Are Changing Constituencies Driving Rising Polarization in the U.S. House of Representatives?

Jesse Sussell, James A. Thomson

For more information on this publication, visit www.rand.org/t/rr896

Published by the RAND Corporation, Santa Monica, Calif.
© Copyright 2015 RAND Corporation
RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.html.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Preface

The level of partisan polarization in the U.S. Congress is higher now than at any time since the Civil War. Although some view polarization as a normal outcome of the political struggle between left and right, it can also have undesirable effects; for example, the "fiscal cliff" and government shutdown scenarios. For this and many other reasons, political polarization has important consequences for the policymaking process. This report explores the issue of whether the increase in polarization can be attributed, at least in part, to growing geographic separation of liberal and conservative voters. It addresses two questions: first, whether the spatial distribution of the American electorate has become more geographically clustered over the past 40 years with respect to party voting and socioeconomic attributes; and second, whether this clustering process has contributed to rising polarization in the U.S. House of Representatives. We find support for both hypotheses and estimate that long-term geographical clustering of voters is responsible for roughly 30 percent of the increase in polarization in the House between the 93rd and 112th Congresses. An important ancillary finding is that socioeconomic variables—including those measuring race, education, income, and urbanicity—are dwarfed by the within-district percentage of married adults as a predictor of local partisanship, as measured by both the party affiliation of the House representative and by the presidential vote share.

This report will be of interest to researchers and policymakers interested in the causes and consequences of congressional polarization, the effects of polarization on the policymaking process, and possible strategies for mitigating polarization.

This report is a product of the RAND Corporation's continuing program of self-initiated independent research. Support for such research is provided, in part, by donors to RAND.

The research was conducted within the Pardee RAND Graduate School. The Pardee RAND Graduate School is unique in American higher education. Founded in 1970 as one of eight graduate programs created to train future leaders in public policy, Pardee RAND is the only program specializing exclusively in the Ph.D. It is also the only program based at a public policy research organization—the RAND Corporation.

For substantive questions about this report, contact Jesse Sussell at jsussell@gmail.com. For more information on the Pardee RAND Graduate School, see www.prgs.edu.

Contents

Preface iii
Figures vii
Tables ix
Summary xi
Acknowledgments xiii
Abbreviations xv

CHAPTER ONE
Introduction 1

CHAPTER TWO
Is Partisan Geographic Clustering of the American Electorate a Reality? 3
The Big Sort: Concepts and Critiques 3
Re-Reconsidering the Clustering Question 4

CHAPTER THREE
Is Geographic Clustering of Voters Driving Rising Polarization in Congress? 13
Method 1: The Regression Discontinuity Model 16
Method 2: The Rescaling Model 18
Method 3: The Multistage Model 18
Findings 19

CHAPTER FOUR
Discussion and Conclusion 25

APPENDIX
Notes and Technical Methods 27

References 45

Figures

2.1. Weighted Standard Deviation of Average PREP Across Counties, 1948–2008 6
2.2. Standard Deviation of PREP Across Congressional Districts, 1952–2010 7
2.3. Distribution of College Attainment Across U.S. Congressional Districts, 1970 (Actual), 2010 (Actual and Hypothetical) 9
2.4. Distribution of Marriage Prevalence Across U.S. Congressional Districts, 1970 (Actual), 2010 (Actual and Hypothetical) 11
2.5. Percentage Change in Congressional District–Level Standard Deviation of College Attainment, Marriage Prevalence, and Median Family Income, 1970–2010 12
3.1. House Polarization Has Been Rising Since the 1970s....................................... 14
3.2. Geographic Dispersion in Presidential Voting and Polarization in Roll-Call Votes in the House of Representatives 16
3.3. DW-NOMINATE Scores and District Election Republican Vote Share, 108th Congress (2003)... 17
A.1. Percentage Change in Population-Weighted County-Level Standard Deviation of College Attainment, Marriage Prevalence, and Average Income, 1970–2010............ 28
A.2. RD-Derived Estimate of the Upper Bound of the Percentage of House Polarization Attributable to Differences in Constituencies, 93rd–112th Congresses (Pooled) 31
A.3. Absence of Discontinuity at 0.5 Vote Threshold for District-Level Covariates, 2010 Midterm Elections 33
A.4. Distribution of College Education Rates by U.S. Congressional District, 1970 and 2010 34
A.5. Actual and Rescaled Distributions for 2010 College Education Rates, U.S. Congressional Districts 34
A.6. Difference in Inter-Party Means of DW-NOMINATE Scores as Predicted by Actual and Rescaled District-Level Covariate Values, 1972–2010.......................... 35
A.7. Variable Effects from Logit Model for “R” 37
A.8. Variable Effects from OLS Model of PREP 37
A.9. Actual Polarization and Polarization Predicted from Model of Constituency Characteristics... 40
A.10. Actual Polarization and Polarization Predicted from Model of Constituency Characteristics (Full Model and Model Restricted to Marriage Covariate Alone) 40
A.11. Difference in Interparty Means of DW-NOMINATE Scores as Predicted by Actual, Rescaled, and Rescaled (Excluding Marriage) District-Level Covariate Values, 1972–2010... 41

