned by an ascending ranking function R on the “right” edge, so that ties are all assigned the *maximum* rank of members of the set. This means that, for 435 congressional districts, the district most strongly aligned with the reference party has rank 1, the three districts tied for the second all have rank 4, and the district with the lowest support for the reference party has rank 435. Thus, when a hypothetical “pointer” district with vote share $h_i + \delta^\circ$ crosses .5 and increases rank, the seat share won by Democrats also increases at that δ° .

Together, the set of $(1 - \mathbf{h}_i, \mathbf{r}_i)$ points define the rank vote curve that corresponds to the seats-votes curve. This is simply a reflection on the x -axis of $(\mathbf{h}_i, \mathbf{r}_i)$. At each observed h_i , the rank either stays the same or increments by one. The strongest-aligned Democratic party districts are in the bottom left, have the smallest rank numbers, and are won with the lowest levels

²³This uses the technique suggested by Jackman (1994) from in the Political Science Computational Library (Jackman, 2015), reimplemented in Python.

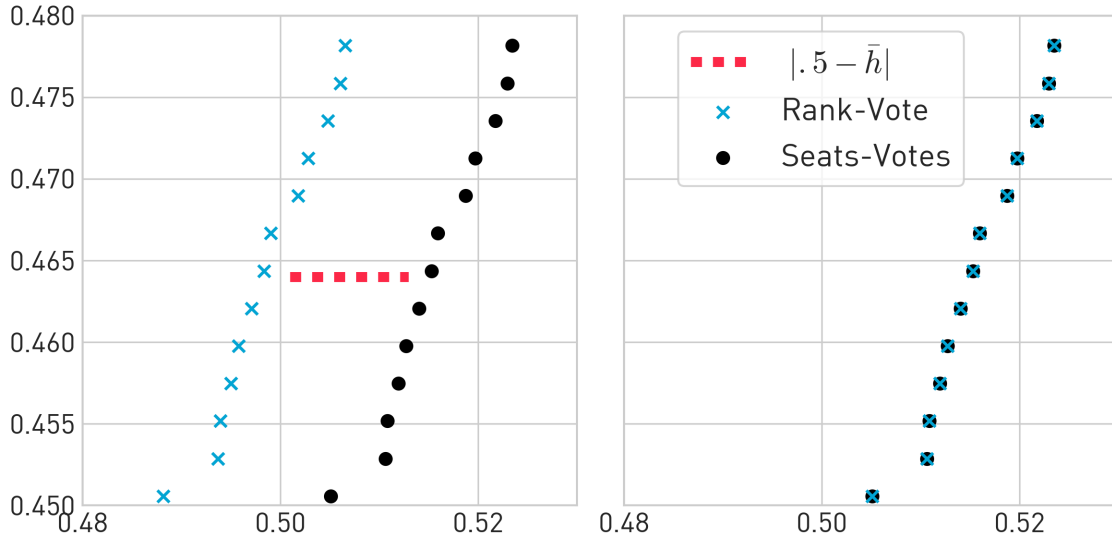


Figure 13. Detail of a uniform-swing Seats-Votes curve in the 2012 House elections alongside a normalized Rank-Vote curve.

of aggregate support; the upper right is the set of districts most aligned with the opposition, have the largest rank numbers, and require that the Democratic party win a large average vote share to flip.

However, if r_i alone were used, then the curve would not have the correct range, varying between 1 and N instead of 0 to 1. In addition, the rank-vote curve is slightly offset from the seats-votes curve in some cases, namely when $\bar{h} \neq .5$. To show this shift, let us first define the rank of the “marginal district,” the district where Democrats win with the smallest margin:

$$r_k = R(\{h_i | h_i > .5\}) \quad (5.2)$$

where $(x - y)_+$ denotes a censored positive difference between x and y .²⁴ Since the maximum rank is assigned among ties, $\bar{s} = \frac{r_k}{N}$ for any δ° where $\bar{h} + \delta^\circ$ maintains the rank r_k . When the rank of the hypothetical $\bar{h} + \delta^\circ$ changes, the seat share changes by $\frac{\#r_k}{N}$, the fraction of all districts that share rank r_k and thus change hands when the marginal rank changes.²⁵ This

²⁴In the Washington example from 2 (data in Table 3, this is district 1. Its rank would be 6, since it is the 6th most Democratic district. In that example, Democrats win $\bar{s} = .6$ at $\bar{h} = .53$).

²⁵Practically, contested district vote shares are rarely exactly equal, so there is often only two sets of ties: the

means any step point on the rank-vote curve in Figure 13 can be re-scaled to match the domain of the seats-votes curve: each seats-votes step point is level with a re-scaled rank-vote step point, $(1 - \mathbf{h}_i, R(\mathbf{h}_i)/N)$.

Finally, the rank-vote curve may not necessarily pass through (\bar{h}, \bar{s}) , but it must have a level set at \bar{s} . In fact, there are only, at most, N possible level sets of seat shares because of the finite number of seats. If a party wins 12 districts, then some district (or set of districts) must be the 12th-most aligned district for that party. Since that district is the marginal district by definition, its level set must pass through .5, and it must be the seat with the closest vote share below \bar{h} . Thus, the marginal district k is the step point to the left of \bar{h} , and can be aligned with the seats-votes curve by adding $\bar{h} - .5$ to \mathbf{h}_k . Since the width of steps is the same between the two curves, adding $\bar{h} - .5$ to the marginal district aligns the rest of the rank-vote curve with the left edge of the seats-votes curve, as shown in Figure 13.

5.3 Generalized Uniform Partisan Swing Methods

Critically, recognizing the seats-votes curve as a transformation of the rank distribution simplifies both the language and the empirics of structural elections analysis. This relationship is easiest to present for the empirical, observed election. But, it is most helpful when considering stochastic methods to model the seats-votes curve. These techniques, sometimes called “generalized” uniform partisan swing methods, are obtained from the uniform swing method of constructing seats-votes curves by relaxing the assumption of strict uniform swing. Given the mapping from rank-votes to seats-votes curves, these techniques can also be seen as leveraging many simulated rank-votes curves to provide a single expected seats-votes curve.

In this vein, stochastic seats-votes modeling methods construct a model for \mathbf{h} , simulate \mathbf{h}° under controlled conditions \mathbf{X}° , and summarize the resulting \bar{s}° given the controlled \bar{h}° and \mathbf{h}° . With the mapping between rank-vote and seats-vote distributions, this is equivalent to summa-

uncontested districts where either the reference party or the opponent receives all recorded votes. All other steps are increments of $\frac{1}{N}$.

rizing the average rank of districts that attain a specific h_i° . Some techniques explicitly model the deviation in simulated elections as if they were random changes from the previous election, and thus partition the variance in outcomes between a “structural” component due to electoral rules, norms, or geographies, and an idiosyncratic component that embodies uncertainty due to estimation. Other methods focus simply on generating accurate process models of \mathbf{h} that produce believable \mathbf{h}° under hypothetical conditions (\mathbf{X}°), and another set focuses on modeling the distribution dynamics of \mathbf{h} , without reference to process justifications. In the discussion that follows, I present two current methods for constructing seats-votes curves in a predictive context and suggest two alternative methods of seats-votes curve construction.

5.3.1 Stochastic Methods for Estimating Seats-Votes Curves

Earlier stochastic models of seats-votes curves focused on estimating the empirical relationship between \bar{h} and \bar{s} , pooling observations over many elections (Tufte, 1973). In contrast, newer methods aim to use the district-level information about the electoral system contained in the full vote share vector (h) to model the relationship between \bar{h} and \bar{s} . The uniform swing method discussed above has its roots in early scholarship that relied heavily on the assumption, since many seats-votes analyses use no explicit stochastic models for \mathbf{h} or swing (Brookes (1960), see also Jackman (1994); Johnston et al. (1999)). And, while the assumption of uniform swing is useful, it is certainly not empirically accurate, as shown in Chapter 4. However, it does appear to be relatively well behaved in recent congressional elections. This regularity drives the pervasive use of normal approximations and is the justification for “generalized” uniform partisan swing. Critically, this assumption makes estimation of the seats-votes curves from district-level data tenable using simple regression techniques. Before wading into the complexities of how this modeling strategy works, I present a simple example using one of the methods considered in this dissertation.

5.3.1.1 The Gelman-King Model

One of the more commonly-used methods to construct seats-votes curves is due to Gelman and King (1994a). First, a model of \mathbf{h} is estimated. Then, a counterfactual/hypothetical election, \mathbf{h}° , is simulated using this model. After \mathbf{h}° is obtained, it is then *adjusted* using the model of swing, δ° . Finally, this results in a single realization of a rank distribution, \mathbf{r}° that is simply the rank vector of the hypothetical outcome, $\mathbf{h}^\circ + \delta^\circ$. As discussed above, this rank vector is sufficient to construct *one* realization of the seats-votes curve implied by the model. Then, many seats-vote curves over many simulations are generated and analyzed. When each of these seats-votes curves are considered together together, any level set of \bar{s}° has a α -sized confidence band: the middle $1 - \frac{\alpha}{2}\%$ range of simulated vote shares for the $\bar{s}^\circ * N$ th-ranked district contains a “level set” of $1 - \frac{\alpha}{2}\%$ of all simulated seats-votes curves. When considering all simulations, this “level set” represents the set of \mathbf{h}_i° that are attained at that rank, \mathbf{r}_i , or seat share \mathbf{r}_i/N . By connecting the quantiles of h_i° across all ranks, an estimate and simulation interval for the entire seats-votes function is attained.²⁶

One example of this method is shown in Figure 14. On the left, a single simulated election is plotted against the observed election’s rank-vote distribution. The observed \bar{h} is plotted in the gold vertical line on the left, and the observed \bar{s} is the point at which that line intersects the black quantile plot. Then, on the right, 1000 seats-votes realizations are plotted in green underneath the observed outcome in black. The 5th, 50th and 95th percentiles of vote shares within the rank are connected by the overlaid gold lines, meaning 90% of all simulated seats-votes curves fall within the gold lines. Practically, these gold lines are formed by connecting the target percentiles within each rank/level set, $(1 - h_i, \frac{i}{N})$, $i = 1, 2, \dots, N$. Critically, *all* hypothetical measures of advantage pertain to quantities that summaries the sets of simulations in green. Likewise, “empirical” measures of advantage pertain only to the black observed rank-vote out-

²⁶While this is not discussed directly in Gelman et al. (2010), the package it explains implements this method when computing quantities of interest.

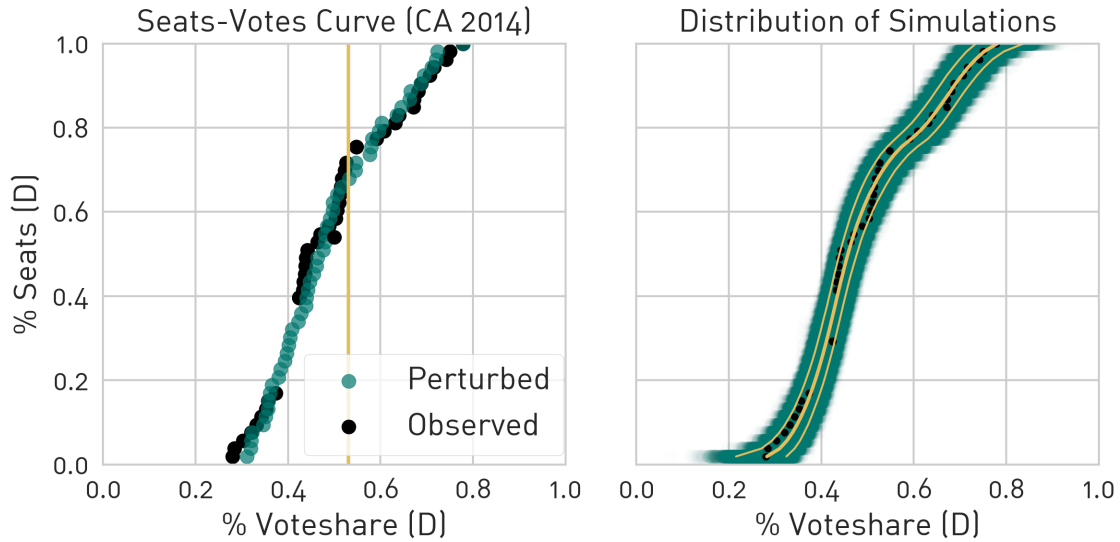


Figure 14. A Seats-Votes curve estimate for the California congressional delegation in 2014. On left, the gold vertical line is the statewide average party vote share. On right, the gold lines connect the 5th, 50th, and 95th percentiles of simulated vote shares within each rank, centered on the median of simulations.

come set. Thus, if spatial misspecification were to affect this model, it must affect the shape of the *simulation envelope* shown in green. It is not enough to simply affect the covariance of the disturbances in the model; it must also generate differently-shaped rank distributions to matter.

Another view of this procedure is shown in Fig. 15. In this example, we are attempting to construct $E[\bar{s}|\bar{h} = .5]$, a very common quantity of interest. This value is used to estimate the median bonus measure discussed in Section 2.3, a measure of partisan advantage. The figure portrays the full simulation envelope (like that shown on the right of Figure 14) on the left. Then, the center plot shows a high-detail focus on the median vote share, $\bar{h} = .5$. In this case, Gelman et al. (2010); Linzer (2012) suggests summarizing the distribution within a search band around $\bar{h} = .5$ instead of directly at the target value. Thus, a small search band, $.5 \pm .005$, is plotted in light blue around .5, and is made clear in the detail at the center of the plot. The distribution of \bar{s}° that fall within within the search band is shown to the right of the plot. So, the rank of districts within this search band are recorded for every simulation. Then, a measure of central tendency

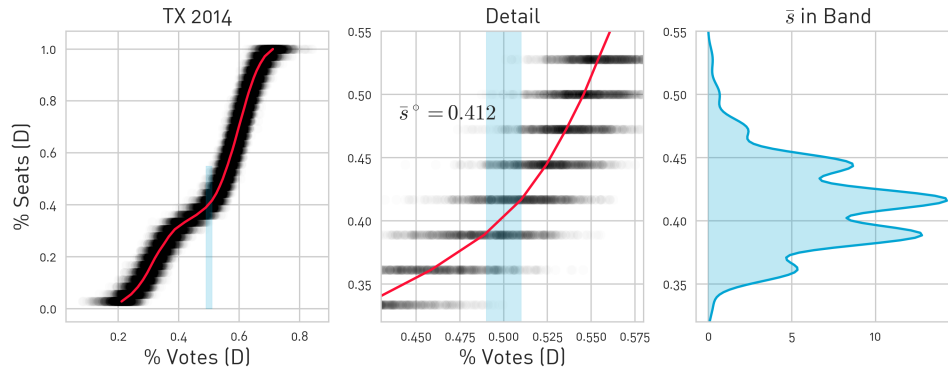


Figure 15. Seats-Votes curve for Texas in 2014. When the share of the vote Democrats win is 50%, the party can expect to win only around 41% of the Texas Congressional delegation. This would correspond to a median bias estimate of $-.18$. Here, this estimate is computed from the average rank of a district that falls within the blue band, divided by N , and is the average value of the distribution on the right of the figure. This distribution is the set of simulated seat shares that fall within the search band and is shown on the right.

for the distribution to the right is used as the estimate of \bar{s}° , and can be used in the measures of advantage outlined in Section 2.3.

Proceeding to a more detailed presentation, the authoritative discussion is provided in Gelman and King (1994a), while the most current implementation derives from work in Gelman et al. (2010). Occasionally, these describe different techniques; when in doubt, the interpretation provided in Gelman et al. (2010) is considered the canonical form, since it is the only existing implementation of the technique. A presentation of the method described there will be provided below. While popular, the method has also been soundly critiqued many times since its inception. I use it here not as the *end-all* method of modeling seats-votes curves, but rather in hopes that using a stable model specification with well-known properties will make the exploration of the new impact statistics simpler. The following discussion focuses on the main estimating concerns as discussed by Kastellec et al. (2008); Gelman et al. (2010) and that are required to generate the style of curve from Figure 14.

First, Gelman and King (1994a) suggest modeling the district vote shares for the reference party in a given year, \mathbf{h}_t , as a function of available electoral conditions in that year, \mathbf{X}_t and the

vote outcome in the previous election, \mathbf{h}_{t-1} if available. No attempt is made to link districts across the redistricting threshold if boundaries change. Thus, every year immediately following a redistricting, only \mathbf{X}_t is available for the model; in all other years, both \mathbf{X}_t and \mathbf{h}_{t-1} are available. In this way, the method suggested by Gelman and King (1994a) and many analyses that follow from it (Gelman and King, 1994b; McKee et al., 2006; McGann et al., 2015, e.g.) conceptualize each districting plan as a single continuous panel. Each decadal redistricting breaks the panel, even if the districts themselves mostly remain unchanged. However, Gelman and King (1994a) do not use typical panel analysis techniques, since their model uses the one-step pooling of \mathbf{h}_{t-1} to predict \mathbf{h}_t . Over a single decade, this results in four separate two-cycle models and one single-cycle model, each with their own parameter estimates. Thus, even though \mathbf{h}_{t-1} is present in most models of \mathbf{h}_t , the model is not a typical autoregressive process model either, since all parameters are considered non-stationary and temporally independent.

Using these models, an estimate of the inherent deviation (σ) and structural uncertainty (λ) are made for each decade. To explain, let the basic model be stated for pairs of elections in time t and $t - 1$:

$$\begin{aligned}\mathbf{h}_t &= \mathbf{X}_t\beta_t + \alpha_t\mathbf{h}_{t-1} + \gamma + \epsilon_t \\ \gamma_t &\sim \mathcal{N}(0, \lambda\sigma^2) \\ \epsilon_t &\sim \mathcal{N}(0, (1 - \lambda)\sigma^2)\end{aligned}\tag{5.3}$$

In this, γ is the structural error component that applies to all realizations within the redistricting decade, ϵ_t is the inherent error component, β_t are the marginal effects of the electoral conditions, and λ is the fraction of the overall variance that is “systemic,” which is assumed to remain constant from year to year. In a single model, γ_t is not separable from ϵ_t without additional information, so their sum is the only thing identified at each individual t . This also means λ is unidentified at any t . In addition, it is unlikely that separate, two-cycle regressions will recover time-stationary σ , though they may be similar.

Thus, a correction/secondary modeling step is suggested. First, to estimate λ and the invariant σ , Gelman and King (1994a); Gelman et al. (2010) estimate λ from the average of the

first-order temporal autoregressive effect, α_t , over all of the two-cycle models, and estimate σ^2 from the average of σ_t^2 . Since the specification for each cycle pair $((t, t - 1))$ has a different estimate of β_t, α_t and σ_t^2 , the average α_t over the entire decade provides the estimate for the fraction of the variance that is “structural,” i.e. the decade random effect in a panel sense. Then, each year’s random effect ϵ_t , derives from the fraction of remaining variance once this is accounted for.

For the purposes of counterfactual simulation and election prediction, these grand estimates are then plugged back into the model in 5.3. But, since γ is now known, a “more precise” counterfactual can be generated, since more information is known about h_t than that available if it were not observed at all. Thus, Gelman and King (1994a) ostensibly suggest a process justification for this model term, but appear to leverage it heavily in order to reduce the variance of counterfactual simulations. The “known” information from observing \mathbf{h}_t that is yielded when simulating \mathbf{h}_t° is contained directly in γ , so only the remaining $(1 - \lambda)\sigma^2$ variance is required. As such, the distribution of a hypothetical/counterfactual district vote share vector, \mathbf{h}_t° , for hypothetical electoral conditions \mathbf{X}_t° conditions on the observed district vote share vector \mathbf{h}_t in time t and is Gaussian:

$$p(\mathbf{h}_t^\circ | \mathbf{h}_t) = \mathcal{N}(\lambda \mathbf{h}_t + (\mathbf{X}_t^\circ - \lambda \mathbf{X}_t) \hat{\beta}_t + \delta_t, (1 - \lambda^2) \sigma^2 I + (\mathbf{X}_t^\circ - \lambda \mathbf{X}_t) \hat{\Sigma}_{\beta_t} (\mathbf{X}_t^\circ - \lambda \mathbf{X}_t)') \quad (5.4)$$

where Σ_{β_t} is the covariance matrix of the β estimates in the model for period t and δ_t is a scalar electoral swing term used to control the magnitude of \bar{h}° . In this model, λ ensures that the counterfactual realization of \mathbf{h}_t° is shrunk towards the observed \mathbf{h}_t . Gelman and King (1994a) emphasizes this similarity by considering the expected counterfactual \mathbf{h}_t° when hypothetical conditions are equivalent to observed conditions, $\mathbf{X}_t^\circ = \mathbf{X}_t$:

$$E[\mathbf{h}_t^\circ | \mathbf{h}_t] = \lambda \mathbf{h}_t + (1 - \lambda) \mathbf{X}_t \hat{\beta}_t + \delta_t \quad (5.5)$$

Thus, the use of this decadal “structural” random effect term allows \mathbf{h}° to condition on \mathbf{h} , reducing its variance using λ . This is made clear when examining the predictive distribution, which has no conditioning on \mathbf{h}_{t+1} because it is unavailable. This takes the form:

$$p(\mathbf{h}^\circ) = \mathcal{N}(\mathbf{X}^\circ \hat{\beta} + \delta_t, \mathbf{X}^\circ \hat{\Sigma}_{\beta} (\mathbf{X}^\circ)' + \sigma^2 I) \quad (5.6)$$

which shows no shrinkage due to λ . However, I use only the counterfactual mode in this dissertation, so the predictive distribution, while implemented, is largely ignored.

This method is defensible, since expanding a first-order autoregressive representation yields a similar expression to that given in 5.3, if α were time-consistent, this would result in a model for \mathbf{h}_t with a persistent error akin to γ :

$$\mathbf{h}_t = \mathbf{X}\beta + \alpha\mathbf{h}_{t-1} + \epsilon = \mathbf{X}_t\beta_t + \alpha(\mathbf{X}_{t-1}\beta_{t-1}) + \alpha\epsilon_{t-1} + \epsilon_t \quad (5.7)$$

But, since this is indeed not a $VARX(1)$ model, α_t are not necessarily bounded between -1 and 1 like typical stable autoregressive coefficients. This means λ has no domain bounds, either. In theory, λ may fall outside of $(0, 1)$, which would force the resulting variance of γ or ϵ to be negative. In all models considered, both λ and α remain firmly in $(0, 1)$, clustering tightly between .6 and .8.²⁷ In addition to potential issues with λ , the substantive effects (β) can vary wildly over time since they are unconstrained. In practice, these effects also tend to be relatively stable over time for the two-cycle models. However, their estimates in the initial model each decade is rather different. This occurs because the mode *with* \mathbf{h}_{t-1} is quite different from the model without it, since \mathbf{h}_{t-1} is often a very good predictor of \mathbf{h}_t .

Three implications are important to note about this modeling strategy. First, since Σ_{β_t} is used in the covariance of \mathbf{h}_t° , the covariance between the electoral processes in \mathbf{X}_t° may induce correlation or heteroskedasticity in the simulated outcomes. In practice, this makes sense, since the vote share won by Democrats in a district with an uncontested Democratic incumbent might be expected to be much less volatile than a contested election with no incumbent. If this is consistently observed, then it will also manifest in the resulting counterfactuals.

Second, the variance of the counterfactual can be decomposed to emphasize how this is a generalized uniform random swing model. Reprising the counterfactual distribution from Eq. 5.4:

$$p(\mathbf{h}_t^\circ | \mathbf{h}_t) = \mathcal{N}(\lambda\mathbf{h}_t + (\mathbf{h}_t^\circ - \lambda\mathbf{h}_t)\hat{\beta}_t + \delta_t, (1 - \lambda^2)\sigma^2I + (\mathbf{X}_t^\circ - \lambda\mathbf{X}_t)\hat{\Sigma}_{\beta_t}(\mathbf{X}_t^\circ - \lambda\mathbf{X}_t)')$$

²⁷A result noted by Gelman and King (1994a) as well.

we can identify a “functional” separation into three components. First, a component from the conditioning with no uncertainty attached. Second, a stochastic component due to the intrinsic uncertainty surrounding the estimation of β_t and its application to a counterfactual electoral condition \mathbf{X}_t° . Third, a stochastic *generalized uniform random swing* term. Stated in descending order:

$$\begin{aligned}
p(\mathbf{h}^\circ | \mathbf{h}_t) = & \lambda \mathbf{h}_t \\
& + \mathcal{N}((\mathbf{X}_t^\circ - \lambda \mathbf{X}_t) \hat{\beta}_t, (\mathbf{X}_t^\circ - \lambda \mathbf{X}_t) \hat{\Sigma}_{\beta_t} (\mathbf{X}_t^\circ - \lambda \mathbf{X}_t)') \\
& + \mathcal{N}(\delta_t, (1 - \lambda) \sigma^2 I)
\end{aligned} \tag{5.8}$$

In addition, note that *any* linear model of \mathbf{h}_t that uses \mathbf{h}_{t-1} as a predictor can be treated in this manner. Thus, this structure is not unique to the Gelman-King formulation, and is instead a property of first-differencing. Without using the repeated estimation and averaging of Gelman and King (1994a), we can state a generic two-cycle model:

$$\mathbf{h}_t = \mathbf{X} \beta_t + \alpha \mathbf{h}_{t-1} + \epsilon \tag{5.9}$$

Then, we introduce an additional $\pm \mathbf{h}_{t-1}$ term, which has no net effect on the model, but allows us to rearrange the model for \mathbf{h}_t into one for δ_t :

$$\mathbf{h}_t = \mathbf{X} \beta + \alpha \mathbf{h}_{t-1} \pm \mathbf{h}_{t-1} + \epsilon \tag{5.10}$$

$$\mathbf{h}_t - \mathbf{h}_{t-1} = \mathbf{X} \beta_t + (\alpha - 1) \mathbf{h}_{t-1} + \epsilon \tag{5.11}$$

$$\delta_t = \mathbf{X} \beta_t + (\alpha - 1) \mathbf{h}_{t-1} + \epsilon \tag{5.12}$$

Thus, the “total” uncertainty for δ_t will be due to uncertainty in estimating β , uncertainty in estimating α , and *inherent error* in ϵ . Again, the breakdown can be applied from Eq. 5.8 and result in the same statement of an explicit model for the covariance of the generalized partisan swing term, δ_t , which will typically be independent of the estimation uncertainty for $\hat{\beta}$ and $\hat{\alpha}$ for specifications in this style.

Third, the use of varying values for $E[\delta^\circ]$ is not required to generate a seats-votes curve *unless* $E[\delta^\circ]$ affects the covariance of $\mathbf{h}_t^\circ | \mathbf{h}_t$. As discussed above, the rank vector for any \mathbf{h}° is

sufficient to construct a seats-votes curve. When simulations are conducted for the observed electoral conditions \mathbf{X} , simulations have $\bar{h}^\circ = \bar{h}$ on average. When shifted by δ° , the δ° applies uniformly to all *simulated* h_i° in the same fashion as the uniform constant swing method. Since $\bar{h}^\circ = \bar{h} + E[\delta^\circ]$ by construction, the expected fraction of seats won at \bar{h}° is the average rank of the marginal district with $h_i^\circ = \bar{h} + E[\delta^\circ]$ over simulations, divided by N . However, depending on the specification, the magnitude of δ° may change the expectation of each realization of any given $h_{i,t}^\circ$. Many specifications for this effect will be explored later in this chapter, to examine whether this can indeed affect the structure of the simulated seats-votes curve.

5.3.1.2 Alternative Model-Driven Seats-Votes Constructions

A model for \mathbf{h} that can generate plausible elections in simulation is a critical concern for seats-votes curves. Refinements of the Gelman and King (1994a) method, such as those developed by Linzer (2012) or McGann et al. (2015, 2016), use substantively different models of electoral structure, leading them to simulate h° in markedly different ways. For brevity, Linzer (2012) will not be discussed at length here, since it incorporates no information about electoral conditions, and instead jointly models turnout and vote share. It provides a useful technique in estimating seats-votes curves for multiparty systems, where techniques for doing so are rarer and more difficult to justify, and in cases where a stochastic model for both turnout and vote share is desired. Its lack of predictive capability means that it is critically limited in the analysis of candidate redistricting plans, however. A simpler post-hoc modeling strategy that uses only information about vote shares and turnout is also available for two-party systems. This technique is based on bootstrapping (Efron, 1982) and serves as a diagnostic method for parametric prediction-capable seats-votes curve models.

5.3.1.3 Modeling Swing as Hierarchical Deviation

In recent work, McGann et al. (2016) presents (and McGann et al. (2015) uses) a strategy to construct seats-votes estimates in linear modeling framework similar to Gelman and King (1994a). Their model, however has a substantially different take on how electoral variability should be modeled in a seats-votes framework. McGann et al. estimate the deviation used for simulations in a seats-votes framework directly:

$$\mathbf{h} \sim \mathcal{N}(\mathbf{X}\beta + \alpha\bar{h}, \tau^2) \quad (5.13)$$

where \bar{h} is the party vote share at some higher level of geography. The analyst has options on how to best pick the aggregating units for \bar{h} : McGann et al. (2016) suggest \mathbf{h} be pooled either over states or nationally, depending on the scale of analysis. In addition, \mathbf{h} is pooled over an entire redistricting decade, making the model as a whole fit a full set of Nt district-years, with \bar{h} then becoming a block-constant Nt vector of the aggregate vote shares for the reference party in year t . As before, \mathbf{X} contains information about the electoral system. Like Gelman and King (1994a), McGann et al. (2016) suggests an ordinal incumbency fixed effect and, depending on the way uncontested elections are handled before model fitting, a centered ordinal fixed effect for contestedness by party. After estimating this model over the pooled decade of data, simulated h° are drawn directly from Eq. 5.13, possibly under counterfactual electoral conditions \mathbf{X}° .

Together, the direct modeling of \mathbf{h} and accommodation of \mathbf{X}° means the McGann method can be compared to the Gelman-King method directly. Notably, McGann et al. (2016) does not provide a direct comparison of their technique against the implementation in Gelman et al. (2010)²⁸. In addition, the development of their method in McGann et al. (2016, ch. 3, Appendix 3B) does not provide the same kinds of model justification provided by Gelman and King

²⁸This may be due to the fact that the program made by Gelman et al. (2010), uploaded to the R package index in 2011, was removed in early 2015 for relying on deprecated functionality. I provided a patch to maintainers at the IQSS in fall of 2015, but the patch was not merged. It is still unavailable from the Central R Archive Network, and the source hosted by the IQSS website is unusable for current versions of R.

(1994a). Discussing the aspects in which the techniques differ and comparing them empirically may provide insight into the strengths and weaknesses of the two techniques and illustrate where the derived quantities of interest may differ.

First, the variance of \mathbf{h}° is modeled exactly as τ^2 . This means observing \mathbf{h}_t provides no additional information about $\mathbf{h}_t^\circ | \mathbf{h}_t$ beyond the estimate of the mean and variance. Flatly, uncertainty about what would happen is not reduced by knowing what did happen. This results from the fact that there is no division of variance into estimation uncertainty (Σ_β) and inherent electoral deviation (σ), and no modeling of the potential temporal autoregressive relationship between subsequent elections outside of the incumbency variate. For this, the authors suggest that decadal pooling rather than using an autoregressive model or a full panel design yields a more robust yet tenable simplification of the complicated Gelman-King method.

This simplicity comes at a steep cost: since the model for \mathbf{h} is indistinguishable from the generator for \mathbf{h}° , the i.i.d. covariance model of \mathbf{h} applies to \mathbf{h}_t° as well, for any time period during the decade. Drawing *new* \mathbf{h}° directly from Eq. 5.13 assumes that any magnitude of swing is equally likely for any district at any time. Electoral conditions, such as contestedness, incumbency, or candidate quality have no effect on the variance of simulated vote shares in each district like they do in the Gelman-King model. On its face, this seems exceedingly implausible, but it remains to be demonstrated that it matters empirically.

On less technical grounds, the two methods express different fundamental conceptions of “swing” in the contexts of a structural election model. McGann et al. include higher-level party vote shares where Gelman-King use previous years’ vote shares. Holding electoral conditions in \mathbf{X} equal, electoral swing is modeled as random fluctuations in districts around the state or national mean in that year. In contrast, the Gelman & King model takes “swing” to mean the changes in district-level vote shares *between elections*, either counterfactual or observed. While both approaches involve some sort of pooling over contiguous districting periods there is no reason to believe that variance within an election around a group mean should correspond to the variance between pairs of elections across time.

5.3.1.4 Bootstrapping

Model-based approaches, such as Gelman and King (1994a) or McGann et al. (2016), are desirable insofar as they stipulate real, contestable theories of how vote shares and electoral swing may be understood. Less process-driven methods focus on modeling the distribution of electoral results directly (Linzer, 2012). One novel technique I have developed in this vein uses bootstrapping to simulate alternative plausible elections directly from observed results. Bootstrapping is a data amplification technique that has been used extensively for model validation and sensitivity analysis (Efron, 1982; Efron and Hastie, 2016). Seats-votes curves are inherently data-limited, since only one vector of \mathbf{h} (and thus only one pair (\bar{h}, \bar{s})) occurs for each election. Bootstrapping to generate plausible counterfactual elections, then, may provide a third method against which the McGann and Gelman-King methods can be compared.

To construct a bootstrapped seats-votes curve, electoral swings are simulated using a pair of sequentially-observed elections (\mathbf{h}_t and \mathbf{h}_{t-1}). Vote share in the previous election is treated as fixed, and alternative \mathbf{h}_t° are constructed by resampling with replacement from the vector of empirically-observed district-level swings, $\delta = \mathbf{h}_t - \mathbf{h}_{t-1}$. The resampled swing vector, δ° , is then added to the previous years' results to generate hypothetical elections in time t . When districts have equal turnout, this ensures:

$$E[\mathbf{h}_t^\circ | \mathbf{h}_{t-1}] = E[\mathbf{h}_{t-1} + \delta^\circ] = \bar{\mathbf{h}}_{t-1} + \bar{\delta} = \bar{\mathbf{h}}_t \quad (5.14)$$

However, when districts have unequal turnout, this does not hold, since \bar{h}_t is a turnout-weighted average of \mathbf{h}_t and δ° is drawn without respect to turnout.²⁹ Since the full rank distribution for any realization is available *regardless* of the value of \bar{h}° , the elections can be analyzed like in the examples at the head of Section 5.3.1.1. This method ensures that the simulated vote share for each district, on average, will have the same dispersion as the overall distribution of swings. In addition, it avoids placing a parametric model on the election outcomes. In a similar fashion

²⁹This also occurs for linear models of \mathbf{h} : if OLS is used, $E[\bar{h}]$ may not be \bar{h} , since all districts are weighted equally. If WLS is used with weights proportional to turnout as recommended by Gelman et al. (2010), this holds.

to the McGann et al. method discussed above, the variance of a district’s simulated vote shares is independent of the political conditions within it, since the swings are sampled without respect to this data. In addition, the bootstrap assigns swing without respect to potential inter-district correlation. This method may also be used to resample for new districting plans. For any new district composed of q pieces of old districts, weight q draws from the swing distribution by a population variate (either raw population, eligible voters, registered voters, etc.) and add to the weighted combination of observed \mathbf{h}_t .

5.4 Comparing Seats-Votes Curves

To get a sense of how different the resulting estimates are for these three methods, the results of the three techniques are compared for California congressional districts and the national congressional seats-votes curve in 2014. The fitted curves for three methods are shown for California in Fig. 16 and for the US as a whole in Fig. 18. The Gelman-King model specification predicts the 2014 Democrat vote share using the 2012 Democrat vote share and political conditions in the district:

$$\mathbf{h}_t = \alpha \mathbf{h}_{t-1} + \beta_0 + \beta_1 \text{incumbent} + \beta_2 \text{uncontested} + \gamma + \epsilon \quad (5.15)$$

where “incumbent” is -1 if the district has an incumbent Republican candidate running for reelection, 1 if the district has a Democratic candidate running for reelection, and 0 if no single incumbent is running. Likewise, “uncontested” is -1 when a Republican runs uncontested, 1 when a Democrat runs uncontested, and 0 when the election is contested.³⁰ Here, elections below 1% or above 99% Democrat share of the two-party vote are considered effectively un-

³⁰This model can get much more complex, using a state fixed effect and a hierarchical model over states when fitting for the national seats-votes curve, but the naive model typically has an R^2 of around .95, so many of the additional covariates simply degrade the quality of the model by inducing collinearity or worsening its parsimony.

contested. When district i is uncontested, its vote share is imputed from a sub regression on the complete data.³¹

In this case, the model is estimated using weighted least squares, with weights proportional to turnout. This is mainly done to ensure that $E[\hat{\mathbf{h}}] = \bar{h}$ when all elections are contested, since \bar{h} is the turnout-weighted average of \mathbf{h} . This also ensures that \bar{h}° generated in simulation are comparable to \bar{h} , conditional on an effective imputation of missing vote shares. For the event that a district's turnout is unavailable, it is imputed from the available data. A side effect of weighting by turnout is that any potential heteroskedasticity due to turnout like that Linzer (2012) is concerned with is resolved. More broadly, heteroskedasticity in the national-level model is likely to follow a significantly different form than that due to turnout alone. To conduct the weighted least squares estimation, a diagonal matrix, T , is used that contains turnout in each district. Then, γ is modeled as a heteroskedastic error term with variance $\lambda\sigma^2T^{-1}$ and ϵ is normal heteroskedastic with variance $(1 - \lambda)\sigma^2T^{-1}$. In this example, data from 2012 and 2014 are used, meaning counterfactual 2014 elections are drawn from the weighted modification of Eq. 5.8 that uses heteroskedastic independent swing rather than homoskedastic independent swing.

In addition, a McGann-style model is fit by pooling elections in 2012 and 2014:

$$\mathbf{h} = \beta_0 + \beta_1\bar{\mathbf{h}} + \beta_2\text{incumbent} + \beta_3\text{uncontested} + \epsilon \quad (5.16)$$

with ϵ being a normal, independent error term with variance σ^2 and similar heteroskedastic turnout variance weights T^{-1} . Using the McGann strategy, the estimated seats-votes curve is significantly different, depending on whether $\bar{\mathbf{h}}$ is the statewide average vote share or the national average vote share. Using statewide vote share in the national model actually increased the discrepancy between the McGann model and the other two models, tending to bias the curve towards Republicans by almost ten seats over the entire curve. At the state level, the

³¹Sensitivity analyses were conducted to determine whether there is a strong impact depending on the location of the cutoff and method of resolving uncontested elections. A subset of these explorations are presented in Section 6.2.

choice made no difference. So, a consistent specification using the national Democrat vote share is compared for both examples.³²

Finally, a bootstrapped seats-votes curve is constructed, where swings are resampled from:

$$\delta^\circ \sim \{\mathbf{h}_t - \mathbf{h}_{t-1}\} \quad (5.17)$$

Simulated 2014 election results are then constructed by $\mathbf{h}_i^\circ = \delta^\circ + \mathbf{h}_{t-1}$. In this case, the bootstrap is unweighted, so all swings are equally likely to be chosen. Akin to the other methods considered here, this results in a spatially-independent swing vector, since the swing at district i is independent from the swing of any of its neighbors. This method also does not respect potential spatial heterogeneity in swing, but a stratified bootstrapping approach might resolve this concern. In addition, for the national curve, all swings, regardless of state, are pooled for the national bootstrap. The bootstrap, in addition to the simulations for the Gelman-King and McGann methods, are run 10,000 times.

5.4.1 Example: California Congressional Districts in 2014

The example for the 2014 seats-votes curve estimate from California is shown in Figure 16. The bands denote the space within the 5th and 95 simulation percentiles for the simulated seats-votes distributions, which typically appear as point clouds along each rank as in the right side of Figure 14. In terms of the lowest average discrepancy, the McGann and Bootstrap methods agree most closely over the full range of values. In comparison, the median simulated election from the Gelman-King method has Democrats winning more seats than either the McGann model or the Bootstrap over the range of competitive elections, where $.45 < \bar{h} < .55$. The difference in median between the simulations (for either the McGann method or the bootstrap)

³²Here, the use of hierarchical or spatial fixed effects has an impact on the national estimated seats-votes curve, though it appears to be smaller than the impact of using the statewide average vote share in each year. State fixed/pooled effects could be used in either linear modeling framework, and bootstrapping could be stratified by state. Notably, using state fixed effects in addition to statewide means in each year induces unacceptable multicollinearity in the design matrix, and so are not presented. More generally, McGann et al. (2016) provides no guidance on this specification question.

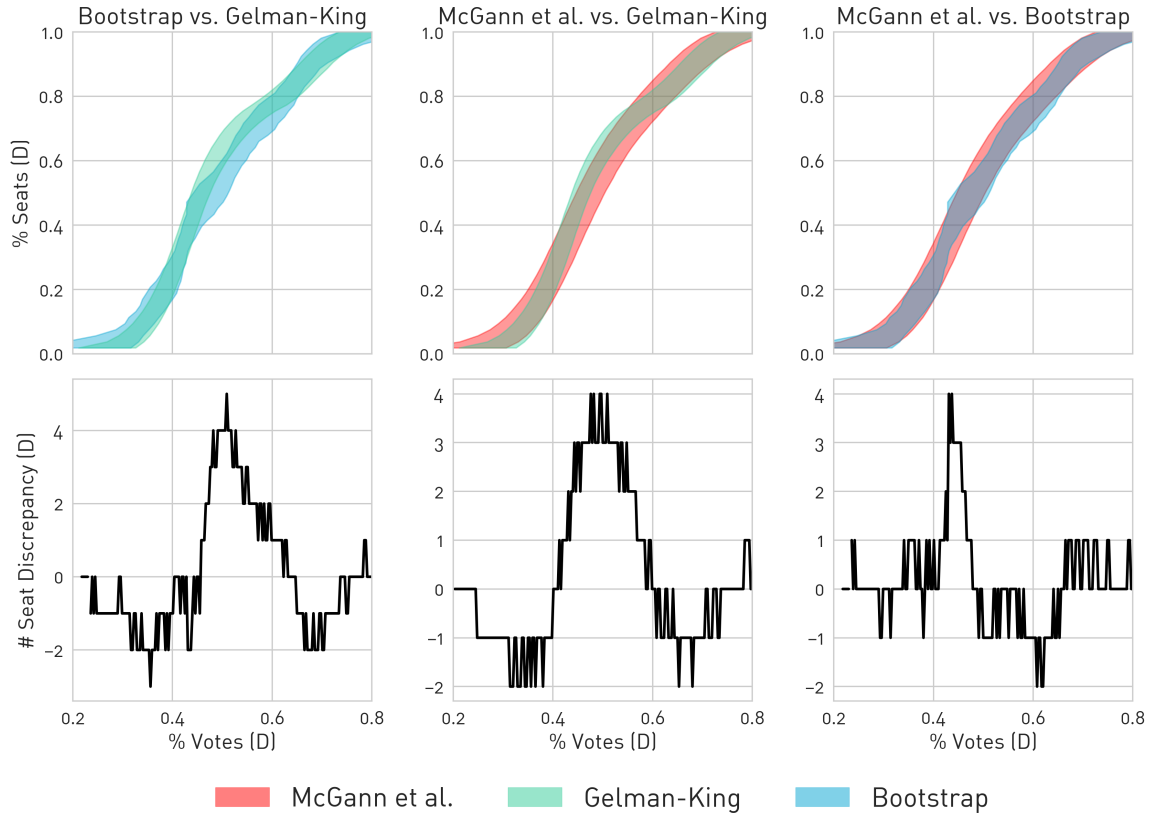


Figure 16. Three methods of estimating the seats-votes curve for California Congressional districts in 2014. The top row is a pairwise comparison between two methods, and the bottom row displays the difference in number of seats awarded as a function of statewide Democratic vote share.

in this range varies between a one- to four-seat difference. Out of fifty-three seats, a four seat difference is nearly 10%, and is a significant discrepancy.

While the middle 90% of simulations overlap over the entire domain for all simulations, the medians of the bootstrap and the Gelman-King curves are not covered by each others' confidence intervals for "tightly" contested elections, where $.48 < \bar{h} < .52$. In fact, the simulation intervals almost diverge entirely at $\bar{h} = .5$. This means that the expected share of the California congressional delegation that Democrats can expect to win under bootstrapping would be exceedingly unlikely in the Gelman-King simulations, and vice versa. At $\bar{h} = .5$, the Gelman-King model would suggest democrats win four more seats, an excess seat share of almost 7%, than

that estimated by the bootstrap model. This range is also the range of maximal disagreement between the Gelman-King and McGann curves. Critically, disagreement is highest in the range where many counterfactual advantage measures focus, suggesting these estimates may be sensitive to the structure of the model used. In contrast, the McGann and Bootstrap methods differ most strongly just below the electoral median, $.4 < \bar{h} < .5$. Even in this range, however, the median simulations in either technique is within the middle 90% range of simulations.

In part, the divergence in median \bar{s}° of the Gelman-King and Bootstrap methods for competitive elections is driven by narrower confidence bands around that range. Again, the largest difference in \bar{s} over \bar{h} for the McGann-Gelman contrast and the Bootstrap-Gelman contrast is four seats. But the simulation intervals are wider for the McGann method in that range, so it is unsurprising that the median Gelman-King simulation remains firmly within the wider simulation envelope for the McGann method.

Importantly, the variance of vote shares attained within each level set is not necessarily the variance of each *district*, as is shown in the furthest right of Figure 17. In Fig. 17, the apparent standard deviation over simulations is computed for three different marginalizations. On the left is the standard deviation of \mathbf{h}° in any given simulation. This is comparable to the deviation in observed vote shares. On the right is the deviation of each district over all simulations. This would be comparable to the deviation of h_i for some district i if many elections were run from the same conditions. On the right, the deviation of vote shares in each rank is provided, which corresponds to the width of the seats-votes curve estimate within each level set.

All methods generate simulations with roughly comparable electoral deviation to the observed election. Each election simulated in the Gelman-King method tends to be slightly less noisy than the observed 2014 election. The bootstrap tends to be more noisy, and the McGann method centers on the empirical variance, but the difference in all three methods is small in this regard. For the heteroskedasticity in the right plot, Gelman-King method explicitly incorporates heteroskedasticity due to turnout. Potential correlation in $\Sigma_{\hat{\beta}}$ also reduces the magnitude of the post-hoc univariate estimator of standard deviation.

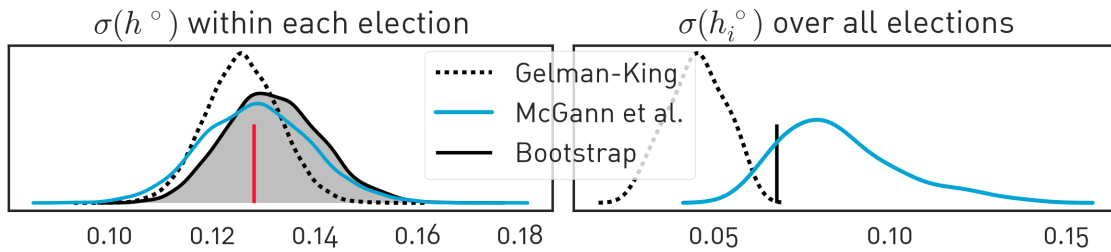


Figure 17. Distributions of variance within and between simulations used to generate Fig. 16. The empirical mean is shown by the vertical line in the left plot. In the right plot, the bootstrap variances are so tightly clustered around .061 that its kernel density is too tall to display on the same scale.