Tables

3.1. Summary of Model Estimates: Contribution of Voter Clustering to Rising Polarization in the House of Representatives, 1972–2010 19
3.2. Republican and Democratic Membership of the House of Representatives, by Year of Election and Quartile of Within-District Marriage Prevalence.................... 20
A.1. U.S. House of Representatives' Mean Democratic and Republican First-Dimension DW-NOMINATE Scores, 1972–2010 29
A.2. OLS Regression of PREP on Constituency Variables, 1972–2008......................... 38
A.3. Marginal Effects from Logistic Regression of R (Indicator for Republican Legislator) on Constituency Variables, 1972–2010 39
A.4. Marginal and Incremental Effects of Individual Characteristics on the Probability of Voting for the Republican Candidate (Logistic Regression), 1972–2008 Presidential Elections (ANES) 42
A.5. Logistic Regression Results: Outcome Is Voting for the Republican Candidate in the Presidential Election (as Opposed to the Democratic Candidate), 1972–2008 43

Summary

Virtually all observers of American politics agree that there is a high degree of polarization between the Democratic and Republican parties in Congress. There is also a general consensus that this interparty polarization has been increasing over time, but much less consensus as to its causes. This paper addresses two questions: first, whether the spatial distribution of the American electorate has become more geographically clustered over the past 40 years with respect to party voting and socioeconomic attributes; and second, whether this clustering process has contributed to rising polarization in the U.S. House of Representatives. We find support for both hypotheses; our findings suggest that long-term growth in the geographical clustering of voters is responsible for roughly 30 percent of the increase in polarization in the House between the 93rd and 112th Congresses.

With respect to the first question, we examined changes over time in the geographic distribution of presidential voting and sociodemographic attributes. Instead of emphasizing changes in the means of distributions (as is typically done), we focused on changes in the standard deviations of the data. Standard deviation measures the overall level of dispersion—the average degree to which individual values in a population differ from the population average. We found that the levels of geographic dispersion of presidential voting, college attainment, median income, and the marriage rate have all risen significantly over the past 40 years. In other words, the extent to which places in America are different from one another with respect to these attributes has increased significantly. These changes were present at two different levels of geography (congressional districts and counties), which suggests that they did not result from gerrymandering.

We used three different statistical models to address the second question. Each of these looked at changes over time in polarization in the House of Representatives (as defined by changes over time in a quantitative measure of the ideology of individual members), and explored the relationship between those changes and within-district constituency changes over time. These models suffered from significant limitations, but also had methodological approaches that were relatively distinct. The results produced by the three models were consistent with one another, estimating that between 23 percent and 31 percent of the growth in House polarization between 1972 and 2010 can be attributed to changes in the distributions of constituencies over time.

An important ancillary finding is that the marriage rate—the within-district percentage of adults who are married—dwarfs other socioeconomic variables (including those measuring race, education, income, and urbanicity) as a predictor of local partisanship, as measured by both the party affiliation of the House representative and the presidential vote share.

Acknowledgments

We are grateful to Sarah Binder (Senior Fellow, Brookings Institution, and Professor, George Washington University) and Kevin McCarthy (Adjunct Staff, RAND) for serving as reviewers of this manuscript. We would also like to thank RAND colleagues Linda Martin, Dick Neu, and Win Boerckel, as well as Jeff Stonecash (Professor Emeritus, Syracuse University) for providing valuable feedback on earlier drafts. The views presented here are those of the authors alone.

Abbreviations

ANES	American National Election Survey—a recurring, nationally representative survey of Americans that focuses on political beliefs and voting behavior.
ATE	Average Treatment Effect—an estimate of the "effect" of a variable of interest on an outcome measure in which the estimate is averaged across all those receiving the treatment.
DW-NOMINATE	Dynamic Weighted Nominal Three-Step Estimation. NOMINATE was the first-generation DW-NOMINATE score, a widely used measure of legislator ideology. The score is measured on a liberal-conservative axis, with –1 indicating "very liberal" and +1 indicating "very conservative."
LATE	Local Average Treatment Effect—an estimate of the treatment effect that is "local" to a subset of the treated population; in the regression discontinuity context, this subset is those in the neighborhood of the discontinuity.
MPR	Nolan McCarty, Keith Poole, and Howard Rosenthal—a group of scholars whose work is cited throughout this document. Poole and Rosenthal invented the procedure to estimate DW-NOMINATE scores.
OLS	Ordinary least squares—a standard statistical technique for examining relationships between two or more variables.
OVB	Omitted variable bias—a concept in econometrics that says estimates of the effect of an explanatory variable on a dependent variable may be biased if other factors that affect the dependent variable are also (a) correlated with the explanatory variable in question and (b) not included in the model.
PREP	The Republican share of the two-party presidential vote, calculated as R/(R+D).
R	Dummy variable indicating a Republican legislator.
RD	Regression discontinuity—a quasi-experimental research design that allows for theoretically unbiased estimation of causal effects.
VRA	Voting Rights Act—the 1965 legislation that prohibited racial discrimination in voting; widely believed to have brought about the collapse of the South as a Democratic stronghold.

CHAPTER ONE

Introduction

Virtually all observers of American politics agree that there is a high degree of polarization between the Democratic and Republican parties in Congress. There is also a general consensus that this interparty polarization has been increasing over time: The ideological gap separating the parties of Tip O'Neill and Gerald Ford in the 1970s may have been large, but it was smaller than the distance between the Clinton Democrats and the Gingrich Republicans in the 1990s, and smaller still than the gulf between the parties of Obama and Boehner today. There is much less consensus, however, as to the causes of this rising polarization. Many authors have noted the role of the so-called "Southern Realignment"—the gradual transition of the Southern congressional delegation (through attrition, replacement, and—in some cases—party-switching) from a Democratic to a Republican stronghold. Others have hypothesized, variously, that rising polarization in Congress may be caused by gerrymandering, rising income inequality, closed primary elections, or poorly structured campaign finance laws.

Some scholars have claimed that rising polarization in Congress has been driven at least in part by changes in the nature and distribution of the electorate.[1] One theoretical model argues that the voting behavior of elected representatives is determined by four factors: the policy preferences of the members themselves, the preferences of the national political party to which members belong, the preferences of within-district constituencies, and the preferences of the within-district subconstituency likely to support the representative.[2] In particular, the notion that a lawmaker's voting behavior is determined (in part) by the preferences of within-district voters has been fairly well supported in the literature.[3] Under the reasonable assumption that this relationship does hold—i.e., on average, conservative districts tend to elect conservative members, and liberal districts liberal members—one possible explanation for rising polarization in Congress is the "Big Sort." This term describes the hypothesis—first proposed by Bill Bishop—that in recent decades, politically like-minded voters have become less diffuse and more clustered as a result of geographic "sorting" along economic, demographic, religious, and lifestyle lines.[4] Since members of the House represent specific geographic regions, clustering of like-minded voters into House districts could contribute to polarization in Congress as members respond to gradual changes in constituency preferences.[5]

---

[1] Jeffrey M. Stonecash, Mark D. Brewer, and Mack D. Mariani, *Diverging Parties: Social Change, Realignment, and Party Polarization*, Boulder, Colo.: Westview Press, 2003; Richard Fleisher and John R. Bond, "The Shrinking Middle in Congress," *British Journal of Political Science,* 2004.

[2] Steven D. Levitt, "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology," *American Economic Review,* June 1996.

[3] Kristina C. Miller, "The Limitations of Heuristics for Political Elites," *Political Psychology,* December 2009.