Both the Gelman-King and McGann models are estimated using weighted least squares, and this provides the McGann method the heteroskedasticity in h_i° shown in the right plot of Figure 17. In contrast, the districts in bootstrapped simulations tend to all have a standard deviation of around .061. In the rank-deviation plot on the furthest right, the McGann and Bootstrap curves both have long right tails, indicating some ranks are much wider than others. However, the bootstrap also has a few very narrow ranks, with the narrowest occurring near $\bar{h} = .36$. The Gelman-King curve tends to be narrower than the McGann curve, and its widest point is also narrower than the widest point in the other two curves.

5.4.2 Example: National Seats-Votes Curve in 2014

A second example, the national seats-votes curves, is shown in Figure 18. In this case, the bootstrap is the closest to the Gelman-King curve, only coming close to divergence at extreme average national Democratic vote. Both the Gelman-King and Bootstrap methods pick up on a distinct undulation in the seats-votes curve present in the uniform partisan swing curve shown in Fig. 12. That is, responsiveness is lower (the curve flattens slightly) as Democratic vote share approaches .5 — in this case $.44 < \bar{h} < .51$ — and then increases in responsiveness again. This means that, in that range, Democrats tend to win less seats than Republicans for every marginal increase in vote share. After they attain a majority of support, this reverses,

meaning Democrats win more seats than proportionate. The shape of this undulation, common in many seats-votes curves, is critical to many estimates of partisan advantage.

Counter to the other two methods, the McGann seats-votes curve has a nearly uniform responsiveness, around 2.2, over the entire domain. While the magnitude of the seat discrepancy is larger in the national case, it is less than a half of the discrepancy in the California case in terms of the size of the legislature. In addition, the shape of the discrepancy as a function of \bar{h} is different in this case, and is dominated by the undulation near $\bar{h} = .5$. In aggregate, the Bootstrap simulates a more favorable curve for Republicans than the other methods, in that it indicates Republicans tend to win more seats with fewer votes than would be expected if the curve were symmetric. The Gelman-King and McGann do not exhibit a clear directional discrepancy towards or against either party.³³

In terms of the variance profiles shown in Figure 19, the simulated elections tend to be much noisier on average than the observed national 2014 congressional election. Every bootstrapped election had a higher variance than the observed election; 99.2% of the McGann elections and 85.7% of the Gelman-King simulations did as well. This means that the conditional variance reduction of the Gelman-King method is working as intended, since it tends to have lower variance than the other methods. The right plot of district heteroskedasticity is actually very similar between the California and national cases; the swings in California had only a slightly larger deviation than the swings nationally, so bootstrapped district deviations converge quickly to .056. Also like the California case, the districts in the Gelman-king simulations tend to have lower deviation and the McGann simulations have higher district deviation with a long right tail. This long right tail occurs in almost all curves' rank standard deviation plots. This reflects the fact that the vote shares covered by the low- and high-ranked districts is much wider than the width of ranks near the middle of the domain. Notably, the bootstrap has the broadest rank-variance

³³It is important to reinforce this is in terms of the discrepancy in the predicted number of seats Democrats win at a given national vote share, not a *bias* towards one party or another embedded in the model. In fact, all methods provide similar estimates of the bias at median in this case. This may not be the case for other bias measures, though.

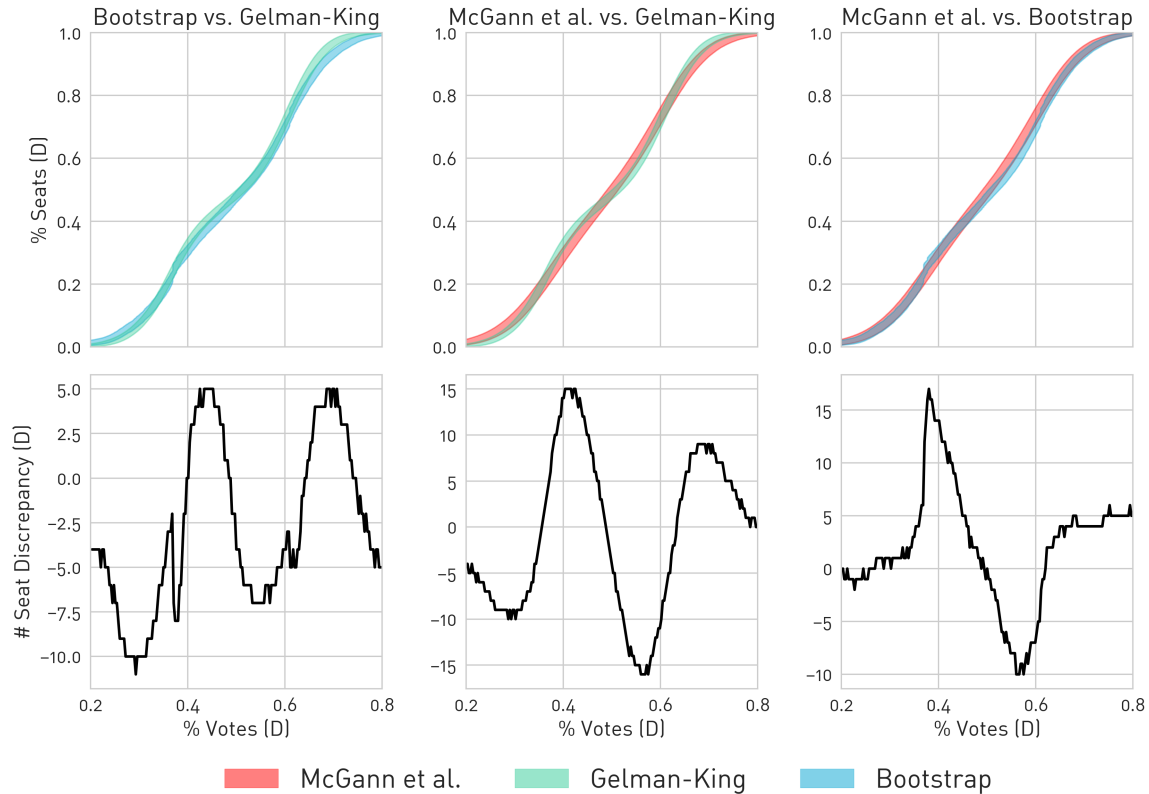


Figure 18. National Congressional seats-votes curve estimates for 2014, with plots of the estimated seats votes curves in the first row and the discrepancy in the expected number of seats won by Democrats given their average vote share at the bottom.

distribution; the Gelman-King and McGann methods both cluster tightly around their average rank deviations, meaning the seats-votes curve estimates tend to have a more consistent width than the bootstrapped curve.

Compared to the difference between the McGann et al. estimate discussed in the previous section, the bootstrap estimates are somewhat closer to the McGann estimates over all quantiles. If, as McGann et al. (2015) suggest, a simpler method of estimating seats-votes curves is useful for post-hoc analysis, bootstrapping the seats-votes curves appears to provide a significantly simpler method that retains the same sense of “swing” used by other authors than the series of regressions suggested by McGann et al. (2016). However, the bootstrap seats-votes method cannot estimate the seats-votes curve for *counterfactual electoral conditions*, since

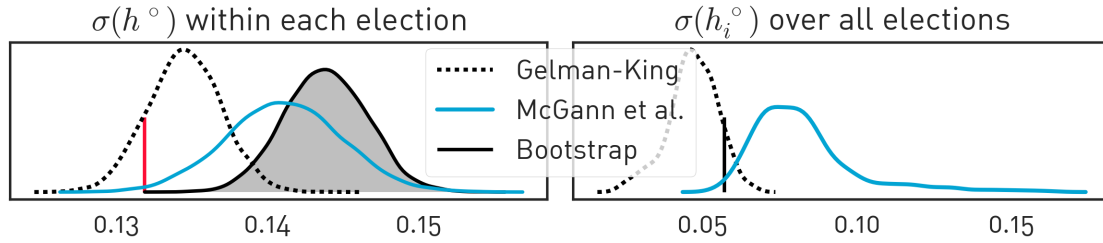


Figure 19. Distributions of variance within and between simulations used to generate Fig. 16. The empirical mean is shown by the vertical line in the left plot. In the right plot, the bootstrap variances are so tightly clustered around .056 that its kernel density is too tall to display on the same scale.

no information about electoral conditions other than outcomes are used in the technique. In that case, McGann et al. (2015)'s technique still may prove a more efficient trade off between simplicity and estimability, as long as the conceptual differences in the operationalization of electoral swing is resolved or considered unimportant by the analyst.

One straightforward extension would be to introduce \mathbf{h}_{t-1} directly into the McGann model, retaining the pooled estimate. The ability to prove predictions or results under counterfactuals is critical for the analysis of candidate redistricting plans, a common use of seats-votes estimation techniques, and the lack of apparent directional bias in either predictive method suggests both may be used, with the McGann method preferred if simplicity is desired.

5.5 Concluding Remarks on Seats-Votes Specifications

Two existing seats-votes modeling frameworks were presented and one novel method discussed. The novel method is based around bootstrapping, and constructs simulated elections by resampling the observed distribution of difference in vote share between years. In general, the bootstrap method reproduced most strongly the Gelman-King model's estimates of the seats-votes curve in the national case study and in California in 2014. Thus, where counterfactual inference or prediction is necessary, we suggest using the Gelman-King model, and where it is not required, using bootstrapping. In general, both of those methods had more ef-

ficient simulation results than the method suggested by McGann et al. (2016). Although the McGann method is marginally simpler to implement than the Gelman-King method, its slightly different semantics and more simple counterfactual generation method make it less desirable for this analysis. However, no existing public implementation of these methods are available, so the bootstrapping implementation (as well as all other implementations of seats-votes modeling frameworks) will be released as free software along with this dissertation.

Further, I note that the discrepancy between the modeling techniques tends to be highest in the range where elections are most competitive, with \bar{h} near .5. This means that most of the observed electoral outcomes, which have an average vote share near .5 in many states, fall within the range where the simple choice of specification may significantly affect the resulting seats-votes curves. In all cases, this means that a rigorous defense of the actual primary model specification, that for vote shares, should be conducted each time the seats-votes curve is constructed. Lastly, the semantics of the model should be rigorously examined; the continuity McGann et al. (2016) argues for is broken by the different model of swing. While they both generate seats votes curves, it is unclear whether swing terms and counterfactuals that depend on them are the same in the McGann specification as in a Gelman-King model.

Chapter 6

ARE GENERALIZED UNIFORM PARTISAN SWINGS SPATIALLY REALISTIC?

The seats-votes curve estimation methods discussed in the previous section all share significant assumptions about how legislative elections operate. Critically, one longstanding assumption in modeling elections focuses on the modeling of congressional elections as stationary spatial processes where many independent electoral contests occur simultaneously. In contrast, many longstanding models of electoral outcomes suggest that swings should be modeled as hierarchically dependent (Stokes, 1965). Recently, spatial multilevel modeling has come to prominence as a method for accounting for the effect of spatial context on voting and partisanship (Levendusky et al., 2008; Gelman et al., 2005; Durch and Stevenson, 2005). Some have even suggested that spatial multilevel techniques are sufficient to predict the bulk of congressional and presidential elections, in addition to constructing simulations that comport well with observed outcomes. This, as some have claimed recently, “solves” the problem of geography and context in electoral modeling (Gelman, 2014). At the very least, multilevel models are now a common general technique for controlling for some degree of spatial heterogeneity and providing within-group dependence for spatial groups (Park et al., 2004; Hersh and Nall, 2015). While this is often not framed explicitly in reasoning about elections as geosocial processes, this evolving standard of practice indicates a gradual embrace of basic spatial reasoning about context sensitivity that was essentially rejected out of hand in earlier times (King, 1996).

Indeed, the introduction of multilevel strategies in redistricting applications and electoral modeling derives from the recognition discussed in Chapter 4 from Gelman and King (1990):

“[m]odeling districts with additional information, such as *spatial correlation or covariates*, if they were available, would probably yield more accurate estimates” (p. 277)

Despite this, models of explicit spatial correlation in election results or inter-year swing are not as well recognized as the importance of controlling for heterogeneity. While the spatial reason-

ing embedded in multilevel models engages with classical quantitative geographic concerns about spatial heterogeneity, its treatment of spatial dependence and potential non-stationarity is not as fluent. Classic spatial multilevel models hardly engage these concerns on their own (Owen et al., 2015).

A pervasive justification of multilevel models in treating spatial heterogeneity is in the case where a spatial fixed effect would be used to treat spatial groups of varying size (Gelman, 2006). By stipulating these spatial fixed effects as having a common underlying distribution, groups with fewer members or noisier groups are shrunk towards the global average effect. In the typical centered multilevel intercept model, this provides group means that are spatially heterogeneous, divided into $j = 1, 2, \dots, J$ groups, with the $J \times 1$ group means α_j distributed normally around a single global intercept μ . Letting an $n \times J$ dummy variable matrix Δ classify n observations into J groups, the typical varying-intercept model for response Y and response-level data X is stated:

$$\begin{aligned} Y &\sim \mathcal{N}(\Delta\alpha_J + X\beta, \sigma^2) \\ \alpha_J &\sim \mathcal{N}(\mu + Z\gamma, \tau^2) \end{aligned} \tag{6.1}$$

where σ^2 is the variance of the response-level component, τ^2 is the variance of the group-level error component, Z is an group-level design matrix, γ are the substantive effects unique to the upper-level, and β are the substantive effects unique to the response level. This results in a “spatially-varying intercept” model, where all observations i in unit j have a common group effect. Stated in a single line, this becomes:

$$Y \sim \mathcal{N}(\mu + \Delta Z\gamma + X\beta, \sigma^2 + \Delta\Delta'\tau^2) \tag{6.2}$$

Notably, the covariance matrix of this model is non-diagonal: $\Delta\Delta'$ produces within-group correlation due to its off-diagonal elements. This results in the response being correlated within groups, but independent across groups.

No formal spatial specification test is available for these kinds of models. But, introducing a state-level hierarchical component into the Gelman-King national seats-votes model for 2016

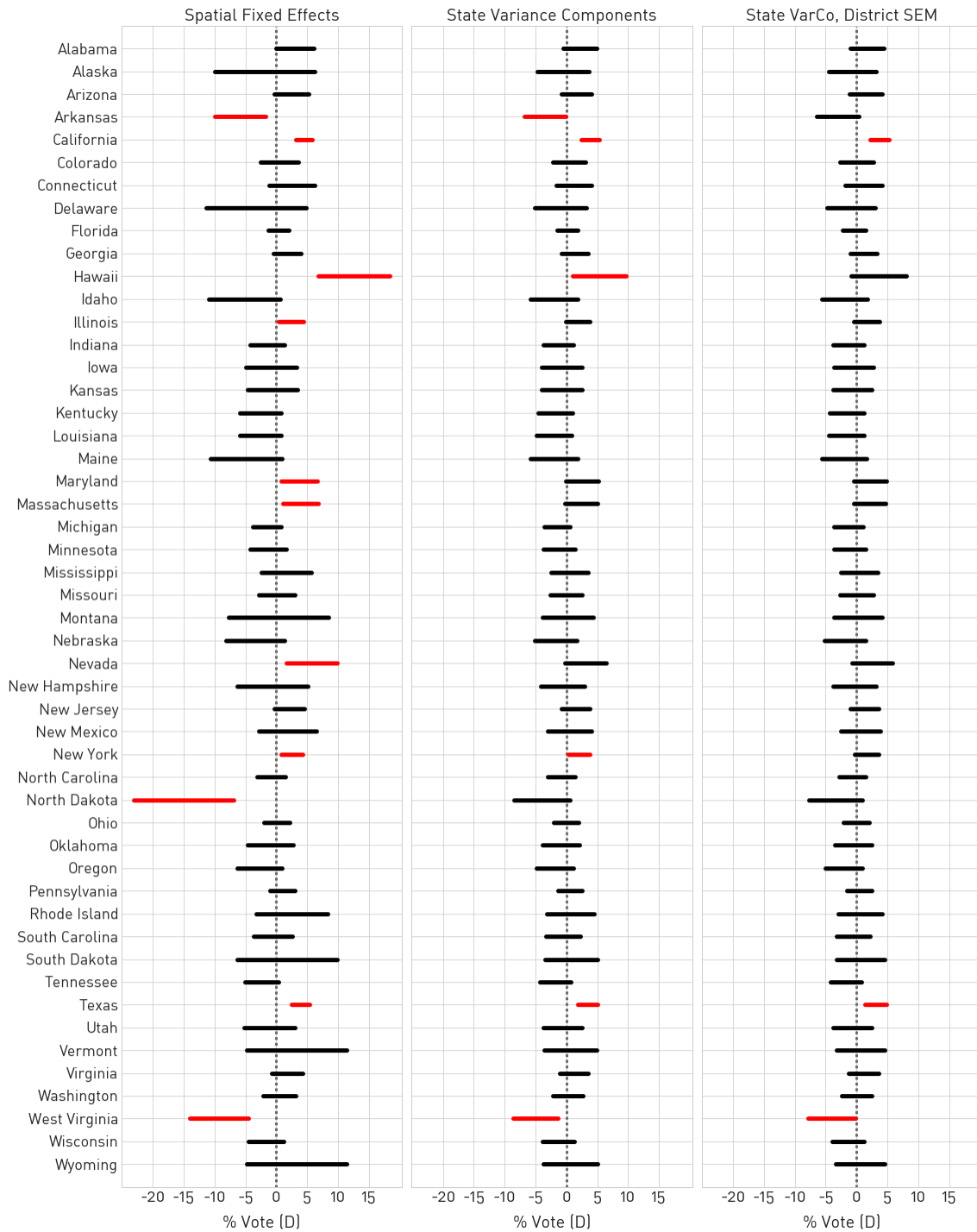


Figure 20. Shrinkage plot showing the band of the fixed effect estimates $\pm 2\sigma$. Effects that overlap with zero are colored black, and effects that do not are colored red. Each column is one different specification to control for spatial heterogeneity.

indicates that most states do not have significantly different distributions of swing, conditional on the rest of the model. This is exemplified in the shrinkage plot in Figure 20, which compares three different models for the 2016 congressional votes shares using incumbency, the 2014 congressional vote shares, and state-level effects. Contrasted in the figure are three specifications; a spatial fixed effect specification which simply includes each state, an uncorrelated state variance components model in the vein of Browne et al. (2006); Leckie et al. (2014), and state *spatial autoregressive* variance components model (Wolf, 2016). What becomes apparent is that, even in the spatial fixed effect model which tends to exaggerate the magnitude of difference between subgroups (Gelman, 2006), most states are not substantially different from the common mean. When *aspatial* hierarchical shrinkage is introduced, more states become indistinct. Finally, when allowing for the possibility that district vote shares might follow a spatial autoregressive error model with a hierarchical state effect, states become even less distinct.³⁴ In addition, these substantially more complicated models do not yield significantly better predictions for \mathbf{h}_{2016} , so their utility in this context is dubious. Finally, in state-specific analyses, these hierarchical components are unavailable anyway, so their introduction is irrelevant for most deployments of seats-votes models. Unless a national (or super-state) structure is retained, stratifying the model using state-level hierarchical effects is useless for processes that occur *within the state*.

While this form of model is useful, many recent explorations of spatial structure in electoral geography focus instead on the prospect of spatial *nonstationarity* in the electoral process. Often, this is done by stating a model with hierarchical *substantive* effects, in addition to or instead of hierarchical intercepts. In this case, some number of covariates, $k = 1, 2, \dots, p$, are separated into l “local” covariates and g global covariates. Then, a model is specified with:

$$Y \sim \mathcal{N}\left(\alpha + \sum_k^l X_k \Delta \beta_{k,J} + X_g \beta_g, \sigma^2\right) \quad (6.3)$$

³⁴This is somewhat unsurprising, as the prospect of spatial autocorrelation smooths the boundaries between groups that multilevel modeling aims to leverage, regardless of the level at which the correlation process is admitted.

Then, each $\beta_{k,J}$ has :

$$\beta_{k,J} \sim \mathcal{N}(\mu + Z_k \gamma_k, \Sigma_{\beta_k}) \quad (6.4)$$

where Σ_{β_k} is often shared over all k processes. This model structure generates β -nonstationarity, since the estimates of β are unique to the group. This modeling structure has been used to explore spatial non-stationarity in the electoral impacts of income (Gelman et al., 2005), race (Hersh and Nall, 2015), and partisanship (Levendusky et al., 2008).

Spatially-nonstationary models with endogenous scales are also used in the analysis of elections and voting behavior. In an attempt to handle the complex interplay between heterogeneity, dependence, and “alchemy” at issue in the ecological inference problem (Anselin, 2000; Anselin and Cho, 2002), geographically weighted regression (GWR) techniques have been used (Calvo and Escolar, 2003; Wing and Walker, 2010). Recent work using GWR to explore smooth nonstationarity in political behavior or electoral dynamics is also promising (Crespin et al., 2011; Clemens et al., 2015), since the formal treatment of nonstationarity by nested “geo-in-geo” hierarchical is limited by the extreme heterogeneity in basic spatial enumeration units available for social scientific research.

Centrally, the pervasive use of exogenous-boundary spatial multilevel models provides for spatial dependence in a limited group-wise sense. In these cases, inter-observation correlation is a matter of membership, not proximity or relation (Owen et al., 2015). Heterogeneity is discrete, bounded, and its spatial structure is known a priori. Some marginal theoretical justifications for these strategies exist. Elections law and candidate contests are bound within states. In addition, group-wise dependence in electoral outcomes may result from national- and state-level coattails effects (Hogan, 2005, e.g.). However, state-specific coattails are quite weak, and both coattails effects can be controlled for with dedicated covariates. Regardless, the formal justification of multilevel structures in terms of substantive arguments about voting behavior and electoral geography is not routine.

In contrast, spatial dependence in electoral outcomes, as well as spatial dependence in inter-year electoral swing is empirically pervasive. To this end, explicit models of spatial dependence have been employed in electoral analysis and voter behavior. Darmofal (2006) pro-

vides a thorough discussion of spatial econometric praxis for questions in political science, and Franzese Jr and Hays (2008, 2007) provide applications and review of the opportunities of explicit spatial reasoning for models in comparative politics. In an analysis of political participation, Cho and Rudolph (2008) identifies endogenous feedback between neighboring areas while using a spatial group effect strategy to control for potential contextual impacts. In a similar vein, Burnett and Lacombe (2012) conducts an econometric specification search for a demographic model of vote choice at the county level, finding significant improvements when dependence between counties is modeled with a local, small-scale spillover process embedded in a spatial error or spatial Durbin error model.

In addition, Monogan (2013) analyzes the structure of immigration policy adoption in US States with conditional autoregressive effects, and suggests these may provide further effective tools to analyze policy contagion (Monogan, 2012). With the advent of various multilevel model specifications that incorporate simultaneous autoregressive spatial dependence and hierarchical heterogeneity (Lacombe and McIntyre, 2016; Dong and Harris, 2015; Wolf, 2016), multilevel group-wise dependence and neighborly spatial dependence — implicated by spatial Markov random fields (Besag, 1974) or simultaneous autoregressive processes (Anselin, 1988) — can be modeled together. In fact, this has already been applied in the modeling of recent British elections (Lacombe et al., 2014). Thus, the analysis of the impact of spatial dependence on estimates of seats-votes curves, either through dependence in vote shares themselves or through dependence in swing from year to year, may improve the validity of elections simulated in seats-votes curve modeling.

6.1 Spatial Misspecification in Vote Share & Swing Models

However, these various specification changes largely do not buy any significant benefit in the context of seats-votes modeling, since the rank distribution often remains the same. In spite of this, the prospect of spatial correlation in swing or vote shares is both empirically justified and theoretically interesting. If swing were correlated in space, then we might observe “pockets”

of volatility, areas that all shift together towards or against the global trend. In addition, the implicit relationship between last years' vote and this years' vote may be spatially-nonstationary, justifying the use of local models. However, since no specification testing regime is available yet to determine whether GWR is appropriate, this comparison is elided for a future project.

There are two clear ways to determine whether spatial dependence is a necessary concern in models of \mathbf{h} or δ° . The first method is to examine the regression residuals for potential spatial structure using classic tests for spatial misspecification. The second would be to examine whether the introduction of a correlated model for \mathbf{h} or \mathbf{ffi} affects the rank distribution.³⁵ In Figure 21, the robust Lagrange Multiplier statistics to identify potential spatial misspecification in model residuals (Anselin et al., 1996) are shown for each two-cycle Gelman-King model specification in the national vote share model considered in Section 5.4. For assuredness, each two-cycle or one-cycle model since 1992 is estimated under two different methods of resolving uncontested elections. On the left, only the districts that are contested in both t and $t - 1$ for the two-cycle models or are contested in election t for the single-cycle models are retained. On the right, uncontested elections are imputed using the first-order autoregressive sub-regressive strategy detailed in Section 6.2. Significant robust LM statistics are shown in red, and non-significant statistics are shown in black. If each two-cycle model is considered independently, almost all models relating only the mutually-contested elections in each cycle exhibit a significant robust LM test. This means that there is statistically significant spatial correlation in the error term, even when admitting the potential for an endogenous spatial lag term for the response. These statistics are large, and many also are significant under a conservative Bonferroni correction. In contrast, the robust spatial lag statistic is not significant as consistently, with only one significant statistic, that for contested elections in 2012. Thus, it is likely that the direct model for vote shares is spatially misspecified, and could benefit from some type of correlated error correction in the model for \mathbf{h} .

³⁵This exploration could be directly conducted with Geographically-Weighted Regression, even though it does not have recourse to formal specification testing.

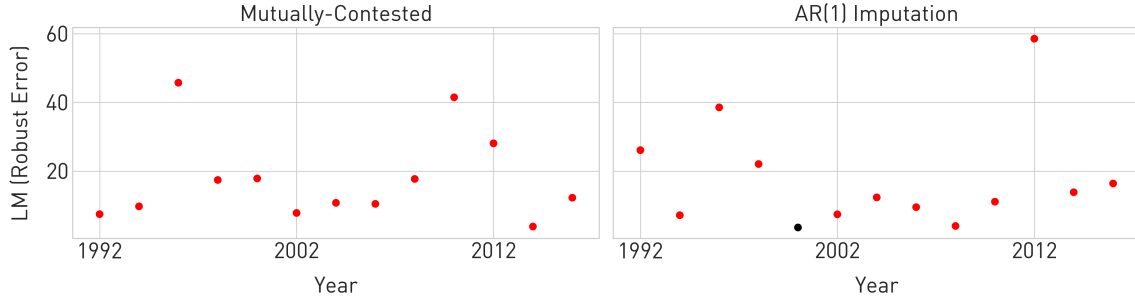


Figure 21. Robust LM Error tests for national Gelman-King models since 1992. On the left, the tests apply to a Gelman-King model with Democratic vote share censored to .25 and .75. On the right, the diagnostics are computed for the AR(1) imputation suggested above.

6.1.1 Generating Spatially-Correlated Electoral Swing

Critically, introducing spatial patterning into δ° or h° opens up many potential model specifications. Correlation in either process can be specified independently. In theory, swing and expected typical vote may differ in the strength of spatial patterning, and different models may be required for either process. Here, I consider specifications with only a single autoregressive term, not allowing for both correlated \mathbf{h} and δ .

First, a spatial lag model could be used for \mathbf{h} , which would result in spatial patterning in both \mathbf{h}° and δ° :

$$\mathbf{h}^\circ = (I - \rho W)^{-1} \hat{\mathbf{h}}^\circ + (I - \rho W)^{-1} \delta^\circ \quad (6.5)$$

This could be achieved within a Gelman-King specification by estimating each submodel using the lag specification and then treating the model artifacts in the same manner. Alternatively, a model with endogenous lag *only* for the swing process can be stated:

$$\mathbf{h}^\circ = \hat{\mathbf{h}}^\circ + (I - \rho W)^{-1} \delta^\circ \quad (6.6)$$

In the McGann model, this would correspond to a spatial error model for the panel of congressional districts. For the Gelman model, this requires specifying a correlation structure for δ° at the end of the modeling process.

Then, a spatial error process for δ° may be useful. In that, the process becomes:

$$\mathbf{h}^\circ = \hat{\mathbf{h}}^\circ + \mu + (I - \rho W)^{-1} \epsilon \quad (6.7)$$

where δ° has now been broken into its constituent parts, μ , the mean swing, and ϵ , the random fluctuations around the mean swing. Note that, in this specification, the expected swing is still μ , a useful property that will be discussed later.

Finally, the variance in Eq. 6.7 should be critically examined. Critically, if ϵ has variance τ^2 , the *apparent* variance of each observation will be *larger* than τ^2 if the autoregressive specification in 6.7 is used. To ensure that spatial correlation exists without inflating the variance of each district's outcomes, one can recast the specification in terms of the correlation matrix, R . First, note that the covariance matrix of δ° under Eq. 6.7 is:

$$\text{cov}(\delta^\circ) = [(I - \rho W)'(I - \rho W)]^{-1} \tau^2 = (F'F)^{-1} \tau^2 \quad (6.8)$$

where τ^2 is simply σ^2 for the McGann model and $(1 - \lambda^2)\sigma^2$ for the Gelman-King specification. In general, the diagonal of $(F'F)^{-1}$ is *greater* than one; this means the apparent simulation variance of δ_i° for any district i will appear *greater* than the variance intended by τ^2 .

To prevent this, a separation strategy inspired by Barnard et al. (2000) is used. This strategy separates the implied correlation matrix from $(F'F)^{-1}$ and the variance parameters intended for the distribution. To do this, let $\text{diag}(M)$ be the square matrix containing the diagonal of M and zeros elsewhere. Then, the correlation matrix R corresponding to $(F'F)^{-1}$ is available using the familiar formula:

$$R = \text{diag}\left((F'F)^{-1}\right)^{-\frac{1}{2}} (F'F)^{-1} \text{diag}\left((F'F)^{-1}\right)^{-\frac{1}{2}} \quad (6.9)$$

The Cholesky decomposition of R can then be used to draw random variates with correlation governed by ρ and variance governed *strictly* by σ^2 . Letting the Cholesky factorization be:

$$R = C_R C_R' \quad (6.10)$$

then,

$$\delta^\circ = \mu + C_R \epsilon \quad (6.11)$$

means that, marginalizing a single district i over many simulations:

$$\text{var}(\delta_i^\circ) = \text{var}(\epsilon) = \tau^2 \quad (6.12)$$

Whereas, without this standardization,

$$\text{var}(\delta_i^\circ) = \text{var}(\epsilon) * (F'F)_{i,i}^{-1} = \tau^2 * (F'F)_{i,i}^{-1} \quad (6.13)$$

with $(F'F)_{i,i}^{-1} > 1$. This apparent variance inflation is important for data generating processes in simulation studies where control of variance is required for precise simulation design, but is not a generally-applicable model specification due to its complexity.

Fortunately, for all of these specifications, the introduction of any spatial correlation does not affect the relationship between the rank-vote and seats-votes curves. First, for the equivalence to hold, it must be the case that the fraction of seats won at swing μ is also the fraction of seats won under no swing, plus the seats where a swing of size μ flips the seat. Since μ is the same for all districts, it does not affect the rank order of \mathbf{h}° . So, the rank of the marginal district at $\bar{h}^\circ + \mu$ is the same as the rank of the district of $\bar{h}^\circ + 0$ nearest to the left side of the level set of $\bar{h}^\circ + \mu$.

In Eqs. 6.6 & 6.5, the net swing applied to all districts is not μ , but $(I - \rho W)^{-1}\mu$, for some value of ρ and mean swing μ . While the expected apparent swing due to this term depends on both μ and ρ , the resulting mean is *still constant over districts* since μ is scalar. So, the marginal district under swing μ is still the district under no swing that sits below $\bar{h} + [(I - \rho W^{-1})\mu]$.

However, this does affect the expected value of \bar{h}° depending on ρ (holding all else constant) so any simulation regime that requires \bar{h}° to be fixed precisely must account for this discrepancy. If $E[\delta^\circ]$ is not strictly constrained in models with endogenous \mathbf{h} or δ° , the responsiveness of the seats-votes curve necessarily changes. The discrepancy between the mean swing parameter, μ , and the expected swing, $E[\delta^\circ]$ drives this change; if $E[\delta^\circ|\mu] > \mu$ for some ρ , then the apparent swing at that value of μ will be larger than the specified μ , and thus $E[\bar{s}^\circ|\mathbf{h}^\circ]$ under that mean swing will be larger than if $\rho = 0$ if any districts fall in the gap $\bar{h} + \mu$ and $\bar{h} + E[\delta^\circ|\mu]$. This reverses when $E[\delta^\circ|\mu] < \mu$.

Whether or not this should be “controlled” (or, indeed, whether the variance inflation and heteroskedasticity induced by the error autoregressive specification) are largely questions of experimental design, and are subjective in terms of what the analyst aims to discover. Plainly speaking, I am interested in whether modeling places as “swinging together” or “swinging apart” changes the substantive interpretation of the resulting seats-votes conclusions. Thus, I simply admit outright that uncorrected endogenous \mathbf{h} or δ° necessarily increases responsiveness due to its inflation of $E[\delta^\circ|\mu]$, since I view this as a constraint of the specification rather than a substantively interesting trait. Instead, if effects show up in the moment-corrected specifications, where $E[\delta^\circ|\mu] = \mu$ always and $var(\delta^\circ) = var(\epsilon) = \tau^2$ always, then the substantive question might be evaluated: “does *spatial dependence* matter?” — not induced mean non-stationarity, heteroskedasticity, or excess variance.

6.1.2 Correlated \mathbf{h} or δ° in Seats-Votes Model Specifications

Fortunately for the status quo of electoral modeling, the simulation of seats-votes curves under these autoregressive/correlated specifications yields nearly identical seats-votes curves as those that assume votes or swing is spatially independent. Thus, while there may be potential spatial misspecification in the Gelman-King model (or many other models of vote shares at the legislative level), the impact this misspecification has on the estimated *seats-votes* curve and eventual quantities of interest appears to be quite slight.

Recalling the robust LM test results from Figure 21, most of the two-cycle models clearly had significant correlation in their residuals. Thus, the 2014 election was selected for a close, intensive simulation study. Across the specifications considered in Section 6.1.1, 10,000 realizations of \mathbf{h}° were constructed for the previous example’s national seats-votes model for 2014. Thus, in Figure 22, the distribution of simulated Democratic vote shares is plotted with respect to the autoregressive parameter. As you move down the histogram matrix, the autoregressive parameter increases, moving from modeling strong negative spatial autoregressive processes to strong positive spatial autoregressive processes. At almost all levels in any of the specifica-

tions, the distributions of simulated results, shown in grey pooled over all simulations, is very close to the observed distribution, shown in red. Except for the case where \mathbf{h} is modeled in a mixed-regressive, spatial autoregressive specification, the distributions are nearly identical across the whole range of the spatial autoregressive parameter.

Since the distributions of \mathbf{h}° tend to appear to be the same as the distributions of observed \mathbf{h} , the rank-vote distributions (and thus the seat-vote distributions) are likely to be the same as well. To examine, this, seats-votes curves are constructed for these simulated elections under the spatially-correlated data generating processes. Then, a similar style of discrepancy-plot as that shown in Figures 16 & 18 is made for each scenario. That is, over the entire range of party vote share, the difference in the expected number of seats is computed. Thus, if the curves for the spatial data generating processes are significantly different over a set of \bar{h}° values than that under the null, the discrepancy plot will provide both the extent and the magnitude at which these plots differ. In addition, we can compare the overlap of quantiles in each level like done between the three different types of seats-votes curve estimation strategies.

The raw discrepancy plot is shown in Figure 23. Each row of the plot corresponds to a simulation batch where the spatial autoregressive parameter is fixed to the same value across specifications. The columns contain each of the four different specifications. The inner y axis labels correspond to the raw number of seats difference between the seats-votes curves estimated from the null process and the spatial autoregressive process in that column. Thus, a negative value indicates that the null seats-votes curve is below the spatial autoregressive seats-votes curve at that level of \bar{h}° by that many seats. A positive value indicates the null is above the spatial autoregressive estimate. Thus, regions of \bar{h}° where discrepancy is large would be regions where the resulting estimates of partisan advantage would differ the most.

As is apparent from the distributional plots, the data generating specification with an endogenous lag for the \mathbf{h} vector is most distinct from the null of no spatial autoregressive effect in any component. In general, this forces the mass of the vote distribution to the tails, which reduces the total number of districts that can be flipped by the parties. Effectively, as the strength of

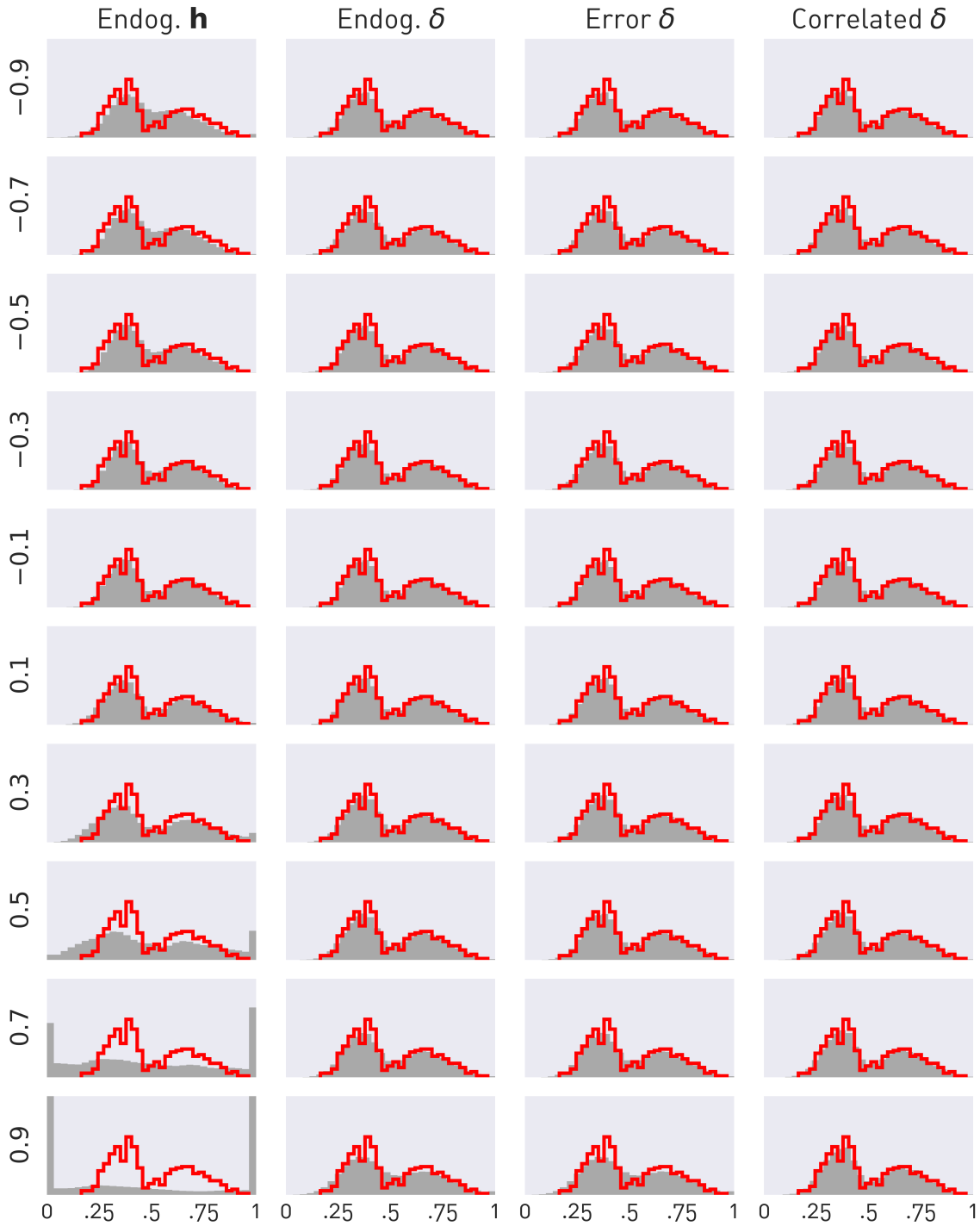


Figure 22. Distributions of vote shares for various process specifications over a range of autoregressive effect size. The autoregressive effect for 2,000 realizations is shown on the y -axis. The grey distribution is the distribution under the simulated data generating process, and the red distribution is the observed vote share distribution for 2014.

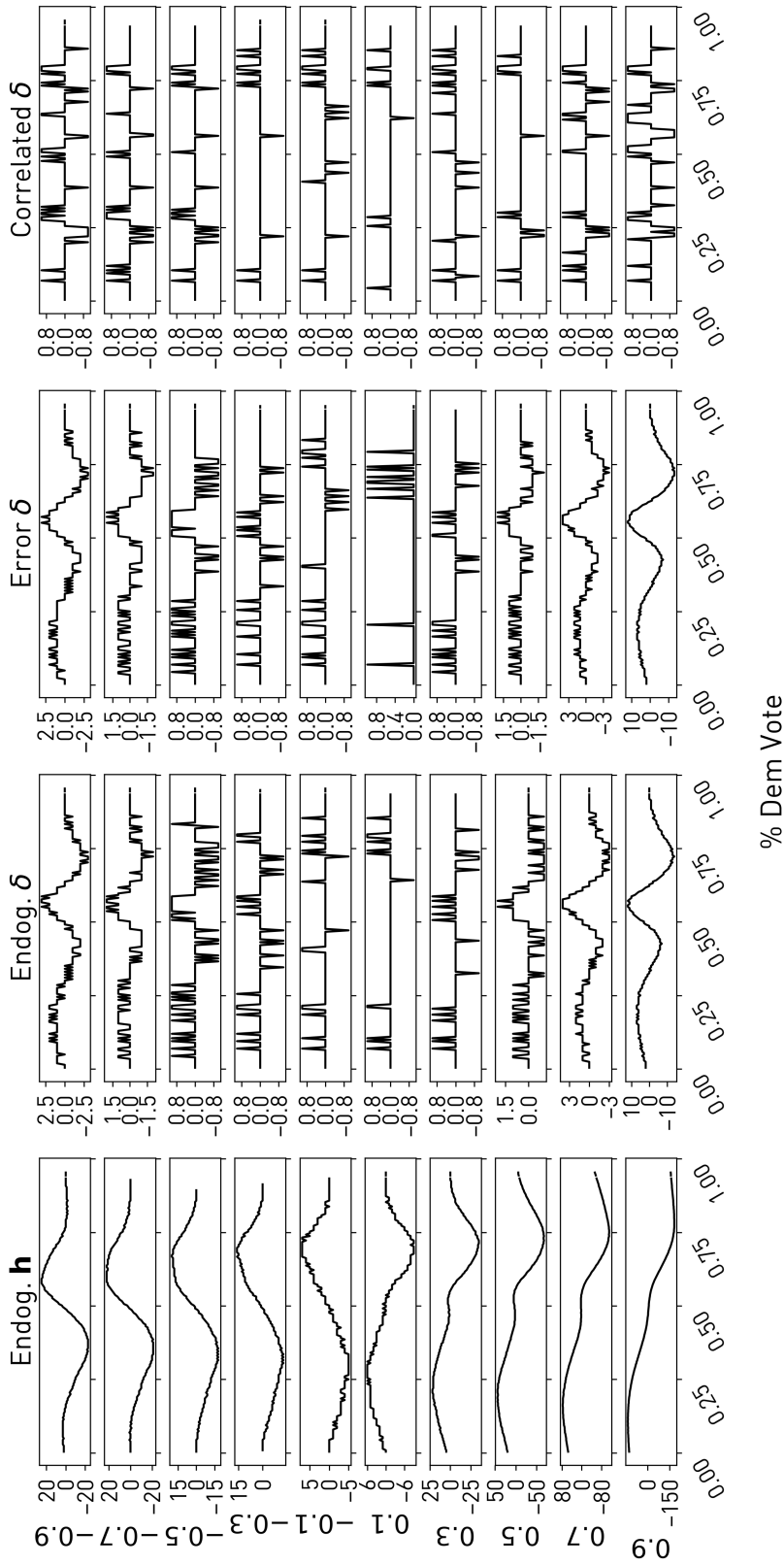


Figure 23. Discrepancy plots for the national congressional seats-votes curves in 2014, estimated for elections of varying spatial data generating processes and autoregressive effect strengths. Each column expresses one of the four spatial autoregressive specifications for \mathbf{h}° or δ° . The outer y label is the value of ρ for the data generating process. The inner y label is the difference in the expected number of seats Democrats win at the level of the popular vote on the x axis. Negative values indicate the null DGP predicts fewer seats at that level of vote than the spatial autoregressive DGP, and positive values indicate that the null assigns more seats than the spatial autoregressive DGP.

autoregressive effect in the endogenous lag of \mathbf{h} model becomes large, districts either become wholly Republican or wholly Democrat, and the rank distribution flattens significantly. Many ties at 1 and 435 result. This causes the seats-votes curve to flatten significantly, since fewer districts are in the center of the \bar{h}° range, so the change in y must be smaller, too. Thus, even at mild levels of an induced endogenous spatial autoregressive effect in \mathbf{h}° , the maximum seat discrepancy is around 6 seats. For seats that are “safe Democrat” on the left of the seats-votes curve (where $\bar{h} \approx .25$), positive ρ results in the null assigning *more* seats to Democrats than would be assigned under the spatial autoregressive simulations. In addition, for the mode of seats that are “safe Republican” on the right of the seats-votes curve (where $\bar{h} \approx .75$), positive spatial autoregressive effects in the endogenous \mathbf{h}° specification results in the null assigning *fewer* seats to the Democrats than the autoregressive process. That is, if \mathbf{h}° is modeled even with a weak spatial autoregressive effect, the seats-votes curve will be “flatter” than the null curve if the spatial autoregressive effect is positive, and “steeper” than the null curve if the spatial autoregressive effect is negative.

Regardless, the introduction of a non-zero spatial autoregressive effect in this specification makes the seats-votes curve more linear, since the bimodality is flattened. Thus, if vote shares were endogenously correlated in this manner, even minor changes in the way \mathbf{h} were modeled would affect anything depending on the seats-votes model. However, tests for model misspecification suggested instead that the specification lay in the error term, δ , so (strictly speaking), this sensitivity is inapplicable here.

The discrepancy in all simulations tends to be lowest around the median, $\bar{h}^\circ = .5$, so bias measures that focus on the median will be robust to this discrepancy. This means that measures like the median bonus & attainment gap discussed in Section 2.3 should be more robust to dependence misspecification. In contrast, a measure like Observed Bonus, which evaluates the curve at \bar{h}° and $1 - \bar{h}^\circ$ with $\bar{h}^\circ = \bar{h}_t$, may catch this discrepancy at its height. And, the slight asymmetry in the discrepancy curve about .5 means that the under-prediction on the left likely will not be balanced by the over-prediction on the right side of the median.

However, for all other specifications, the discrepancy in expected number of seats won by

Democrats is marginal for all but the most extreme values of ρ . However, for the cases where the shape of the discrepancy curve can be identified, it tends to remain the same over the varying levels of autoregressive effect strength. Thus, when $\rho = -.9$ or when $\rho = .7$, the resulting discrepancy between either the endogenous lag specification for the swing term δ° or the nuisance dependence specification for δ° both result in similarly-shaped discrepancy curves with similar magnitudes. In these scenarios (and all scenarios with ρ between these values) the estimated discrepancy between the null is with ± 3 seats. The form holds approximately the same for both specifications, with only the amplitude of the discrepancy curve affected by the value of ρ . Practically speaking, this means that most commonly-encountered values of ρ , between 0 and .7 or so, result in a maximum discrepancy of three seats between the null and the autoregressive model. This discrepancy is quite small with respect to the total 435 seats being assigned, and would not affect the resulting advantage estimates in a noticeable way. In addition, none of the results for the correlated δ specification, which keeps the apparent variance in simulation fixed to σ regardless of the value of ρ , differ from the null simulation model by more than one seat.