Our aim is to test this two-part hypothesis: Is the clustering described by Bishop in fact occurring? And if so, is it contributing to polarization in the House of Representatives? In Chapter Two, we provide evidence to support Bishop's hypothesis by showing that clustering across congressional districts has gradually increased along several lines—specifically income, education, and marriage. In Chapter Three, we present results from three analytical models designed to test the hypothesis that this clustering has contributed to growing polarization in the U.S. Congress since the mid-1970s. We conclude with a discussion of our findings and the implications they have for continued polarization and gridlock in Congress.

---

[4] Bill Bishop with Robert Cushing, *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart,* New York: Houghton Mifflin, 2008. Bishop uses the term *sorting* to describe this phenomenon; we use the term *clustering* to avoid confusion—other scholars have used the term *sorting* to describe the nongeographic process of conservatives increasingly identifying with the Republican party and liberals increasingly identifying with the Democratic party.

[5] See "Legislator Ideology as a Function of Constituency Attributes" in the Appendix for a simple model of how legislators' ideology might be modeled as a function of constituency attributes.

CHAPTER TWO

Is Partisan Geographic Clustering of the American Electorate a Reality?

The Big Sort: Concepts and Critiques

A number of authors have addressed the question of how the American electorate has changed in recent decades, and the subject remains a matter of some dispute. Yet books with starkly different titles that imply starkly different conclusions—Morris Fiorina's *Culture War? The Myth of a Polarized America,* and Alan Abramowitz's *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*—seem to agree on three key points: First, the electorate as a whole is not as polarized as Congress; second, the politically engaged portion of the electorate is more polarized than the voting population as a whole; and third, polarization within this group is growing.[1] Politically engaged liberals have increasingly moved into the Democratic Party and politically engaged conservatives into the Republican Party. Abramowitz shows that polarization within the electorate has increased and attributes this change to a significant increase over time in the fraction of voters who are politically engaged.

Fiorina and Abramowitz largely omit consideration of geography, which we believe provides a related but distinct framework for thinking about how the American electorate has changed over time.[2] Imagine that between 1972 and 2008, liberals and conservatives (or Democrats and Republicans) became increasingly clustered in different regions of the country: liberals in such places as San Francisco and Brooklyn, and conservatives in places like Orange County, Calif., and Kansas. Even in the absence of any changes to national-level averages in partisan ideological self-identification, this type of clustered electorate would be very different from one in which the spatial distribution of liberals and conservatives was relatively even.

This clustering phenomenon is precisely the type of change that Bill Bishop claims has occurred in the United States over the last several decades. In his 2008 book *The Big Sort*, Bishop argues that liberal and conservative voters in the United States have become increasingly spatially isolated from one another. His principal analysis in support of this claim is a comparison of county-level returns in the presidential election over time. To demonstrate that counties are becoming increasingly internally homogenous politically, Bishop divides counties into *landslide* and *competitive* counties, where *landslide* is defined as a margin greater than 20 percent for either the Republican or Democratic candidate. He makes the argument that clustering is occurring by comparing maps of the 1976 and 2004 presidential elections for the 48 contiguous states, with each county coded as being either "Landslide Democrat," "Landslide Republican," or "No Landslide."

---

[1] Morris P. Fiorina, Samuel J. Abrams, and Jeremy C. Pope, *Culture War? The Myth of a Polarized America*, 3rd. ed., New York: Pearson Education Inc., 2010; Alan I. Abramowitz, *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*, New Haven: Yale University Press, 2010.

[2] Fiorina does include a chapter on the "Red State/Blue State Divide" and concludes that among all voters, voters in Red states did not have significantly different political attitudes from those in Blue states—a striking finding, given the growing polarization in the Senate.

Bishop's evidence of an increase in clustering is the obvious and dramatic increase between the two elections in the number of landslide-classified counties.[3] The Southern realignment is obvious, as the blue band crossing the "Bible Belt" in 1976 had turned almost completely red by 2004. Also noteworthy in Bishop's maps is the fact that all but one of the coastal counties of California was competitive in 1976. By 2004, almost all the counties north of Orange County were solidly behind the Democratic candidate. Similarly, aside from a few rural areas, all Kansas counties were competitive in 1976. Almost none were in 2004, with most Kansas counties having become solid supporters of the Republican candidate. This general trend is consistent with the observation that the number of "Swing States" has declined in presidential elections. California, for example, was considered a swing state until the 1990s.