The extent to which the change in covariance specification affects the width/height of quantiles or the magnitude of the variance for level/vertical sets of the seats-votes curve is also relevant. The width/height of the seats-votes curve contributes directly to the variance of relevant advantage measures. Thus, if the autoregressive specifications change the dispersion of the curve across its domain, it may affect the dispersion of the estimated advantage statistics. Thus, the relative difference in standard deviation of \bar{s}° at each \bar{h}° is reported Figure 24, with the formula:

$$\frac{\sigma(\bar{s}_0^\circ) - \sigma(\bar{s}_A^\circ)}{\sigma(\bar{s}_0^\circ)}$$

where k_0 pertains to the statistic k from the null process. Each set of \bar{s}_0° is constructed from realizations that fall within a small search band around \bar{h}° , like that discussed in detail of the Texas example from Figure 15. While it does appear that the change in variance is non-random. For the endogenous lag of \mathbf{h} specification, the variance tends to inflate sharply around $\bar{h}^\circ = .5$, with variance either deflating or staying nearly the same when outside of part of the vote

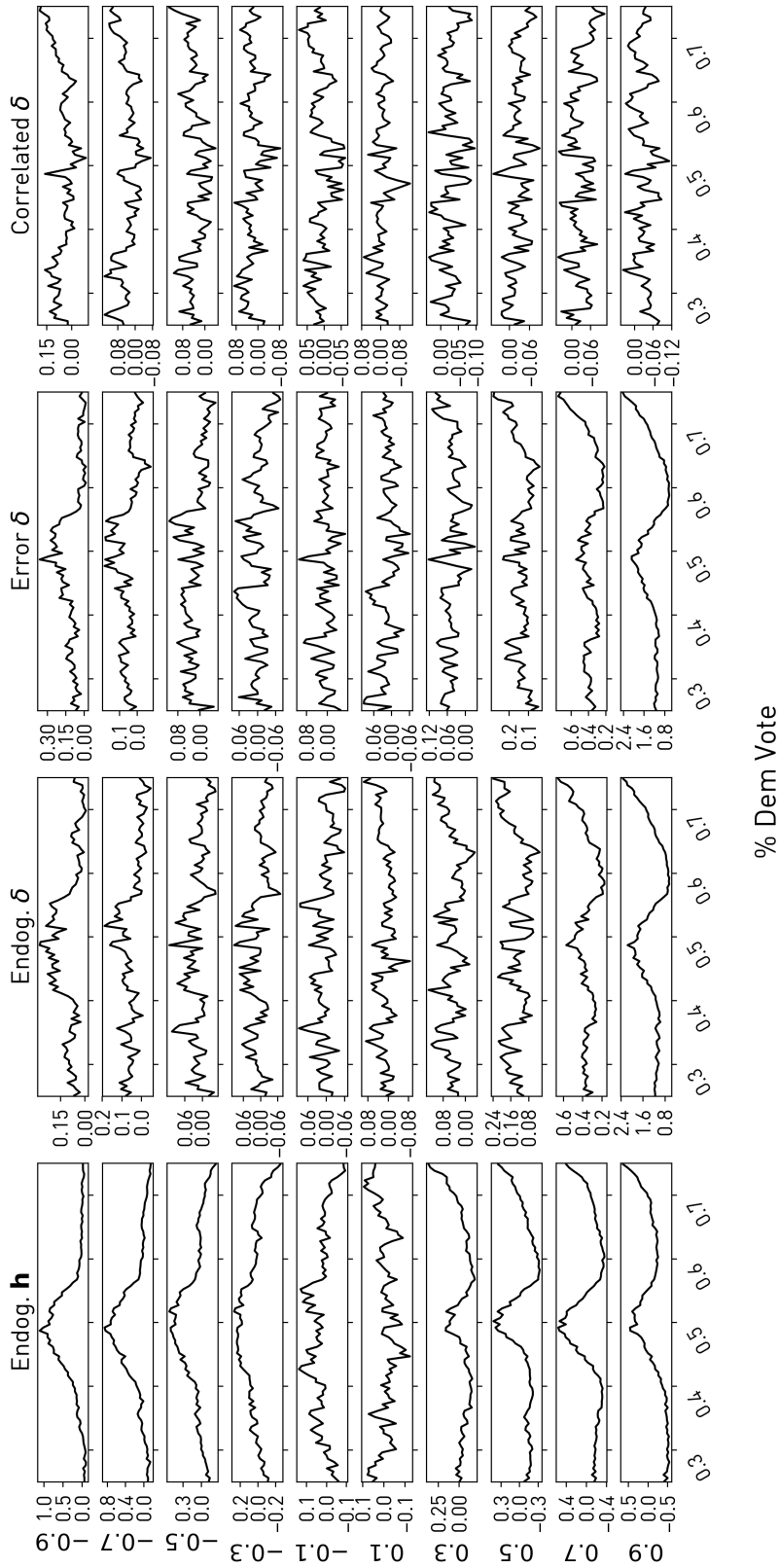


Figure 24. Deviation discrepancy plots for the national congressional seats-votes curves in 2014, estimated for elections of varying spatial data generating processes and autoregressive effect strengths. Each column expresses one of the four spatial autoregressive specifications for h° or δ° . The outer y label is the value of ρ for the data generating process. The inner y label is the relative percent difference in the standard deviation of seat shares at the level of vote shown in the x axis. Positive values indicate that the middle percentiles of the seats-votes simulation envelope are taller (i.e. more dispersed) under the spatial autoregressive DGP than under the null; negative values indicate a tighter distribution when spatially-autoregressive than under the null.

share distribution. Semantically, this means that the endogenous \mathbf{h} specification increases the range of ranks possible for a district that Democrats might barely win, meaning that that district is much less likely to be pivotal (or nearly-pivotal) to a legislative majority, since its rank is much more variable. This same general behavior occurs for the endogenous lag specification for δ and for the nuisance spatial autoregressive specification for δ , albeit only manifesting clearly when the autoregressive effect is very large. Otherwise, when ρ is not too large in these specifications, the autoregressive specification tends to inflate the vertical width of the seats-votes curve by under 10%, and also occasionally narrows the vertical width. In contrast, the model with spatially-correlated δ that fixes the apparent variance tends to deflate the variance with ρ is large and positive, and inflate the variance when ρ is large and negative. For most of the realizations, though, the correlated model fluctuates around inflating/deflating the variance by under 6%, a quite marginal change in vertical width.

6.2 Validating the Imputation Model

Uncontested elections pose a significant problem to post-hoc estimation of the seats-votes curve and prediction. Critically, they cannot be treated simply as outliers, since they represent substantively interesting parts of the process and contribute to any hypothetical seats-votes curve. Dropping uncontested elections entirely may also bias estimates of partisan advantage (Niemi and Fett, 1986; Tufte, 1973; Niemi and Deegan, 1978), since these districts are not counted as “won” by any party.

In the dataset constructed for this dissertation from Kollman et al. (2016) and King (1994), “truly” uncontested elections, where the two-party vote share is exactly zero or one, comprise around 14% of all elections between 2016 and 1992. The distribution of these vote shares is shown on the left of Figure 25. In addition, around a quarter of elections are “nominally contested,” with candidates winning by more than 75% of the vote. This can be seen on the left side of Fig. 25. Since 1992, republicans nominally contest many Democrat-held districts, where $.75 < h_i < 1$, but few Republican districts are nominally contested, with Democratic vote

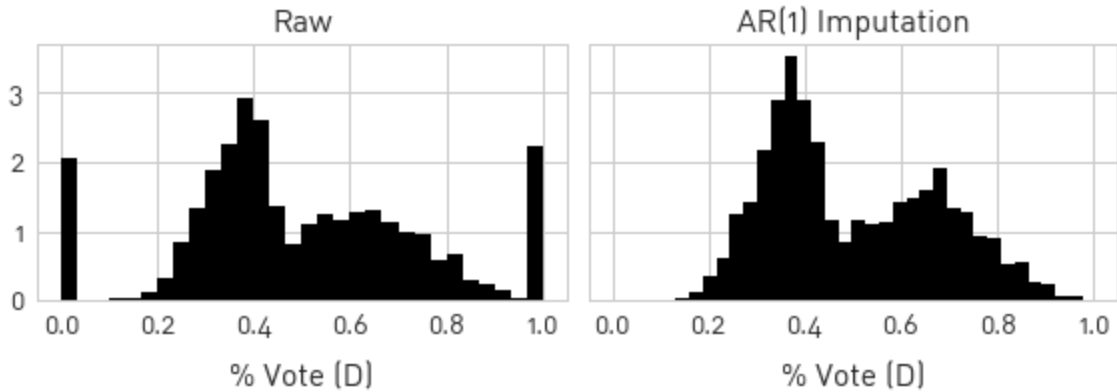


Figure 25. Distribution of Congressional district vote share for Democrats, pooled since 1992.

shares below $0 < h_i < .25$. The number of “fully” uncontested districts by the two parties is roughly equal, though.

Uncontested elections are too common to ignore outright, so many techniques to handle uncontested elections exist. One method common in seats-votes analysis is to estimate the vote share for uncontested districts as if they were contested. Often, this is done using an auxiliary regression or single imputation. McGann et al. (2016) use presidential election results at the congressional district level to predict the baseline partisanship of a district in a panel. This is motivated by the assumption that split ticket voting is rare, so congressional margins should appear similar to presidential margins if all seats were contested. This strategy is a useful, simple method that yields reasonably accurate results, as will be shown below. However, this data augmentation approach is less useful than one that imputes directly from the available data.

6.2.1 Common Imputation Strategies

To impute uncontested elections directly from the available data, Gelman and King (1994a) use a single-imputation strategy under a first-order temporal autoregressive model. Uncontested districts in a given year are imputed from the model fit to the available data in that year.

	Two	Three	Four	Five
1990s	43	11	6	0
2000s	48	13	7	3
2010s	35	6	0	0

Table 7. Count of congressional districts with uncontested runs of the target length in each decade.

Since the Gelman-King specification relies on h_{t-1} to predict t over a sequence of two-cycle models, this imputation can take two forms in practice. One form uses only *mutually-contested* districts, those that are contested in both both the current and the previous cycle. For the mutual strategy, a model for election t is fit using districts that were contested in both t and $t - 1$, meaning both h_t and h_{t-1} are informative. Then, uncontested districts in t are predicted as out-of-sample cases. However, this approach loses efficiency with serially contested elections. If an election has been uncontested for two periods, then the previous vote share does not reflect the characteristics of the district when contested.

To remedy this, a recursive propagation strategy is available. This recognizes that an imputation for t may require imputation for $t - 1$, which itself may require imputation for $t - 2$, and so on. Thus, imputed values are propagated forward and used in the next time period. First, an imputation is conducted for the first available time period using all available data. Then, the next time period uses the full vector of h_{t-1} , both imputed and observed, to fit a model for h_t . Since all h_{t-1} are either contested or imputed, all uncontested districts in h_t have an informative h_{t-1} value.

Serially-uncontested districts comprise a small portion of the dataset, as shown in 7. Out of 5616 general congressional election contests in the contiguous US since 1992, 126 have been uncontested by a major party for at least 2 sequential elections. This comprises just shy of 2% of the total. In addition, this fraction tends to be relatively constant over the decades since 1992. In the redistricting decade from 1992 to 2000, 43 districts went uncontested for two sequential elections, again around 2%. In the period from 2002 to 2010, 48 went uncontested for two sequential elections. Over the three Congressional elections since 2012, 35 districts

have been uncontested in two sequential elections, around 2.5%. Notably, three districts were contested by only one of the major parties during the 2000s redistricting decade. Alabama's 6th district, represented by Spencer Bacchus, was never contested by Democrats; Florida's 17th district, represented by Kendrick Meek, and Massachusetts's 8th district, represented by Mike Capuano, were never contested by Republicans in that decade.

6.2.2 Presidential Imputation is Slightly Superior

To compare the quality of the various imputation strategies, a k -fold crossvalidation was conducted, and the results are shown below. In a set of test years, a fifth of all available \mathbf{h}_t was sampled, censored to appear uncontested, and recovered using one of the three methods discussed above; imputation from presidential results, mutually-contested imputation, and recursive forward propagation. Since recursive imputation may be sensitive to the sequence of \mathbf{h}_t drawn in the k -fold crossvalidation, 2000 replications of 5-fold crossvalidation were conducted, resulting in around 10,000 imputation passes for congresses since 1992.

For the mutually-contested and recursive forward propagation strategies, cases where h_t are available were randomly sampled and censored to appear uncontested. Periods immediately following a redistricting period were omitted, since they pose no difference between the two approaches. This means that three periods, 1992, 2002, and 2012, are omitted from the comparison. In both cases, the model used for imputation is

$$\mathbf{h}_t = \alpha \mathbf{h}_{t-1} + \beta_0 + \beta_1 \text{incumbency} + \epsilon \quad (6.14)$$

where the \mathbf{h}_{t-1} contains the raw vote shares in the mutually-contested case, but may contain imputed values for the recursive forward propagation.

Using the presidential imputation strategy, periods immediately following redistrictings are available. However, due to the limits of data availability for presidential elections by congressional district, the comparison is restricted to congressional elections after 2004, using presidential results since 2000. The imputation model also includes an incumbency variable and

a year fixed effect. However, due to data availability restrictions, the presidential imputation strategy is broken up into three distinct model sets. First, congressional elections from 2006 to 2010 are predicted using presidential election results from 2000, 2004, and 2008, meaning the imputation model is stated:

$$\mathbf{h} = \beta_0 + \Delta\gamma + \beta_1\text{incumbency} + \beta_2p_{2000} + \beta_3p_{2004} + \beta_5p_{2008} + \epsilon \quad (6.15)$$

where \mathbf{h} is the vector of contested district results pooled over 2006, 2008, and 2010, Δ is the set of dummy variables classifying district-years into the year fixed effects γ , p_t is the presidential vote share in the congressional district in time t , and ϵ is an independent and identically-distributed error term. In this case as well, incumbency is an ordinal effect. For congressional elections in 2012 and 2014, presidential vote shares from 2008 and 2012 are used:

$$\mathbf{h} = \beta_0 + \Delta\gamma + \beta_1\text{incumbency} + \beta_2p_{2008} + \beta_3p_{2012} + \epsilon \quad (6.16)$$

with \mathbf{h} now being pooled over 2012 and 2014. Due to inter-censal redistrictings between 2012 and 2016, congressional elections in 2016 are imputed separately, using presidential election results from 2008, 2012, and 2016:

$$\mathbf{h}_{2016} = \beta_0 + \Delta\gamma + \beta_1\text{incumbency} + \beta_2p_{2008} + \beta_3p_{2012} + \beta_4p_{2016} + \epsilon \quad (6.17)$$

As in the recursive and mutually-contested imputation methods, a 5-fold crossvalidation strategy is conducted with 10,000 passes for each time period. This results in a comparatively larger number of replications for each congress, but the accuracy statistics in crossvalidation stabilize quickly with the number of passes, so this difference in typical number of passes per congress is not itself significant.

To accommodate the differing sets of data availability, imputation accuracy will be presented both over all available years and in each election year. If an imputation strategy is not effective, the out-of-sample prediction after censoring \mathbf{h}_{it} will be far from the true \mathbf{h}_{it} . In addition to raw prediction accuracy, the classification accuracy is also analyzed. This ensures that an imputation method produces likely vote shares *and* seat winners. The two forecast accuracy

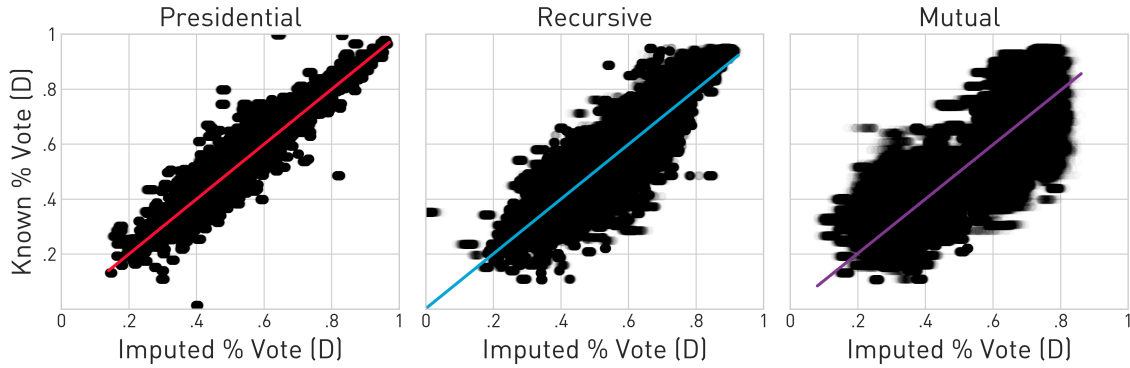


Figure 26. Imputations versus known vote shares over 10,000 crossvalidation passes

	RMSE	SMAPE	FDWR	FRWR	Accuracy	<i>N</i> Replications
Presidential	0.0544	0.0444	0.0307	0.0396	0.9297	4,550,000
Recursive	0.0645	0.0505	0.0422	0.0372	0.9207	4,627,037
Mutual	0.0833	0.0674	0.0494	0.0323	0.9182	4,627,037

Table 8. Summary of accuracy & precision for imputation over all available data. “FDWR” stands for “False Democrat Win Rate”, and “FRWR” for “False Republican Win Rate.”

statistics, the symmetric mean absolute percent error (SMAPE) and root mean squared error are defined for N observations in a forecast \mathbf{f} and observed vector \mathbf{k} :

$$SMAPE = \frac{1}{N} \sum_i \frac{|\mathbf{f}_i - \mathbf{k}_i|}{|\mathbf{f}_i| + |\mathbf{k}_i|} \quad RMSE = \sqrt{\frac{\sum_i^N (\mathbf{f}_i - \mathbf{k}_i)^2}{N}}$$

These measures are provided in Table 6.2.2, alongside the false party win rate, or the percentage of cases the imputed vote share suggested that the party wins when the observed vote share indicates the party lost. This is computed for both Democrats and Republicans. Finally, the “accuracy” is the percent of times the winner of the imputed election was the observed winner. Finally, all this occurs over a K -fold cross-validation, so the number of cross-validated predictions for each type is shown in the last column.

First, the scatterplot of the imputations is shown in Figure 26, and full results for the total crossvalidation runs is shown in Table 6.2.2. Over all instances, the presidential imputation method has significantly lower root mean square error and symmetric mean absolute percent-

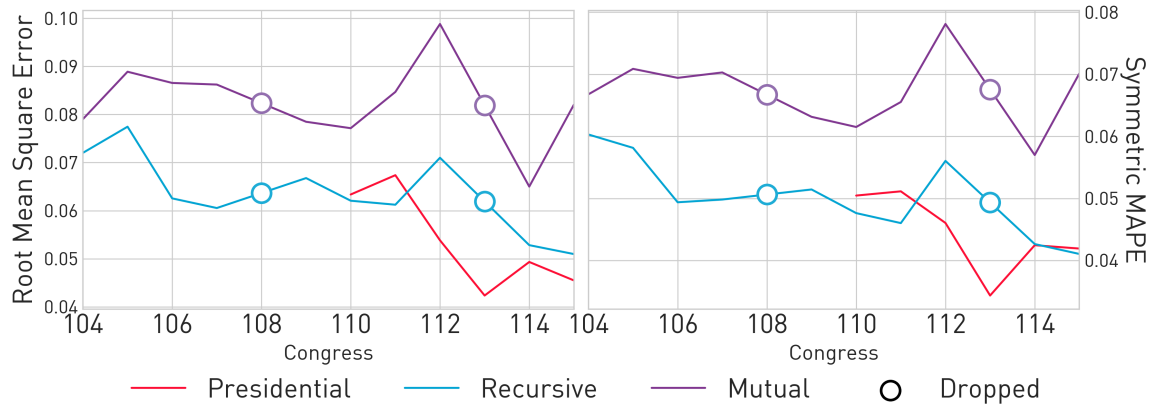


Figure 27. Root Mean Square Error and symmetric MAPE for the three imputation methods over all Congresses since 1994. Circles represent elections where the crossvalidation was unavailable due to the lack of a predecessor election.

age error, followed closely by the recursive forward imputation. Finally, the mutually-contested strategy consistently has the highest prediction error.

When slicing the imputations by congress, this trend also holds. The comparison of imputation accuracy by congress are shown in Figure 27. In general, the recursive and mutual imputation strategies move in the same direction over congresses, with the recursive method consistently dominating the mutually-contested strategy. For the range of Congresses available, the presidential imputation strategy has lower error in all but the 110th and 111th Congress, in 2006 and 2008. The misclassification rates by congress are shown in Figure 28. There, the presidential imputation strategy is almost always more accurate than the other two approaches, with the recursive forward propagation typically misclassifying about as many seats as the mutually-contested strategy. The misclassification rate for some elections, mainly elections to the 104th and 112th Congresses, implies that around an eighth of the winners in elections to that congress are not correctly predicted by the imputation methods.

The weakness of the mutually-contested strategy is also tangible when considering cross-validation cases for districts within each congress. For example, the root mean square error for each district in each congress is shown in Figure 6.2.2. In general, the presidential imputation

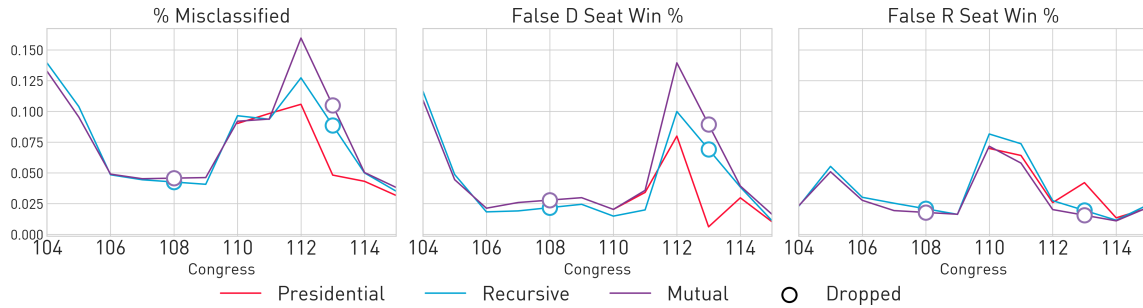


Figure 28. Classification Error rates by Congress. Here again, circles represent elections where the imputation method was unavailable.

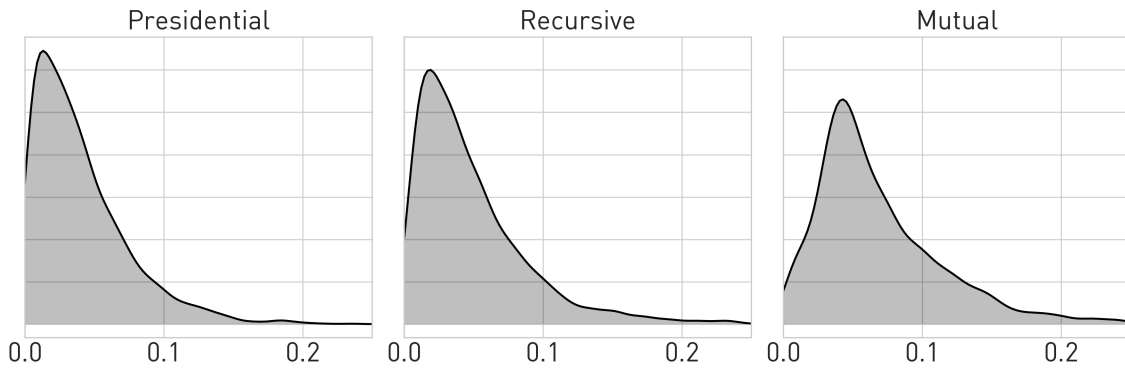


Figure 29. Distribution of Root Mean Square Errors (RMSE) for each district in each Congress.

method is consistently more accurate for most districts in most congresses, but the recursive imputation method fares nearly as well. The mutually-contested tends to have larger RMSE by district as well.

This also holds for the percent of all correct classifications-per-district-year, but the differences in classification accuracy between methods are all under a single percentage point. Critically, around six percent of district-year outcomes are *never* classified correctly in any imputation method. In addition, a contest is almost always classified correctly as a Democrat/Republican win or is almost never classified correctly: between 93% and 98% of contests, regardless of imputation method, are either correct 99% of the time or incorrect 99% of the time.

Thus, when presidential results are available, they tend to produce more accurate estimates of observed vote share, and thus may provide more plausible hypothetical values for uncontested elections. If presidential data is not available, recursive imputation is nearly as effective. Both approaches have clear, sharp disadvantages, though. Estimating a hypothetical contested vote share for uncontested elections using presidential returns at the congressional district level suffers strongly from data availability, since it requires an augmenting. Using the recursive forward-propagating method, while nearly as accurate, cannot be easily used for elections immediately following redistrictings, unless successor districts are identified. Regardless, a choice between the two appears to be a lateral move: no appreciable gain in classification or prediction accuracy is obtained from using one over another, although both have lower error than the mutually-contested imputation strategy. Likewise, it is also important to note that the true “accuracy” of these methods are unobtainable, since the anticipated vote share for an uncontested district *if it were contested* is an inherently counterfactual, unobservable quantity.

6.3 Conclusion

When considering the potential for spatial misspecification in the electoral models that drive seats-votes constructs, it is likely that models for vote shares at the congressional level require some form of correction for spatial dependence. However, this correction likely does not significantly impact the resulting model for the seats-votes curve if the “correct” treatment is in the error term for the vote share model. Corrections for spatial heterogeneity only result in some states with significantly-different vote share or swing distributions. These effect structures complicate the model significantly while contributing little to its predictive performance or model accuracy, and so likely do not constitute a useful innovation in their own right. In addition, since many analyses (such as the novel ones conducted later in this dissertation) occur at the state level, the use of a super-state hierarchy is moot.

If the correction for spatial dependence in the vote share model is of the mixed-regressive,

spatial autoregressive form,³⁶ the shape of the estimated seats-votes curve, both its slope and width, may change significantly. This would, in turn, impact the measures of partisan advantage defined in §2.3 computed about those seats-votes curves. In addition, the assumption of generalized uniform partisan swing, that change in vote shares over time can be efficiently modeled by an independent and identically distributed random effect, is empirically unsound. In truth, electoral swing is highly spatially patterned, while the distribution of swings in *most* states are nearly identical. However, introducing a correction to this assumption does not significantly change the properties of the estimated seats-votes curves, unless spatial dependence is quite strong. In that case, the exact specification of the spatial autoregressive structure for the swing term is critical. If a class simultaneous autoregressive specification (either the lag or error form) is used, the change can be detectable, even when the autoregressive effect is small. However, when considering a novel variance-consistent specification in simulations, the magnitude of the autoregressive effect does not significantly change the estimated seats-votes curve from that estimated under the null. Thus, it may be the case that heteroskedasticity induced by introducing spatial autoregressive structures is truly driving the change in the two alternative specifications for the partisan swing term.

More generally (and more directly), this dissertation suggests that spatial misspecification in the vote models or counterfactual models driving seats-votes models may affect the resulting estimate of the seats-votes curve. But, the size of the impact depends greatly on the model specification. A classic econometric testing regime would suggest a specification that has no large impact for most of the commonly-encountered spatial autoregressive effect sizes. In contrast, in other electoral data (for the United States Congress or abroad), the specification search may settle on a different model or have a different spatial autoregressive effect strength. Thus, this simulation-driven exploration of the impact of spatial effects in seats-votes models builds intuition about a wide variety of the models discussed in the seats-votes modeling literature. In general, it seems that seats-votes curve estimates should be robust to moderate spatial autore-

³⁶also known as the “spatial lag” or “endogenous lag” form

gressive effects or spatial misspecification in large samples *if and only if* the misspecification is of a special form. This means a rigorous specification search should *always* be conducted, since it is very likely that the underlying model requires a spatial correction, and this correction may, depending on its form, significantly affect results.

Chapter 7

LOCALIZING PARTISAN SYMMETRY MEASURES

Measures of partisan bias have long been used in attempting to characterize the fairness of electoral systems. From the theory and measures discussed in Chapter 2, many different analyses could be conducted, but all would *statewide* analyses. That is, all of the measures of partisan bias in Chapter 2 work by first estimating a seats-votes model like those considered in Chapter 5, simulating many elections at each partisan advantage measure's reference position, and then characterizing those simulations in some way. These models are estimated for the entire state under study, and the bias measures refer to the statewide tilt of the congressional districts. No individual district scores are available.

Thus, only the *geometric* measures of boundary manipulation provide “local” indications of which districts may be gerrymandered. This is clear in popular and legal discussions attempting to identify gerrymandering. Often, allegations about gerrymandering characterize the overall slant of the districting plan, and then attempt to argue that specific oddly-shaped districts generate that advantage. This pattern is present in many of the recent lawsuits and analyses of gerrymandering in after the 2010 redistricting discussed in Section 2.1, but chiefly in the *Cooper v. Harris*, *League of Women Voters v. Detzner*, and *Whitford v. Gill* cases.

However, the use of different standards of evidence to argue about the partisan advantage as a whole and which districts generate this bias means that weirdly shaped districts might not end up contributing significantly to any given statewide bias measure. While the analysts presume that strangely-shaped districts drive the unexpectedly large bias measures, it might be the case that *other* districts impact the measure more strongly, whose shapes are considered more regular or intelligible. Again, as discussed before, the mere presence of strange boundaries does not indicate boundary manipulation, since human habitation and community character does not necessarily admit regular polygonal tilings. But, intuitively speaking, many of the districts singled out for legal review or popular reproach have strange boundaries. Thus,

strange shapes are neither necessary nor sufficient for partisan advantage, but they are what's available to most. As such, the use of boundary regularity is suspect, but this is not always recognized. Regardless, shape measures are the most wide-spread "score" used to assess how likely it is any individual district is gerrymandered, while bias scores at the state level are used to identify statewide advantage.

In this sense, Chapters 7 and 8 will be about the *forensics* of gerrymandering; detecting the potentially-illicit districts that impact statewide bias measures. This exhibits a local/global divide, as currently no local measure of partisan advantage is available to assess districting plans, whereas geometric measures are often only used in a local context. A "good" forensic, then, should be a statistic with some sort of control for "significance" and should provide a standardized method to compare between states or across time. The significance filter should avoid identifying a *fixed fraction* of bad districts, like a top- k percentile-based filter might. Instead, a district should only be singled out as gerrymandered if its effect on plan-wide bias is both large relative to its variance and is consistent over time. Below, I develop a method to define such a statistic for any measure of statewide partisan bias using classic methods in model criticism.

I am interested in local measures of district impact for three reasons. First, I am interested in examining the set of districts identified by these techniques, if they "work" so-to-speak. Whether they "work" is not necessarily indicated by them singling out the same districts that have been identified as gerrymandered by the legal system. This is for two reasons. First, the courts and most analyses up to now have used the same inconsistent global/local measures, suggesting that individual districts that impact statewide bias may not be the ones identified in the past. Second, the impact measures may select an entirely different (but meaningfully-consistent) set of districts. That is to say, the districts which courts have identified as gerrymandered (using potentially flawed or inconsistent empirics) should not be used as the sole truth against which these measures are assessed. If the impactful or influential districts have a "typology," a consistent interpretation or set of interpretations, then the intuition about these impact measures can be built independently of whether the measures identify the same districts that the courts or interest groups select. I am also interested in whether one class of these impact measures,

those deriving from classical leverage measures of regression influence are distinct from another class, those deriving from jackknife techniques. Critically, if the model-based measures of district influence all agree, then it may be the case that the simplest way to create local scores for district partisan impact is to analyze the source model's leverage matrix. This would indicate that the peculiarities of each bias measure are effectively moot when it comes to which districts impact their values, since partisan impact would be a function of observation influence.

Further, I am interested in determining whether these impact measures relate to the *geometric* measures of boundary manipulation. Since geometric measures are currently the most common "local" district discriminant, the relationship between the measures of boundary manipulation and impact on partisan bias is critical. Proponents of boundary manipulation measures suggest they *may* identify districts that create advantage, but the arguments linking identified boundary manipulation to advantage directly are weak and fiercely contested. It may be the case that some districts with irregular boundaries do not significantly impact partisan advantage in a state. It also may be the case that strangely-shaped districts *do* tend to impact partisan bias in states. While aggregate arguments exist (Altman, 1998b, e.g.), the lack of a viable local measure of partisan advantage has left the direct relationship unexamined.

Finally, I am interested in the distribution of effect sizes for its own sake. One question that surfaces from interviews that will be discussed Chapter 9 is a trade-off: should practitioners attempt to make each district as fair as possible versus making a plan fair *in aggregate*? Often, respondents suggested that each district's "fairness" was essentially its competitiveness, so this question became about whether all districts should be competitive, or whether a congressional delegation should avoid large majorities. Many did not consider the fact that having many competitive districts may magnify the extent to which a states' aggregate representation does not reflect the partisan preferences of its electorate. Thus, I am curious as to the existence of two types of plans with nonzero advantage: what I call "balanced" plans and "accumulative" plans. Balanced plans are plans where district impact measures might cluster on both sides of zero with large magnitudes. But, altogether, the plan is not significantly biased. Accumulative plans would be plans where most districts have a small impact on statewide bias, but the bias

is present nonetheless. This speaks to questions of district design and the way representation might be balanced in states to achieve aggregate representativeness versus competitiveness in each district.

7.1 Classical Leverage Measures for Seats-Votes Models

I suggest two ways to examine the impact individual districts have on the general partisan bias scores. First, and most straightforward, the seats-votes models discussed in Chapter 5 tend to reduce to a standard linear regression (or weighted linear regression) to predict vote shares for a reference party given some set of political or demographic covariates. Thus, examining the structure of the *influence* statistics for that model may indicate which districts are playing an outsized role in defining the characteristics of the plan in aggregate. Influence analysis is a foundational subfield of model criticism (Carota et al., 1996), and is related to many other styles of residual analysis in regression work (Atkinson and Riani, 2012; Chatterjee and Hadi, 2009; Cook and Weisberg, 1982; Belsley et al., 2005). Typically, the analysis of influence involves identifying or estimating some score for each of the observations that characterizes how important observations are to a model. One measure relevant to linear models is an observations' leverage. Formally, an observation's leverage on a linear model is a measure of how far any given point i covariate vector, X_i , is from the center of the point cloud in X -space. If the point has high leverage, it means that the point represents part of the vector support of the underlying information in X which few other observations in the dataset also span. Thus, the "distance" from the remaining points means that the regression line of best fit will fit the point as strongly informative. For general linear models, leverage is constructed through the *hat matrix* for the given model specification (Hoaglin and Welsch, 1978).

Extending the analysis of Gelman-King models from the previous chapter, it is possible to identify the hat matrix for any single model for a given election. Thus, it may be the case that high-leverage districts are the districts that strongly affect the partisan advantage of a district. Typically, h_{ii} represents the i th element of the diagonal of the leverage (or *hat*) matrix, and

corresponds to the “leverage” of observation i . For a linear regression with variance weight matrix \mathbf{V} , the hat matrix is:

$$\mathbf{H} = \mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \quad (7.1)$$

Thus, for any of the model specifications in the Gelman-King seats-votes framework, the leverage of each years' model can be computed and analyzed.

However, as often noted, high-leverage points are not necessarily *influential*, since their removal may provide essentially the same estimated relationship between the response and covariates. This idea of *removal* is critical to the idea of *influence*, which is a much larger, informal concept in model criticism. A point is influential when its absence is “noticed,” for some operational, formal definition of “notice.” Whether or not an observation's absence is “noticed” depends on the model property the analyst examines. Since influence is dependent on the model property, many different kinds of influence measures exist, and arbitrary new ones created through the focus on empirical influence functions (Mallows, 1975).

Explored initially in the first edition of Belsley et al. (2005) in 1980, commonly-used influence measures are constructed *focusing on* a given model property. The difference-of-fits (DFFITS) statistic is a diagnostic providing the standardized change in predictions for observation i when \mathbf{X}_i is omitted from the regression. For a response vector \mathbf{y} with elements $y_i, i = 1, 2, \dots, N$, the difference-of-fits statistic for observation i is:

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)i}}{\hat{\sigma}_i \sqrt{h_{ii}}} \quad (7.2)$$

where $k_{(i)}$ is the statistic k from a model where observation i has been removed, whereas k is from the model estimated with all observations. Substantively, this statistic provides the difference in the prediction of the fitted y_i value when observation i is treated as an out-of-sample prediction. Another commonly-used influence measure, the difference-of-betas (DFBETAS) statistic, provides the extent to which the removal of i influences the estimate of β :

$$DFBETAS_i = \hat{\beta} - \hat{\beta}_{(i)} = \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_i\mathbf{e}_i}{1 - h_{ii}} \quad (7.3)$$

Here, e_i is the residual for i in the full regression and \mathbf{X}_i again is the i th row of the design matrix \mathbf{X} .

For the purposes of this dissertation, I focus on the Cook's D , or distance statistic (Cook and Weisberg, 1982). Cook's D is a joint influence measure, describing the extent to which an observation affects all p marginal effects in a given regression.

$$D_i = \frac{\mathbf{e}_i' \mathbf{e}_i}{\hat{\sigma}^2 p} \left(\frac{h_{ii}}{(1 - h_{ii})^2} \right) \quad (7.4)$$

An observation that substantially changes the generating process for the election might be an observation that also substantially affects the bias statistics derived from that process. Since the bias statistic is insensitive to the substantive interpretation of a β effect in the Gelman-King models for a given decade in a given state, the Cook's Distance provides an effective computationally-simple measure of overall influence an district may exert on the seats-votes curve. Importantly, Cook's D is a standardized distance with common rules-of-thumb on what an extreme value is. While there is no formal distributional testing for the statistic *per se*, one rule of thumb common in statistical practice is to consider an observation "influential" when it is $D_i \geq 2$. More generally, Cook's distances that are markedly larger than the rest of the distances in the set of observations should be considered suspect. Thus, we can not only examine the high leverage districts in Gelman-King models, but also use the Cook's D as a significance filter. This avoids the issue with many of the common geometric forensics that have no significance filter.

However, Cook's D or h_{ii} may not be consistent over time. This is magnified by the use of two-cycle models like the Gelman-King specification. Since one model is fit in each year, the leverage & residual (and the D) may be different for each district in each model. Thus, I will examine whether or not Cook's D or h_{ii} fit the criteria outlined for our statistics above. While I expect them to perform fine within a given year, there is no guarantee that the measures will be able to characterize the influence a district has *consistently* over time. In addition, a full-decade model (such as the one suggested by McGann et al. (2016) considered in Chapter 5) could be used, since leverage and influence measures for some specifications of longitudinal models are available (Tan et al., 2001). To hone discussion, I focus only on the Gelman-King models, but

leverage and Cook's D -style statistics could be identified for any parametric model specification for \mathbf{h} where leverage statistics are identified.

7.2 Local Measures of Partisan Impact: Jackknifing the Plans

Fortunately for the original authors of various deletion statistics, many of the parameters of interest for influence analysis are available *without* having to estimate the global model and N deletion models. This is primarily because the parameters of interest are related directly to \mathbf{H} , and each of the influence statistics can be sufficiently characterized as transformations of full-model statistics and elements of the leverage matrix. Since most analysts are interested in identifying the influence of \mathbf{X}_i on direct model properties like β , \hat{y}_i , or both, many analyses of influence do not actually require the evaluation of the model with \mathbf{X}_i removed. And, while the measures of leverage or Cook's D are available to examine the influence districts have on the regression underlying the seats-votes model, direct parameters of the vote model are not at interest here.

In fact, the measures of partisan advantage from Chapter 2 are what the impact analysis requires. They are not "direct parameters" of the vote share model, and have no convenient expression in terms of the leverage matrix. Indeed, since they pertain to the seats-votes curve, itself estimated from *many* sets of simulations from the model, characterizing the influence of each district in \mathbf{X}_i on a given advantage measure would likely not have a formal expression.

However, the original intuition behind influence analysis is still available: which observations generate markedly different estimates when they are missing from the analysis? If this can be computed, then their *presence* in the full ensemble of size N , conditional on the remaining $N - 1$ observations, can be estimated. This is the concept behind jackknife estimators, suggested by Efron (1982). Jackknife estimators, conceptually, take a given estimate of interest, $\hat{\theta}$, and compare it to the set of estimates $\hat{\theta}_{(i)}$ constructed when each observation i , $i = 1, 2, \dots, N$, is removed. Jackknifing has a long history in statistical practice. Efron and Gong (1983) demonstrate that the jackknife is related to the empirical influence function dis-

cussed by Mallows (1975), where the “influence” of an observation is the empirical impact of its removal directly on the quantity of interest.

Jackknifing is commonly-used to derive estimates of standard errors or confidence intervals that are robust to collinearity or poor specification; however, they are much more broadly used in machine learning contexts to improve or assess model sensitivity in chains of analyses. Here, I propose to use a similar approach: by jackknifing the districting plan, we can identify which districts significantly influence a given bias score, which is the final statistic obtained from a chain of analysis. The removal of the single district and re-simulation of advantage, then, provides an estimate for how that district affects the entire analytical pipeline. If the district is not influential, then its impact will not propagate. I call these district partisan impact statistics, or simply *district impact statistics*.

Although they do not parcel out the fraction of a bias score that each district contributes (Anselin, 1995), the local impact measures *do* allow the analyst to determine which the districts impact statewide advantage. In addition, it allows for the filtering out of districts that have no significant impact on advantage. It allows for the consistency of influence to be characterized over time. The measures admit uncertainty in the estimate and a characterization of the “significance” of the result using the statewide variance and jackknifed simulation variance. Most importantly, the district impact statistics work *directly* in terms of the measure used to characterize statewide bias. So, the impact measure for each district uses the same conceptual model as the statewide bias estimate and its impact is estimated on the same terms. This is exactly what is desired by policymakers when attempting to infer which districts drive the plan-wide estimates of advantage. Thus, with these local statistics, we characterize the political impact of *each district in the plan* in a manner that is consistent with the statewide estimate, resolving precisely the inconsistency of using partisan measures statewide but using geometric measures locally/district-wise.

I characterize this influence directly using a deletion & re-estimation strategy, since the impact cannot be stated in terms of the observation leverage for arbitrary bias statistics.

Algorithm 3 Consider a partisan advantage statistic for a district plan with N districts in election t , denoted b_t . Let \hat{b}_t be the point estimate of b_t constructed from a set of L replications, $\mathbf{b} = b_1, b_2, \dots, b_L$, each made over a batch of simulated elections in the N districts. Let the observed conditions in election t be \mathbf{X}_t . Each element of \mathbf{b} is constructed using the a simulation algorithm $\mathcal{A}(\mathbf{X}_t^\circ)$, where \mathbf{X}_t° is a scenario under which the bias will be evaluated. It may be true that $\mathbf{X}_t^\circ = \mathbf{X}_t$ but it is not necessary. Finally, let $\mathcal{A}(\mathbf{X}_t^\circ)$ involve a model of vote shares estimated from the observed outcomes & electoral conditions in time t , $\mathcal{M}(\mathbf{h}_t, \mathbf{X}_t)$. A set of N influence measures corresponding to each district can be constructed:

1. Estimate $\mathcal{M}(\mathbf{h}_{t(i)}, \mathbf{X}_{t(i)})$, the results from election t in $N - 1$ districts, having omitted district i , $i = 1, 2, \dots, N$.
2. Construct the bias of the state plan as if district i were not included in the plan. This involves L replications of elections in the $N - 1$ districts using simulation regime $\mathcal{A}(\mathbf{X}_{t(i)}^\circ)$.
3. Store the set of L deletion bias statistics, $\mathbf{b}_{t(i)}$.
4. Increment i and return to step 1 until all districts have been evaluated.

This yields N sets of “deletion statistic simulation distributions,” $\mathbf{b}_{t(i)}$, each with L independent, identically-distributed replications.

Given the “statewide” bias simulation distribution \mathbf{b}_t and each of the N deletion statistic simulation distributions $\mathbf{b}_{t(i)}$, these distributions must be compared. First, however, note that each b_l is independent of other b_k for $k, l \in \{1, 2, \dots, L\}$. This is because, in any one problem configuration (statewide or deleting i), each realization from $\mathcal{A}(\cdot)$ is independent from every other realization, thus elements of \mathbf{b}_t are independent from one another. In addition, since the data generating process does not change during the simulation runs, elements of \mathbf{b}_t must be identically distributed. This ensures that, *between* elements in \mathbf{b}_t and elements in $\mathbf{b}_{t(i)}$ (or between elements of $\mathbf{b}_{t(i)}$ and others in $\mathbf{b}_{t(j)}$, $i, j \in \{1, 2, \dots, N\}$), there is no correlation *even though* the data generating processes share at most $N - 2$ members. Conceptually, this occurs because we analyze statistics *generated from* the model, not the models themselves; these realizations are “strongly” independent, since the simulation regime is stateless between

replications. Further, since the model does not change during replications for a given jackknife of i , $b_{(i)}$ are identically distributed. In addition, \mathbf{b} is unordered, since the replications from $\mathcal{A}(\mathbf{X}_t^o)$ could occur in any order. Thus, analyzing \mathbf{b} and a permutation of \mathbf{b} yields the same results. This means that the sets of \mathbf{b}_t and all N $\mathbf{b}_{t(i)}$ cannot be *paired*, so a specific set of distributional analysis methods must be used instead.

Before proceeding, though, it is also important to reiterate that the analysis is *not* focused on identifying which districts' removal benefits Democrats or Republicans, i.e. shifts the distribution of bias statistics left or right in absolute terms. Rather, *given* that statewide bias estimates may advantage Democrats or Republicans, identifying which districts *increase* bias or *decrease* bias is the more central concern. Thus, it is not the raw value of the deletion bias estimate that matters, it is the position relative to the statewide estimate. If a district is removed and *increases* bias, then the plan is *more fair* than it would be if that district were not present, regardless of the winner of the seat or the tilt of the state as a whole. However, I will discuss districts at first brush as moving the state more towards Democrats or Republicans, but whether these increase or decrease bias depends on the statewide advantage estimate.

Further, it is important to note that this strategy does not reapportion the omitted individuals back to the districting plan. At the district-level, it is reasonable to consider the impact of a *district* to be some estimate of advantage of that particular spatial-social configuration of voters. Thus, the fact that those voters are omitted and *not* redrawn into a new districting plan is precisely the point of these influence statistics. The district, its candidate, and its voters each exert an influence on the statewide advantage experienced by all. While sufficient controls may be placed on candidate properties, it seems unlikely that sufficient ecological inference can be conducted to construct *population countfactuals* more fine-grained than the shifts in aggregate partisan support considered in Chapter 5 & 6. Regardless, the omission of the district, its voters, and its candidate is intentional here, and the re-drawing of an $N - 1$ plan that includes the voters omitted from the jackknife would obscure the point of this analysis.

7.3 Interpreting and Comparing Impact Distributions

To enable the analysis of jackknifed impact distributions, three testing procedures are available. To ensure consistency over elections in a decade, the tests can be adapted in three ways. First, one could test in each t and only accept consistently-significant results using an $N \times T$ multiple comparison correction. Second, one could simply pool \mathbf{b}_t over all $t = 1, 2, \dots, T$ in a decade to provide a single omnibus statistic for each of the N districts for a pooled sample of size $T \times L$. Third, one could construct a stratified (or hierarchical) estimate, admitting that the mean might be different in each t . This would pool the *differences* between statewide and jackknife distributions in each time period and use an appropriate pooled estimate of variance. Any method should yield consistent results, so long as the statewide and deletion distributions are stable over t . I will only define the stratified and pooled forms of the effect size statistics discussed in Section 7.3.3.

7.3.1 Nonparametric Difference in Distribution Tests

First, each set of realizations might be judged to be distinct from the state or from one another using some kind of distributional ANOVA test. This can be done in many ways (Elliott and Hynan, 2011), but I choose a standard Mann-Whitney U test between the statewide distribution and each deletion distribution. I choose this *instead of* other types of nonparametric distribution tests (such as the Kruskal-Wallis) since I am only interested in determining whether each deletion distribution is different from the statewide distribution, not whether each distribution is distinct from one another. This means that there are only N direct comparisons instead of all $\frac{N(N-1)}{2}$ pairwise comparisons accounted for by Kruskal-style nonparametric ANOVA, since we are only concerned with identifying whether a district is distinct from the statewide distribution, not from every other distribution. Thus, multiple comparison corrections are still required, but the Bonferroni correction is not as severely conservative in this case as it would be if all pairwise comparisons were analyzed. In case studies using Q measures in S scenarios, I will

use an $N \times T \times S \times Q$ correction, where N is the number of districts and T is the number of decades. In most cases, significance is not marginal, so the choice of including a correction for the number of scenarios or measures in the Bonferroni factor is not critical.

7.3.2 Binomial Sign Testing for Deletion Distributions

The binomial sign test is a method to determine whether or not a treatment shifts a sample in a consistent direction. In this case, the removal of district i has a *consistent* effect if its removal tends to benefit one party. This would allow the analyst to filter out which districts have an inconsistent impact on a bias measure from those that consistently either increase or decrease the statewide bias estimate. This can be done in a two-tail specification (simply looking for “consistent influencers”) or can be done in a single-tail fashion (looking at whether a specific district consistently increases bias). This is done first by computing the number of realizations in $\mathbf{b}_{t(i)}$ that are below the statewide point estimate:

$$n_{t(i \leq)} = \sum_l^L \mathcal{I}(\mathbf{b}_{t(i),l} \leq \hat{b}_t) \quad (7.5)$$

where \mathcal{I} is the indicator function which is 1 when its argument is true and zero otherwise. Then, $n_{t(i \leq)}$ is distributed binomially with population parameter $n = L$. Testing the undirected hypothesis pair, that $H_0 : \hat{p}_{t(i)} = .5$ versus $H_A : \hat{p}_{t(i)} \neq .5$ provides a two-tailed test identifying whether the removal of district i has a *consistent* influence on \hat{b}_i in terms of the direction of the effect. In concept, $n_{t(i \leq)} / L$ is the fraction of cases where the removal of district i benefits Republicans, since I have designated Democrats as the reference party arbitrarily. When the fraction of realizations benefiting Republicans relative to the statewide estimate is distinctly smaller than .5, the removal of i benefits Republicans more than Democrats, and the opposite when flipped. When pooling all samples over a decade, $t = 1, 2, \dots, T$, then $n_{.(i \leq)}$ is binomially distributed with $n = T \times L$, and is computed from all $\mathbf{b}_{t(i)}$ pooled over the decade. The α level for the pooled test under a Bonferroni correction in time t is simply $\frac{\alpha}{N}$, since N comparisons are made between the deletion distributions and the statewide distribution, each distribution having

$T \times L$ observations. For the unpooled comparison strategy, the Bonferroni correction is $\frac{\alpha}{T \times N}$ since T sets of N tests are conducted on samples of size L .

7.3.3 Effect Size Estimation for Deletion Distributions

In addition to sign-consistency, the effect of removing district i can be estimated directly. Typical effect size estimators in the statistical literature can be adapted to this context, modeling each $\mathbf{b}_t, \mathbf{b}_{t(i)}$ as two distributions separated by a treatment, removing district i from the districting plan. One estimate of the effect of removing district i from the plan (for each N) would be:

$$d_{(i)} = \frac{\hat{b}_{t(i)} - \hat{b}_t}{s^*} \quad (7.6)$$

$$s^* = \sqrt{\frac{\sigma(\mathbf{b}_t)^2 + \sigma(\mathbf{b}_{t(i)})^2}{2}}$$

This estimator is a Cohen's d -style effect estimate and also bears similarity to measures of deletion residuals. In this case, it measures the difference in means between the statewide & jackknifed distributions, and divides by an estimator of the shared deviation of the two distributions. A similar nonparametric estimator is suggested by Grissom and Kim (2012) as:

$$d_{r(i)} = \frac{U_{(i)}}{N(N-1)} \quad (7.7)$$

where $U_{(i)}$ is the Mann-Whitney U statistic comparing $(\mathbf{b}_t, \mathbf{b}_{t(i)})$. I will focus on the Cohen's d -style estimator for simplicity. Here, interpreting the effect size from the d estimate is done using the typical rules-of-thumb common in statistical literature. I break down the difference using three tiers of effect magnitude. A marginal effect is one between .25 and .5, a moderate effect is one between .5 and 1, and an effect size greater than 1 is considered a large effect. The sign of the d retains the original meaning, so that if the state is biased towards Democrats (\hat{b}_t is positive) then a negative d statistic indicates the district i increases the bias towards Democrats.

7.4 Advantage Impact, Model Leverage, and Model Influence

First, I engage in a few case studies to examine the properties of the jackknife impact measures. Then, I examine whether or not these statistics relate well with each model's Cook's D & leverage. In the next chapter, I will compare impact statistics directly to the geometric measures, examining many potential multivariate relationships in exploratory regression & correlation analyses. In all cases, the jackknife measures are constructed from $P = 1000$ replications of $K = 1000$ simulated elections. This yields a set of P statistics for each district, and a set of P statistics for the full state, reflecting a full simulation load of 1000×1000 elections in each state. In most cases where a point estimate is required for the deletion distribution, the effect size estimate from Eq. 7.6 will be used. This means that any point estimate is an expression of the district's impact *relative* to the statewide bias. Thus, even when statewide bias is negative, a positive effect size indicates that the state becomes *more* Democratic when the district is removed/more Republican when the district is present, conditional on the rest of the districts in the plan. Thus, $N + 1$ sets of P jackknife impact statistics (as well as the N difference-in-means point estimates) will be analyzed. In addition, N of the "classic" model influence measures, such as the observation residual, leverage, and Cook's D , are available. Relationships between the impact measures and these other measures will be conducted using the difference-in-means statistics, since the *change* in point estimate reflects the impact of the deletion on the model.

7.4.1 General Properties about Impact Measures

I conducted preliminary study using the dataset developed in Chapter 3 for eleven states in two decades: Arizona, California, Illinois, Michigan, New York, North Carolina, Ohio, Pennsylvania, Texas, Washington, and Wisconsin in 2000 and 2010. I focus on two decades for California and a single decade in Wisconsin in the following sections. In general, there is a strong negative relationship between impact measures and Democratic vote share in all cases and nearly

all times, meaning that districts that are won by Republicans tend to benefit Republicans when included in the districting plan. In some rare cases, this relationship is reversed, in that the presence of districts won by Democrats actually benefits Republicans. In theory, this is possible, since measures like the efficiency gap or the observed bonus measure may catch districts whose intent is to pack partisans. If packing is operant, then districts that pack are intended to disadvantage the party that wins the district. So, a district won by Democrats whose deletion *strongly benefits* Democrats indicates a district packing Democrats, but most districts benefit the party that wins them.

In addition, there *is* a distinct difference between plans that are balanced versus plans where most districts have marginal impact on an unbiased statewide measure. However, this is highly contingent on the measure type. Some measures are strongly bimodal, as will be discussed for California. The strongly-bimodal measures track the partisanship of the district. However, these measures may be more or less separated into partisan clusters depending on the decade under analysis. Thus, it does seem to be the case that plans can be characterized as “balanced,” where many districts provide significant advantage but balance each other out statewide, or “consistently neutral,” in that all districts have very small impact on statewide bias. But, this characterization tends to be strongly influenced by the measure type, since some measures are much more strongly bimodal than others under a given simulation scenario.

In addition, simulation scenarios strongly affect the impact statistics. Depending on whether the summaries of impact are constructed from simulations under observed conditions or with no incumbents, most districts shift from being “impactful” to having no clear impact. This highlights the importance of controlling for incumbency in these analyses, but does not resolve the normative argument about *whether* it makes more sense for the analysis of advantage to be about scenarios that are *never experienced*. I believe it makes sense to compare both scenarios, the simulations under observed conditions and simulations about the no-incumbency elections

(which never occur), but priority should be given to districts identified in the no-incumbent analysis.³⁷

7.4.2 Case Study: California, 2000 & 2010

California's congressional districting plan from 2000 was thoroughly-discussed in popular literature as an incumbency-protecting gerrymander (Fan et al., 2015). Critique of the 2010 districting plan has been much less pointed, often suggesting the plan exhibits less bias. While both plans were drawn by the same style of commission, changes in electoral politics in California may have changed the balance of power on the commission and its institutional culture. In addition, the adoption of a so-called blanket or top-two primary, where the top two candidates of any party get selected in the primary election to run in the general election, likely changes the resulting measurement of the congressional districting plan. The blanket primary applies to all elections under the 2010 plan and a traditional partisan primary applies to all elections before 2000.

Thus, direct comparisons over the redistricting event cannot necessarily declare that the commission's new plan is solely responsible for the change in performance of the new system; both rules changes and district line changes may have impacted the scores. While only a few districts in California ever see the practical *result* of the blanket primary result in single-party general elections—a single-party general election—these can be treated as uncontested elections in the same style as standard elections or the primary election can be used to assess an effective two-party vote. In this analysis, the single-party general elections are treated as uncontested, and the two-party contested vote share imputed according to the methods in Section 6.2. I *do not suggest* that the change in structure from one plan to another is solely caused by the new district lines.

³⁷Many of the interviewees in Chapter 9 referred to incumbency when mentioning skepticism of partisan advantage measures.

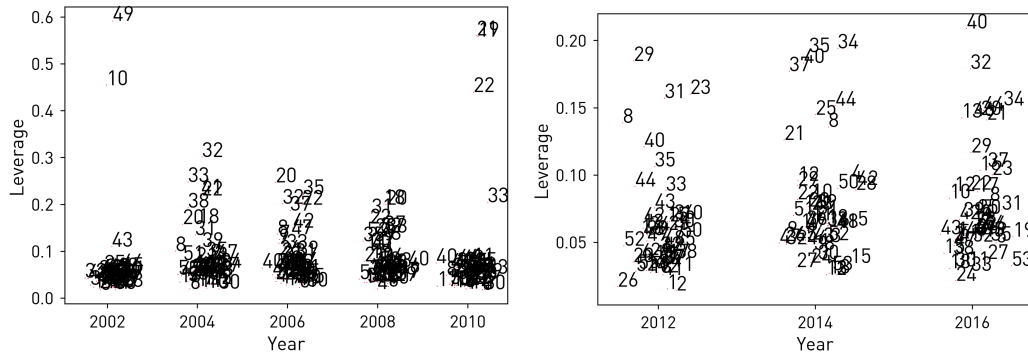


Figure 30. Annotated scatterplots of the leverage for observations in each Gelman-King-style model in California, from 2002 to 2016. The labels are the district number, and have been jittered to improve visibility within each year.

7.4.2.1 Leverage

The leverage of each district in the various single- and two-cycle Gelman-King models of California congressional districts is shown in Figure 30. In general, the observations with high leverage do not persist over time. While some districts do have consistently large leverage (such as district 40 in the latter decade), they often are not *unusually* large, in that they are significantly larger than any other observations' leverage. Districts 10 and 49 in 2002 have exceedingly high leverage, but they immediately return to the fold for later years. Adding to the fact that leverage cannot solely identify influence and no rules of thumb or statistical testing exist for identifying unusually extreme values, leverage is ineffective at determining which districts might be impactful.

7.4.2.2 Cook's Distance

First, consider the distributions of Cook's D in each year shown by Figure 31. Here, the district numbers are plotted each year with the Cook's distance on the y -axis. Thus, we see that some districts consistently have large Cook's distances in the first decade, such as district 20 (showing up as distinct in 2004, 2008, and 2010) and district 19 (showing up as distinct

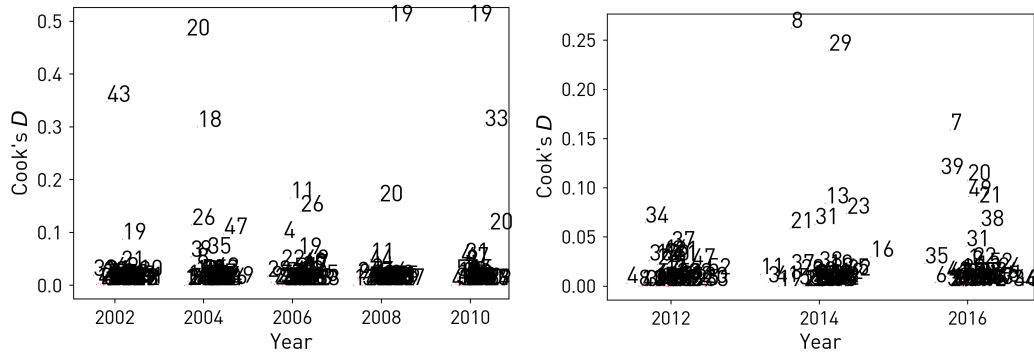


Figure 31. Annotated scatterplots of the Cook's distance for each Gelman-King-style model in California, from 2002 to 2016. For the 2000 decade plot, the axes have been truncated; district 19 in 2008 has a Cook's distance of .74 and in 2010 of 8.19. The labels are the district number, and have been jittered to improve visibility within each year.

in 2002, 2006, and extremely distinct in 2008 and 2010).³⁸ The fact that the district returns such a large margin for an open seat is likely what drives the large Cook's distance in the 2010 model. District 20, immediately south of district 19 & containing parts of Kings, Kern, and Fresno counties, was also a safe district for Democrats. It consistently returned a Democratic representative, but typically with a much lower (sub 10%) margin when contested. Both districts went uncontested a single time during the decade, with district 19 being considered "influential" in 2008, when it was uncontested. In the second decade, some districts have consistently large Cook's distances, such as district 31 & district 21, but the districts with extreme Cook's d values tend to not persist over each year.³⁹

³⁸District 19, which contained parts of Tuolumne, Mariposa, Madera, Stanislaus, and Fresno counties, was consistently won by Republicans during the 2000 decade, with Republicans winning with margins in the two-party vote of around 15 percent when contested and was uncontested in 2008. Even when the Republican incumbent (George Radanovich) retired in 2008, the district returned a 30 percent margin for the newcomer (Jeff Denham). Electoral margin refers to the difference in vote shares between the winning and losing vote shares between the largest two parties. Thus, a 30% margin reflects a two-party vote split around 65/35.

³⁹ District 31 encompasses a portion of San Bernardino county near north-west Los Angeles. It was won by a Republican incumbent (Gary Miller) in 2012 with a margin of 12%. The seat became open again in 2014, when a Democrat won by a margin of 4% and won re-election in 2016 with a margin of around 15%. District 21 encompasses much of the same area as the previous plans' district 20 did. District 21 has been consistently represented by a single Republican (David Valadao) who ran and won in the open contest in 2012 and returns a consistent margin of around 14%.

Describing any of these districts as “gerrymandered” based on their consistently large (but not ever extreme) Cook’s distances seems premature. Further, the most extreme districts tend to hold for only a single election. Sometimes, this single election is the election in which we know the least about the district: when it is uncontested and \mathbf{h}_i imputed. Since we seek a forensic that can characterize whether a district has a *consistent impact* over the decade, the Cook’s distance of a district in two-cycle models is also too unstable to distinguish any districts as influential over a decade.

7.4.2.3 Impact Measures

For analyzing the impact measures, I will step through three types of analyses outlined in Section 7.3. The first is a nonparametric distributional difference test, designed to indicate where and when the impact statistics differ from the statewide distribution. The second is to identify when each district’s removal (or presence) specifically benefits one party. The final is a measure of both the size and direction of impact when a district is removed. I also consider the pooled and stratified forms of the effect estimate towards the end of the chapter.

To briefly recapitulate the five advantage statistics shown in Chapter 2 that are used in the remainder of the chapter:

- The efficiency gap reflects the difference in parties’ “wasted votes” as a percentage of all votes cast (Eq. 2.5).
- The simple efficiency gap is the difference from the observed \bar{s} and \bar{h} and the line with a slope of 2 through (.5, .5), derived from the efficiency gap when all seats have identical turnout (Eq. 2.6).
- The attainment gap reflects the expected smallest \bar{h} at which the party wins the smallest feasible majority (Eq. 2.9).
- The bonus at median reflects the difference in expected seat share when parties both win 50% of the vote (Eq. 2.8).

- The observed bonus reflects the difference in expected seat share received by Democrats when they win \bar{h} and the expected seat share received by Republicans if they were to win \bar{h} (Eq. 2.7).

Since these measures each reflect different types of advantage, they may consider different districts as impactful or disagree about who benefits. This means that some measures may be nonzero while others are indistinguishable from zero. In addition it may also mean that some measures are negative (biased Republican) and others biased positive (towards Democrats) for the same state. My concern is with the development of the jackknife localization method rather than with the validation of these statewide statistics directly, so potential disagreement between measures themselves is not at issue in this dissertation. However, disagreement about the district *impact* does occur and is relevant to this analysis.

Exploratory distribution analysis and visualization for California is conducted in the chapter appendix, 7.6. This includes both examining the distribution of each of the 53 districts' impact statistics with respect to the statewide distribution, and an examination of the relationship in the effect estimate across measures. If further detail is desired about the structure and relationship of these measures to one another (and the performance of statewide indices more broadly), refer to this appendix.

Otherwise, let us proceed informed by a few general concepts. First, the relationship between the impact statistics' effect size estimates indicates that the two decades tend to have quite different distributional structure, with the 2000 plan having much more bimodal distributions than the 2010 plan. This indicates the 2000 plan was "balanced" around its average bias, which tends towards Democrats in most measures and years. In contrast, the partisan separation is not as stark in the 2010 plan. Second, there is a clear positive moderate to strong rank correlation in many years between measures. Almost all rank correlations between measures (except for the attainment gap) had significant rank correlation. Further, the variance of district realizations within a year tends to be quite stable. The jackknifed distributions tend to have similar dispersion to the statewide distribution. This dispersion is more different between years

than within any year, meaning that *all* of the jackknife distributions within a year are more similar to one another than they tend to be to another years' simulations. So, jackknife distributions tend to fall quite close to the statewide distribution, exhibit similar dispersion within years, and tend to concentrate into peaks that clearly benefit Democrats or Republicans.

7.4.2.4 Are the Jackknife Distributions Meaningfully Distinct from the Statewide Distributions?

For each jackknife distribution, I conduct a Mann-Whitney nonparametric distribution dominance test to identify whether it is distinct from the distribution of statewide bias statistics. The Mann-Whitney test examines the pair of distributions and indicates whether the two rank distributions constructed from the source distributions are likely to share a common distribution. In Figure 32, the significance of a p-value at the .05 α level (with a Bonferroni correction) for the Mann-Whitney dominance test is shown. A black cell indicates that the district had a significantly-different distribution than the state in every year after correcting for multiple comparisons. What becomes immediately apparent is that the median bonus and the efficiency gap recover every district as "distinct" from the statewide distribution when incumbents are used in simulation. But, the measures disagree about which districts are distinct. Further, the incumbent simulations tend to identify more districts than the simulations without incumbents. This illustrates a claim made in Gelman and King (1994a), that the partisan advantage of a districting plan is strongly dependent on the structure of incumbency advantage in that state, and that failure to control for incumbency may make partisan bias appear. Simulating the expected advantage *absent incumbency* gives us one way to do this, but the measures thereof are inherently unrealistic in the sense of McGhee (2014): the characterization of partisan advantage in a never-realized "no incumbent" scenario may not reflect the *experienced* partisan advantages that the system provides. Regardless, what this makes apparent is precisely the magnitude of difference between simulations under observed conditions and simulations in the no incumbent counterfactual.

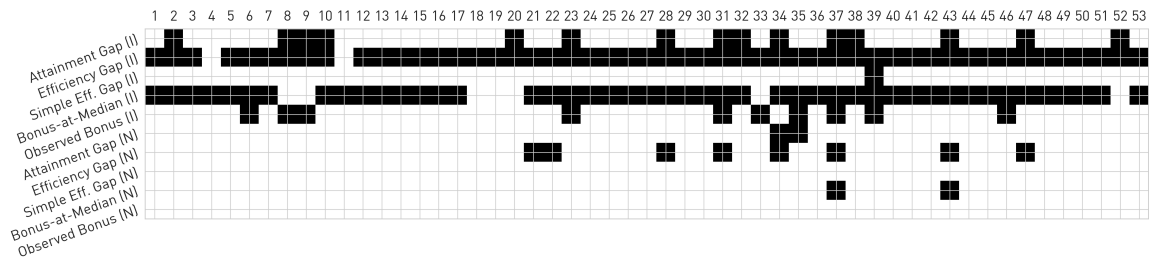


Figure 32. Significance results for the Bonferroni-corrected Mann-Whitney U distribution dominance tests for California congressional district impact measures in the 2002-2010 congressional elections. If the cell is black, it indicates that district (marked on the horizontal axis) had a significantly different deletion distribution than the statewide distribution for the given statistic (marked on the vertical axis) under a Bonferroni correction factor $f = N \times T \times Q \times S = 53 \times 5 \times 5 \times 2$. Measures suffixed with “(I)” are simulated with the observed incumbency structure and measures suffixed with “(N)” are simulated with no incumbents.

For the 2010 plan, the analogous significance diagram is shown in Figure 33. This test set uses a smaller correction factor, since it entails only 3 years of elections (rather than 5). In general, fewer districts are selected as significantly different from the statewide distribution in the 2010 decade than the 2000 decade, which implies that the 2010 redistricting provides a political advantage that is less sensitive to any individual district. Again, the no-incumbent simulations tend to find fewer districts than the incumbent cases. District 21 appears particularly egregious, selected by three measures as distinct in the incumbent simulation case and by two measures in the no-incumbent case. More generally, it seems the jackknife measures *do* differ substantially from the statewide distribution, so characterizing *how* they differ would be helpful.

7.4.2.5 Examining Influence Direction

To characterize how the jackknife distribution is different from the statewide distribution, I focus first on the binomial sign test procedure outlined in Section 7.3.2. The results of the sign test for the 2002-2010 congressional elections is shown in Figure 34, and for the 2012-2016 congressional elections is shown in Figure 35. In these figures, a cell is colored to correspond to

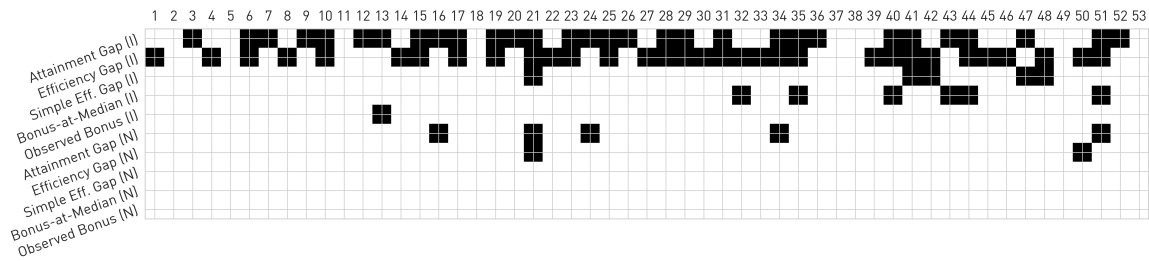


Figure 33. Significance results for the Bonferroni-corrected Mann-Whitney U distribution dominance tests for California congressional district impact measures in the 2012-2016 congressional elections. If the cell is black, it indicates that district (marked on the horizontal axis) had a significantly different deletion distribution than the statewide distribution for the given statistic (marked on the vertical axis) under a Bonferroni correction factor $f = N \times T \times Q \times S = 53 \times 3 \times 5 \times 2$, since there are only three observed elections since the 2010 redistricting. Measures suffixed with “(I)” are simulated with the observed incumbency structure, and measures suffixed with “(N)” are simulated with no incumbents.

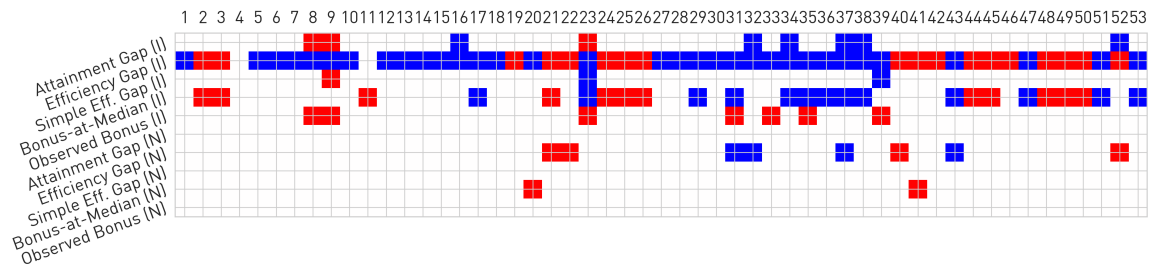


Figure 34. Directionality results for the Bonferroni-corrected binomial sign tests for California congressional district impact measures in the 2000-2010 districting plan. If the cell is white, it indicates that the district did not consistently move the plan towards either party. If the cell is red, the removal of the district shifted the advantage towards Republicans in every year & if blue, Democrats. Inverting this partisan relationship would provide the impact of *including* the district in the plan.

the districts’ consistent partisan impact. If a cell is white, the district has no consistent impact on that statistic. If a cell is red, the district’s jackknife distribution is substantially more Republican than the statewide distribution; if blue, the jackknife distribution is substantially more Democrat, at a Bonferroni-corrected $\alpha = .05$ level. Thus, this characterizes how the district affects the state *when removed*. The reverse of this relationship is the impact of the district’s *inclusion* conditional on retaining the rest of the districts. In this sense, a district whose removal benefits Democrats is one whose deletion distribution stochastically dominates the statewide distribution and is also one whose presence benefits Republicans given the rest of the plan.

In Figure 35, corresponding to the 2002-2010 elections, most districts are classified consistently if they are classified at all. There are notable exceptions to this, however. The efficiency gap differs from the rest on districts 8 & 9, finding their removal shifts the state towards Democrats when all other measures find them negligible or shift-Republican. All measures classify district 23 as having an impact in the incumbency case, but disagree about the direction of influence. The difference in reasoning about how each district affects the impact statistics provides a useful way to understand how the measures operationalize advantage differently. So, I discuss the 9th district in detail.

7.4.2.6 Digging Deeper Into Impact

Examining the 9th district during 2002-2010, the district was a strong Democratic district with a notable incumbent, Barbara Lee. Critically, note that these directionality measures refer to the *deletion* distribution; this means that the districts' *inclusion* in the plan has the reverse partisan effect. One might anticipate that a California delegation *without* Barbara Lee would be a more Republican delegation, and so the *inclusion* of her district should benefit Democrats. However, the efficiency gap impact statistic suggests instead that the inclusion of Barbara Lee's district wastes a significant amount of Democratic votes.

Recall that the efficiency gap from Equation 2.5 is driven by three factors: the total votes cast, the lost vote cast for losers, and the excess vote cast for winners who do not need them. The inclusion (removal) of any district can only increase (decrease) the first and increase (decrease) the two latter factors. The latter two factors will also apply to *one* party for each district, since the winner will change the excess vote and the loser will change the lost vote. The difference between a districts' excess and lost vote provides the *differential* in that district (numerator of Equation 2.5) and its turnout affects the denominator of Equation 2.5. Marginally, the presence of a district won by Democrats will (on average) bias towards Democrats when $.5 < \mathbf{h}_i \leq .75$, since votes cast for losing Republicans outnumber the excess vote won by Democrats. When $\mathbf{h} > .75$, the presence of district benefits Republicans, since they cast

fewer votes than excess Democrats waste in districts at this vote share range. Thus, Lee's district, with $h_i \approx .85$ for the entire decade, typically has a differential benefiting Republicans in simulations, since its simulated Democratic vote share lies mainly in the 80s. Put simply, the efficiency gap suggests the district wastes too many Democratic votes, so Democrats are better off without Lee's district.

Further, in a state with a system-wide bias towards a party, the inclusion of a "neutral" district (whose differential is zero) will reduce the statewide efficiency gap, since it increases the denominator of Eq. 2.5 while keeping the numerator constant. Since California's plan during the 2000s exhibited a statewide Democratic bias, the presence of Lee's district pulls the efficiency gap towards Republicans with *both* its Republican vote waste differential and increase in total vote. Thus, it is *also* a de-biasing district: its presence makes the plan more balanced by reducing the statewide Democratic advantage.

In contrast, the attainment gap suggests that Lee's district is a boon for Democrats, since it provides a safe seat in the congressional delegation. When Democrats win a bare majority of the California congressional delegation, the district tends to be a part of the majority. Thus, its presence lowers the fraction of votes required to win a bare majority, and removing it drops a safe Democratic seat in the typical minimal majority.⁴⁰ Along similar lines, the removal of Lee's district benefits Republicans according to the observed bonus measure: if Democrats were to win as many votes as the Republicans do statewide for congress, they still tend to win Barbara Lee's seat in simulations. So, its removal would harm Democrats during the tables-turned counterfactual.

In addition, it is important to note district 23 during the 2000s. This district is selected by all of the directional statistics as having an impact on the state, although measures disagree whether its existence benefits Republicans or Democrats. Critically, district 23 was a highly irregularly-shaped district, stretching in a thin strip along the coast from Ventura to north of San Luis Obispo. That the district shows up as directionally-influential in all statistics is notable,

⁴⁰Further, the attainment gap is likely influenced slightly by seat integrality. Dropping Lee's district reduces the number of districts by 1, but does not decrease the minimal majority, which is 27 in either the full or jackknifed case.

especially since it is the only district to be selected by all of the measures in the incumbent simulation scenario. However, the district is not selected by *any* of the non-incumbent simulation sets, meaning its partisan impact may be solely due to incumbent effects.

In light of this, measures are influenced by different types of districts. However, the efficiency gap and the median bonus measures *never* disagree on their directional classifications for districts in elections from 2002 to 2010: if a district is identified as having an impact with consistent direction by both measures, they never disagree on the direction the jackknife impacts the state. In addition, only one district identified by the median bonus statistic is not identified by the efficiency gap. There is an even split of districts identified as shift-Democrat and shift-Republican in most of the measures under the incumbent simulations, although the observed bonus only identifies districts whose removal benefits Republicans. In the no-incumbent simulations, no one district is identified as directional by any two measures. The attainment gap, simple efficiency gap, and observed bonus measures identify no districts as having a directional impact when incumbent effects are removed. Notably, the efficiency gap identifies some districts as having a partisan impact under the no-incumbent case, and the direction of this impact is the same as when incumbents are used in simulation. Since the impact is observed regardless of incumbents, these districts confer a durable partisan advantage.

For the second decade shown in Figure 35, the picture is quite different. Many more shift-Republican districts are identified. This means that the plan has more districts whose *inclusion* benefits Democrats conditional on the rest of the plan. The attainment gap identifies these types of districts exclusively, regardless of the incumbency controls. In addition, the median bonus measure never identifies a California congressional district in the 2012-2016 elections as directionally impactful, even though the Mann-Whitney tests suggest that some districts are distinct from the statewide distribution.⁴¹

Notably (again) district 21 is a shift-Republican district in both incumbent and no incumbent

⁴¹This is likely because the removal increases the variance of the outcomes, which would cause the deletion distribution increase its mass into extreme ranks of the pooled distribution against which the *U* test is constructed. Thus, it is possible for two distributions to be distinct in rank sets but not directionally-different (i.e. a non-equality, non-dominance relationship)

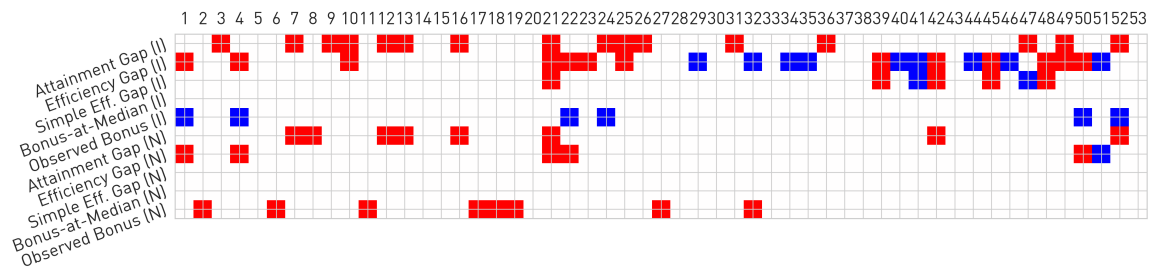


Figure 35. Directionality results for the Bonferroni-corrected binomial sign tests for California congressional district impact measures in the 2010-2020 districting plan. If the cell is white, it indicates that the district did not consistently move the plan towards either party. If the cell is red, the removal of the district shifted the advantage towards Republicans in every year & if blue, Democrats. Inverting this partisan relationship would provide the impact of *including* the district in the plan.

simulations. District 21 is Devin Nunes’s district, a reliably-Republican district with a margin of around 15% (when contested). Its removal benefits Republicans, making the plan less biased overall (since the statewide bias distribution indicates a bias towards Democrats in all statistics). Alternatively, one can interpret its *inclusion* given the rest of the plan as benefiting Democrats. Classifications of the district impact differ often between the efficiency gap and the observed bonus measures; each district identified by the observed bonus measure as shift-Democrat is identified by the efficiency gap as shift-Republican or no impact. In addition, the attainment gap and efficiency gap again select some districts that have consistent impact on partisan advantage in the state, regardless of the inclusion of incumbents. Like in the 2000s, these are likely gerrymanders, since they strongly and consistently influence partisan advantage. Lastly, only one district, district 51, is identified as shift-Democrat in the no incumbent simulations. Since this is a safe Democrat (majority-minority) district, its inclusion tends to create a safe Democratic district in simulations. This gives it a rather large number of wasted votes, picked up by the efficiency gap. All the remaining districts identified in the no incumbent scenario are shift-Republican, or identified as districts whose inclusion provides advantage to Democrats.

7.4.2.7 Which Districts are Truly Beyond the Pale?

When it comes to the magnitude and direction of impact a district has on the system, I move to consider the effect estimates. This presents a similar view to the binomial sign test, but now characterizes both who benefits & how much they benefit. If the cell is colored, then the effect is “larger than marginal,” so that $|d| > .25$. Then, effects with $|d| > .5$ are colored slightly darker, and effects where $|d| > 1$ are darker still. I first consider each election year separately, illustrating the patterns in each years’ impact estimates. Then, I will compute a pooled- and stratified-effect estimator, which groups observations by decade and computes an effect over the entire decade. The exact method by which this pooling or stratification is presented, and then the pooled analyses conducted.

7.4.2.8 Yearly Impacts in California Since 2000

In the yearly analysis mode, a few things become clear about the effect size estimates. These insights echo the conclusions from the nonparametric and sign-only analyses. The efficiency gap and bonus at median effectively take the same perspective on most districts in most years, and the gist of the classification for each district provided by either measures is the same in all cases. In addition, the measures *also* conform exactly to the partisan winner in each district. Thus, reading the chart across in 2002 for either the efficiency gap measure or the median bonus measure provides the losing party in that district. Since these are deletion effects, the removal of the district benefits the opposing party; blue districts were districts won by Republicans so their removal helps Democrats. In general, the classification provided by the efficiency gap is strictly partisan, and nearly useless in terms of how the impact measure “filters” a given district: most districts have consistently large effect sizes to be considered not “marginal” within each year.

The attainment gap, though, appears to reflect something substantively different from the other measures. The determination of its directional effects do not simply follow the winner in

the district. Some of the districts identified as shift-Republican for the efficiency gap/median bonus are identified as shift-Democrat by the attainment gap, but *not all* are (e.g. districts 31,32); likewise, some shift-D districts for the efficiency gap/median bonus impact statistics are shift-R for the attainment gap (e.g. districts 4,52), but not all (districts 19, 11). Notably (again) the observed bonus measure tends to agree with the attainment gap impact estimate in most cases, although the effect size is smaller. In the no-incumbent simulations (Figure 37), a similar story holds: some effect estimates for the attainment gap are nearly the reversed estimates for the efficiency gap/median bonus again, and some are not.

For the 2010 decade simulations shown in Figures 38 & 39, there tends to be slightly better agreement between the attainment gap and efficiency gap, but the correspondence between the efficiency gap and the median bonus breaks down. In addition, fewer districts have a consistently-large effect; most districts (at some point) have a negligible impact measure. While many districts are identified in the simulations under observed conditions, no districts are identified under no incumbent conditions for all years⁴² and one district is identified as having a large but inconsistent impact (district 24, attainment gap). Regardless, depending on a persons' chosen measure of bias, the impact statistics provide a consistent directional characterization of a district's impact on the partisan advantage of a districting plan over a decade. So, measures tend to either be all shift-R, all shift-D, or fade between one color and white. It is rare for a district to have a district be both shift-R and shift-D in one decade, but this occasionally does happen (e.g. district 20). While each measure tends to be consistent over time, they are often disagree with one another. Thus, the impact measures considered in this dissertation appear to be *internally consistent*, repeatedly characterizing a district with respect to its own operationalization of advantage, but they are not *externally consistent*, agreeing on how districts impact a plan across measures.

⁴²Although, considering only years where both h_t and h_{t-1} are available changes this. These are the elections *not* immediately following a redistricting.

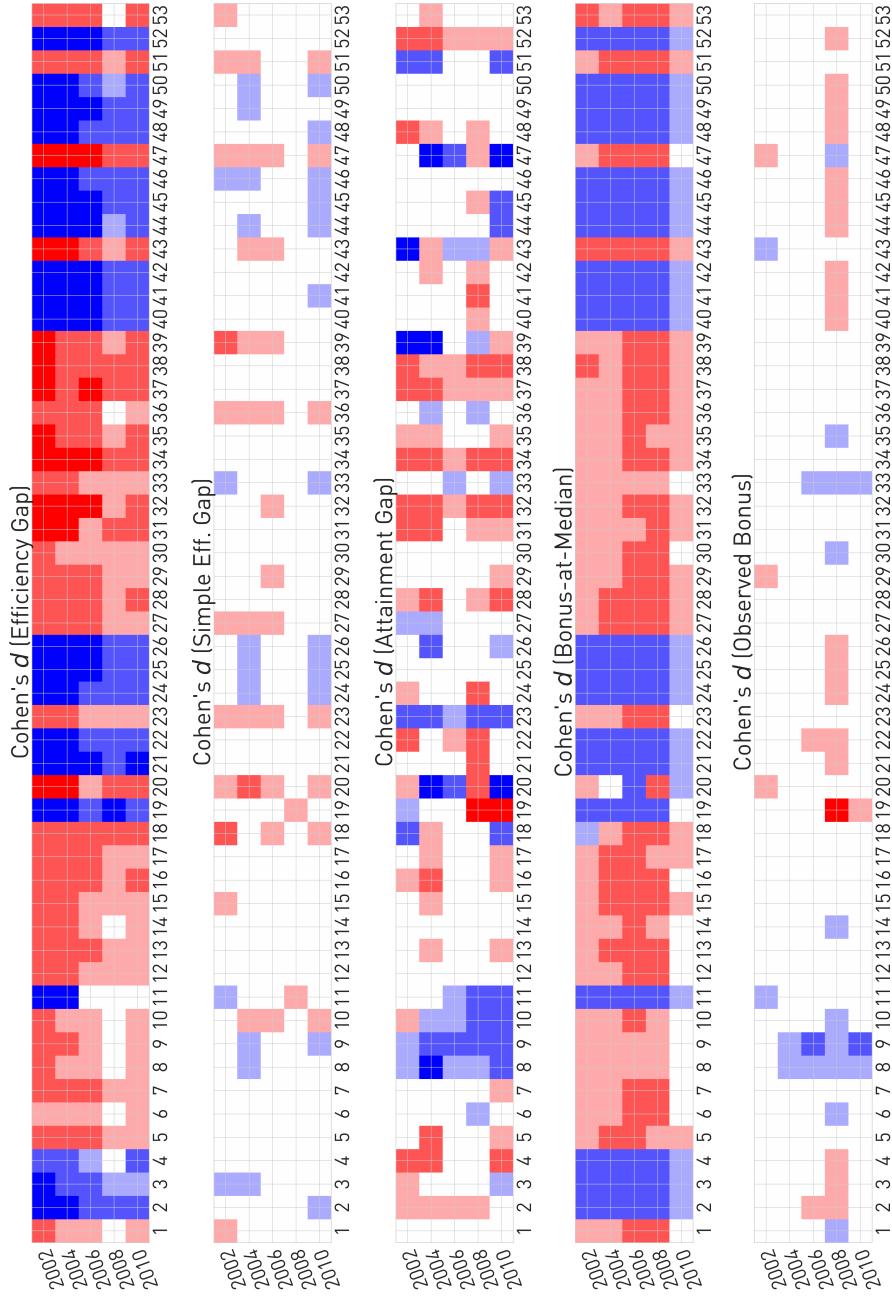


Figure 36. Effect estimate for the five measures during elections in simulations under observed conditions for the 2000 districting plan in California. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue benefit Democrats. The direction of benefit *conditional on inclusion* is the reverse.

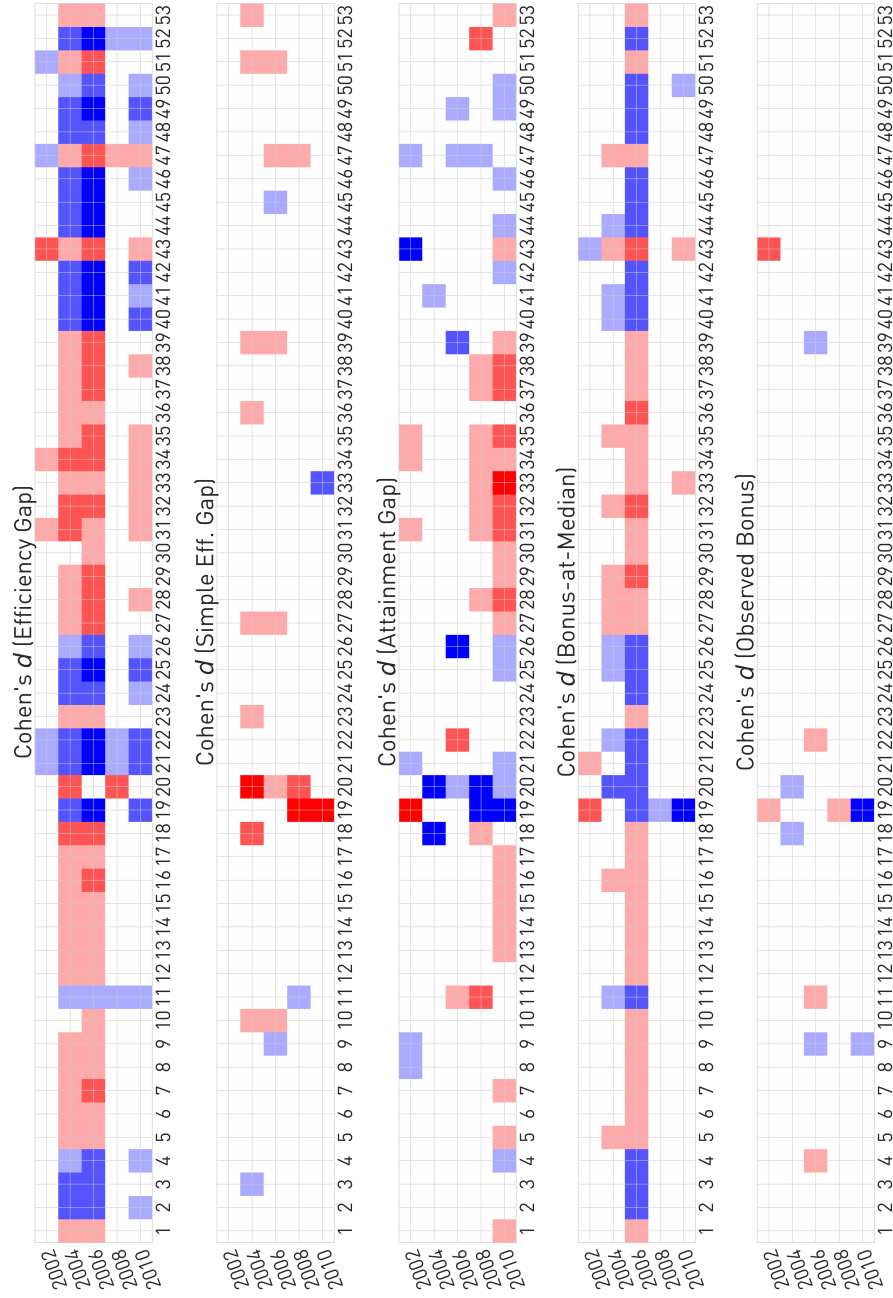


Figure 37. Effect size for the five measures during elections in simulations with no incumbents for the 2000 districting plan in California. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue Democrats.

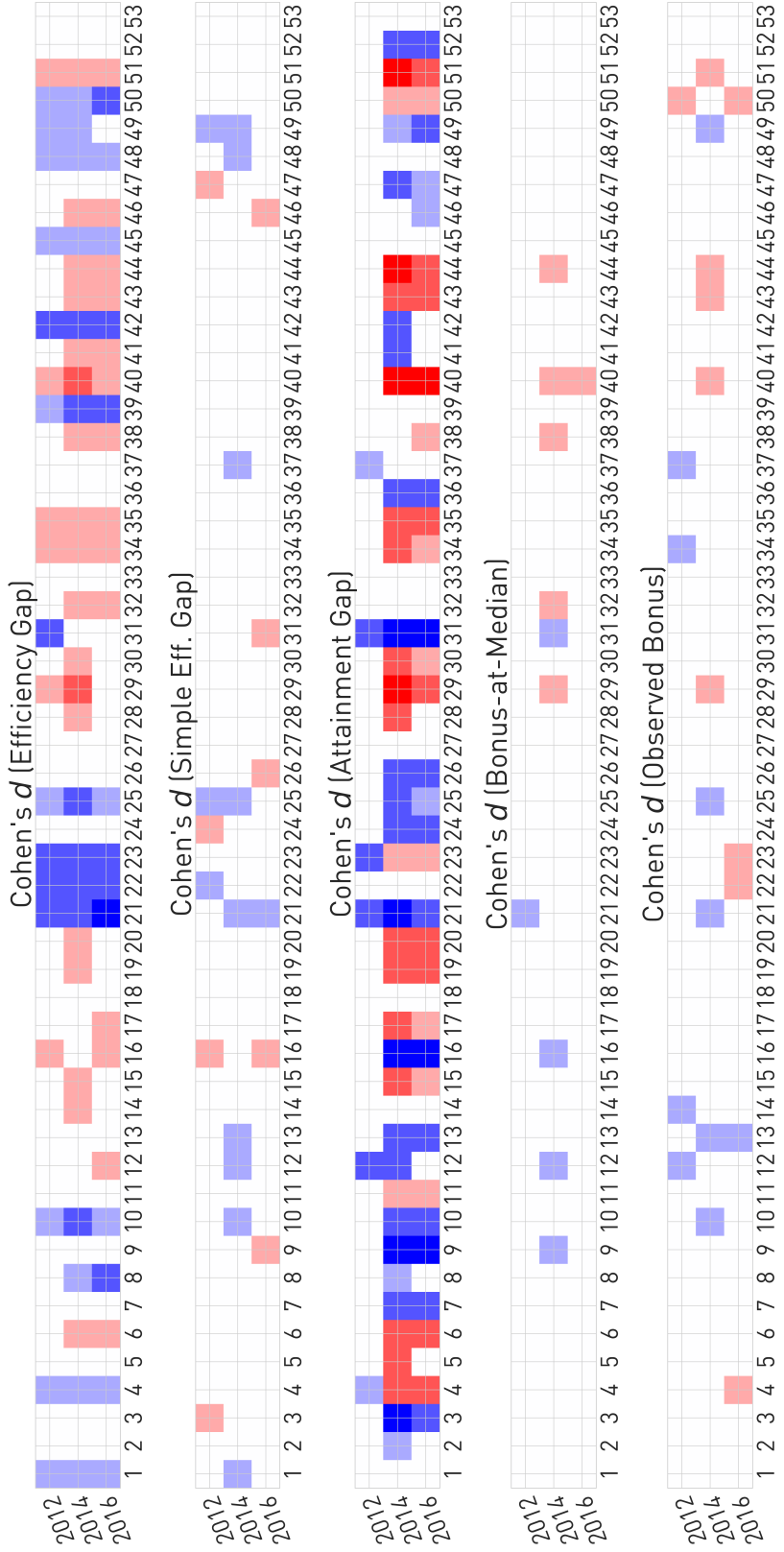


Figure 38. Effect size for the five measures during elections in simulations under observed conditions for the 2010 districting plan in California. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue Democrats.

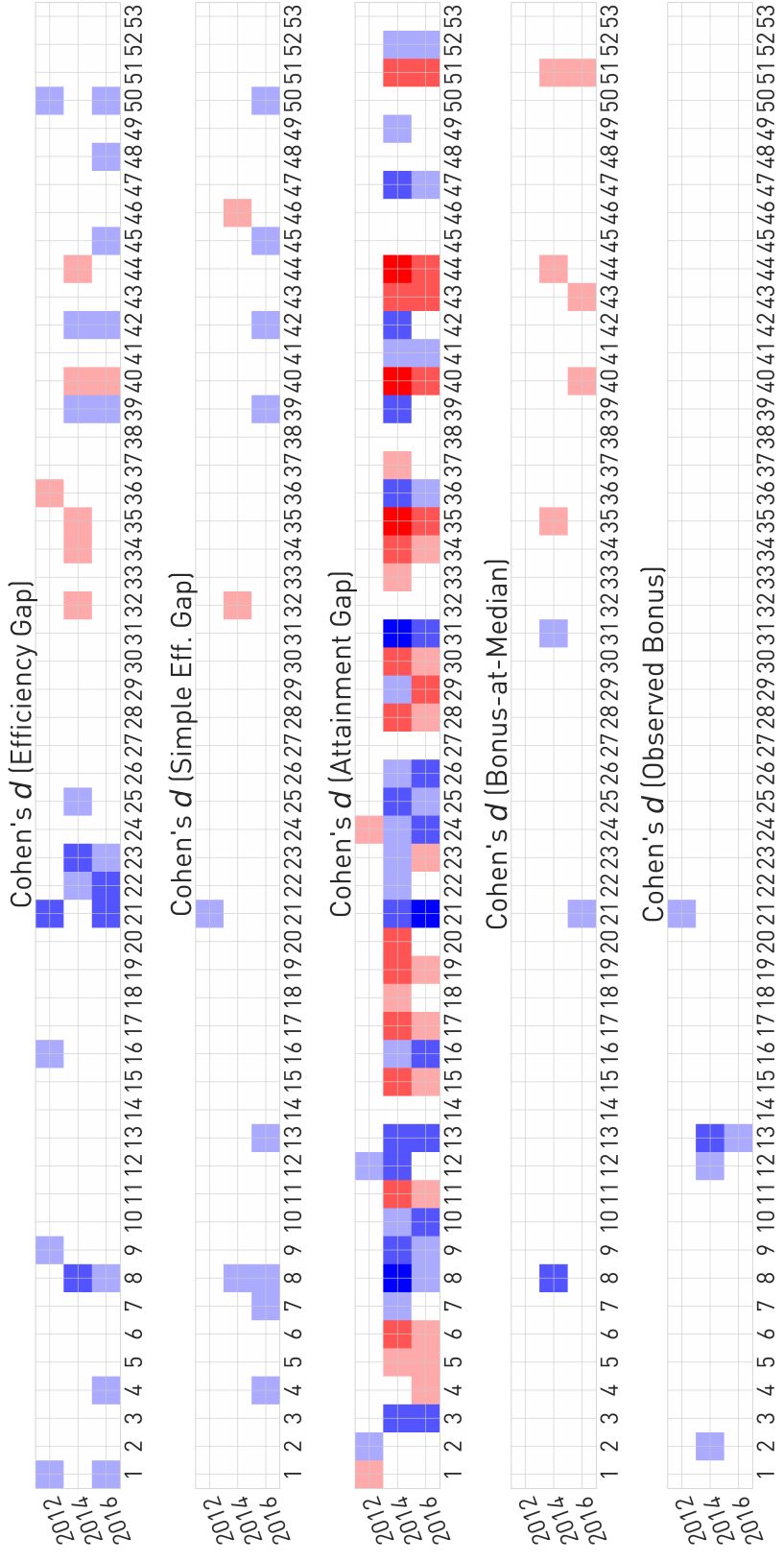


Figure 39. Effect size for the five measures during elections in simulations with no incumbents for the 2010 districting plan in California. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue Democrats.

7.4.2.9 A Grand Decadal Estimate of District Partisan Impact

To reduce the number of comparisons and make the conceptual picture clearer, I examine a “grand” estimate of effect size. This provides a single estimate of district impact over the entire decade, rather than an estimate in each year. I see two ways to construct this full-decade estimate of district impact. One method is through pooling. The pooled estimate is of the same form as the estimator in Eq. 7.6, with the statewide and deletion distributions pooled over all years in a decade. This simply treats the T distributions with K replications of the deletion statistic as if it were a single $T \times K$ distribution of statistics. Thus, the grand decadal means and variances are used to compute a grand decadal effect. This ignores the fact that each years’ mean (and districts’ impacts) are distinct, reducing the magnitude of the effect estimate.

In contrast, a stratified estimator takes into account the temporal differences while incorporating information together across years. Stratification acknowledges the fact that each year may have a distinct mean, and that a consistent estimate of the effect size should be made relative to that years’ mean (given an appropriate correction for the variance over years). One such stratified estimator of effect size is a modification of Eq. 7.6 using within-year differences and a stratified estimator of variance:

$$d_{(i)} = \frac{\sum_t^T (\hat{b}_t - \hat{b}_{t(i)}) / T}{s^*} \quad (7.8)$$

$$s^* = \sqrt{\frac{\sum_t^T (\sigma(\mathbf{b}_t)^2 + \sigma(\mathbf{b}_{t(i)})^2)}{2T}}$$

This provides the overall effect of removing the district within a decade, accounting for the fact that each years’ mean may not be equal. Essentially, it uses a grand mean of a districts’ difference-in-means for the entire decade, divided by the square root of the average pooled variance estimator from the standard Cohen’s d . Using this stratified “mean-of-differences” effect size estimate, the stratified scores are perfectly linearly related to their unpooled counterparts,

but the stratified effect sizes are more extreme, since the difference in each year's mean is accounted for.⁴³

Thus, I present both the pooled and stratified effect size estimates for districts in the 2000 plan (Figure 40) and the 2010 plan (Figure 41). In this, it again becomes clear that the efficiency gap and median bonus impact measures essentially recover all districts and the removal (inclusion) of each district benefits the party that lost (won) the district in effectively all cases. Again, the relationship between the attainment gap and efficiency gap is inconsistent. Regardless, the stratified/pooled method allows us to identify a few districts that are impactful on the statewide scores. In the 2000s, district 19 again becomes a district-of-interest. The district, represented consistently by Republicans who win with margins at-or-above 30%, has an impact on the statewide bias measure for more than one measure in both incumbent and no-incumbent simulations, indicating it is impactful. In addition, district 20 is also identified as a potential impact district, with two measures identifying it in no-incumbent simulations.

In the 2010s, districts 21, 40, and 44 become notable. Again identified as consensus impactors in previous steps, these districts are further filtered out when they are identified by multiple measures regardless of pooled/stratified, incumbent/no-incumbent simulation structures. District 21's removal consistently benefits Democrats.⁴⁴ California's 40th and 44th districts both are marked as districts whose removal benefits Republicans.⁴⁵ Thus, for nearly all measures, the districts' inclusion in the plan provides a strong benefit for Democrats, and their removal

⁴³A more involved method to do this comparison while capturing potential covariance between years would be to model the temporal dimension hierarchically (Kruschke, 2013). This would induce further shrinkage on the stratified estimate, likely placing it somewhere between the pooled and stratified values according to the variance of that difference.

⁴⁴Represented by Republican David Valadao, the district is one of the few districts that split between the congressional and presidential vote in 2016 *and* 2012. The district returns reliable Democratic majorities in state and national races, but Valadao won the open contest in 2012 with around a 15% margin, and wins re-election with nearly the same.

⁴⁵ These districts are both Democrat-only districts; since the blanket primary and redistricting both occurred from the 2012 elections onwards, the districts have not had a two-party contest since they were drawn.

strongly benefits Republicans. Altogether, the districts that “fall out” of the analysis as significant would be district 19 in the 2000s and districts 21, 40, and 44 in the 2010 plan.

7.4.2.10 Relationship Between Classical Influence and Impact in California Since 2002

Notably, district 19 in the 2000s plan and district 21 in the 2010 plan have larger-than-typical classical leverage/impact measures as shown in Figure 30. District 19 is a strongly-outlying, with a Cook's distances so large that its value required truncation to display in Figure 31. However, district 21 does not have remarkably high leverage or large Cook's distances in any of the years during the 2010 plan. Further, the fact that *some* districts selected by the classical measures are also impactful on the advantage measures does not help characterize those high-leverage/outlier districts with no impact. Thus, I show the relationship between the classical impact measures and the jackknife effect estimates in Figure 42 for the 2000s and Figure 43 for the 2010s. The relationships between measures are decidedly weak; while some of the pairwise correlations pooled over the entire decade are statistically significant, all are of marginal or weak magnitude ($\tau < .25$). Some of the correlations *within* years are significant and modestly large ($\tau < .5$), but they're not consistent over time. The closest to consistent significance comes from the relationships between the simple efficiency gap and model residuals; in the 8 elections under study, the relationship is significant ($p < .05$) in six of them, varying in strength between .19 and .55. In addition, each measure has a nonzero rank correlation with the Cook's D in two out of three years in 2010s, with some pairing 2012 & 2014 and others 2014 & 2016. Regardless, the aggregate relationship between the new measures of partisan impact and classical sensitivity measures is weak & inconsistent, suggesting they provide different information about the state's districting plan.⁴⁶

⁴⁶An exploratory regression analysis was also carried out relating each of the classical measures to the set of 5 impact measures. In general, the relationships were quite weak, but it was notable that the *magnitude* of the effect size tended to be marginally more related to classical influence/leverage measures than the raw estimate.

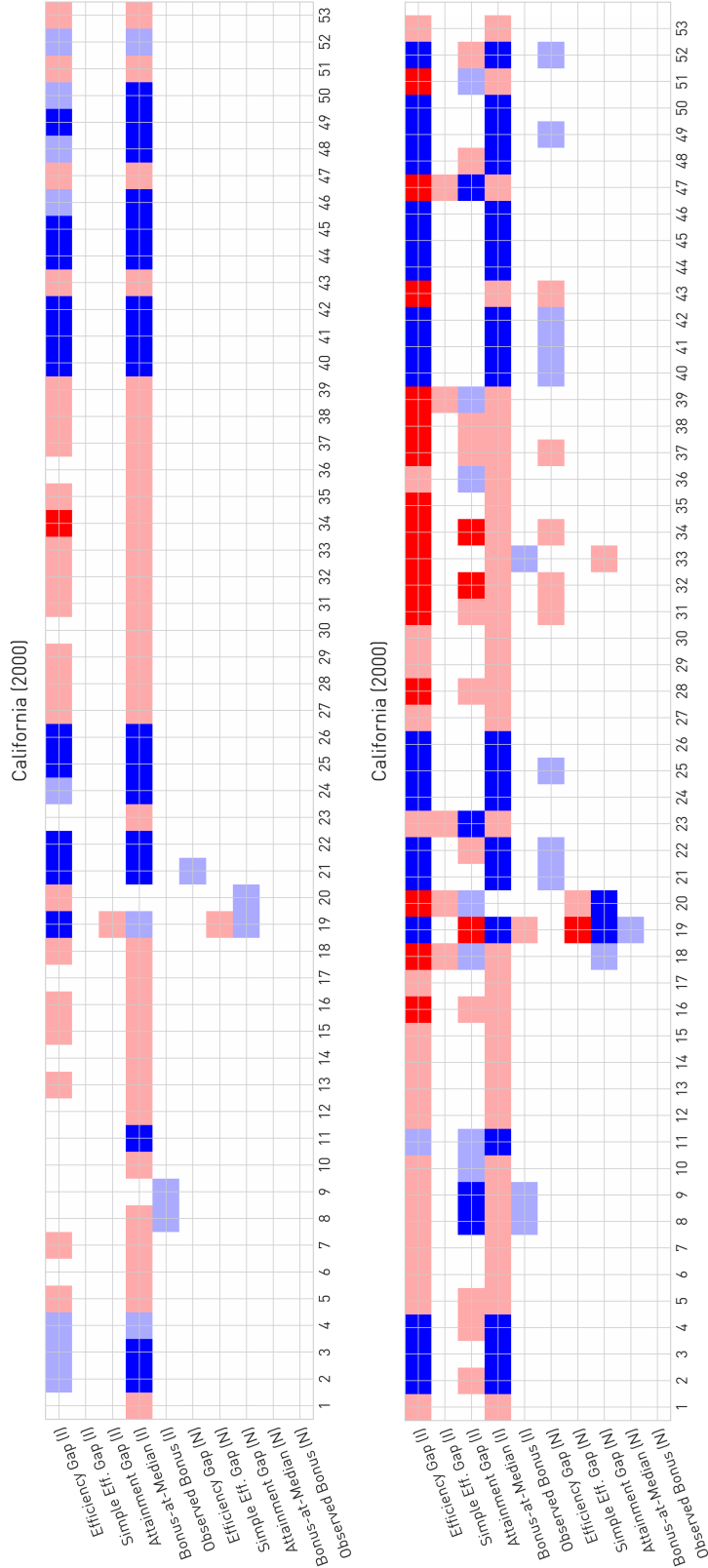


Figure 40. Grand effect size estimates for the five measures during the 2000 districting plan. Pooled estimates are on top and stratified estimates on bottom. Measures ending in "(I)" indicate simulations under the observed conditions and "(N)" with no incumbents. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue Democrats.

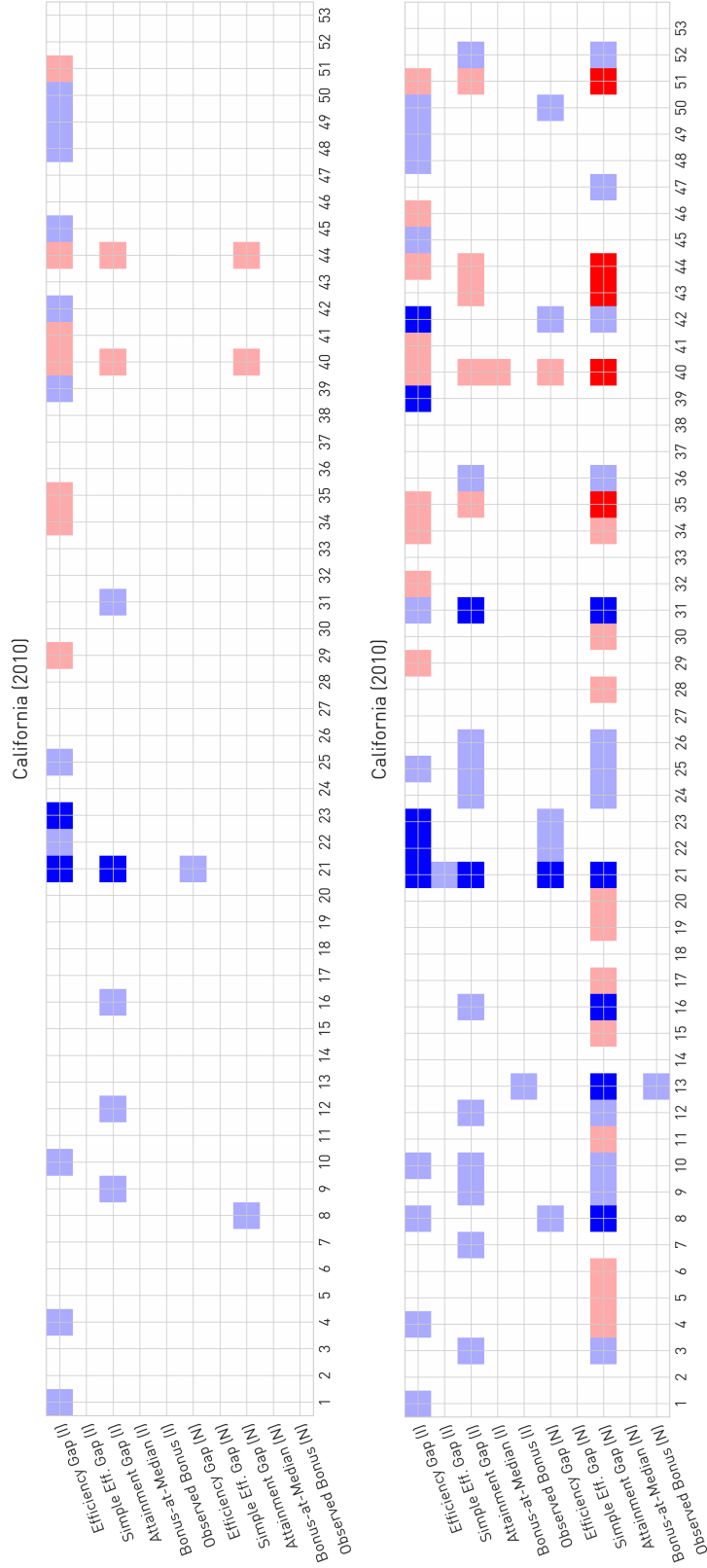


Figure 41. Grand effect size estimates for the five measures during the 2010 districting plan. Pooled estimates are on top and stratified estimates on bottom. Measures ending in "(I)" indicate simulations under the observed conditions and "(N)" with no incumbents. The darkest colors have a $|d| > 1$, with the second tier having $|d| > .5$ and the third $|d| > .25$. District-years with small effect sizes ($|d| < .25$) remain white. Red districts are districts whose removal benefits Republicans and blue Democrats.

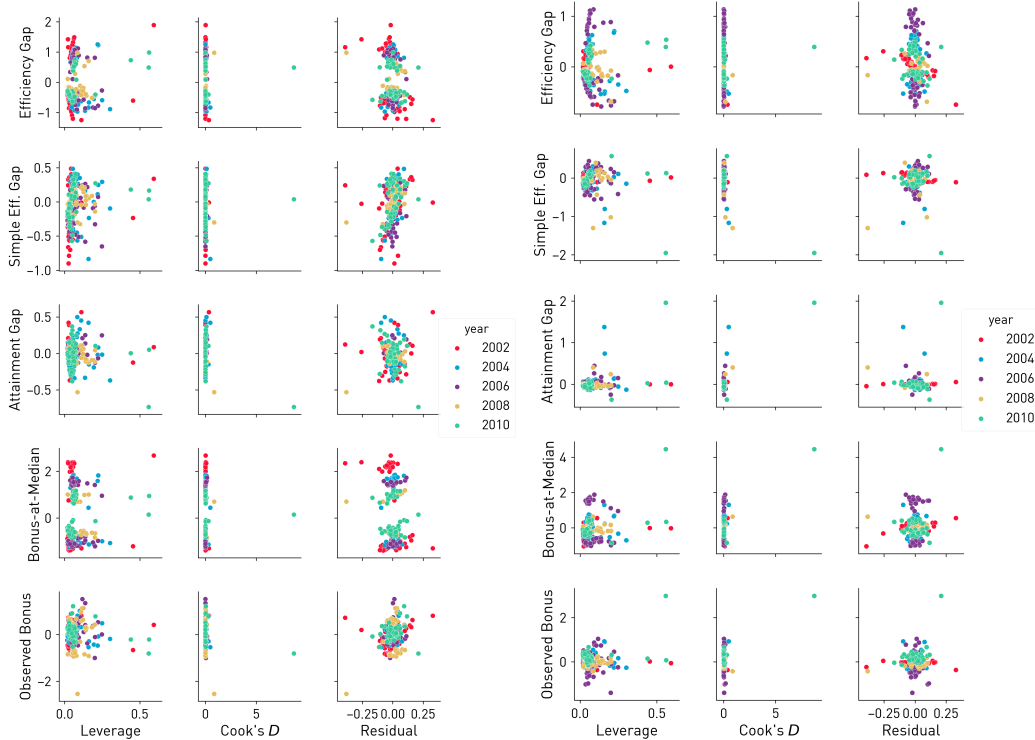


Figure 42. The relationship between the classic measures of model structure and jackknifed measures of district impact in elections held using the 2000 districting plan. On the left, the relationship for realizations under observed conditions are plotted; on the right, simulations with no incumbents are shown.

7.4.3 Case Study: Wisconsin, 2010

Before drawing general conclusions, I present a smaller case study, that of Wisconsin’s congressional districts since the 2010 redistricting. The districting plan currently under litigation in *Whitford v. Gill* involves the state legislative districts generated for Wisconsin in the 2010 redistricting. Critics allege that the plan creates a strong, durable advantage for Republicans using the efficiency gap measure of McGhee (2014). In the nearly 100-member state assembly, the modeling techniques used above in the California case study become even further empowered, since nearly 100 districts are available in each cycle to fit, the model can be much more robustly specified. A relatively thorough analysis conducted by Jackman (2017) examines sets of state legislative districts. in order to develop a threshold value for which “extreme” efficiency gaps

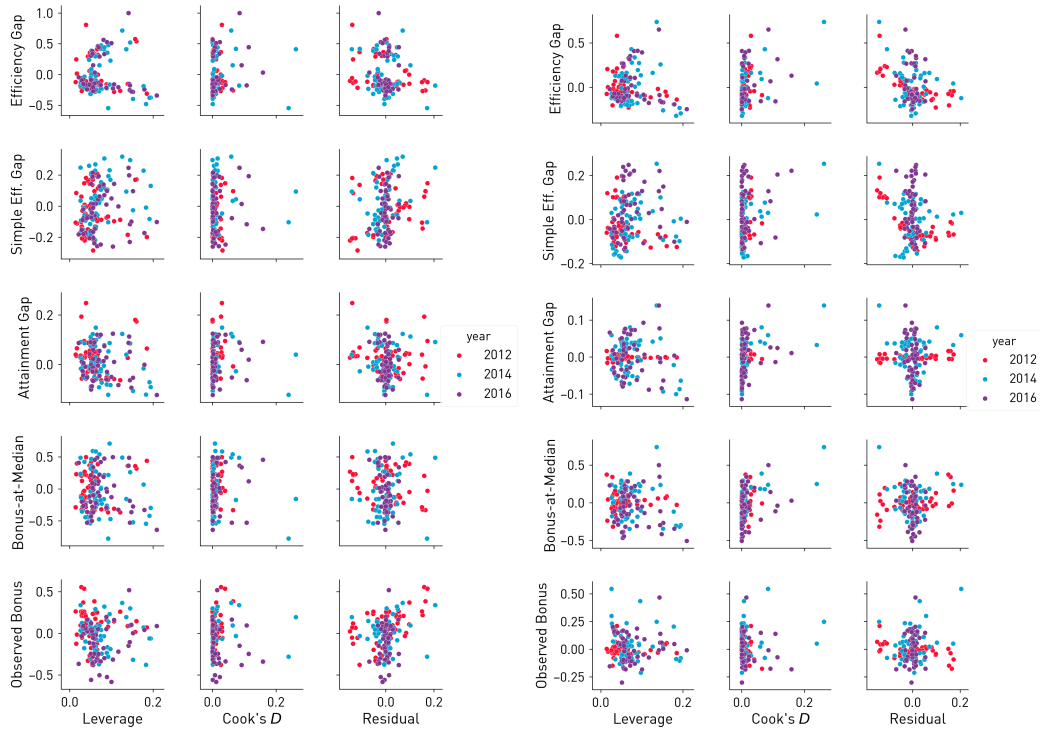


Figure 43. The relationship between the classic measures of model structure and jackknifed measures of district impact in elections after 2010. On the left, the relationship for realizations under observed conditions are plotted; on the right, simulations with no incumbents are shown.

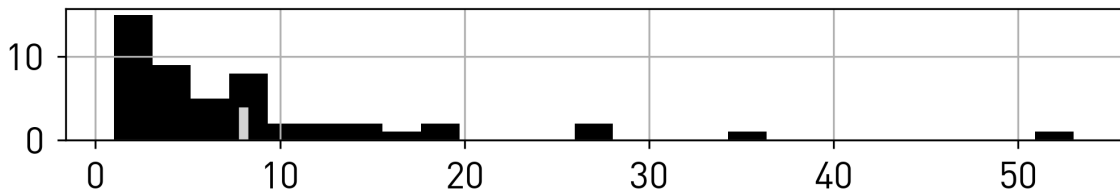


Figure 44. Apportionment histogram showing the number of districts in each state for the 2010 redistricting. California, out on the far right, has 53 congressional districts. Marked by the grey line slightly above the median, Wisconsin's 8 districts makes it a decidedly small- n problem, but places it well above the number of observations available in some states.

can be differentiated from non-extreme gaps at the state legislative level. However, allegations about the congressional district plan have also been made, namely that the congressional map is also tilted unacceptably towards Republicans. In what follows, I analyze that plan.

However, a typical barrier to seats-votes analysis in many states' congressional district plans is that the number of districts may be quite small. For illustration, the apportionment histogram showing the number of districts in each state is provided in Figure 44. California has the most congressional districts (53), while most states have under 10 districts. Wisconsin has eight districts, two more than the national median. This means that each year's distributional model for Wisconsin congressional elections only takes into account eight observations.

What I am interested in identifying with this example is whether the procedure I have suggested above is powerful enough to identify differences *even in small numbers of districts* when the underlying vote model fit is strong. Part of the benefit of the Gelman-King approach is that the counterfactual simulation distribution (Eq. 5.4) has a lower variance than the predictive distribution (Eq. 5.6), since the counterfactual is shrunk towards the observed result. In general, this leads to more precise counterfactual simulations, as discussed in Chapter 5, and the use of the previous' years election returns in the two-cycle model structure provides quite good model fit. In fact, the R^2 for the \mathbf{h}_t models in 2012, 2014, and 2016 are .82, .95, and .99, respectively. Thus, the models have incredibly strong fit. This is a function of the stability of congressional elections, since \mathbf{h}_{t-1} is an exceptionally strong predictor of \mathbf{h}_t . Notably, the two-cycle models, for 2014 and 2016, do markedly better than the first, single-cycle model. This occurred in California as well, but to a smaller degree. Thus, despite the fact that the models are only over 8 observations, the simulations of elections results are quite precise.

What this *does* mean is that the leverage indicators for the model are somewhat uninformative. With only 8 observations, all observations will have relatively high leverage. In addition, it is nearly impossible to identify an outlier in 8 observations using a Cook's distance (or a similar style of measure). However, the new impact measures will accurately incorporate the strength of the effect *and* the variance of the underlying measure given that only 8 observations are

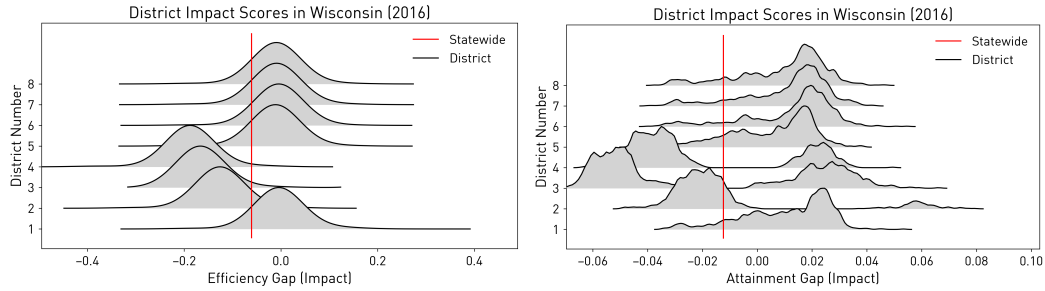


Figure 45. Ridgeline plots for the attainment gap and efficiency gap in the 2016 congressional elections in Wisconsin.

available to structure the simulations. This reinforces the utility of these measures over the deterministic leverage measures.

Here may be important to identify that the bonus measures are dependent on the granularity of the seat share. Since the expected seat share in an 8-member delegation has only 9 possible values, the expected difference in seat shares also has a finite number of possible values. This increases quickly with respect to the number of seats, but becomes quite noticeable below around 20 seats. Here, the observed and simulated bonus measures have four or five unique realization values from simulation. Taking their *average* presents a point estimate that is not empirically possible. Gelman et al. (2010) summarize seat-denominated bias measures using their means/variances, and so present a smoother picture that disguises the integrality. But, since you can only win an integer number of seats, using the mean and standard deviation provides artificial precision or granularity to the estimate. In contrast, summarizing the distributions in terms of medians and their quantiles provides empirically-attainable summaries. The integrality is reflected in the “lumpiness” of the kernel density plots in Figure 53, although the smoothing inherent to KDE presentations disguise the inherent integrality for the realizations of seat-share differences. These issues do not affect the measures that operate in terms of votes, such as the attainment gap or the efficiency gap, since vote shares and totals are effectively continuous. Although individual vote counts are integral, there are so many votes at the congressional district level as to make their counts or shares nearly continuous.

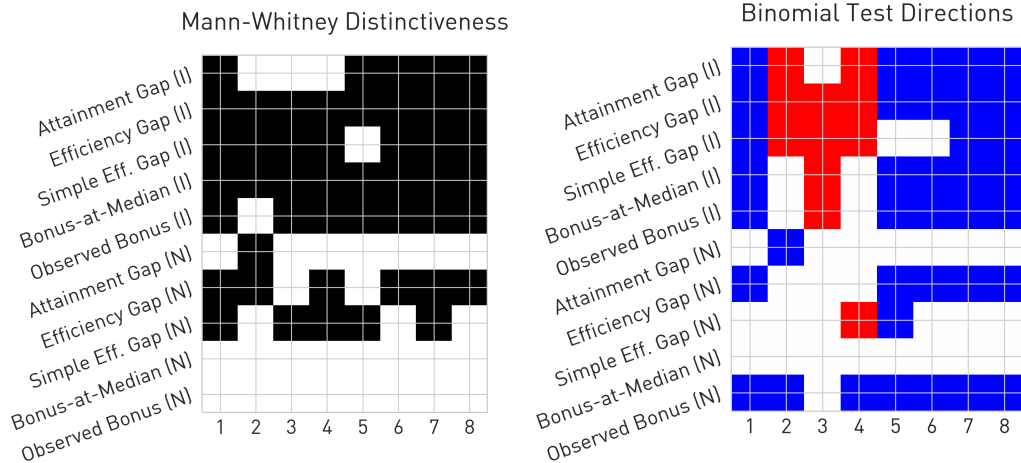


Figure 46. Distinctiveness and directionality tests for the Wisconsin districts since 2010, in the style of Figures 52 & 33, using the same style of Bonferroni-corrected significance filter.

Interestingly, the bimodality for the attainment gap is generated *solely* from the impact of three districts in the Wisconsin plan. That is, the impact distributions for districts 2,3, and 4 are bimodal, but no other distributions are. This anomaly is visualized in Figure 45, where only the impact distributions for districts 2-4 are bimodal, and the rest are not. This bimodality does not manifest in other measures, such as the efficiency gap. In the efficiency gap, the impact distributions do have some mass beyond their mode, but it does not manifest in the distinct way that the attainment gap does. Indeed, in most measures, the relationship between the shape of the statewide distribution and the shape of the subdistributions is relatively consistent; all distributions have the same general silhouette. However, the statewide attainment gap impact distribution appears quite directly to be a composition of the impact distributions added together. The prominent two modes in the statewide distribution appear to come from the masses of districts 2,3,4 and districts 1, 5-8's modes. Regardless, the deletion distributions are clearly distinct from the statewide distributions, and would result in different inferences about the bias of the plan. So, the impact measures so far function as desired in this small- N case.

Since there are only eight districts, it is simple to represent the choropleth in the same space as a full table of effect estimates. Thus, I present the full table of effect size estimates in

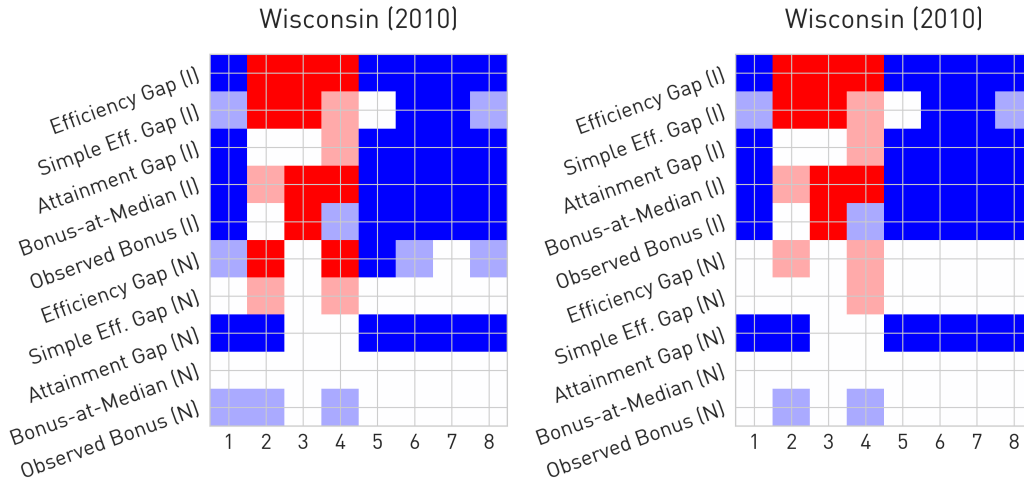


Figure 47. Effect estimate choroplots for Wisconsin during the 2010 redistricting cycle in the style of Figure 38. On the left is the stratified estimate, on the right is the pooled estimate. Again, the darkest colors have $|d| > 1$, the middle range have $|d| > .5$, and the lightest have $|d| > .25$. Red cells are districts whose removal significantly benefits Republicans, and blue benefit Democrats.

Table 9, and the corresponding choropleth is in Figure 47. Entries in Table 9 that are *negative* indicate districts whose removal benefits Republicans, or whose presence, conditional on the rest of the plan, benefits Democrats. Since the plan tends to have an aggregate Republican bias overall, districts with a negative sign in the plot are “centering” districts, whose presence reduces aggregate bias if they were not present. Districts that have a positive sign (whose removal benefits Democrats/inclusion benefits Republicans) exaggerate the bias in the state.

In this case, the impact scores break down consistently and directly along partisan lines for all measures. However, not all district effects are equal. Examining the choropleth, we see that districts 2,3,4,5, and 7 are inconsistently characterized, either having small effect sizes for a given measure or having opposing-sign indices for a given measure. Thus, districts 1, 6, and 8 are singled out as consistently partisan, and strongly so. Districts 6,7 have effect estimates larger than 1 for all measures, meaning that each measure identifies the district as strongly compounding the statewide Republican advantage, conditional on the rest of the districts in the plan. In this case, districts 1,5,6, and 8 might be singled out as districts whose impact consistently increases Republican advantage in the state overall, since their stratified

Measure	1	2	3	4	5	6	7	8
I Efficiency Gap	1.223	-1.566	-2.066	-2.124	1.119	1.251	1.164	1.179
Simple Eff. Gap	0.348	-0.665	-1.072	-0.368	0.162	0.577	0.539	0.432
Attainment Gap	0.897	-0.034	-0.110	-0.314	0.652	0.880	0.876	0.896
Median Bonus	1.045	-0.440	-1.151	-0.602	1.030	1.055	1.071	1.063
Observed Bonus	1.295	-0.160	-1.340	0.309	1.239	1.309	1.294	1.326
N Efficiency Gap	0.287	-0.603	-0.008	-0.876	0.516	0.354	0.145	0.290
Simple Eff. Gap	0.110	-0.416	-0.129	-0.456	0.248	0.246	0.094	0.161
Attainment Gap	1.180	0.860	0.109	0.250	0.784	1.435	1.059	1.263
Median Bonus	0.210	0.071	-0.192	-0.070	0.112	0.220	0.183	0.163
Observed Bonus	0.256	0.382	-0.089	0.368	0.120	0.234	0.244	0.196

Table 9. Grand effect estimates for the removal of districts from the 2010 Wisconsin plan. Positive numbers denote that Democrats benefit from the removal and negative numbers indicate Republicans benefit from removal. The inverse of the effect estimate provides the impact of *including* the district, conditional on the rest of the districts in the plan. Thus, if a deletion effect is positive/Democrat, the district's *inclusion* increases advantage for Republicans. Like the choroplots, the "I" block involves incumbents and the "N" block does not.

estimates without incumbency are picked up by both the efficiency gap and the attainment gap. A more skeptical take might use instead the directionality results in Figure 46, only admitting districts as "biasing" if their inclusion pushes the plan towards Republicans (deletion is shift-Democrat) in all measures. This would identify districts 1, 7, and 8 as consistently assisting Republicans. Regardless, districts 1 and 8 show up in many analyses as biasing districts, those whose presence in the plan magnifies the advantage of the distinctly Republican-biased plan.

The classical influence/leverage measures are shown in Figure 48. Again, since there are only eight observations within each year, most observations tend to have higher leverage. However, district 3 shows up as having a distinct Cook's distance in 2012 and 2014. Further, district 6 has a large Cook's distance in 2014, and district 1 shows up in 2016. This provides us with no indication as to the direction which the district shifts the plan, but the partisan impact measures in this case suggest these might be interpreted simply as biasing in the direction of whoever won the district in that year.

Regardless, the size of these measures does not provide an estimate with any level of uncer-

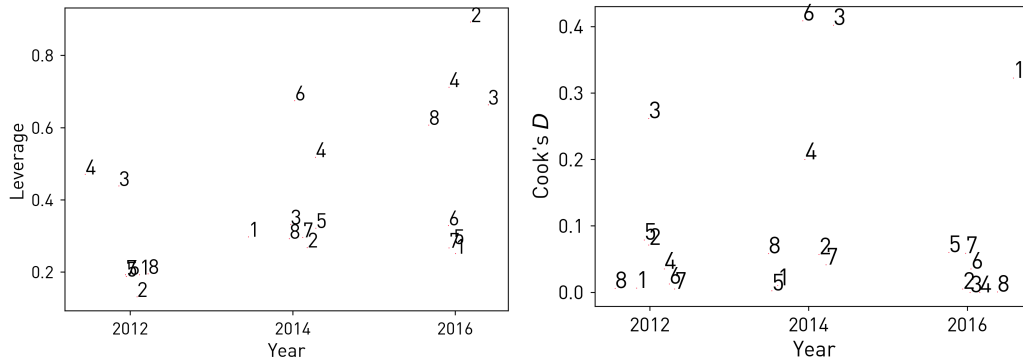


Figure 48. Jittered annotated scatterplots for model leverage and Cook's Distance in the Gelman-King models for 2012, 2014, and 2016.

tainty or stochastic theory, and cannot differentiate between the advantage the district confers due to incumbency versus without incumbency, as the simulation-based jackknife impact measures do. The correlations between the impact and influence/leverage measures (when pooled over all years) are never statistically significant. When computed within each year, the absolute value of the impact measures are only weakly correlated with leverage. This holds the same as in the California case, suggesting that only impact magnitude and leverage have a consistent weak relationship. Since effect magnitude is not enough to identify whether a district reduces or increases bias, the leverage measure is again insufficient in this case, even though there is consistent agreement among all of the jackknife impact measures with the partisan control of the district in Wisconsin, which was not the case in California.

Thus, overall, the districts singled out for further examination would likely be districts 1, 6, 7, and 8. These districts are reliably Republican districts in the state, but none are the *most* Republican (by typical margin in Congressional races or by Cook's Partisan Voting Index). The Speaker of the House, Paul Ryan, represents district 1, so it is unsurprising that the district consistently benefits Republicans, as it is likely drawn to provide a safe district with sufficient representation from the party regulars to provide the speaker with consistently large margins, but avoid over-packing Republicans into the district, which might fuel primary challengers. Of the remainder, districts 6 and 7 have had an incumbent occupant since 2012, and a Republican

succeeded a Republican incumbent in district 8 in 2016. They tend to have margins in the 20 to 30% range, with candidates winning around 60% of the vote. Thus, the districts are quite “safe” when accounting for an estimate 8-to-10 point incumbent advantage in the state.

7.5 Conclusion

In sum, the impact measures can be used to provide distinct indicators of which districts strongly affect bias statistics. These impact measures, derived from jackknife theory, are mostly consistent over changes in the counterfactual specification, and the identification of fairness under a given counterfactual should be possible, regardless of the counterfactual. While the measures tend to disagree with one another, this reflects disagreement that is embodied in the statewide estimates. Thus, no novel disagreement is uncovered by the impact analysis, only the intensification of lurking ones.

What is *not* answered here is whether it is right to assess partisan advantage under conditions that are *never observed* versus under conditions *as experienced*. At least in the California case, many of the districts’ outsized effect estimates fell away when the effect of incumbency was removed from simulations. While California will likely never experience an election where no incumbents run for congress (and indeed never has), the fact that the behavior of the bias measures under these two conditions is so starkly different means that the sensitivity of the *study of advantage itself* to the counterfactual simulation design requires critical thought about the intended use, theory, and practical implications of this study. Fortunately, some districts are clearly impactful regardless of incumbency.

In the Wisconsin case, the impact statistics demonstrated utility in a (statistically) small- N scenario. However, as noted by Tam Cho (2017) and visualized in Figure 44, the statistical analysis of districting plans will *always* be hampered by smaller N than Wisconsin’s 8 congressional districts. Thus, while the process of generating impact scores, identifying impactful districts, and understanding the typologies of what makes them impactful is available for Wisconsin, what constitutes unacceptable partisan bias in a system with two districts? These

measures clearly make large-sample assumptions, and the routine use of mean/variance summaries (rather than quantile summaries) in software and other results (Gelman et al., 2010; McGann et al., 2016) obscures this fundamental validity issue. I address this here by considering and expressing when the analysis is fundamentally constrained by integrality, and the consistent use of quantile-based summary, which avoid representing the bias statistics as continuum measurements.

As far as a workflow to deploy these impact statistics, the measures and the jackknifed impact scores do seem to be sufficient to both identify some districts as impactful and filter out districts that are not. Further, they clearly identify districts that a standard analysis of model leverage *would not* identify unambiguously. These districts have meaningful types, and these types can be understood consistently with respect to the statewide bias measure with which they are specified. While any identification is contestable, the general technique used to *construct* the impact scores can be applied to any model or statewide bias measure.⁴⁷ Analysts interested in coherent detection of biased plans and districts within those plans that generate the bias can follow a relatively straightforward analysis workflow:

1. Specify and estimate a model of the elections and a simulation strategy to generate new elections under plausible conditions.
2. Examine the model and the simulation strategy for potential misspecification issues, either in the specification of the counterfactual simulation process or in the source model itself.
3. Estimate statewide effects.
4. *Regardless* of unambiguous statewide effects, examine the jackknife impact distributions to identify whether certain districts have an outsized influence on the districting plan.

The last two steps are critical; as *Florida v. Detzner* discusses (and mentioned previously in Section 2.1), a challenge to a districting plan should be consistent; a challenge to an entire plan should have specific egregious districts in mind, and a challenge to individual districts should

⁴⁷I also examine the component-wise contribution of districts to the efficiency gap estimator in Appendix 7.6.

demonstrate that their presence results in an overall plan with unacceptable bias. Since most of the measures in California demonstrate a likely non-zero advantage for Democrats, districts like district 40 in 2010 are implicated as “impactful” districts since their presence pushes the plan to be *further tilted* towards the advantaged party. Districts like district 19 in 2000 (or district 21 in 2010) have a typical impact that *increases* bias with respect to the party who enjoys the full-state advantage. Thus, while their removal is impactful, they work to push the plan’s total advantage towards the center and would likely not be “gerrymandered,” only impactful.

7.6 CHAPTER APPENDIX I: A FIRST BRUSH WITH IMPACT DISTRIBUTIONS

For analyzing the impact measures, it is first helpful to characterize the distributions of realizations. These distributions are shown for all advantage measures in all years since 2002 in Figure 49 & 51. In these plots, the left column is the summary of realizations under the “observed” electoral conditions that year, so that $X^{\circ} = X$. The right column is the summary of realizations under an electoral counterfactual where no incumbents run. This means the resulting “No Incumbent” distributions can be interpreted more nearly as “raw” partisan advantage, since the effects of incumbency have been estimated and removed in each year through simulation. In all images, a black vertical peg is placed at zero. If distributions overlap significantly with zero, then the estimated advantage should be interpreted as marginal, regardless of the magnitude of its point estimate.

McGann et al. (2016) suggest lowering the significance threshold to provide an indication of how “certain” it might be to detect bias at a given α . I do not suggest doing that in this dissertation. This would interpret the fraction of realizations of simulated effects on the “right side” of zero (i.e. the side that the point estimate is on) as as if this fraction reflected a true posterior probability about the sign of the bias estimate. In truth, the bias estimates have no explicit distributional model. Thus, I will avoid this language of pseudo- p -values and express everything directly in terms of the simulation quantiles.

7.6.1 Distributions for California since 2002

Note the distributions in Figure 49. First, see that most of the jackknife statistics lie close to the observed statewide statistics. This indicates that the jackknifing estimates are well-conditioned. The jackknifed estimate distributions should (in large part) nearly the same as

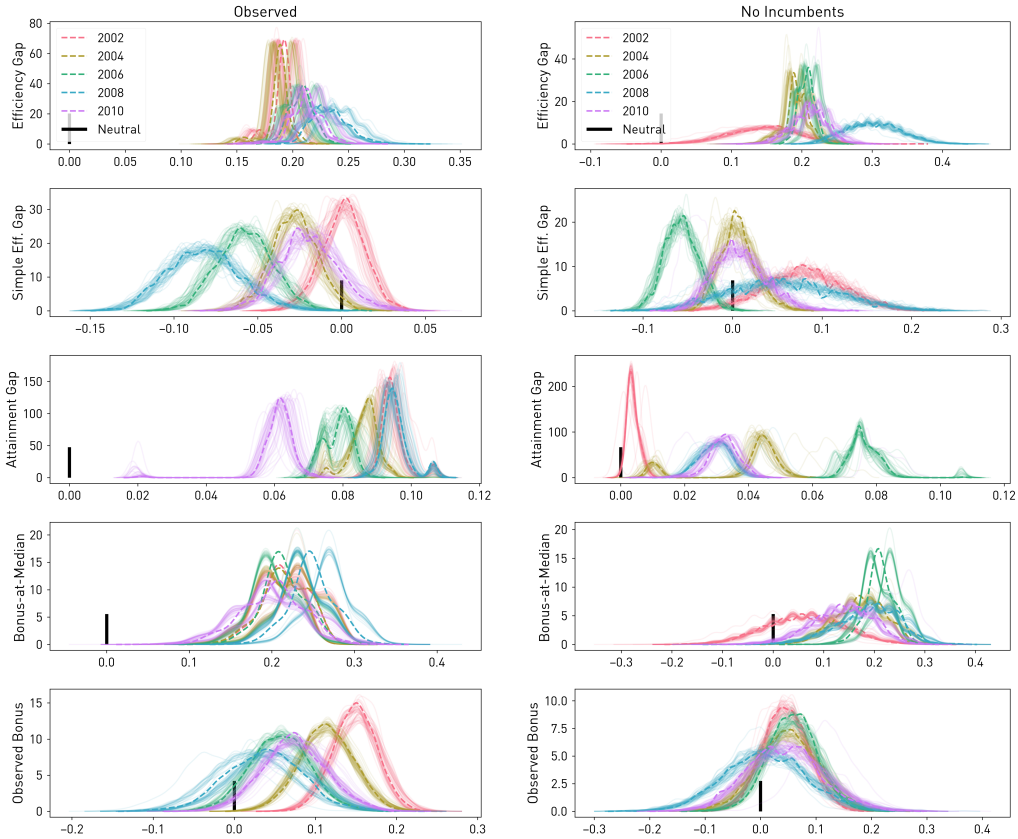


Figure 49. Distributions for impact statistics for California Congressional elections after the 2000 redistricting. The saturated dashed line indicates the distribution of all P full-plan statistics. Each thin line represents one set of P jackknifed impact distributions. The counterfactual where no incumbents run is plotted shown on the right, and the left shows simulations conducted under observed conditions.

the statewide estimates if *most* observations are not very influential. Further, the impact of any one district within a 53-district plan should be relatively small. This is what occurs.

In addition, all measures indicate that the state plan in the 2000s was biased towards Democrats except for the simple efficiency gap, which suggests instead that the plan was biased towards Republicans. For the simulations with incumbents, all of the efficiency gaps, attainment gaps, and bonus-at-median realizations are nonzero; the observed bonus in 2004 and 2002 is nonzero, but the remaining years in the decade have significant mass below zero. This means (by the observed bonus measure) bias decreased over time in the state, so the gerrymander was not “durable.” In a similar vein, the simple efficiency gap drifts increasingly

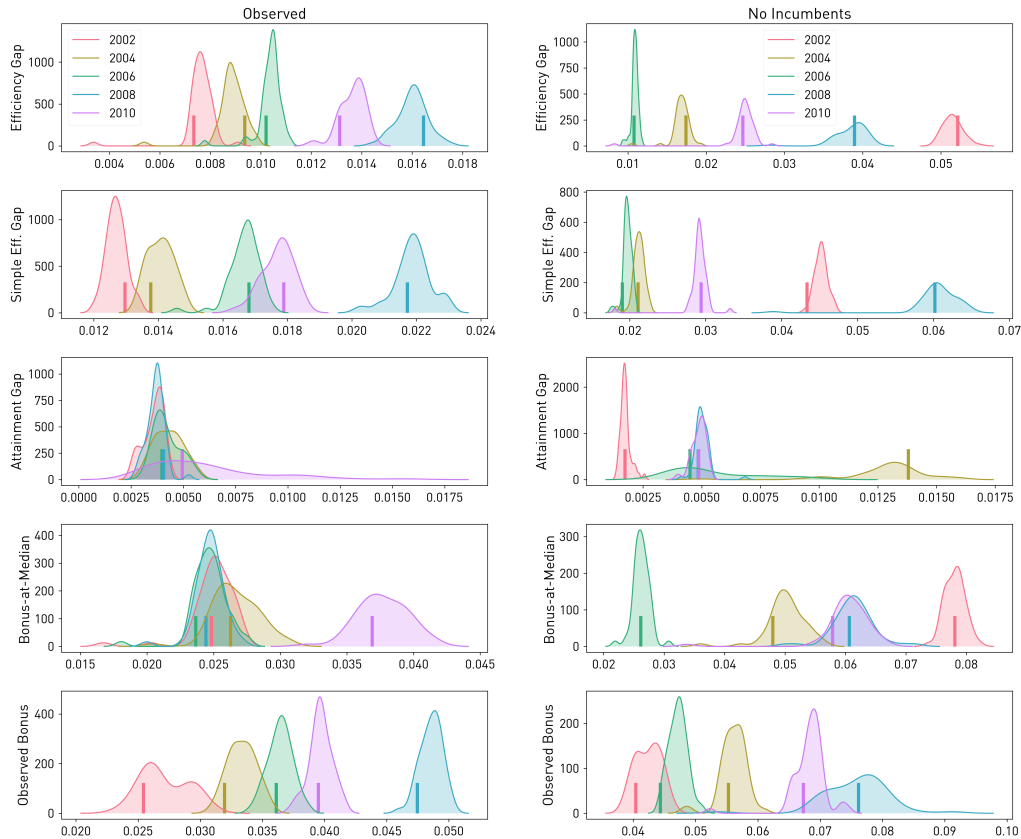


Figure 50. Distributions of the standard deviations for simulated distributions of each statistic after the 2000 redistricting. The kernel density estimate reflects the distribution of deviations for the deletions and the vertical peg is the standard deviation of the statewide realizations.

towards Republicans from a “neutral” value, finding the state plan biased towards Republicans in 2006 and 2008.

Accounting for the presence of incumbency affects the bias estimate distributions substantially. In most cases, the removal of incumbency in simulations causes shrinkage towards zero. However, for 2002 in the simple efficiency gap, the measure actually *grows* in magnitude in some years. The attainment gap & efficiency gap realizations shrink substantially, but all remain distinct from zero. The 2002 bonus at median becomes indistinct from zero, with nearly 25% of realizations as negative and the remainder positive. All observed bonus distributions become indistinct from zero in the no-incumbent case. This means most that, in most years, under

most measures, there is somewhere between a slight and a significant Democratic advantage in the California congressional districting plan since 2000.

In addition to the distributions of effects, I show the distributions of realizations' standard deviations in Figure 50. We see that the deviance of the statistic is consistent within each year. Most statistics cluster around the statewide deviation estimate (shown in the colored vertical peg), indicating that the realizations from the jackknifing process tend to be about as dispersed as the full data realizations. The move from simulation as observed to simulating with no incumbents tends to increase the variance of the advantage statistics in this simulation set.

Moving to the realizations from the post-2010 redistricting decade, the distributional realizations are shown in Figure 51. In this instance, all distributions indicate either no advantage or significant advantage for Democrats. Notably, the distribution of attainment gap statistics has two collections of mass; one around 2% (indicating Democrats can expect to win a bare majority of the California congressional delegation with around 48% of the votes while Republicans tend to require 52%), and one around zero. In addition, the efficiency gap statistic indicates a durable Democratic bias, but its simplified version suggests no bias in the state over simulations. The bonus-at-median statistics also suggest a Democratic bias consistently over all years. Finally, the observed bonus suggests an observable Democratic bias only in 2014. When simulating with no incumbents, most distributions again shrink towards zero. In this case, the efficiency gap realizations remain positive and, in 2012 and 2016, increase. In addition, the 2012 simple efficiency gap estimate increases substantially, becoming positive and distinctly nonzero. The bonus statistics both become indistinct from zero. Finally, the attainment gap statistics all shrink towards zero, with 2016 becoming the only one remaining nonzero when incumbency is removed.

The standard deviations for the 2010 realizations are shown in Figure 52. Again, the distributions of deviances tend to be similar within years and distinct between years, although the deviance tends to be much more similar between years in this set than in the 2000s. However,

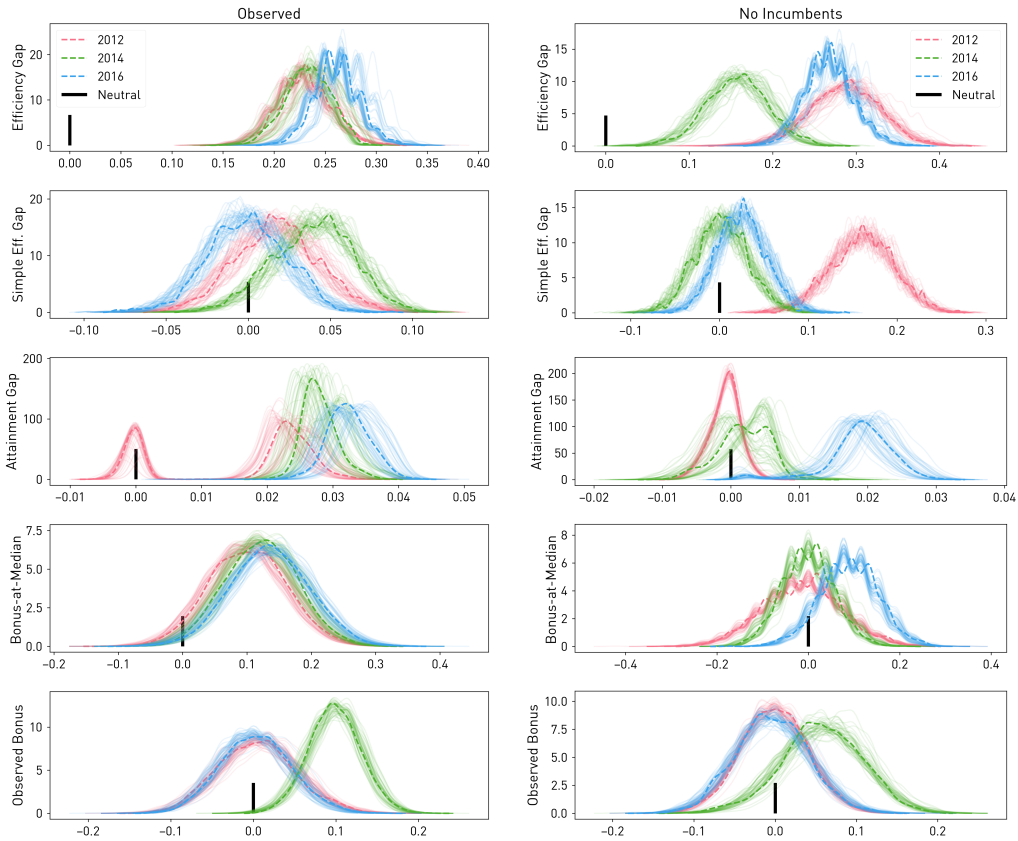


Figure 51. Distributions for impact statistics for California Congressional elections after the 2010 redistricting. The saturated dashed line indicates the distribution of all P full-plan statistics. Each thin line represents one set of P deleted-plan statistics. The counterfactual where no incumbents run is plotted shown on the right, and the left shows simulations conducted under observed conditions.

the statewide deviations tend to be slightly smaller than the average jackknife distributions' deviation, which is especially clear in 2016 deviations for the case with no incumbents. Regardless, the assumption that realizations have common variance in each year seems to be tenable given the similarity of the simulation deviances.

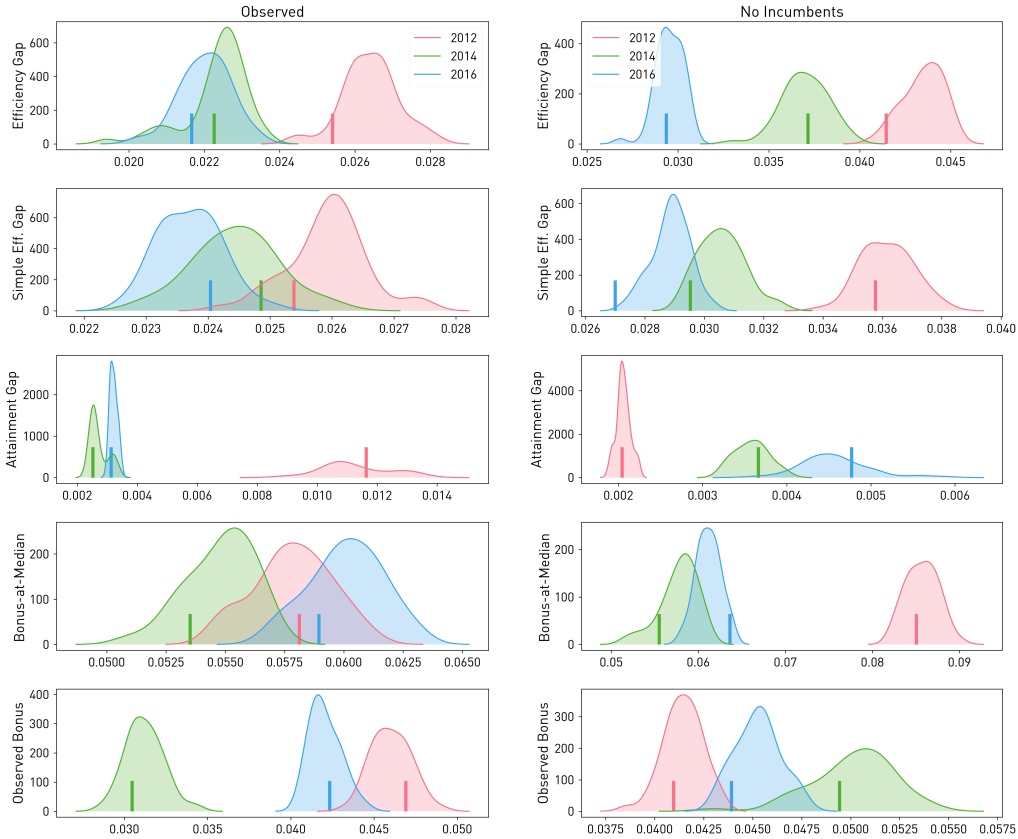


Figure 52. Distributions of the standard deviations for simulated distributions of each statistic in California after the 2010 redistricting. The kernel density estimate reflects the distribution of deviations for the deletions and the vertical peg is the standard deviation of the statewide realizations.

7.6.2 Distributions for Wisconsin since 2012

Distributions of the statewide and jackknife impact measures for Wisconsin since 2012 are shown in Figure 53. In it, the left side demonstrates simulations under the observed conditions and the right shows simulations with no incumbents. In general, the state shows a strong bias to Republicans under the observed conditions, but not unambiguously so. For example, the vast majority of simulations for the efficiency gap under observed conditions result in Republican bias, with 75% of simulations showing Republican bias in 2012, 96% showing it in 2014 and 2016. The observed bonus and simple efficiency gap show a Republican bias in around

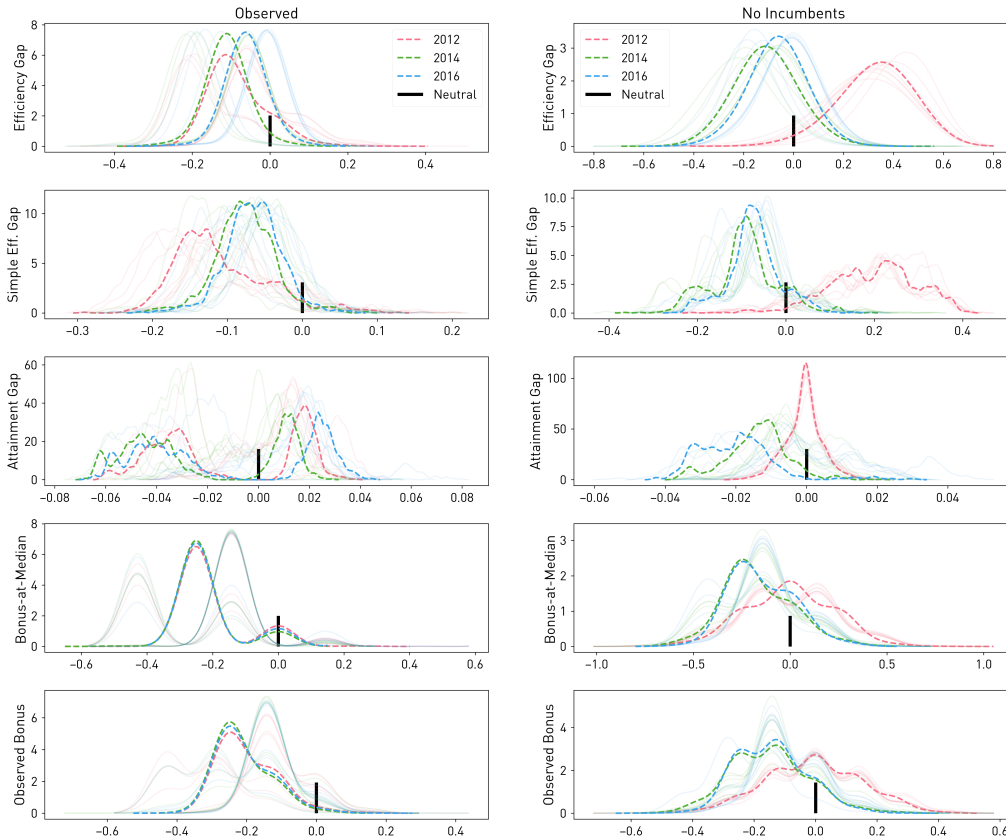


Figure 53. Distributions of statewide and jackknifed effects for Wisconsin after the 2010 redistricting. On the left are simulation results under observed conditions. On right, simulations with no incumbents are shown. This is akin to the plots shown in the Chapter Appendix, Section 7.6.

95% of simulations in all years. The median bonus is Republican in around 85% of simulations in all years. The attainment gap, however, is nearly evenly split, showing a Republican bias only around 60% of the time, and has a decidedly bi-modal distribution of votes required for democrats to win 5 or more seats in Wisconsin. This means the bias detected in the state legislative district plan is *likely also* present in the congressional district map, but to varying degrees of certainty depending on the measure. Under no-incumbent conditions, many simulations also result in a Republican advantage, but this is strongest in the cases where the R^2 is the highest, in 2014 and 2016. In addition, the efficiency gap measures indicate a nonzero

Democratic advantage in 2012 when incumbents are removed. Finally, the bonus measures are quite inconclusive, moving strongly to the center when simulating under no incumbents.

7.7 CHAPTER APPENDIX II: INTER-IMPACT CORRELATIONS IN CALIFORNIA SINCE 2002

To examine how distinct the impact measures are from one another, I conduct two analyses. First, I provide the scattermatrix of pairwise relations for both the 2000s districting plan and the 2010 districting plan for the estimated effect size using Eq. 7.6.⁴⁸ The simulations for the 2000s are shown in Figure 54 and those for the 2010 districting plan are shown in Figure 55. The plots show the results using simulations under observed conditions on top and under no incumbent simulations on the bottom. In addition, the plot hues are stratified by election year, and the Kendall's τ rank correlation for each year (in display order) is shown in the upper right, with p -values in parenthesis. What becomes clear first is that some statistics, namely the efficiency gap and median bonus statistics, are strongly bimodal and decompose into Republican and Democrat district clusters. In addition, some years are more strongly bimodal than others, with the 2000s having more strongly bimodal impact scores than any case in the 2010s. The rest of the measures tend to not exhibit this stark bimodality, regardless of year.

Second, the rank correlation between the efficiency gap and the bonus at median is the strongest and most consistent of all measures. The rank correlation between the efficiency gap and the median bonus is around .6 for each year in 2002-2010 and is around .4 for each year in 2012-2016. In the no-incumbent case, the correlation tends to be even stronger in the years where both \mathbf{h}_t and \mathbf{h}_{t-1} are available. This casts doubt on whether the measure is actually strongly distinct from the median bonus measure McGhee (2014) intends to critique.

⁴⁸Direct comparisons of the simulation vectors is inapt since the ordering of the simulations is arbitrary.

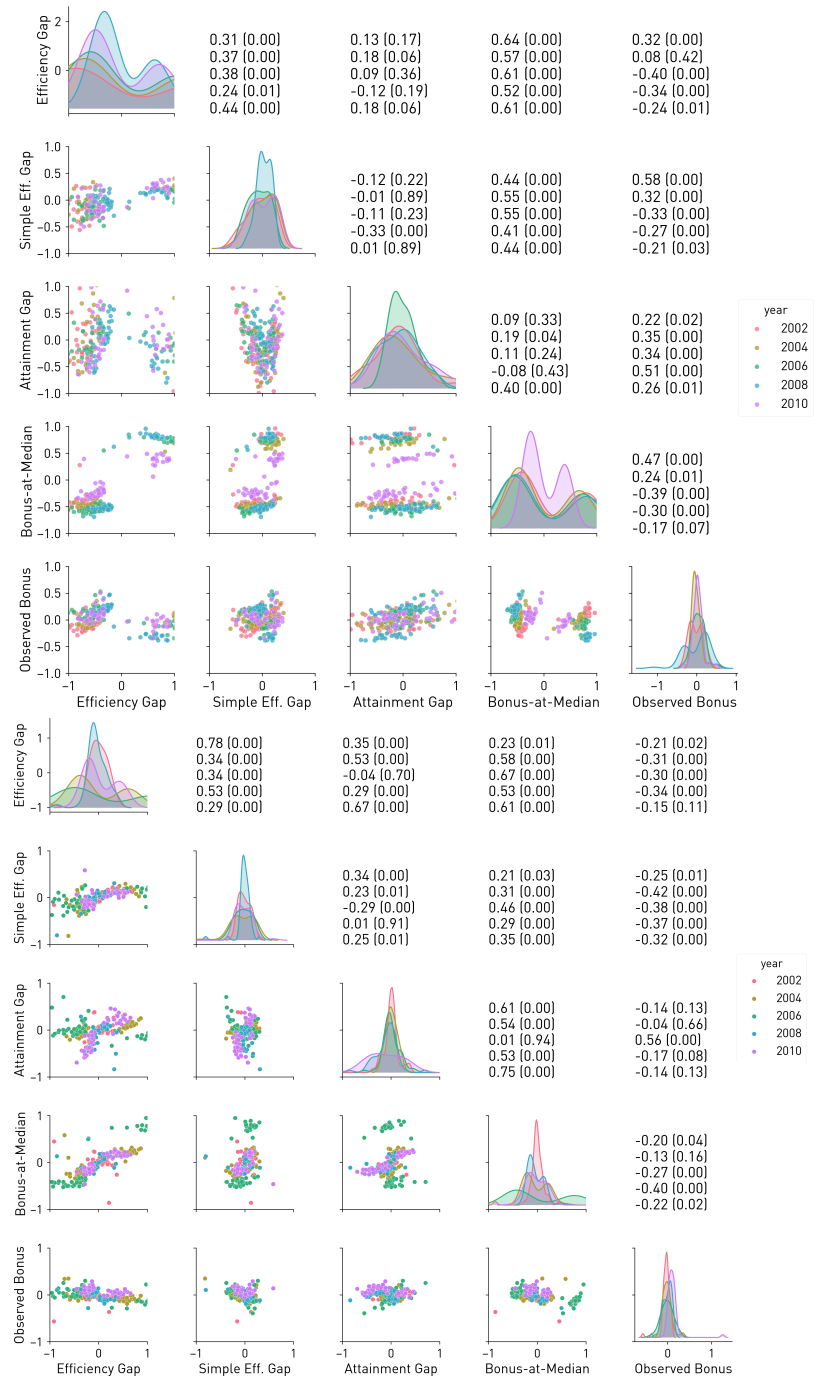


Figure 54. Scattermatrix relating all California districts' deletion effect estimate in the 2000s districting plan. Hue is the year in which the deletion effect is estimated. The top figure shows the measurements under observed conditions and the bottom figure shows the simulations with no incumbents. The upper triangle of each figure shows the Kendall's rank correlation for that pairwise relationship in each year (in display order).

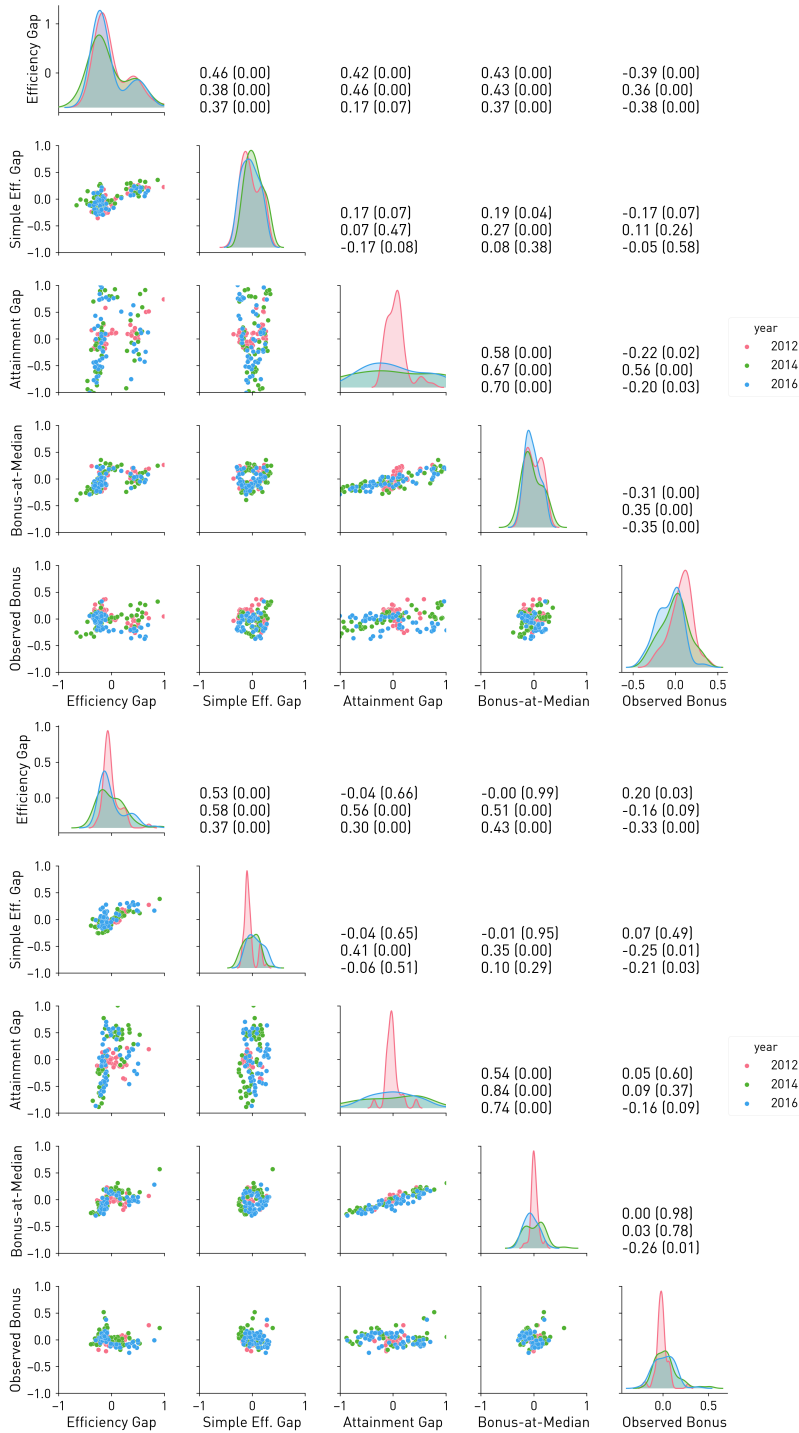


Figure 55. Scattermatrix relating all California districts' deletion effect estimates for the 2010s districting plan. This figure is styled akin to Figure 54.

Chapter 8

MANIPULATION AND LOCAL ADVANTAGE

In past studies (Fan et al., 2015, e.g.) and some theories about gerrymandering (Pildes and Niemi, 1993), district shape is taken to be a likely indicator of a district conferring advantage. In principle the geometric measures discussed in Chapter 2 indicate when a districts' boundary is manipulated; districts whose boundaries are highly irregular are assumed to be drawn that way for partisan advantage or racial animus. Many measures of shape regularity exist, and new methods are developed frequently. However, these measures often have high correlation with the few, standard measures of shape regularity discussed in Chapter 2, and it is difficult to find studies in the grey literature that use measures not discussed by Young (1988) or that are not classified in the taxonomy provided by Altman (1998a).

Part of the reason why boundary irregularity measures are both illustrious and suspicious are because they are difficult to connect directly to measures of partisan advantage. Further, boundary irregularity is symmetric; a boundary is shared between a district and its neighbor, so irregularity in one part of a district's boundary influences irregularity in a neighbors' boundaries. Altogether, this means that boundary irregularity can be used both to argue gerrymandering *for* the group in the district and *against* the group in the district, depending on the evidence. While it is *simple* to construct the relationship between the measures and the size of electoral margins (or consistency of a districts' partisanship over a decade), these measures are not explicit measures of partisan advantage. Indeed, while studies on the partisan impact of compactness preferences are not new (Altman, 1998b; Chen, 2013), they do not fit into a *causal* framework; that is, does boundary irregularity *cause* partisan bias?

Answering this question in total is difficult, and I cannot do so here. I can, however, demonstrate that while boundary regularity and shape measures have partisan import, they do not identify districts with significant partisan impact. Put another way, while boundary regularity and partisan impact are related, the set of high-impact districts are not always the set of

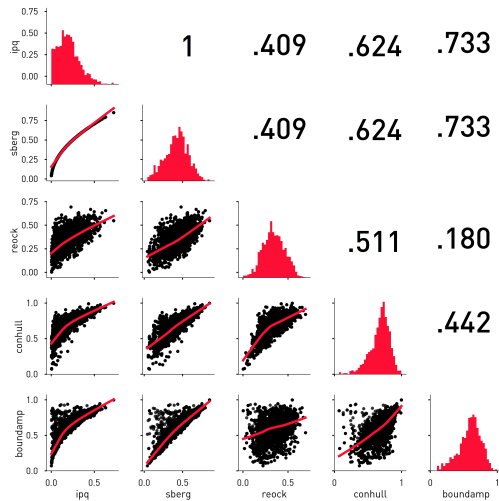


Figure 56. Scattermatrix of the shape measures used in this dissertation. Notably, all measures have significant rank correlation (shown in the upper triangle of the matrix), and in the case of the IPQ and Schwartzberg measures, perfect rank correlation. This pools over all states under analysis and all times.

weirdly-shaped districts. To show this, I conduct two analyses. First, I examine the relationship between geometric scores and the decade stratified effect estimate under observed conditions from Chapter 7. If boundary regularity were causal for advantage, there should be a detectable relationship between the two. In addition, analyzing the stratified effect estimates in simulations with no incumbents would indicate the extent to which this relationship changes depending on incumbency.

8.1 Shape and Impact

Across the eleven states where exploratory analysis was conducted, it appears that there is no consistent verifiable relationship between the shape measures and the political impact measures like there is between the impact measures and partisanship. In general, the impact measures are not well-predicted by geometric scores, even when accounting for fixed effects by state decade, or incumbent/no-incumbent scenario. Further, the shape measures all correlate

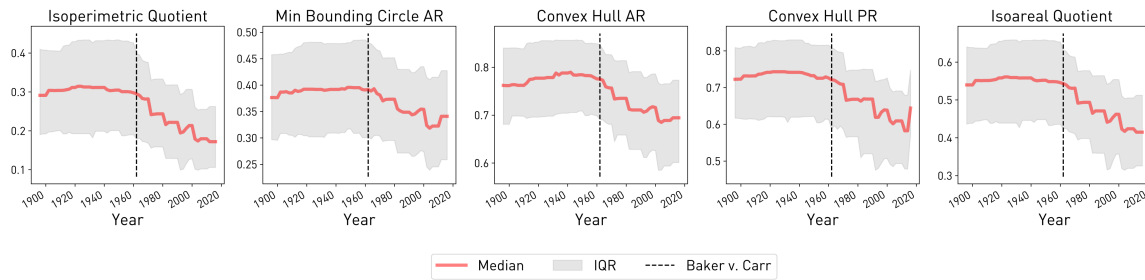


Figure 57. Shape measures considered in this dissertation over the whole time period contained in the dataset from Chapter 3. The dashed black line denotes the *Baker v. Carr* “one person, one vote” ruling which dramatically changed the landscape of districting law and constraints on district design.

well with one another, as shown in Figure 56, suggesting that they may not identify distinct traits about a district. That is, the fact that the perimeter-sensitive (IPQ, Schwartzberg, & Boundary Amplitude) measures correlate well with areal measures (Reock & Convex Hull Areal Ratio) suggests that the supposed differences in the measures are slighter than proponents suggest.

Moving to their relationships with the impact measures, I provide three analyses. In addition to the pairwise correlation measures plotted in Figures 58 & 59, I conducted a set of exploratory regressions to examine the multivariate relationships between the shape regularity measures and the models. Last, I will examine the set of districts selected as “historically bad” and examine their partisan impact. This follows from the typical use of compactness as an exploratory filtering diagnostic, a *decision rule* about which districts are bad and which are not. Thresholding arguments about compactness and bizarre shapes derive largely from academic discussion around *Shaw v. Reno*, but their use as a method with quantifiable accuracy has not been studied. Thresholding was considered by Ansolabehere and Palmer (2015), but I use a different cutoff. Here, I use the distribution of *all* measures observed in the state to construct a critical α value specific to that state, rather than arbitrarily selecting the compactness measures of the original “gerrymander” as the reference district. That is, I pick those districts in the 2000s and 2010s which are in the lowest 1% of compact districts ever seen in that state. The use of a floating threshold presents an interesting problem, since district compactness has been

declining since the Baker v. Carr decision, as shown in Figure 57. Some also contend that this decline in compactness is due to computerized districting algorithms, in that the redistricting process now is easier to game. It is unlikely that the distinction between the two is possible. Regardless, the use of a historical percentile-based selection criteria means that more districts will be selected as bad in modern times than before 1960. Stratifying the sample into pre- and post-1960 components would also be a viable method, but I will select only the worst 1% of districts by a given measure for the decision rule analysis.

8.2 Shape is Hardly Related to Impact

First, consider the results from exploratory regressions shown in Table 10. In these cases a model to predict effect estimates (or their magnitude alone) was fit on the effect estimates generated under observed conditions for all states and decades under study. The model used, for each partisan impact measure, is:

$$\text{impact} = \Delta_{\text{state}} + \Delta_{\text{decade}} + \mathbf{X} + \epsilon \quad (8.1)$$

where Δ_{factor} represents the dummy matrix classifying N observations into the J groups denoted by the factor and \mathbf{X} contains the geometric statistics, and ϵ is a normal independent and identically distributed error term. This model is fit only on the stratified effect estimates under observed conditions for the decades under study. Thus, I fit five models, one for each impact measure type. In addition, I conduct a regression on only the *effect size*, disregarding sign:

$$|\text{impact}| = \Delta_{\text{state}} + \Delta_{\text{decade}} + \mathbf{X} + \epsilon \quad (8.2)$$

The collinearity between the Schwartzberg and IPQ measures doubles the condition number of the regression, but I retain both measures because the models with both covariates have significantly better fit, and both are variables of interest. All models have significant F -statistics. But, the models of effect estimates (Eq. 8.1) have exceptionally poor fit. While the models relating the effect size alone (Eq. 8.2) have much better fit, it is difficult to suggest they are

well-fit to the data. Regardless of whether the effect or simply the effect magnitude is analyzed, the models simply do not indicate a strong relationship. Further, none of the shape statistics are good predictors across more than one impact measure. That is, no geometric measure has a consistent impact on the impact measures, either when considering their effect size or their estimate directly. The only consistently-significant effects across the impact measures are the state fixed effects and decade fixed effect, meaning some states and decades have substantially different baseline effect sizes. Not all states have significant fixed effects, however. Regardless, while there is an improvement in R^2 when moving from predicting effect size to simply predicting the magnitude of the impact, the effect of geometric measures remains too noisy to contribute to predictions about partisan impact.⁴⁹ In an absolute sense, the R^2 of any of these models is weak enough to suggest that *if* these measures of boundary manipulation are causal for partisan advantage, then there must also be a significant amount of extraneous noise or additional factors at play. More generally, this compounds the already significant doubt in the literature that strange shapes cause partisan bias.

Moving to a less structured analysis of pairwise correlations, the scattermatrix of relationships between the impact measure effect estimate and vote share & geometric measures is shown in Figures 58 & 59. When examining only the pairwise relationships between measures and compactness, we see a few things. First, note that the first column in each scattermatrix is the vote share, and the remaining columns are the geometric measures. Second, lines/points in blue are the stratified effect estimate for 2010 in each state, and in red are the 2000 decade estimate. In cases of mid-decade redistrictings (such as Texas in 2003) I have simply taken the average effect over the decade.⁵⁰ Since each facet shows a different state, this implicitly

⁴⁹A hierarchical model was also fit to examine whether the relationship between geometric measures and impact was different over states. This model did not improve on the fit from the non-hierarchical model, and none of the hierarchical effects were significant.

⁵⁰A more detailed analysis of the bias due to intercensal redistricting in Texas is provided by McKee et al. (2006). The existence of a single election before the intercensal redistricting would make an impact analysis more difficult in this case. In addition, since the redistricting is generally acknowledged to have targeted specific districts for reevaluation and those districts have easily-identified successor districts, simply averaging over this revision yields a conceptually consistent (if not adjusted) panel.

	R ²	BoundAmp	ConHull	IPQ	Reock	S'berg
Eff.						
Efficiency Gap	0.104	0.229 (0.771)	-0.011 (0.416)	-1.067 (1.322)	0.481 (0.493)	2.028 (1.568)
Simple Eff. Gap	0.058	-0.316 (0.310)	-0.276 (0.167)	-0.274 (0.531)	-0.004 (0.198)	1.192 (0.630)
Attainment Gap	0.091	0.473 (0.565)	0.027 (0.305)	0.709 (0.968)	0.329 (0.361)	-0.677 (1.149)
Bonus-at-Median	0.091	1.093 (0.580)	0.272 (0.313)	1.313 (0.994)	0.798 (0.371) *	-1.869 (1.179)
Observed Bonus	0.052	0.777 (0.618)	-0.008 (0.334)	2.272 (1.059) *	0.769 (0.395)	-2.550 (1.256)
Mag.						
Efficiency Gap	0.495	-0.735 (0.348) *	-0.344 (0.188)	0.517 (0.597)	-0.340 (0.223)	0.446 (0.708)
Simple Eff. Gap	0.433	0.000 (0.162)	-0.087 (0.087)	0.007 (0.278)	-0.013 (0.104)	-0.026 (0.329)
Attainment Gap	0.307	0.533 (0.368)	0.393 (0.199)	0.567 (0.631)	0.084 (0.235)	-1.679 (0.748) *
Bonus-at-Median	0.604	-0.116 (0.277)	0.040 (0.149)	-0.474 (0.474)	-0.131 (0.177)	0.291 (0.563)
Observed Bonus	0.656	-0.137 (0.298)	-0.138 (0.161)	-0.424 (0.510)	-0.074 (0.190)	0.269 (0.605)

Table 10. Marginal effects and standard errors for the estimated impact of geometric scores on partisan impact measures. Each row is its own regression, with the regressand on the far left. The top half of the table focused on predicting the effect estimate, and the bottom half predicts only the effect size. All regressions had significant F statistics, but many of the geometric factors were not useful in predicting the partisan impact.

acknowledges that the differences between states dominate the variance between these relationships, and the difference between decades is the second-most critical factor governing these relationships. Corresponding correlation matrices are reported in Figures 60 with a .05 significance filter, with the correlation matrices for each decade separated by a thick black line; the 2000 estimates are on the left and the 2010s on right.

The first thing that becomes clear from these scattermatrices and correlation visualizations is that the relationships between many of the geometric statistics is mainly inconsistent. Where it *is* consistent, it is weak. Democratic vote share tends to be negatively correlated with the impact measure, when it is correlated at all. The cases where it is *positively* correlated with impact are mainly in the observed bonus measure, and sometimes also in the attainment gap. Vote share does have an opposing correlation direction between measures in California (both decades), New York (2000s), Ohio (2010s), and North Carolina (2010s). All of these states have had allegations of partisan packing, suggesting that the plans may have too many districts with overwhelmingly-Democratic support. While statewide bias in California and New York benefits Democrats, statewide bias estimates in Ohio and North Carolina are either inconclusive or benefit Republicans. The fact that the observed bias and attainment gap measures are the only measure to flip in their relationship to vote share suggests that they may have more unique properties as indicators of district impact. This difference also manifests in Chapter 7, where the districts identified by the efficiency gap/median bonus tended to be the same, and attainment gap occasionally opposed, whereas the observed bonus was often too weak to single out any districts using the same cutoffs as the rest of the measures.

Geometric measures (if they are correlated at all with the impact measure) tend to be positively correlated. This means that districts that are more compact tend to benefit Democrats. The only state where this holds for all measures in a decade is in Michigan in the 2010 plan, where all geometric measures are significant and positively correlated with efficiency gap and observed bonus impacts. In Illinois and Texas in 2010, four of the five shape measures were

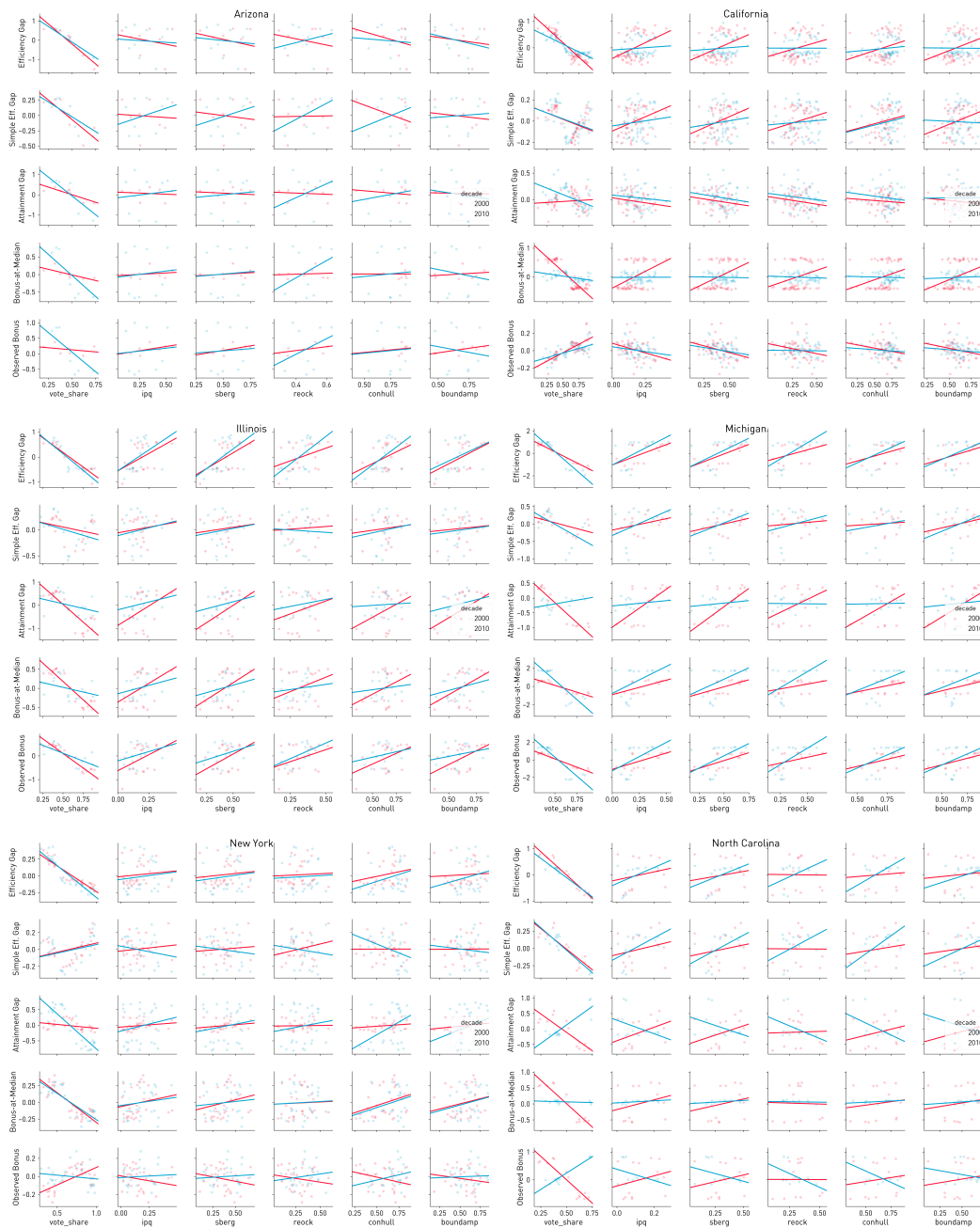


Figure 58. Scattermatrix relating impact measures and geometric measures for six out of the 11 states under study. The impact measures appear on the vertical axis, and Democrat vote share & the geometric measures on the horizontal axis.

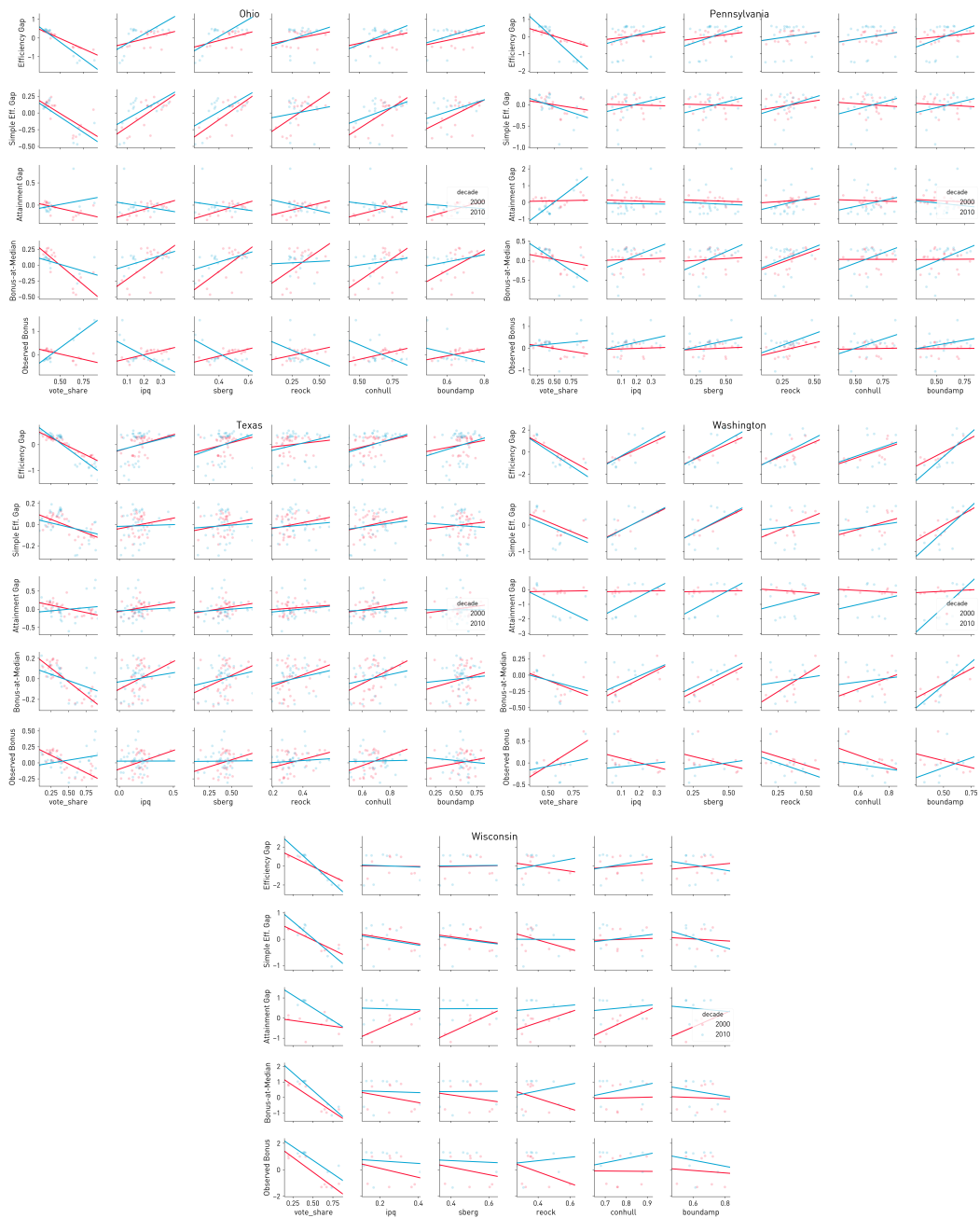


Figure 59. Scattermatrix relating impact measures and geometric measures for the remaining five out of 11 states under study. The impact measures appear on the vertical axis, and Democrat vote share & the geometric measures on the horizontal axis.

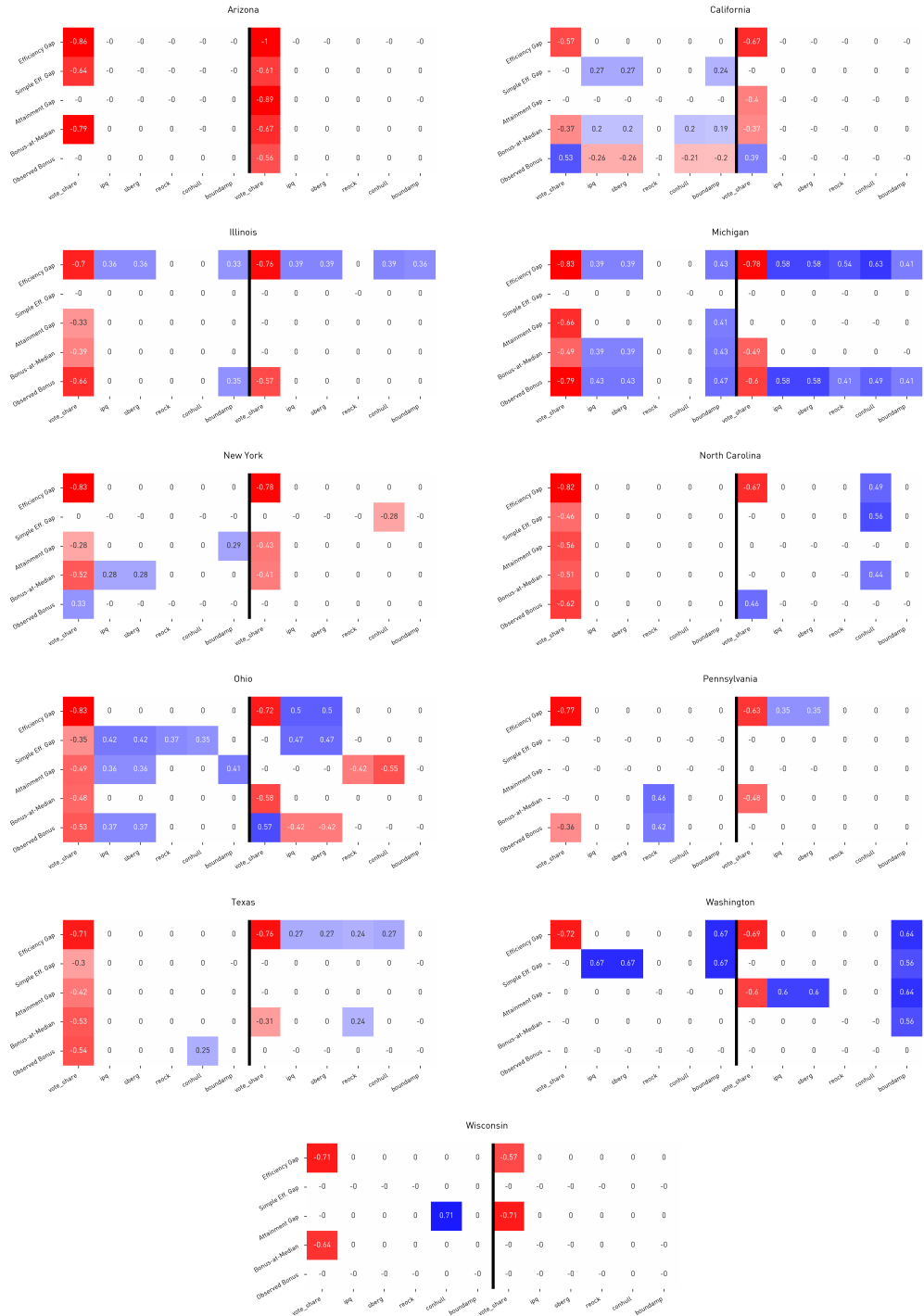


Figure 60. Correlation matrices that correspond to the comparison shown in Figures 58 & 59. The matrix on the left of the vertical black line in each subfigure shows the relationship for the 2000s and on right, the 2010s.

positively correlated with the efficiency gap impact score. In some cases, the relationship between geometric measure and impact measure appears to reverse between decades. This occurs for the relationship between IPQ and Schwartzberg measures to the observed bonus in Ohio, while no shape-impact correlation is significant during the 2000s for North Carolina's congressional districts. Critically, these are *also* where the relationship between impact and vote share reverses between those decades, indicating again that the properties of the observed bonus measure in cases of partisan packing may be significant. In most cases, however, the relationship is only slightly different in slope between the two decades. Both the correlation and the change in correlation between decades is marginal more often than not, and any argument that geometric measures provide any consistent information about partisan advantage is not supported.

8.3 Compactness as a Decision Rule

Lastly, the results shown in Tables 11 and 12 contain the 30 districts selected as in the worst 1% of districts in the state by a given compactness measure. Many districts were selected by more-than-one measure. The measures which the district is "historically bad" in are contained in the column titled Measure(s); SB denotes the Schwartzberg measure, R for Reock, CH for convex hull areal ratio, and BA for boundary amplitude. Since the IPQ and Schwartzberg measures are perfectly rank-correlated, districts fall in nearly the same percentiles between the two distributions, out to rounding precision. Thus, every district that is identified as historically bad in the IPQ measure is also historically bad in the Schwartzberg measure. Only one district, Illinois district 17 in the 2000s, is selected by the Schwartzberg and boundary amplitude measures and not the IPQ. We see no district selected *only* by IPQ/SB, since they are effectively identical in percentiles. This occurs since the IPQ is *right* on the boundary of the class, whereas the Schwartzberg measure falls below. In addition, we see that the measures tend to work together, with perimeter-focused measures selecting together (IPQ/SB & BA) and area-focused measures (CH,REO) selecting together. This is *in spite* of the fact that most of the measures

have significant aggregate correlation (noted in Figure 56)). Thus, studies of aggregate shape-party correspondence do not tell the whole story.

What is striking is that most of the districts selected by shape criteria are safe Democratic districts, according to their average decade vote share. Thus, the “worst” districts according to their geometric scores tend to be won by overwhelming Democratic margins. Assuming this is *casual* however, is fallacious; as shown in Section 8.2, shape does not correspond well to impact. This means that even though ill-compactness *does* tend to correlate with partisan lean, just because a district is weirdly shaped and won by Democrats does not mean the district is gerrymandered *to favor Democrats*. The impact measure (and statewide bias) must be considered. More generally, most of the districts selected by the shape measures are very safe districts, with margins of victory at-or-above 10%. The most competitive selected district, Michigan 1 in 2010, had an average vote share of .461, a Republican win by 8%.

Going further, districts that would be picked up as “historically bad” districts according to their compactness *do* occasionally have large impact scores. Take, for example, the second row of Table 11, Ohio’s 9th district in the 2010 plan. Its stratified efficiency gap impact estimate (under observed conditions) is about -1.2. Thus, by that measure, the district’s inclusion in the Ohio plan provides a significant benefit to Democrats. In contrast, the observed bonus impact is around 1, suggesting its inclusion provides a significant seat bonus to Republicans at the observed level of support for Democrats in Ohio. The district is won by Democrats (with around 70% of the vote) and statewide average Democrat vote share in Ohio is 43%. Plus, the state experiences a marginal Republican advantage, although (depending on the measure and the year) this advantage is not strongly distinguishable from zero in simulations. Recall that the observed bonus compares the seat discrepancy between what Republicans were observed to win and what Democrats would be predicted to win if their average support was the observed Republican support, 57%. Since OH-9 is quite safe, it likely does not make a difference to either the simulations at the observed vote or the simulations at the counterfactual 57% Democratic average vote, being reliably Democrat. Notably, district 23 from California in 2000 was picked up by both the partisan analysis and the compactness analysis. As mentioned in Chapter 7,

it is picked up as directional in the partisan analysis, but its effect estimate is not overly large. Thus, like Ohio's 9th district in the 2010 plan, the district is impactful but not the most impactful in the plan.

Only two districts selected by these compactness measures have stratified effect estimates in simulations from observed conditions larger than .25 for all measures.⁵¹ The two districts are Washington's 2nd district in 2010 and Michigan's first district. WA-2's presence benefits Democrats unambiguously, with consistently large negative removal impact scores. MI-1 likely benefits Republicans, although the attainment gap suggests its presence actually tends to benefit Democrats marginally. Notably respondents in the surveys discussed in Chapter 9 mentioned district 2 as a particularly interesting district; they attempted to make the district competitive but, in the view of some practitioners, failed to do so. MI-1, containing the upper peninsula of Michigan and northern portions of the state, has been continuously represented by Republicans during the 2010s, but has *also* been consistently competitive. It had an open seat in 2016 when incumbent Dan Benishek retired, and was replaced by a Republican. Notably, the district has an occasionally-decisive third-party contingent of Libertarians and Green party voters who, in 2012, had more than enough vote share between the two to flip the major party result. Thus, it is surprising that such a competitive district is picked up by the impact measures as likely shift-Republican *and* by geometric measures as being irregular. In theory, though, the district fits the reverse of the profile respondents from Washington described WA-2 to be: a theoretically-competitive district that consistently returns to a single party likely due to incumbency advantage. Further, in both cases, the districts are won by the party who likely benefits statewide. It is unlikely that either district is *drawn* for the candidate that occupies the district, but the fact that these two districts do share the same typology is notable.

Inverting this structure, only one district is selected by *all five shape measures* as uniquely ill-compact for its state, Pennsylvania district 1 in the 2000s. It has marginal partisan impact when measured by the attainment gap or the efficiency gap, and again there is disagreement on

⁵¹Note that the stratified estimates are strictly larger than the pooled estimate, so stratified estimates are the *most likely* to be large if significant.

its direction of effect. The remaining impact estimates are not large. This reinforces insight from the the regressions from Section 8.2 which suggest that the effect magnitude alone may be easier to predict than jointly predicting the magnitude and direction. Regardless, though, the least regularly-shaped districts clearly do not have large political impacts across all measures, although many of districts selected by three-or-more measures do have large negative efficiency gap impacts (indicating their removal benefits Republicans/presence benefits Democrats). This would comport with the intuition that compactness constraints tend to favor Republicans (Altman, 1998b), since removing the districts by a consensus on poor compactness tends to benefit Republicans' electoral efficiency. However, the fact that these districts are not consistent in the other measures indicates that this benefit is ambiguous.

I anticipated that majority-minority districts may be selected as most strongly impactful or most bizarrely shaped due to the discussions from racial gerrymandering practice and literature. Since race and political alignment are typically quite strongly-linked, I anticipated deeply Democratic majority-minority districts to be selected as "gerrymandered" according to their shape measures. However, just considering the districts in 2010, six of the 17 districts selected by the shape measures are majority-minority districts.⁵² This is around 35% of the selected districts; for reference, majority-minority districts tend to compose around a third of the Congress. Thus, majority-minority districts are not markedly over-represented in this selection of ill-compact districts. Further, none of the ill-compact districts selected that *were* majority-minority had consistently large partisan impact scores. Therefore, it seems unlikely that weirdly-shaped *majority-minority* districts have an outsized impact on partisan bias, which may be one empirical implication of the theory of gerrymandering by expressive harm.(Pildes and Niemi, 1993). This provides a further skeptical light (furthering Ansolabehere and Persily, 2015) to a difficult-to-measure argument at the intersection of partisan and racial gerrymandering study. Thus, the analysis of the intersection of the partisan measures and racial data seems a promising avenue

⁵²Here specifically I mean plurality-white, which is slightly different from majority-single-minority.

to explore, considering that the aggregate relationship between the “bizarre” shapes focused on by expressive harms arguments appear to be relatively unrelated to the partisan impact scores.

EGap	S.EGap	AGap	MBonus	OBonus	IPQ	S'berg	Reock	ConHull	BA	Measure	#	State	Year
-0.440	0.136	0.416	0.087	0.047	0.035	0.187	0.123	0.389	0.384	All	1	PA	2000
-1.204	-0.322	-0.034	-0.138	1.104	0.027	0.163	0.042	0.276	0.601	¬ BA	9	OH	2010
-1.066	0.200	-1.027	-0.551	-1.399	0.035	0.188	0.172	0.328	0.372	¬ R	4	IL	2000
-1.329	-0.468	-0.164	-0.289	0.846	0.056	0.237	0.456	0.650	0.309	IPQ,SB,BA	3	OH	2010
-1.487	-0.051	1.342	-0.342	0.666	0.038	0.195	0.117	0.491	0.437	IPQ,SB,R	1	PA	2010
-0.251	-0.145	0.133	-0.363	0.087	0.033	0.183	0.394	0.570	0.261	IPQ,SB,BA	10	CA	2000
0.345	0.121	-0.148	0.078	-0.168	0.010	0.101	0.209	0.404	0.181	IPQ,SB,BA	3	NY	2000
0.457	0.173	-0.564	0.290	0.313	0.039	0.198	0.266	0.459	0.315	IPQ,SB,BA	7	PA	2010
-0.873	0.131	-0.446	-0.216	-0.368	0.050	0.225	0.225	0.420	0.385	IPQ,SB,BA	4	IL	2010
-0.230	-0.410	-1.203	-0.342	-0.063	0.056	0.237	0.299	0.396	0.397	SB,BA	17	IL	2000
-0.888	-0.354	0.300	-0.012	0.826	0.029	0.171	0.070	0.249	0.522	R,CH	12	NC	2010
-0.077	-0.051	-0.061	-0.239	0.122	0.052	0.227	0.073	0.266	0.599	R,CH	28	NY	2000
-0.742	-0.221	0.030	-0.098	0.019	0.054	0.232	0.097	0.365	0.531	IPQ,SB	35	TX	2010
-0.502	-0.218	-0.206	-0.241	-0.128	0.070	0.264	0.102	0.457	0.567	R,CH	9	OH	2000
-1.176	-0.677	-2.093	-0.440	-0.443	0.011	0.105	0.226	0.462	0.327	IPQ,SB	2	WA	2010

Table 11. First set of districts selected as worst 1% of districts by shape measure. The # field provides the number of measures for which the district is in the worst 1% of that states' distribution, from 1 to 5.

EGap	S.EGap	AGap	MBonus	OBonus	IPQ	S'berg	Reock	ConHull	BA	Measure(s)	#	State	Decade
0.790	0.219	0.321	0.578	0.662	0.004	0.066	0.249	0.377	0.114	IPQ,SB	3	NC	2000
0.687	0.284	-0.430	1.766	1.130	0.029	0.171	0.181	0.430	0.461	IPQ,SB	1	MI	2010
-0.285	-0.197	0.223	-0.372	0.098	0.024	0.155	0.044	0.165	0.527	IPQ,SB	23	CA	2000
-0.783	-0.292	-0.382	-0.249	-0.222	0.126	0.355	0.164	0.520	0.634	R	15	TX	2010
-0.769	-0.206	-0.700	-0.543	-0.680	0.035	0.187	0.111	0.346	0.414	R	12	NC	2000
-0.848	-0.287	-0.201	-0.232	-0.151	0.015	0.122	0.175	0.510	0.340	R	34	TX	2010
0.260	0.082	-0.327	-0.077	-0.154	0.038	0.194	0.172	0.499	0.446	R	14	TX	2010
-1.515	-0.096	0.068	-1.103	-1.648	0.082	0.287	0.204	0.390	0.540	CH	14	MI	2010
0.361	0.085	0.091	0.132	0.096	0.074	0.272	0.226	0.421	0.468	CH	2	TX	2010
-0.422	-0.355	-0.241	-0.465	-0.444	0.115	0.339	0.162	0.468	0.625	CH	6	OH	2000
0.104	-0.096	-0.665	0.007	-0.106	0.098	0.313	0.186	0.435	0.564	CH	12	PA	2010
-0.085	0.044	-0.137	-0.178	0.133	0.023	0.152	0.107	0.311	0.347	CH	8	NY	2000
-1.199	-0.100	-0.137	-0.219	0.092	0.044	0.210	0.193	0.430	0.381	CH	33	TX	2010
-0.296	-0.213	0.008	-0.365	-0.404	0.050	0.223	0.218	0.537	0.357	BA	12	PA	2000
-0.969	-0.449	0.817	-0.175	0.627	0.082	0.287	0.234	0.582	0.459	BA	13	OH	2010

Table 12. Second set of districts in 2000 and 2010 as worst 1% of all districts by shape measure.

8.4 Conclusion

In general, shape measures are weakly related, if at all, to a district's impact on partisan advantage in any of the eleven states under study. In addition, it is not necessarily the case that strangely-shaped districts have a significant partisan impact. Further, it may be the case that theoretical arguments about compactness's inability to constrain gerrymandering (Humphreys, 2011) are empirically validated, since districts with poor compactness scores are not necessarily impactful, and districts that are impactful are not necessarily ill-compact. This is a clear-sighted conclusion, in that these results simply validate these arguments.

Another concurrent perspective may be that compactness and advantage are not strongly linked because districting officials may have a hard time making precise predictions about what the impact of their drawn lines will be. That is, maybe gerrymanders are ineffective. Since effective gerrymandering involves complex predictions about the interaction of future candidates, electorates, and district lines, it is possible that uncertainty about the characteristics of marginal districts results in lasting advantage despite intent. This perspective was offered by some redistricting officials in Washington, discussed in Chapter 9. These people suggested that, despite their best efforts, some districts that were drawn to be competitive ended up as non-competitive, returning a consistent partisan majority since their inception and initial open-seat election.

Uncertainty about the role of future incumbents and the candidate recruitment plays a large role in the characteristics of a plan after its enactment, but sometimes are not easily foreseen by linedrawers. Once a quality candidate establishes themselves in a new district, it can be quite difficult to oust them until they retire (often at the cusp of a districting boundary again). Alternatively, the aggregate lack of relationship between shape and partisan impact could be because districts that are "ecologically meaningful" to voters are not necessarily going to be compact. Regardless of these concerns, it is the case that a clear, unambiguous temporally- or spatially-consistent relationship between these measures does not exist.

Instead, we find that there is a fleeting, weak positive relationship, in that districts whose presence benefits Republicans tend to be more compact than districts whose presence ben-

efits Democrats at some times, in some states, for some measures. This is a precise, direct district-level comparison that corroborates the aggregate, approximate findings of Altman (1998b) and Chen (2013), who suggest that the presence of individual district compactness constraints tends to generate plans biased towards Republicans in general. Thus, if there were a relationship, it should manifest in this analysis. This is the *first* time the relationship between shape and impact on partisanship can be *directly* measured at the district level. It appears that the shape of districts does not consistently relate to the districts' impact on partisan advantage.

Chapter 9

SOCIAL, HUMAN, AND POLICY FACTORS

9.1 The General State of the Literature

A significant amount of academic work on identifying gerrymandering, both partisan advantage estimation and detecting boundary boundary manipulation, has yielded a traditionally-robust literature, full of healthy debate and criticism. However, many studies discussed in the preceding pages were developed *by academics for academics*, so to speak. While there is an exchange between academics who study redistricting and the actual redistricting process, this only manifests in the grey literature surrounding cases. This grey literature, a combination of university-sponsored technical reports, *amicus curiae* briefs, and reports to redistricting commissions is massive. While some work does eventually hop from grey (Godfrey et al., 2005) to white (Grofman and King, 2007), it is difficult in some cases to even call those works derivative or successor works, since the process of preparation for academic publishing often radically changes the content and substance of the work.

This is not necessarily a bad thing. However, it does result in a redistricting literature that is somewhat uni-directional in its exchange: papers are written and techniques provided to policymakers. Education and outreach campaigns by motivated academics actively lobby redistricting officials to use their specific measure or method of analysis. The fact that the grey literature is so large speaks to the fact that there is an intense need for a “routine” analysis method for these kinds of claims. While certain outlets like *Election Law Journal* routinely publish high-quality scholarship on quantitative/forensic analysis of gerrymandering and are the host of paper-writing competitions on the topic,⁵³ the penetration of gerrymandering as an academic topic into the mainstream of either political geography or political science has declined.

⁵³From where many studies discussed here follow (McGann et al., 2015; Chen and Rodden, 2015; McDonald and Best, 2015; Nagle, 2015; Cho and Liu, 2016a; Wang, 2016; Arrington, 2016)

Title	Prominent	Recent	# Results
<i>Political Science Quarterly</i>	Vickrey (1961)	Lo et al. (2013)	65
<i>Quarterly Journal of Political Science</i>	Chen (2013)	Chen (2013)	7
<i>American Journal of Political Science</i>	McCarty et al. (2009)	McCarty et al. (2009)	61
<i>Journal of Politics</i>	Niemi et al. (1990)	Friedman and Holden (2009)	196
<i>Annals of the AAG</i>	Morrill (1973)	Fan et al. (2015)	28
<i>Progress in Human Geography Review</i>	None	None	8
<i>Electoral Studies</i>	Grofman et al. (1997b)	Chen and Cottrell (2016)	59
<i>Election Law Journal</i>	Grofman and King (2007)	Nagle (2017)	101
<i>Legislative Studies Quarterly</i>	Young (1988)	McGhee (2014)	14
<i>Political Geography</i>	Johnston (2002)	Yoshinaka and Murphy (2009)	87
<i>Geographical Analysis</i>	None	None	12
<i>The Professional Geographer</i>	None	None	7
<i>Political Analysis</i>	Linzer (2012)	Linzer (2012)	19

Table 13. Popular journals and the frequency/prominence of gerrymandering study.

Indeed, a cursory survey of journals is provided in Table 9.1. I record popular journals in geography, political science, and specific electoral politics journals. Searching for publications on “gerrymandering” in these journals on Google Scholar occasionally yields a significant amount of text hits, but few dedicated studies by keyword, subject classification, or title. In the table, “prominent” articles are the highest cited articles from that journal whose main subject is mainly concerning gerrymandering (either forensic detection or analysis of its suggested impacts on voter or legislator behavior). The “recent” article is the most recent hit from that journal judged to be primarily dealing with gerrymandering. The “number of results” is simply the number of articles returned that mention gerrymandering in some way. If a journal has *no* article detailing mainly with redistricting, I record it as “None.” While there are many full-text hits in many journals, this only indicates that gerrymandering is a phenomenon often mentioned in passing. This is especially true for areas of geographic methodology, where basic research in regionalization heuristics or measures often touts the mitigation of gerrymandering (e.g. Li et al. (2013, 2014a,b)) but whose applications typically do not entail a detailed study of the issue beyond a measure demonstration (Fan et al., 2015). For the papers that treat gerrymandering as the main event, much of the “recent” work outside of *Election Law Journal* is somewhat dated. In general, the recent cycles obtain close to the finish of the previous redistricting cycle, average time since the most recent article across all these journals being 4 years. Again, while the preprint literature⁵⁴, the grey literature, and paper-writing contests held for *Election Law Journal* are all quite robust, work in the mainline journals is infrequent.

9.2 Interview Results

In addition to this infrequency, personal communication with local stakeholders and political operatives in Arizona led me to believe that many of the investigations conducted in the academic literature (i.e. not in the grey literature) *about* redistricting did not involve redistricting

⁵⁴The *Social Science Research Network* is one notable community.

officials personally. Thus, I attempted to ground the statistics & study developed in previous chapters in the needs, perspectives, and affordances of the redistricting process. To do so, I targeted four redistricting commissions for engagement in a set of confidential, anonymized interviews. Commissioners, their support staff, interested non-governmental participants, and journalists were all interviewed. For the interviews, participants in California, Arizona, Iowa, and Washington were targeted. In general, the commissioners in these states are not considered “public figures;” as such, their contact information is occasionally difficult to track down. While the commissioners are interested in participating in social research in the process, they are under no legal or political obligation to do so. In fact, due to the conflictual nature of the 2010 redistricting process, which occurred at the height of the Tea Party movement and spawned at least one Supreme Court case from the states targeted for study, I anticipated recruitment to be difficult.

Therefore, I proceeded through snowball sampling. Individuals with publicly-listed contact information were targeted first. These were never the commissioners themselves. I made sure to consistently and actively assure that interviews were confidential and anonymous, and that the identity of all participants would be protected. Due to the assurance of confidentiality and anonymity, participants often felt comfortable in either providing the contact information for fellow participants or, in some cases, reached out themselves to recruit new candidates. Then, when it came time to enlarge the sample, I would suggest that the candidate provide more/different names if they provided individuals I had interviewed before. Just in the process of generating this sample, a significant amount of information was apparent about the culture and structure of the process in the various states.

All states targeted were “non-traditional” redistricting states with “independent” commissions. However, the legal structures for each state’s commission are different. In addition, the context and objectives of each commission is also distinct, and the character of the processes where interviews were conducted were markedly so. While all states must follow the federal requirements for district legality, the individual state processes and political cultures result in significantly distinct local experiences of the redistricting process. In addition, statutory

constraints on the information that can be considered in drawing lines changed the nature of the discussion in a the states analyzed. Below, I will summarize the processes of the targeted states, outline the snowball sample, and provide insight to the specifics of each states' interviews. At the close of this chapter, I will draw some general conclusions from the interviews that were conducted, and aim to make explicit the trends I note in the interview results.

9.2.1 Iowa

Iowa's lines are approved with final authority by the legislature, but the process typically is dominated by a nonpartisan redistricting advisory body and a bipartisan legislative commission. The nonpartisan component is a bureaucratic organ within the states' Legislative Services Agency, and the bipartisan 5-member secondary committee is composed of four members appointed by the legislature and a chair selected by the legislative appointees members. Ostensibly, the nonpartisan body exercises ultimate authority about the boundaries, with the bipartisan body providing input on non-statutory inputs to the process. However, all requests for interviews made from March to August by phone and email to multiple parties in the legislative services agency were successfully transmitted but unanswered, so no interviews were conducted.

9.2.2 California

The California bipartisan redistricting commission is a 14 member commission with members who represent either major party or ostensibly represent no party. This commission is larger than that selected in any other state and has been extensively studied. The 2010 redistricting was the first time the commission was used. However, no commissioners nor officials involved in the process responded to requests for interviews during the solicitation period from March to August. Many different contact avenues were tried. A few initial participants expressed interest in April of 2017, but never eventually consented to participating in the interview. They also seemed loath to reveal their participation, so when they declined full participation, they

provided no new candidates. While generalizing about the process from these few initial contacts, it seems that the set of candidates who were interested in participating the most felt under pressure to conceal their participation lest they be viewed as “leaking” information about the internals of the process. This attentiveness to information security may also be heightened by the recent release of *Ratfucked: The True Story Behind the Secret Plan to Steal America’s Democracy*, a book by David Daley. Every candidate solicited for interviews in California mentioned the book, and expressed a concern that their material would be used against them or the patron under which they worked (such as executive directors, partisan staff leaders, or interest group leaders). Thus, while the assurance of confidentiality and presence of an NSF grant award page led many to be comfortable entertaining talking to me, none, in the end, judged the risks to their personal or professional stature worth the risk to engage in a candid interview about the process. This was both frustrating and confusing, since the interview protocol provided to participants notes:

- participants can pick and choose questions they want to answer while continuing participation
- participants can terminate interview at any point
- participants can retroactively remove consent to use the conversation transcripts or recordings
- the identity of participants is not recorded in the artifacts generated from the interviews, so the identity of participants cannot be reconstructed from interview artifacts.

Regardless of these assurances and further assurances to candidates about secure methods of interviewing, none eventually consented to participate.

9.2.3 Arizona

In Arizona, this risk-averse behavior was also present. In the staff and stakeholders surrounding the five member committee, participation was similarly difficult to elicit. The Arizona

Independent Redistricting Commission (AIRC) is composed of four individuals identified with the two major parties (two from each) and a single nonpartisan chair. The commission is only intended to be independent of the legislature, and not necessarily a *nonpartisan* commission, as many participants were eager to state. In addition to partisan balance, the commission was geographically balanced, with at most two commissioners from any county in the state. The 2010 redistricting was the second time the AIRC drew lines, the first time being in the 2000 redistricting.

The wariness in participation may be driven by the high-profile litigation around the states' congressional and state legislative plans, or the cases contesting the redistricting commission in general. In addition, nearly all candidates in Arizona *also* brought up *Ratfucked*. However, in this case, an entire chapter in the book was devoted to illustrating allegations of bias in the independent commission, so the concerns were either from having participated in that project or having seen how participation in that project turned out for others. Candidates in my study were concerned about participating in another skewering, and every candidate interviewed expressed notions of betrayal and discomfort at how the process in Arizona was portrayed in *Ratfucked*. These notions of betrayal and discomfort were similar to those mentioned by candidate-participants in California, but was immediate and experienced rather than prospective and risk-avoiding. One participant went so far as to describe the book as a sequence of unfair hit-pieces to fit a revisionist narrative about the Democratic loss in 2014 that caught fire after the 2016 presidential election among despondent liberals looking for one of many "structural" explanations for Clinton's (& local Democrats) 2016 loss. In general, participants were highly sensitized to this narrative, and the experience of the book seemed to feed a difficulty of recruiting new candidates.

Unlike the process in California, however, this was one of *many* aspects of significant partisan consternation in the state. Many candidates noted the political climate of the redistricting process was heated, with significant participation in public comment periods and active individualized protest of the commissioners by various interest groups. While Arizona was not unique in having public comment and testimony in their redistricting process, many participants felt the

process was uniquely partisan and was fixated on the behavior of individual commissioners in particular. Where participants were able, they suggested that the previous redistricting process in 2000 was not as significantly divided, and expressed both surprise and dismay at the character of the 2010 public comment phase.

Nevertheless, participants were eventually recruited. All candidates felt that the process was not supported well by the legislature, and identified significant gaps in technical support for the technical process of exploratory drawing & plan assessment. In many cases, they felt that they understood *what* was required, but felt ill-equipped to delve into the process. The latency between requesting a candidate map and receiving it was too large, and the staff dedicated to the commission was too small to handle the number of requests suggested. The introduction of *partisan* staff, by the accounts of participants, occurred late in the process, only after consistent dissatisfaction by the partisan commissioners about their ability to engage in exploratory analysis. This concern was unique among states where candidates were solicited.

In addition to this concern about technical support, the participation of an interest group focused on improving competitiveness across Arizona districts tended to drive the conversation much more strongly towards generating universally-competitive districts. Many in the process seemed most dissatisfied with the existence of safe seats in general, rather than the partisan balance of the congressional delegation or state house. All participants interviewed felt strongly that competitiveness was a strong component of a “fair” redistricting plan. Almost every candidate defined competitiveness as the margin size of a district, and most put the boundary somewhere between 3 and 5%. That is, districts where the margin of victory is greater 5% were nearly always considered uncompetitive by participants in Arizona.

Notably, the concern about competitiveness reflected a concern of the elections in each district and not over the composition of the state delegation as a whole. That is, the relationship between the congressional delegation and the popular vote for Congress was less important than the fact that as many districts as possible had small-margin winners. Throughout discussions, this concern focused explicitly on the desire of many participants to have *responsive* politics, in the sense of encouraging turnover, national partisan investment, and an effective *general*

election check on the state's members of Congress. Competitiveness was *not* understood in a systemic or structural sense, whereby control of the *overall* majority of the statehouse or congressional delegation mapped well to the states' relatively consistent low popular vote margins. In addition, when asked, the value of having senior representatives sitting in important structural committees in the House of Representatives was less interesting to participants than the prospect that individuals could remove their representation at will.

Thus, the redistricting community interviewed in Arizona exhibited a strong belief in ensuring responsive delegatory representation, and significantly rejected or considered unimportant factors related to the accrual of individual representative power, such as seniority, committee assignment, or even the provenance of a successor district in which an incumbent could run. Notably, this apparent resistance to district boundary consistency contrasted with the *actual* consistency in district boundaries and survival of many incumbents from the 2000 plan to the 2010 plan. While interviewed candidates suggested the commission succeeded in creating competitive districts when it was attempted, many also acknowledged that there were only three "true" seats in which a competitive district could be drawn. The constraining factor was often cited as partisan and ethnic geographies of the state, that the desired "truly" competitive map would never be spatially feasible.

Finally, a subgroup of interviewed participants suggested that the focus on competitiveness was a clearly partisan strategy. This group of participants viewed the fact that Democrats or Democrat-aligned interest groups tended to participate in the public comment period and advocate for competitiveness indicated that the competitiveness was only to *increase* Democratic representation. For these participants, competitiveness was a sham argument used to instead attempt to bias the plan towards Democrats. This group tended to be concerned that the supposedly-competitive districts were drawn using a metric of partisanship focused on past election results, rather than on registration statistics. Many viewed the use of past elections to extrapolate baseline partisanship of a new district as suspect, and instead suggested that competitiveness be understood in terms of the registration discrepancy in a district. This distinction is not novel in the academic literature (embodied often by Kousser (1996)'s critique of Gelman

and King (1994a) by focusing on registration discrepancy in California & resurfaced by McGhee (2014)'s critique of excess seat measures in general), but these discussions among commissioners, partisans, and staff apparently occurred without recourse to the academic study. While a single analysis was discussed that resembled similar to Kousser (1996), the fundamental contestedness of these measures was never resolved, and many analyses provided multiple characterizations of the anticipated Democrat/Republican percent vote of the candidate districts. While this was not a formal stochastic model of anticipated partisanship of the new districts,

Beyond the focus on competitiveness, conflict about what "communities of interest" meant was apparent among participants. Some viewed them as "natural" geographic communities that had both an intrinsic value and legal requirement to be preserved. Other participants suggested that they were created to justify the creation of Democratic safe districts at the expense of Republican safe districts. That is, a subgroup of participants claimed that the delineation of communities of interest was a political tool used to influence which parts of which districts should form safe Democrat districts, and that the focus on communities of interest was neither consistent spatially or procedurally. While there is an ongoing question about what communities *are* of interest and to *whom* they should be interesting (Webster, 2013), this discussion among commissioners was much more immediate: since race & ethnicity is nearly indistinguishable from party, identifying and protecting communities of interest *as if it were* a separate standard from identifying partisan-consistent communities, rankled a few participants.

Finally, a desire for the process to change was expressed by a subset of participants. Many expressed dismay that the commission was so small, and that partisans on the commission did not have a formal or legally-sanctioned staff separate from the commission staff. Most suggestions of larger commission stemmed from a belief that a larger committee would reduce the pressure on individuals in the commission, and particular attention was paid on how to reduce pressure on the chair of the commission. The focus on district-specific competitiveness was identified as a great strength of the model used in Arizona, and many largely considered the commission succeed in its stated goals, regardless of the individual's perspective on how the *process* of drawing the plan may have been canted. In part, this may be due to the fact that the

competitive districts drawn by the commission have indeed tended to flip between Democratic and Republican representation since 2012.

9.2.4 Washington

Washington's commission is distinctly a *bipartisan* commission. A similar 5-member panel is selected from an applicant pool is approved. However, the "nonpartisan" chair of the commission is a *nonvoting* member, meaning that the commission is split between two Democratic and two Republican appointees. This means that the plans for the commission must be approved with at least one opposing-party vote. Many interviewed in the process identified this as a critical strength of the structure of the commission, and claimed that the comity between members derived primarily from this structure. This was also the second time the commission was used, with the first time being the 2000 redistricting.

Acquisition for Washington participants was significantly easier than that for any other state. Individuals were eager, active in their consent, and seemed strongly personally interested in participating in the research. Only one participant mentioned the *Ratfucked* book; no individuals mentioned feeling betrayed, disheartened, or maligned by popular media investigations of redistricting processes. As such, many of the participants were unconcerned with the confidentiality and anonymity of the interview procedures; many eagerly suggested I mention that they participated in the interviewing process as social proof to improve the study's acquisition rate. Many even suggested doing so across states after I had briefly mentioned issues with acquisition in California and Arizona, indicating that many participants felt somewhat of a kindred relationship to redistricting officials in general across state lines. No other states' participants, either full or in part, actively suggested this.

In addition, the Washingtonian participants seemed to exhibit a remarkable level of partisan comity, with offhand remarks from self-identified partisan stakeholders, commissioners, and staff indicating a generally high level of trust and respect in the opposing party's competence. Individuals were largely satisfied with the technical support provided for the commission, and

many suggested that the nonpartisan staff's use of standardized styles and measures in the artifacts of redistricting, the maps & analyses used to actually discuss plans in public comment and internally, helped level the playing field. Nonpartisan staff interviewed also highlighted this role; chief among the concerns identified by the individuals was the importance of serving *equally* and ensuring that discussions and support was provided to both parties equally.

Another notable facet of the redistricting process in Washington was the actual working structure of the commission. Many participants noted that the commissioners paired off, working by twos to construct Congressional and state legislative districts that had bipartisan support *at their genesis*. While participants identified a similar initial phase of providing candidate maps of how commissioners may wanted to have partitioned the state, this procedure of pairing up between parties and working on the plan was not mentioned in other states. In addition, the nonpartisan staff seemed invested in ensuring that these working dyads had sufficient support for their custom needs, interests, and concerns.

Washington participants also noted a similar contest about measures of expected partisanship for the drawn districts. However, the use of partisan registration information (like done in Arizona) is legally unavailable to the Washington commission. Thus, instead of debate between using registration figures and past election results, the debate occurred about *which* races in *which* years might provide a good enough precinct baseline to extrapolate an expected partisanship of new districts. Partisans disagreed about which elections should be used, and the different measures were also propagated in many analyses like done in Arizona. Possibly, the fact that the difference in the *choice within elections* is a smaller difference than the *choice between elections and registration* led to a smaller contested surface on this front.

As far as the chief objectives of the redistricting were concerned, competitiveness was a critical concern for the commission. Depending on the participant, Washingtonians identified anywhere from one to three potential swing districts that the commission attempted to construct. While many felt the commission succeeded in its more abstract goals, many individuals interviewed suggested the commission was not successful in drawing competitive districts in some cases. They cited the emergence of high-quality candidates in the 2012 election as the gener-

ative process; once strong candidates emerge, the combination of incumbency advantage and candidate quality was suggested to both suppress effective challengers *and* the recruitment of future potential successful challengers. Regardless, the commission viewed the districts drawn as significantly fair and representative of a bipartisan consensus. No candidate interviewed suggested any modifications be made to the redistricting process in Washington. All candidates seemed to have high faith in the procedures, and that the next redistricting period would yield similarly-high confidence districts.

In a way, the concern about “successful” or “unsuccessful” attempts to construct swing districts struck me as a “chicken-and-egg” concern; if the first election in an open district is won by a high-quality candidate (over another supposedly high-quality candidate), then whether the district can be considered “swing” or not becomes a question of how effectively the high-quality candidate can develop a personal vote, not an intrinsic property of the district. If it weren’t for the good candidate, would that district revert to swing if the candidate’s personal vote disappeared? I attempted to inquire about this from Arizona participants about the “swingness” of Arizona’s Congressional District 2, as well. A few enlightening discussions with participants did not resolve this fundamental question of whether the “swing” district that the commissions attempted to create was about the composition of the electorate or the anticipated results in future elections. In more casual language, one participant averred the idea that composition entailed a district “likely to swing” because the population was weakly partisan-identified and other districts might be “liable to swing” because the observed election results did not reflect the true partisan nature of the eventual district. The extent to which this is a meaningful distinction may be relevant to explore in future analyses.

9.3 Commonalities between California, Arizona, and Washington Participants

In all states where participants were solicited, there were public comment periods available for members of the general public to remark on candidate plans or submit their own draft districts. However, those successfully interviewed in Washington and Arizona both expressed a

somewhat divided notion of the function of these public comment sections. Those in Washington noted significant involvement of academics, but suggested that their involvement was ineffective. In particular, persistent and dedicated individuals who attended nearly every public comment period occasionally came off as too professorial. Many felt that this severely hindered the efficacy of the lobbying, and suggested that interested academics should *both* speak at public comment and lobby commissioners individually if they have a merit-based quantitative argument about how things might be done better. What all suggested was particularly ineffective was the use of the public comment period as a time to “educate” both the public and commissioners. They suggested that the most useful transfer of knowledge in these public comment periods was *subjective* and occurred from what interviewees believed were members of the general public or a specific protected group.

Regardless, though, a subgroup of the participants in Washington suggested the public comment period was more effective as a legitimization of the commission’s later line drawing activities than as a period of the public generating new constraints and demands for the eventual redistricting plan. In this way, the model for the public comment period in the redistricters’ perspective appeared to be more akin to local zoning boards, where the policy is not *directly set* in community meetings, but many likely courses of action are discussed with affected communities in order to identify potential improvements and judge public reaction. A similar type of skepticism about the public comment period in Arizona was expressed by participants, but the skepticism was engendered by what those participants considered to be its use as a method for partisan operatives to air or instigate grievances with the commission. There was no significant Arizonan analogue to the academic-political exchange described in Washington. While both states did ascribe significant value and learning to these public comment periods, both groups suggested that the direct decisions about which lines eventually were drawn and which were made chiefly in consultation with legal counsel and along statutory or constitutional criteria, rather than members of the public or academia. Participants felt that these main constraints left very little room to act on public suggestions or demands or address the concerns brought forth by concerned academics.

Critically for this dissertation, however, *no participant interviewed in any state recalled a single measure of partisan advantage being used to assess candidate plans*. While these measures are critical in *post hoc* analysis of the districting plan, no participant recalled any of the measures as having been relevant or useful in the 2010 process. In this sense, participants in both states suggested that the focus of drawing and discussion was on *each district*, negotiating its spatial extent, population composition, inclusion of specific communities, rather than the *aggregate properties* of the plan as a whole, such as a reflection of the aggregate preferences of the electorate. Where relevant, the focus on these specific districts was often in terms of their expected partisanship, competitiveness, or status as a majority-minority district.

9.4 Practitioner Beliefs About Construct Validity

Despite this, a few participants stated they were familiar with a few common advantage measures, but suggested that the measures are politically unrealistic. By this, the familiar participants were not concerned with their *model-driven* nature⁵⁵, but instead suggested that the measures were either incomplete, incorrectly-premised, or simply incorrect. For the individuals not familiar with existing measures, many found the premises, assumptions, or operative concepts of these measures as irrelevant or unhelpful. Finally, many found the idea of uncertainty to be critical for the measurement of expected partisanship, but seemed resistant to consider or examine how uncertainty about partisan advantage might be incorporated into the process.

To assess these attitudes, I interviewed participants about the “reference” conditions required to construct each estimate of advantage. Like the ideal shapes of geometric measures, these are the conditions under which the advantage measures make a determination about the size and direction of bias. These are discussed in Section 2.3, but are repeated here for clarity. First, seat bonus measures require the assessment of an electoral counterfactual, when the party average vote share $\bar{h} = 1 - \bar{h}^o$. The difference in the share of seats the reference party

⁵⁵a common objection in the model-focused academic literature

wins when it wins \bar{h} and how many the opponent wins when they win \bar{h} provide the seat bonus. Some analyses construct an out-of-sample prediction for the special case when $\bar{h} = .5 = 1 - \bar{h}$. Second, the attainment gap measure requires the assessment of an electoral counterfactual where the reference party wins the barest majority, or $.5 \geq \bar{s}^\circ \geq .5 + \epsilon$, for sufficiently small ϵ . Then, discrepancies in the expected vote share required to win \bar{s}° for parties reflects advantage. Third, the the standard efficiency gap was discussed with practitioners. This measure requires that parties waste an equal number of votes, which relies both on the full vector \mathbf{m} and vote share vector \mathbf{h} . The so-called “simplified” version of this measure, assuming turnout is constant across all districts, was not mentioned in interviews. When prompted to consider these abstract electoral realizations, participants seemed to be interested in reasoning about the aggregate representativeness of a plan. However, many immediately expressed skepticism of the warrants required to get to that state upon “return[ing] to Earth,” as one participant described. Thus, it remains to be seen whether or not these forensics would be used, regardless of how well-specified they might be made.

9.4.1 Premises of Symmetry Measures & the Attainment Gap

For both Washington and Arizona, the congressional delegations very nearly map to the popular vote breakdowns for the state congressional delegations. As such, many individuals interviewed in either state thought it possible for majority control of the delegation to flip to minority control of the delegation, so measures focused on the fixity of a congressional delegation were deemed quite reasonable. Thus, questions about the votes required to win the barest of majorities, the reference condition for the attainment gap, the smallest possible majority in the congressional delegation, was seen as intrinsically plausible. No individual suggested the state would not see this division in the congressional delegation, but did suggest such a division would be rare in the statehouse, where Republicans tend to win more-than-minimal majorities.

However, many participants also suggested that the percentage of vote required to win that barest majority was not meaningful unless contextualized. They suggested that, in reality,

each years' vote share was highly dependent on the unique properties of legislators and that election cycle. Participants tended to concede that this might be modeled statistically, but none were confident that an analysis they had seen had been done in a way they found satisfactory. Thus, while the attainment gap reference condition was viewed as intrinsically plausible, some suggested it might be uninformative for the lived reality of the political system in each state. Regardless, participants often did agree with the principle of the attainment gap: in general, if parties can win majorities with increasingly smaller percent of the aggregate vote, partisan gerrymandering is likely.

In addition, many considered it plausible that the total raw congressional vote might split evenly between parties. However, no individuals believed that excess or deficit seats when the congressional vote split evenly at 50% would indicate advantage or unfairness outright. Most thought that unless the excesses were flagrant, this split-at-50% reference scenario was unhelpful. In all cases, individuals were skeptical about how that relates to concerns about incumbency, personal vote, and the mechanics of campaigning. Again, while all participants understood that these factors may be addressed statistically, most were not interested in *relying* on this treatment to conduct potentially-legally-binding assessment of partisan advantage. In the case that these factors could be controlled for statistically, participants still felt the difference in seats at an even vote split was suspect, depending on the spatial distribution of those votes. In addition, most participants considered it to be implausible that parties would flip in their level of aggregate congressional district vote shares during the time the congressional districts they participated in drawing were in use.

9.4.2 Premises of Wasted Votes

When identifying whether “wasted” votes matter or should be considered “wasted,” the wording of the question became incredibly relevant for individuals. Participants were divided on whether they were comfortable with calling *any* votes “wasted” or “ineffective.” Many more participants were comfortable discussing whether voters would be *satisfied* with the outcome, in

the sense that the person they voted for won the election. These are subtly different concepts. A “wasted” vote (by the measure in McGhee (2014)) is one cast *in excess of victory* or one cast *for a losing candidate*. Participants were much more comfortable considering only the votes cast for losing candidates as somehow ineffective, and almost all participants felt that the two components should be treated as distinct.

During interviews, some participants suggested that losing votes were *more wasted* than votes cast in excess of victory would be, and asked if a weighted efficiency gap had been considered. The semantic arguments about what type of vote is more or less effective focused on the idea of the electoral mandate; if a candidate wins some strong (but not overwhelming) victory, the votes cast in excess of victory communicate information to the candidate that suggests a reward for effective representation. Thus, those votes are less wasted than votes cast for losers, who accrue no mandates. Centrally, those who tended to endorse this idea, that losing votes are more wasted than those cast in excess of victory, were also mainly interested in those cast for an *overwhelming loser*. Thus, in general, they considered “hopeless” votes, those cast for candidates who lost by a margin of around 8% or larger, as the least effective, if they were willing to entertain differences in vote efficacy at all.

As far as how these recommendations can be translated directly into efficiency gap calculations, a few critical challenges are presented. First, it is unclear whether using a weighted efficiency gap would ever resolve the arbitrariness practitioners identified as undesirable about the measure. In addition, it is unlikely that a consensus weight could be found, despite the remarkable regularity in what individuals considered “hopeless” elections. It is possible that, due to the differences in the distributions of parties’ margins of victory detailed in Figure 25 from Chapter 5, weighting the distributions as the practitioners suggest would result in a shift *towards* Republicans in the values of the bias measures, since Democrats tend to win with larger margins of victory (and thus there may be more “hopeless Seattle Republicans”, as one respondent suggested). Alternatively, since Republicans tend to win with smaller margins, it may be that the bulk of *losses* Democrats experience fall just outside of the 8-10% margin window suggested by many practitioners. Further analysis would be required to identify the effect

of weighting on the aggregate measures (and the resulting local impact scores). Regardless of this concern about weighting, participants tended to agree that the fundamental idea of the efficiency gap was valid: if one party wastes a larger amount of votes than another, the system may be biased.

However, a subgroup of respondents in Washington did feel that the use of turnout information or registration discrepancy is invalid on its face for defining advantage. In those cases, participants suggested two different reasons for its inadmissibility. First, objections tended to place turnout into a different “category” of measure than partisanship, and suggested that discrepancies in the numbers that turnout between districts should simply never be included in redistricting decisions, since that gets too close to fixation on “registered” voter population breakdowns. In this perspective, a district’s turnout is endogenous to the boundaries of the district; a cohesive community may be easier to organize to show up to vote, so using a prediction or anticipation about turnout before the district is drawn is invalid. Second, objections suggested that turnout is a direct, partisan-relevant indicator, but that using it to measure advantage focuses on the wrong variable. These objections suggested that “large” and “small” districts are equally likely to provide bias, and that discrepancies in the sum of waste by party misrepresents the effects of advantage. Instead of focusing on how one party may win more easily than another, many participants suggested that vote waste measures fall apart due to discrepancies in political cultures in the urban and rural districts. In spite of this (and further attempts at clarification), further clarification from participants of the mechanism driving this was not available. The respondents did suggest that this was not a necessary conclusion and could change in the future, but I remain unclear as to what the participants in this subgroup meant.

Chapter 10

CONCLUSION

Partisan gerrymandering is both critical to understand and detect. Methods to model partisan advantage have long provided one method to identify when a legislative district plan may place an undue burden on one party. This dissertation improves the state of partisan advantage analysis in many ways. Through the construction of a novel dataset, exploratory analysis of electoral data, a formal specification analysis and simulation study, new insights about electoral modeling are available.

10.1 Empirics

First, the exploratory characterization of the structure of swing definitively answers disagreement in the political science literature about the empirical structure of electoral swing in legislative and presidential elections in the United States. Electoral swing, the change in vote share between elections, is *neither* strictly uniform nor spatially independent, but this depends on the frame of analysis. Even when accounting for state or regional heterogeneity, spatial dependence is both strong and significant in presidential elections by county. In the case of the legislative elections, spatial dependence in swing is much weaker, and may become totally marginal when controlling for spatial heterogeneity at the regional or state level. The “neighborhood” of swing clusters tends to be somewhere above the state but below the Census division, so modeling spatial heterogeneity in congressional elections at the state level may not be an appropriate scale of treatment. While the legislative model of elections used in Gelman and King (1994b) is spatially-misspecified, the correction of this misspecification yields very small differences for the shape and structure of the simulated seats-votes curves that are used in the analysis of partisan advantage. This may not hold for other types of electoral models, however,

so a thorough specification analysis of the source electoral model should always be conducted when attempting to characterize electoral advantage.

10.2 Electoral Model Specification

Thus, it seems that, while the discussion about spatial effects in electoral analysis around Gelman and King (1994b) is apt, it is not substantial — no major differences were observed using a variety of models of spatial dependence in vote shares or electoral swing at a national level. Further, I find that accounting for a spatially-dependent multilevel model of electoral outcomes provides a different shrinkage structure for spatial effects than that observed in a typical multilevel model, but this difference does not appear to be substantial in terms of model fit or effect interpretation in most cases. In general, the empirical effects of modeling spatial dependence in electoral outcomes at the congressional level in the US appears slight. Accounting for the mean adjustment and variance inflation/de-syncing involved in common simultaneous autoregressive model specifications, only a slight difference in estimated seats-votes curves was observed when allowing partisan swing to be a spatially-correlated random effect with constant variance and fixed mean in simulation. Once controls on the variance and mean were removed, the seats-votes curves began to become significantly more different. Thus, in general, it seems that the *spatial dependence* in swing or vote shares alone is not sufficient to induce large changes in the seats-votes model after controlling for the potential induced heteroskedasticity, variance inflation, and mean inflation.

This issue has further import for spatial research and simulation design. Many studies in spatial analysis stipulate a data generating process (possibly of the simultaneous autoregressive types examined here) and suggest that, as spatial dependence increases or decreases, statistics or models behave in unexpected ways. What I did not appreciate (and I wonder whether is not appreciated more broadly in the literature) is that the choice of this specification has impact beyond simply inducing a pattern in the observed response (or the model error term). Indeed, as discussed in the Chapter 6, the use of a spatially-autoregressive error

model may result in a covariance matrix with a non-constant diagonal, regardless of whether heteroskedasticity is intended (see also Rey and Dev, 2006) While the fact that mixed regressive, spatial autoregressive model specifications affect the mean response is well known in the literature, the fact that a SAR-Error specification induces both heteroskedasticity *and* larger marginal variance is not well-known in the spatial simulation literature, and is often never mentioned in simulation studies. While the movement from spatial covariance to spatial correlation complicates the specification, it allows for a more precise control. Further, it allows the analyst to identify whether is it spatial dependence causing claim *A*, or simply spurious induced variance/heteroskedasticity due to the specification of the simulation design. In this instance, this led me to identify that no, spatially-dependent electoral swing is not significantly distinct from independent electoral swing when it comes to generating seats-votes curve estimates, even though the map of dependent swing will be quite different from that of independent swing.

10.3 Local Indicators of Partisan Impact

Using these electoral models, a novel measure to identify district influence on partisan advantage was developed. I developed measures based on the essential idea of *influence* measures. These measures focus on the effect of removing one district from electoral simulations for N districts used to estimate partisan advantage. Given that a district is removed, the partisan bias for the $N - 1$ districts is estimated. Then, the difference between the estimates for the full (statewide) set and those for the $N - 1$ (deletion) set can be compared using many different types of distribution comparison methods. The analyses can be reversed to provide conclusion about the *inclusion* of a district, conditional on the rest of the districts. Thus, a district i whose removal benefits Democrats is one whose inclusion benefits Republicans. If a state has an aggregate Republican bias, then the presence of i would make the situation less neutral. Since these measures work directly in terms of a given measure of partisan bias, the conclusions about the statistics are sensitive to the measure of bias used. Therefore, the technique I

develop is applicable for many existing measures of partisan bias in a legislature or delegation, and can be used in a large variety of circumstances.

I suggest the use of three different methods: a nonparametric non-directional dominance test, a binomial sign test, and effect size estimation. Each of these uses “more” information to provide an estimate of the partisan impact of districts. The nonparametric test only differentiates the statewide and deletion estimate distributions, suggesting that a district is either “impactful,” show by its distribution being distinct from the statewide distribution. This method provides no indication of the direction of difference, and cannot identify whether a district makes the state plan more or less biased. The binomial sign test provides an indication of the *direction* in which a district influences the statewide bias. Thus, it can characterize which districts consistently benefit Democrats, which benefit Republicans, and which do not consistently benefit either party. However, the binomial sign test provides no measure of the strength of benefit. This is done using the effect size estimates. In this method, a standard effect size estimator (such as Cohen’s d or a nonparametric equivalent) can be used to compare the statewide and deletion distributions. This assumes that the “treatment” is the removal of the district.⁵⁶ This provides an estimate of both the direction and the magnitude of impact, controlling for the variance of the simulations in both distributions. Further, the method can leverage a full decade of data in a variety of ways, allowing for the identification of districts consistently over an entire redistricting decade *without* estimating a panel for the decade.

Using these methods for five different measures of partisan advantage, I find that a few districts in California and Wisconsin provide a marked advantage for Democrats/Republicans during the 2000s and 2010s. Further, I find that the 2010 and 2000 districting plans differ markedly in terms of the structure of district impact in California. The 2000s plan tends to be much more strongly partisan, with each district having a strong partisan impact depending on the measure of advantage used. For the Wisconsin plan, I identified three of the eight congressional districts in Wisconsin as contributing markedly to partisan advantage. The identified districts often (but

⁵⁶Or, measured in reverse, that the treatment is the inclusion of the district.

not always) benefited the party that won the district. Depending on the measure of advantage used, the conclusion about who benefits changed. All impact measures behaved differently from the classical measures of leverage and influence, which characterized the relationship between the districts and the model of vote shares used to generate the measures of partisan advantage.

In all cases, the choice whether to analyze the electoral system in simulations where incumbents run versus simulations where no incumbents run generated significantly different results. When analyzing simulations where no incumbent ran, measures were much smaller and district impact much more slight. Though biasing districts were still identifiable, the signal was much weaker. Often, districts which were selected in a simulation analysis under observed conditions had no impact when incumbents were removed. This means that any district-specific measure of partisan advantage will be sensitive to the size of incumbent advantage.

The existence of districts that are impactful in both scenarios is heartening, but not sufficient to suggest the choice has no effect. Thus, where possible, I suggest that the distributions in observed conditions be analyzed, since the processes of redistricting, strategic retirement, and incumbent advantage are both difficult to disentangle and of dubious use *to* disentangle. If a district is drawn to create favorable conditions for those in power and generates a partisan bias in doing so, then incumbent advantage and impact on bias may be directly, mutually-implicated. Since the analysis of a plan (and its rectification) must deal with districts and incumbents together, the analysis of bias in simulations under observed conditions speaks directly to the way districts function to both build and shape partisan strength in a districting plan. The separation through analysis of simulations under conditions with no incumbents seems to be an extreme counterfactual, since no election has been held under these conditions. Thus, making decisions about districts under an extreme scenario seems premature.

10.4 Shape and Advantage

This dissertation finds that previous studies of the relationship between compactness and partisan advantage are upheld by new measures of partisan impact. In general, no consistent relationship exists between shape measures that purport to identify boundary manipulation and the jackknife measures of impact on partisan bias scores. Further, using geometric to select districts that are ill-shapen and likely manipulated selects districts with no consistent relationship to partisan impact. That is, oddly-shaped districts are just as likely to be non-impactful as impactful. The districts selected by compactness rules are districts whose presence benefits Democrats, reinforcing the observation that compactness constraints tend to benefit Republicans in the US. Thus, in aggregate and in specific, there seems to be no useful relationship between common measures of boundary manipulation and measures of the impact a district has on partisan bias scores. Thus, arguments to the contrary (e.g. Fan et al., 2015) should be resisted.

10.5 Interviews

In soliciting and analyzing interviews with redistricting officials, stakeholders, participants, and scholars, a few conclusions become clear. Over thirty individuals in four states were contacted for interviews, and eventually sixteen individuals between two states consented to participate. Many more participants consented in Washington than in Arizona, and no individuals consented to be interviewed in California or Iowa. Across all states a few things are clear. First, since many stakeholders are private individuals in states with non/bi-partisan citizens' commissions, the technological infrastructure to contact commissioners after the fact is somewhat underdeveloped. There is often no consistent way to get a hold of commissioners or staff involved in the process after the fact, since the state agencies responsible for drawing maps often are drawn down quickly after the commission completes work. It seems that, in many cases, the link between a states' professional services divisions responsible for Geographic

Information Systems more generally and the redistricting commission are not well-integrated, since the commissions are impermanent. Further, in the interviews conducted in Washington and Arizona, it is critical that commissioners have access to dedicated nonpartisan technical support staff *and* dedicated partisan support staff. That is, there must be both a “neutral” map-maker who constructs representations of the consensus of the commission (or for a nonpartisan chair) and dedicated (distinct) partisan staff that draw partisan candidate maps. In the district generation phase, it seems critical for the partisan staff to work closely with partisan-aligned commissioners conducting responsive exploratory regionalization. Without dedicated staff provided by the party for their commissioners, a significant amount of computational issues were exposed directly to commissioners.

When it came to public hearings in Washington and Arizona, the hearings served more as discursive legitimation of the commission’s deliberations rather than a period where specific ideas about districts or plans were incorporated. Commissioners suggested that public comment is, by in large, not the place to lobby the commission about preferred shapes, metrics, fairness, or representativeness. While practitioners seemed to suggest that the contentiousness of the process was different between the two states, it remains unclear as to whether this was due to institutional design or political culture. Public comment in Arizona occasionally resembled protests of the commission, and the litigation surrounding the AIRC presented a distinct character to that experienced in Washington. Further, the Washington participants often suggested that the fact that the commission was an even-number of individuals with no non-partisan voting chair (i.e. that the commission was bipartisan, not non-partisan) accounted for the distinct comity of the Washington commission. In addition, some Arizonans suggested that having publicly-identified official mapmakers from the parties may resolve concerns about the impartiality of the mapping firm. Regardless, it seems that protecting both the commission and commissioners from allegations of nefarious wrongdoing and aligning the incentives of the commission to produce consensus maps seem most important for a successful commission.

10.6 Avenues of Further Work

There are a few avenues of study I believe would be fruitful to explore given the results here. The first would be to examine whether the conclusions about partisan impact hold constant over different types of model specifications, like the shapes of the seats-votes curve estimates broadly do in Chapter 5. Where these models may disagree on the statewide estimates, direct empirical comparisons of the models' results on partisan advantage and an analysis of how sensitive these models are to each district could be helpful. Since the leverage statistics may take a different form in a model like that used by McGann et al. (2016) than that used in Chapter 7, the impact statistics may comport better with leverage or influence in that model. Alternatively, if they continue to be distinct from leverage/influence, then the method can be shown to be more robust to specification. An interesting use may be in the Gaussian Mixture model of Linzer (2012), which would also open the way to deletion impact analysis of multiparty systems.

Second, it may be interesting to conduct a more thorough analysis of the different typologies of extreme impactful districts. While this dissertation examines a case studies in Chapter 7, a comprehensive analysis of the typologies of districts that are rated as impactful is not conducted. This is partially because context, i.e. identifying the idiosyncrasies as to why a given measure may flag a given district in a decade, is quite thick and the construction of these typologies itself requires meta-analysis of the traits of districts and the deletion statistics. However, since the five measures occasionally disagree on which districts are impactful in any given state context, this could be helpful to both shed light on the properties of each measure itself. While Tam Cho (2017) provides a cogent analysis of the efficiency gap measure of McGhee (2014) in empirical settings, similarly detailed analyses tend not to be conducted for other measures when they are used in the literature (McDonald and Best, 2015; Arrington, 2016, e.g.). Thus, using deletion statistics as one avenue of a thorough analysis of the measures may be helpful. This would complement an analysis at more scales as well, such as for state legislative districts or aggregating vote and turnout predictions from a precinct-level model, rather than congressional districts focused on here.

More broadly, using this type of deletion/removal analysis to conduct sensitivity analysis for derived properties in spatial model criticism may be useful. While the use of bootstrap influence measures is quite old (Leamer, 1983), the interest in spatial econometrics has been less-than-apparent. Regardless, these kinds of measures are at the forefront of research (Harris et al., 2017) and may yield significant results for the analysis of spatial systems beyond those discussed here in specific electoral spatial systems. Since the methodology driving this analysis is quite general, it may be used in a host of problems where model sensitivity to region is relevant or important to estimate quantities not covered by classical leverage methods. Going forward, using jackknifing, simulation, and analysis in terms of direct quantities of interest (King et al., 2000) should be at the forefront of making meaning of geographic models in spatial social science.

The fact that interview participants could not recall partisan advantage measures being used in the redistricting process indicates how strongly out of touch this literature is with the politics as practiced. While the wake of the redistricting seems to generate a more academic interest, as litigation about the results of districting propagate upwards through the legal system, it may be of interest in the future to provide actual documents for professional end users to consult in the districting processes. These would likely take the form of handbooks that both technical teams and commissioners could consult during the process that explain and codify academic perspectives on these issues. Further, much of the existing academic work on these measures and types of analyses are short articles; very few book-length studies of this process exist. While McGann et al. (2016) is one recent study, its section on modeling is both short and light on formal detail. Thus, I hope to conduct more work on the empirical sections of this dissertation and publish as a series of articles or a monograph about these models specifically and the use of advantage statistics in this style. Finally, concerns about construct validity of these measures is borne out throughout the analysis; while the local measures do consistently identify biased districts, they occasionally disagree. Since the local measures will inherit all the flaws of the statewide measures, further validation work of the statewide measures (i.e. significant statistical, empirical, simulation study in the vein of Tam Cho (2017)) is critical. While

the new local measures of partisan advantage developed here are useful, their reliance on the statewide measure validity means that it is important to be sure about the structure and behavior of the statewide measures of advantage.

REFERENCES

- Abramowitz, A., Alexander, B., and Gunning, M. (2006a). Don't blame redistricting for uncompetitive elections. *PS: Political Science & Politics*, 39(01):8790.
- Abramowitz, A. I., Alexander, B., and Gunning, M. (2006b). Incumbency, redistricting, and the decline of competition in US House elections. *Journal of Politics*, 68(1):7588.
- Altman, M. (1998a). *Districting principles and democratic representation*. PhD thesis, California Institute of Technology. No. 00038.
- Altman, M. (1998b). Modeling the effect of mandatory district compactness on partisan gerrymanders. *Political Geography*, 17(8):9891012. 00028.
- Altman, M. (2002). A Bayesian approach to detecting electoral manipulation. *Political Geography*, 21(1):3948. 00005.
- Altman, M., Amos, B., McDonald, M. P., and Smith, D. A. (2015). Revealing Preferences: Why Gerrymanders are Hard to Prove, and What to Do about It. *Preprint Available at SSRN*, 2583528.
- Angel, S., Parent, J., and Civco, D. L. (2010). Ten compactness properties of circles: measuring shape in geography. *The Canadian Geographer*, 54(4):441461. 00030.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.
- Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2):93115.
- Anselin, L. (1996). The Moran scatterplot as an exploratory spatial data analysis tool to assess local instability in spatial association. In Fischer, M. M., Scholten, H. J., and Unwin, D., editors, *Environmental Modeling with GIS*, page 454469. Oxford University Press, Oxford.
- Anselin, L. (2000). The Alchemy of Statistics, or Creating Data Where No Data Exists. *Annals of the Association of American Geographers*, 90(3):586592. 00000.
- Anselin, L. and Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section: Spatial fixed effects and spatial dependence. *Pap. Reg. Sci.*, 92(1):317.
- Anselin, L., Bera, A., Florax, R. J. G. M., and Yoon, M. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26:77104.
- Anselin, L. and Cho, W. K. T. (2002). Spatial Effects and Ecological Inference. *Political Analysis*, 10(3):276297. 00119.
- Anselin, L. and Rey, S. J. (2014). *Modern Spatial Econometrics in Practice, a Guide to GeoDa, GeoDaSpace, and PySAL*. GeoDa Press, Chicago, IL.
- Ansolabehere, S. and Palmer, M. (2015). A Two Hundred-Year Statistical History of the Gerrymander. Paper presented at the Congress and History Conference, Nashville, TN.

- Ansolabehere, S. and Persily, N. (2015). Testing Shaw v. Reno: Do Majority-Minority Districts Cause Expressive Harms. *NYUL Rev.*, 90:1041.
- Archer, J. C. (1988). Macrogeographical versus Microgeographical cleavages in American Presidential Elections: 1940:1984. *Political Geography Quarterly*, 7(2):111125.
- Archer, J. C. and Taylor, P. J. (1981). *Section and party: a political geography of American presidential elections, from Andrew Jackson to Ronald Reagan*, volume 4. Research Studies Press Limited.
- Arrington, T. S. (2016). A Practical Procedure for Detecting a Partisan Gerrymander. *Election Law Journal*, 15(4):385402.
- Atkinson, A. and Riani, M. (2012). *Robust diagnostic regression analysis*. Springer Science & Business Media.
- Austen, B. (1978). A comment on Malcolm Mackerras. *Politics*, 13(2):342344.
- Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10:12811311.
- Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*, volume 571. John Wiley & Sons.
- Bensel, R. F. (1987). *Sectionalism and American political development, 1880-1980*. University of Wisconsin Press.
- Bernardo, J. M. (1996). The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, 4:111122.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192236.
- Bishop, B. (2009). *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.
- Bivand, R. and Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63(18):136.
- Bochsler, D. (2016). The strategic effect of the plurality vote at the district level. *Electoral Studies*.
- Boyce, R. and Clark, W. (1964). The Concept of Shape in Geography. *Geographical Review*, 54(4):561572.
- Brinkhoff, T., Kriegel, H.-P., Schneider, R., and Braun, A. (1995). Measuring the Complexity of Polygonal Objects. In *ACM-GIS*, page 109.
- Brookes, R. H. (1960). The Analysis of Distorted Representation in Two-Party Single-Member Elections. *Political Science*, 12(2):158167.

- Browne, W. J., Draper, D., and Others (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3):473514.
- Browning, R. X. and King, G. (1987). Seats, Votes, and Gerrymandering: Estimating Representation and Bias in State Legislative Redistricting. *Law & Policy*, 9(03):305322.
- Burnett, W. and Lacombe, D. J. (2012). Accounting for Spatial Autocorrelation in the 2004 Presidential Popular Vote: A Reassessment of the Evidence. *The Review of Regional Studies*, 42(1):75.
- Butler, D. and Van Beek, S. D. (1990). Why not swing? Measuring electoral change. *PS: Political Science & Politics*, 23(02):178184.
- Calvo, E. and Escolar, M. (2003). The local voter: A geographically weighted approach to ecological inference. *American Journal of Political Science*, 47(1):189204.
- Calvo, E. and Rodden, J. (2015). The Achilles Heel of plurality systems: geography and representation in multiparty democracies. *American Journal of Political Science*, 59(4):789805.
- Carota, C., Parmigiani, G., and Polson, N. G. (1996). Diagnostic Measures for Model Criticism. *Journal of the American Statistical Association*, 91(434):753762.
- Carson, J., Crespin, M. H., Finocchiaro, C. J., and Rohde, D. W. (2003). Linking congressional districts across time: Redistricting and party polarization in Congress. In *Annual Meeting of the Midwest Political Science Association*.
- Chambers, C. P. (2010). A Measure of Bizarreness. *Quarterly Journal of Political Science*, 5(1):2744.
- Chatterjee, S. and Hadi, A. S. (2009). *Sensitivity analysis in linear regression*, volume 327. John Wiley & Sons.
- Chen, J. (2013). Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quart. J. Polit. Sci.*, 8(3):239269.
- Chen, J. and Cottrell, D. (2016). Evaluating partisan gains from Congressional gerrymandering: Using computer simulations to estimate the effect of gerrymandering in the US House. *Electoral Studies*, 44:329340.
- Chen, J. and Rodden, J. (2015). Cutting through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders. *Election Law Journal*, 14(4):331345.
- Cho, W. K. and Liu, Y. Y. (2016a). Toward a Talismanic Redistricting Tool: A Computational Method for Identifying Extreme Redistricting Plans. *Election Law Journal*, 15(4):351366.
- Cho, W. K. T. and Gimpel, J. G. (2012). Geographic information systems and the spatial dimensions of American politics. *Annual Review of Political Science*, 15:443460.
- Cho, W. K. T. and Rudolph, T. J. (2008). Emanating political participation: untangling the spatial structure behind participation. *British Journal of Political Science*, 38(02):273289.

- Cho, W. T. and Liu, Y. Y. (2016b). Toward a Talismanic Redistricting Tool: a fully balanced computational method for identifying extreme redistricting Plans. *Election Law Journal*, 15(4):351366.
- Clemens, A. C., Crespin, M. H., and Finocchiaro, C. J. (2015). The political geography of distributive politics. *Legislative Studies Quarterly*, 40(1):111136.
- Cliff, A. D. and Ord, J. K. (1973). *Spatial autocorrelation*. Pion, London.
- Coleman, S. (2014). Diffusion and spatial equilibrium of a social norm: voting participation in the United States, 19202008. *Quality and Quantity*, 48(3):17691783. doi:10.1007/s11135-013-9873-x.
- Converse, P. E. (1966). The concept of a normal vote. *Elections and the political order*, page 939.
- Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.
- Crespin, M., Darmofal, D., and Eaves, C. (2011). The political geography of congressional elections. In *Annual meeting of the Midwest Political Science Association*, volume 31. 00001.
- Curtice, J. and Steed, M. (1986). Proportionality and exaggeration in the British electoral system. *Electoral Studies*, 5(3):209228. 00057.
- Darmofal, D. (2006). Spatial econometrics and political science. *Society for Political Methodology Working Paper Archive*: <http://polmeth.wustl.edu/workingpapers.php>.
- De Berg, M., Cheong, O., Van Kreveld, M., and Overmars, M. (2008). *Computational geometry: algorithms and applications*. Springer.
- Dong, G. and Harris, R. (2015). Spatial autoregressive models for geographically hierarchical data structures. *Geographical Analysis*, 47(2):173191.
- Duque, J. C., Anselin, L., and Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science*, 52(3):397419.
- Durch, R. M. and Stevenson, R. (2005). Context and the Economic Vote: A Multilevel Analysis. *Political Analysis*, 13(4):387409.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.
- Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):3648.
- Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*, volume 5. Cambridge University Press.
- Elliott, A. C. and Hynan, L. S. (2011). A SAS[®] macro implementation of a multiple comparison post hoc test for a KruskalWallis analysis. *Computer methods and programs in biomedicine*, 102(1):7580.

- Fan, C., Li, W., Wolf, L. J., and Myint, S. W. (2015). A Spatiotemporal Compactness Pattern Analysis of Congressional Districts to Assess Partisan Gerrymandering: A Case Study with California and North Carolina. *Annals of the Association of American Geographers*, 105(4):736753.
- Franzese Jr, R. J. and Hays, J. C. (2007). Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis*, page 140164.
- Franzese Jr, R. J. and Hays, J. C. (2008). Interdependence in comparative politics: Substance, theory, empirics, substance. *Comparative Political Studies*, 41(4-5):742780.
- Friedman, J. N. and Holden, R. T. (2009). The rising incumbent reelection rate: What's gerrymandering got to do with it? *The Journal of Politics*, 71(2):593611.
- Fryer, R. and Holden, R. (2011). Measuring the Compactness of Political Districting Plans. *The Journal of Law and Economics*, 54:493. 00000.
- Gallagher, M. (1991). Proportionality, Disproportionality, and Electoral Systems. *Electoral Studies*, 10(1):3351.
- Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, 48(3):432435.
- Gelman, A. (2007). Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut? *Quarterly Journal of Political Science*, 2(4):345367.
- Gelman, A. (2014). How Bayesian Analysis Cracked the Red-State, Blue-State Problem. *Statistical Science*, 29(1):2635.
- Gelman, A. and King, G. (1990). Estimating the electoral consequences of legislative redistricting. *Journal of the American Statistical Association*, 85(410):274282. 00081.
- Gelman, A. and King, G. (1993). Why are American presidential election campaign polls so variable when votes are so predictable? *British Journal of Political Science*, 23(04):409451.
- Gelman, A. and King, G. (1994a). A Unified Method of Evaluating Electoral Systems and Redistricting Plans. *American Journal of Political Science*, 38(2):51454. 00179.
- Gelman, A. and King, G. (1994b). Enhancing Democracy Through Legislative Redistricting. *American Political Science Review*, 88(03):541559.
- Gelman, A., King, G., and Thomas, A. (2010). Judgelt II: A Program for Evaluating Electoral Systems and Redistricting Plans. 00010.
- Gelman, A., Shor, B., Bafumi, J., and Park, D. (2005). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Poor State, Red State, Blue State: What's the Matter with Connecticut*.

- Gerring, J., Palmer, M., Teorell, J., and Zarecki, D. (2015). Demography and democracy: A global, district-level analysis of electoral contestation. *American Political Science Review*, 109(03):574591.
- Ghitza, Y. and Gelman, A. (2013). Deep Interactions with MRP: Election Turnout and Voting Patterns among Small Electoral Subgroups. *American Journal of Political Science*, 57(3):762776.
- Gibson, J. G. (1992). Measuring electoral change: look before you abandon swing. *PS: Political Science & Politics*, 25(02):195198.
- Giles, M. W. and Hertz, K. (1994). Racial threat and partisan identification. *American Political Science Review*, 88(2):317326.
- Gimpel, J. G. and Schuknecht, J. E. (2009). *Patchwork nation: Sectionalism and political change in American politics*. University of michigan Press.
- Godfrey, H., Courville, C., and Nelson, J. A. (2005). Brief of Amici Curiae Professors Gary King, Bernard Grofman, Andrew Gelman, and Jonathan N. Katz in support of neither party.
- Griffith, D. A. and Arbia, G. (2010). Detecting negative spatial autocorrelation in georeferenced random variables. *International Journal of Geographical Information Science*, 24(3):417437.
- Grissom, R. J. and Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications*. Routledge.
- Grofman, B. (1983). Measures of Bias and Proportionality in Seats-Votes Relationships. *Political Methodology*, 9(3).
- Grofman, B. (1985). Criteria for Districting: A Social Science Perspective. *UCLA Law Review*, 33(1):77184.
- Grofman, B., Brunell, T., and Campagna, J. (1997a). Distinguishing Between the Effects of Swing Ratio and Bias on Outcomes in the US Electoral.
- Grofman, B. and King, G. (2007). The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after *LULAC v. Perry*. *Election Law Journal*, 6(1):235.
- Grofman, B., Koetzle, W., and Brunell, T. (1997b). An Integrated Perspective on the Three Potential Sources of Partisan Bias: Malapportionment, Turnout Differences, and the Geographic Distribution of Party Vote Shares. *Electoral Studies*, 16(4):457470.
- Gudgin, G. and Taylor, P. J. (1979). *Seats, votes, and the spatial organisation of elections*. Pion.
- Haining, R. (2003). *Spatial data analysis: theory and practice*. Cambridge Univ Pr.
- Harbers, I. (2016). Spatial effects and party nationalization: The Geography of partisan support in Mexico. *Electoral Studies*.

- Harris, P., Brunson, C., Lu, B., Nakaya, T., and Charlton, M. (2017). Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spatial Statistics*, 21:241261.
- Hasen, R. L. (2013). Foxes, Henhouses, and Commissions: Assessing the Nonpartisan Model in Election Administration, Redistricting, and Campaign Finance. *UC Irvine L. Rev.*, 3:4671281.
- Hawkes, A. (1969). An Approach to the Analysis of Electoral Swing. *Journal of the Royal Statistical Society, Series A*, 132(1):6879.
- Hawley, G. and Sagarzazu, I. (2012). Where did the votes go? Reassessing American party realignments via vote transfers between major parties from 1860 to 2008. *Electoral Studies*, 31(4):726739.
- Hersh, E. D. and Nall, C. (2015). The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records. *American Journal of Political Science*.
- Hill, T. (2010). *Redistricting and the U.S. House of Representatives: Illuminating Electoral Bias with the Brookes Method*. PhD thesis, Massachusetts Institute of Technology, Boston, MA.
- Hoaglin, D. C. and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, 32(1):1722.
- Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325334.
- Hogan, R. E. (2005). Gubernatorial coattail effects in state legislative elections. *Political Research Quarterly*, 58(4):587597.
- Humphreys, M. (2011). Can Compactness Constrain the Gerrymander? *Irish Political Studies*, 26(4):513520.
- Jackman, S. (1994). Measuring Electoral Bias: Australia, 194993. *Br. J. Polit. Sci.*, 24(03):319357.
- Jackman, S. (2014). The predictive power of uniform swing. *PS: Political Science & Politics*, 47(2):317321.
- Jackman, S. (2015). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*. Department of Political Science, Stanford University, Stanford, California. R package version 1.4.9.
- Jackman, S. (2017). Assessing the Current Wisconsin State Legislative Districting Plan. Technical Report 62, United States District Court.
- Johnston, R. (1982). The Geography of Electoral Change: An illustration of an Estimating Procedure. *Geografiska Annaler, Series B*, 64(1):5160.

- Johnston, R. (1983). Spatial Continuity and Individual Variability: A Review of Recent Work on the Geography of Electoral Change. *Electoral Studies*, 2(1):5368.
- Johnston, R. (2002). Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering. *Polit. Geogr.*, 21(1):131.
- Johnston, R. (2005). Anglo-American Electoral Geography: Same Roots and Same Goals, but Different Means and Ends?*. *Prof. Geogr.*, 57(4):580587.
- Johnston, R., Manley, D., and Jones, K. (2016). Spatial Polarization of Presidential Voting in the United States, 1992-2012: The Big Sort Revisited. *Annals of the American Association of Geographers*, 106(5):10471062.
- Johnston, R., Rossiter, D., and Pattie, C. (1999). Integrating and decomposing the sources of partisan bias: Brookes' method and the impact of redistricting in Great Britain. *Electoral Studies*, 18(3):367378.
- Johnston, R., Rossiter, D., and Pattie, C. (2005). Disproportionality and bias in US Presidential Elections: How geography helped Bush defeat Gore but couldn't help Kerry beat Bush. *Political Geography*, 24(8):952968.
- Johnston, R. J. (1979). *Political, electoral, and spatial systems: an essay in political geography*. Clarendon Press.
- Kastellec, J. P., Gelman, A., and Chandler, J. P. (2008). Predicting and Dissecting the Seats-Votes Curve in the 2006 U.S. House Election. *PS: Political Science & Politics*, 41(01).
- Katz, R. S. (1973a). Rejoinder to Comment by Donald E. Stokes. *American Political Science Review*, 67(03):832834.
- Katz, R. S. (1973b). The attribution of variance in electoral returns: An alternative measurement technique. *American Political Science Review*, 67(03):817828.
- Kayser, M. A. and Lindstädt, R. (2015). A cross-national measure of electoral competitiveness. *Political Analysis*, 23(2):242253.
- Kendall, M. G. and Stuart, A. (1950). The law of the cubic proportion in election results. *The British Journal of Sociology*, 1(3):183196.
- King, G. (1989). Representation through legislative redistricting: A stochastic model. *American Journal of Political Science*, page 787824. 00088.
- King, G. (1994). Elections to the United States House of Representatives, 1892-1992. Inter-University Consortium of Political Science Research. <http://doi.org/10.3886/ICPSR06311.v1>.
- King, G. (1996). Why context should not count. *Polit. Geogr.*, 15(2):159164.
- King, G. and Browning, R. X. (1987). Democratic Representation and Partisan Bias in Congressional Elections. *American Political Science Review*, 81(4):1251.

- King, G. and Roberts, M. E. (2015). How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It. *Political Analysis*, 23:159179.
- King, G., Rosen, O., Tanner, M., and Wagner, A. F. (2008). Ordinary economic voting behavior in the extraordinary election of Adolf Hitler. *Journal of Economic History*, 68(04):951996.
- King, G., Tomz, M., and Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *Am. J. Pol. Sci.*, page 347361.
- Kollman, K., Hicken, A., Caramani, D., Backer, D., and Lublin, D. (2016). Constituency-level elections archive. Center for Political Studies, University of Michigan.
- Kousser, J. M. (1996). Estimating the Partisan Consequences of Redistricting Plans simply. *Legislative Studies Quarterly*, 21(4):521541.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. *Journal of Experimental Psychology*, 142(2):573603.
- Lacombe, D. J., Holloway, G. J., and Shaughnessy, T. M. (2014). Bayesian estimation of the spatial Durbin error model with an application to voter turnout in the 2004 presidential election. *International Regional Science Review*, 37(3):298327.
- Lacombe, D. J. and McIntyre, S. G. (2016). Local and global spatial effects in hierarchical models. *Applied Economics Letters*, 23(16):11681172.
- Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, 73(1):3143.
- Leckie, G., French, R., Charlton, C., and Browne, W. (2014). Modeling Heterogeneous Variance-Covariance Components in Two-Level Models. *J. Educ. Behav. Stat.*, 39(5):307332.
- Levendusky, M. S., Pope, J. C., and Jackman, S. (2008). Measuring District-Level Partisanship with Implications for the Analysis of U.S. Elections. *The Journal of Politics*, 70(3):736753.
- Lewis, J. B., DeVine, B., Pitcher, L., and Martis, K. C. (2013). Digital Boundary Definitions of United States Congressional Districts, 1789-2012.[Data file and code book].
- Li, W., Chen, T., Wentz, E. A., and Fan, C. (2014a). NMMI: A Mass Compactness Measure for Spatial Pattern Analysis of Areal Features. *Ann. Assoc. Am. Geogr.*, 104(6):11161133.
- Li, W., Church, R. L., and Goodchild, M. F. (2014b). The p-compact-regions problem. *Geographical Analysis*, 46(3):250273.
- Li, W., Goodchild, M. F., and Church, R. (2013). An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems. *International Journal of Geographical Information Science*, 27(6):12271250.
- Linzer, D. A. (2012). The Relationship between Seats and Votes in Multiparty Systems. *Political Analysis*, 20(3).

- Lo, J. et al. (2013). Legislative responsiveness to gerrymandering: Evidence from the 2003 Texas redistricting. *Quarterly Journal of Political Science*, 8(1):7592.
- MacEachren, A. M. (1985). Compactness of geographic shape: Comparison and evaluation of measures. *Geografiska Annaler. Series B. Human Geography*, page 5367.
- Mackerras, M. (1973). The Swing: variability and uniformity. *Politics*, 8(1):234241.
- Mackerras, M. (1976). Uniform Swing: analysis of the 1975 election. *Politics*, 11(1):4146.
- Mackerras, M. (1978). Rejoinder to Campbell Sharman. *Politics*, 13(2):339342.
- Mallows, C. L. (1975). On some topics in robustness. Technical report.
- Mandelbrot, B. B. (1967). How long is the coast of Britain? Statistical Self-similarity and fractal dimension. *Science*, 156(3775):636638.
- McCarty, N., Poole, K. T., and Rosenthal, H. (2009). Does gerrymandering cause polarization? *American Journal of Political Science*, 53(3):666680.
- McDonald, M. D. and Best, R. E. (2015). Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases. *Election Law Journal*, 14(4):312330.
- McDonald, M. P. (2006). Drawing the line on district competition. *PS: Political Science & Politics*, 39(01):9194.
- McGann, A., Smith, C. A., Latner, M., and J. Alex, K. (2015). A Discernable and Manegable Standard for partisan Gerrymandering. *Election Law Journal*, 14(4):295311.
- McGann, A., Smith, C. A., Latner, M., and Keena, A. (2016). *Gerrymandering in America: The House of Representatives, the Supreme Court, and the Future of Popular Sovereignty*. Cambridge University Press, 1st ed edition.
- McGhee, E. (2014). Measuring Partisan Bias in Single-Member District Electoral Systems. *Legislative Studies Quarterly*, 39(1):5585.
- McKee, S. C., Teigen, J. M., and Turgeon, M. (2006). The Partisan Impact of Congressional Redistricting: The Case of Texas, 2001-2003. *Social Science Quarterly*, 87(2):308317. 00000.
- Miller, A. H. (1979). Normal Vote Analysis: Sensitivity to Change Over Time. *American Journal of Political Science*, 23(2):406425.
- Miller, P. and Grofman, B. (2013). Redistricting Commissions in the Western United States. *UC Irvine L. Rev.*, 3:637. 00006.
- Monogan, J. (2012). The Fifty American States in Space and Time: Applying Conditionally Autoregressive Models to State Politics. *Unpublished Manuscript*.
- Monogan, J. E. (2013). The politics of immigrant policy in the 50 US states, 2005-2011. *Journal of Public Policy*, 33(01):3564.

- Morrill, R. L. (1973). Ideal and reality in reapportionment. *Annals of the Association of American Geographers*, 63(4):463477.
- Nagle, J. F. (2015). Measures of Partisan Bias for Legislating Fair Elections. *Election Law Journal*, 14(4):346360.
- Nagle, J. F. (2017). How Competitive Should a Fair Single Member Districting Plan Be? *Election Law Journal*, 16(1):196209.
- Niemi, R. G. and Deegan, J. J. (1978). A Theory of Political Districting. *The American Political Science Review*, 72(2):13041323.
- Niemi, R. G. and Fett, P. (1986). The Swing Ratio: An Explanation and an Assessment. *Legislative Studies Quarterly*, 11(1):7590.
- Niemi, R. G., Grofman, B., Carlucci, C., and Hofeller, T. (1990). Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. *The Journal of Politics*, 52(04):11551181.
- O'Loughlin, J., Flint, C., and Anselin, L. (1994). The geography of the Nazi vote: Context, confession, and class in the Reichstag election of 1930. *Annals of the Association of American Geographers*, 84(3):351380.
- Open Geospatial Consortium (2010). Simple Features Specification. Technical report, Open Geospatial Consortium. <http://www.opengeospatial.org/standards/sfa>.
- Openshaw, S. and Taylor, P. J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, 21:127144.
- Owen, G., Harris, R., and Jones, K. (2015). Under examination: Multilevel models, geography and health research. *Progress in Human Geography*, 40(3).
- Park, D. K., Gelman, A., and Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: state-level estimates from national polls. *Political Analysis*, page 375385.
- Pildes, R. H. and Niemi, R. G. (1993). Expressive Harms, "Bizarre Districts," and Voting Rights: Evaluating Election-District Appearances after *Shaw v. Reno*. *Michigan Law Review*, 92(3):483.
- Polsby, D. D. and Popper, R. D. (1991). The Third Criterion: Compactness as a Procedural Safeguard against Partisan Gerrymandering. *Yale Law Policy Rev.*, 9(2):301353.
- Poole, K. T. and Rosenthal, H. (1987). Analysis of Congressional Coalition Patterns: A Unidimensional Spatial Model. *Legislative Studies Quarterly*, 12(1).
- Rasmussen, J. (1964). The disutility of the swing concept in British Psephology. *Parliamentary Affairs*, 18(4):442452.

- Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., and Farrell, C. R. (2006). *The segregation profile: investigating how metropolitan racial segregation varies by spatial scale*.
- Reock, E. C. (1961). A Note: Measuring Compactness as a Requirement of Legislative Apportionment. *Midwest Journal of Political Science*, 5(1):70. 00059.
- Rey, S. J. and Anselin, L. (2007). PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*, 37(1):527.
- Rey, S. J. and Dev, B. (2006). -convergence in the presence of spatial effects. *Papers in Regional Science*, 85(2):217234.
- Rocha, R. R. and Espino, R. (2009). Racial Threat, Residential Segregation, and the Policy Attitudes of Anglos. *Political Research Quarterly*, 62(2):415426. 00058.
- Rodden, J. (2010). The Geographic Distribution of Political Preferences. *Annual Review of Political Science*, 13(1):321340. 00044.
- Rose, R. (1991). The ups and downs of elections, or look before you swing. *PS: Political Science & Politics*, 24(01):2933.
- Sastry, N., Pebley, A. R., and Zonta, M. (2002). *Neighborhood definitions and the spatial dimensions of daily life in Los Angeles*.
- Schwartzberg, J. E. (1965). Reapportionment, gerrymanders, and the notion of compactness. *Minn. L. Rev.*, 50:443.
- Sharman, C. (1978). Swing and the two-party preferred vote: a comment on Malcolm Mackerras. *Politics*, 13(2):336339.
- Shelley, F. M. and Archer, J. C. (1995). The volatile south: A historical geography of presidential elections in the south, 1872-1992. *Southeastern Geographer*, 35(1):2236.
- Shin, M. E. and Agnew, J. (2007). The Geographical Dynamics of Italian Electoral Change, 1987-2001. *Electoral Studies*, 26:287302.
- Skyum, S. (1990). A Simple Algorithm for Computing the Smallest Enclosing Circle. *DAIMI PB*, (314):19.
- Stephanopoulos, N. (2013). The Consequences of Consequentialist Criteria. *UC Irvine Law Review*, 669. U Chicago, Public Law Working Paper No. 423.
- Stephanopoulos, N. O. and McGhee, E. M. (2015). Partisan Gerrymandering and the Efficiency Gap. *University of Chicago Law Review*, 82:8311753.
- Stokes, D. E. (1965). A variance components model of political effects. *Mathematical applications in political science*, 1(1):6185.
- Taagepera, R. and Shugart, M. S. (1989). *Seats and votes: The effects and determinants of electoral systems*. Yale University Press.

- Tam Cho, W. K. (2017). Measuring Partisan Fairness: How Well Does the Efficiency Gap Guard Against Sophisticated as well as Simple-Minded Modes of Partisan Discrimination? *University of Pennsylvania Law Review Online*, 166(1):2.
- Tan, F. E., Ouwens, M. J., and Berger, M. P. (2001). Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(3):271284.
- Thomas, A., Best, N., Lunn, D., Arnold, R., and Spiegelhalter, D. (2004). GeoBugs user manual. *Cambridge: Medical Research Council Biostatistics Unit*.
- Thomas, A. C., Gelman, A., King, G., and Katz, J. N. (2013). Estimating Partisan Bias of the Electoral College Under Proposed Changes in Elector Apportionment. *Statistics, Politics and Policy*, 4(1):113.
- Tufte, E. R. (1973). The Relationship between Seats and Votes in Two-Party Systems. *The American Political Science Review*, 67(2):540.
- Vickrey, W. (1961). On the prevention of gerrymandering. *Political Science Quarterly*, 76(1):105110.
- Walker, K. E. (2013). Political Segregation of the Metropolis: Spatial Sorting by Partisan Voting in Metropolitan Minneapolis-St Paul: POLITICAL SEGREGATION IN MINNEAPOLIS-ST. PAUL, MINNESOTA. *City & Community*, 12(1):3555. 00000.
- Wang, S. S.-H. (2016). Three Practical Tests for Gerrymandering: Application to Maryland and Wisconsin. *Election Law Journal*, 15(4):367384.
- Warf, B., editor (2006). *Encyclopedia of Human Geography*. SAGE Publications, Inc., 1 edition.
- Warf, B. and Leib, J. (2011). *Revitalizing electoral geography*. Ashgate, Burlington, Vt. 00004.
- Weaver, J. B. and Hess, S. W. (1963). A Procedure for Nonpartisan Districting: Development of Computer Techniques. *The Yale Law Journal*, 73(2):288. 00164.
- Webster, G. R. (2013). Reflections on current criteria to evaluate redistricting plans. *Political Geography*, 32:314. 00004.
- Wentz, E. A. (2000). A shape definition for geographic applications based on edge, elongation, and perforation. *Geographical Analysis*, 32(2):95112.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3-4):434449.
- Wilson, R. J. (1978). *The Impact of Modernization on British Columbia Electoral Patterns: Communications Development and the Uniformity of Swing, 1903-1975*. PhD thesis, University of British Columbia, Vancouver, BC, Canada.
- Wing, I. S. and Walker, J. L. (2010). The Geographic Dimensions of Electoral Polarization in the 2004 U.S. Presidential Vote. In Paez, A., Gallo, J., Buliung, R. N., and Dall'erba, S., editors, *Progress in Spatial Analysis, Advances in Spatial Science*, page 253285. Springer Berlin Heidelberg. 00003.

- Wolf, L. J. (2016). A Gibbs sampler for multilevel variance components model with separable spatial effects. Technical report, University of Chicago Center for Spatial Data Science.
- Yoshinaka, A. and Murphy, C. (2009). Partisan gerrymandering and population instability: Completing the redistricting puzzle. *Political Geography*, 28(8):451462.
- Young, H. P. (1988). Measuring the compactness of legislative districts. *Legislative Studies Quarterly*, page 105115.