Bishop's findings have been disputed by some members of the academic community. In particular, Samuel Abrams and Morris Fiorina have criticized the "Big Sort" along several dimensions, one of which was the choice of only two presidential elections to demonstrate an increase in geographic clustering.[4] They argue that the decision to use data from only two elections introduced two major flaws to the analysis: First, presidential election returns are vulnerable to the effects of different candidates on the outcome; for example, Republican Gerald Ford may not have attracted the same voters as George W. Bush (in 2004) or John McCain (in 2008). Second, the choice of beginning and ending points exaggerated the differences because "1976 was the low point for the percentage of the population residing in landslide counties in the post–World War II period and 2004 the high point."[5]

A final limitation of Bishop's map-based approach is that strictly visual analysis tends to disproportionally weight counties with large land areas and small populations. For example, Elko County, in the northeast corner of Nevada, and Kings County, New York (Brooklyn) were both classified as landslides in the 2008 presidential election—Elko County for McCain and Kings County for Obama. The population of Kings County is more than 50 times greater than the population of Elko County (2.5 million to 50,000)—but the large land area of Elko County is easy to find on any U.S. map while Kings County is essentially invisible.

Re-Reconsidering the Clustering Question

We propose an alternative method for testing for the existence of clustering in the electorate, which is to calculate and compare the population-weighted standard deviation of the percentage of votes for the Republican presidential candidate across counties for elections since World

---

[3] Bishop, 2008.

[4] Samuel J. Abrams and Morris P. Fiorina, "The Big Sort That Wasn't: A Skeptical Reexamination," *PS: Political Science and Politics*, April 2012.

[5] Abrams and Fiorina, 2012.

War II.[6] Standard deviation is a measure of the dispersion of the distribution of a variable—the average degree to which individual values in a population differ from the population average.

Conceptually, a distribution in which most observations are extreme (e.g., 90 percent or 10 percent Republican) will have a higher standard deviation than a distribution in which most observations are close to the distribution's center (e.g., 55 percent or 45 percent Republican). An increasing standard deviation over time would thus indicate a growing geographic dispersion of Republican voting (and, by extension, of Democratic voting) measured across counties, and would therefore be consistent with Bishop's clustering hypothesis.

This approach addresses the criticisms already described. Instead of comparing results from two elections, we examine data for each of 16 presidential elections in the postwar period: Cherry-picking is impossible when data from all years in the relevant period are examined. Using population weighted–data resolves the third problem: Our analysis correctly assigns a higher weight to Kings County and a lower weight to Elko County.

Abrams' and Fiorina's first and most important criticism—that presidential voting data are an inherently flawed measure of preferences because of candidate-specific effects—remains valid. It is difficult to remedy this shortcoming because presidential voting data are the only measure of political preference that is consistently available both historically and at a relatively small unit of geography (counties or county equivalents) with complete coverage for the entire country. However, we note that the degree of concern with the validity of this measure is probably reduced when 16 elections are compared, instead of just two: The confounding effect of candidate-specific effects is lessened when a steady trend is evident across multiple decades of data.

Figure 2.1 shows the results of our analysis: the population-weighted standard deviation of the two-party Republican vote share in the presidential election (PREP) for every election between 1948 and 2008.

Two points stand out. First, Abrams and Fiorina were correct to describe the 1976 election as an anomaly. The geographic dispersion of PREP was at its postwar low in that election. Second, there is clearly an upward trend since 1976 or 1980 and especially since 1992.

Because our ultimate focus is on explaining rising polarization in the House of Representatives, we now shift our unit of analysis from counties (Bishop's original focus) to congressional districts: Figure 2.2 replicates Figure 2.1 at that level of geography, starting in 1952. Congressional districts are similar to counties in some ways—they are both medium-sized geographies, larger than census tracts but smaller than states. On the other hand, there are 3,140 counties and only 435 districts in the United States. Another key difference is that while county boundaries are almost entirely static over time, congressional district boundaries are redrawn every 10 years (with states gaining or losing seats) following the decennial census. Perhaps the most important difference is that congressional districts are intended to have roughly equivalent populations, while counties are not. For example, the sparsely populated state of Montana has 56 counties but only a single congressional district.

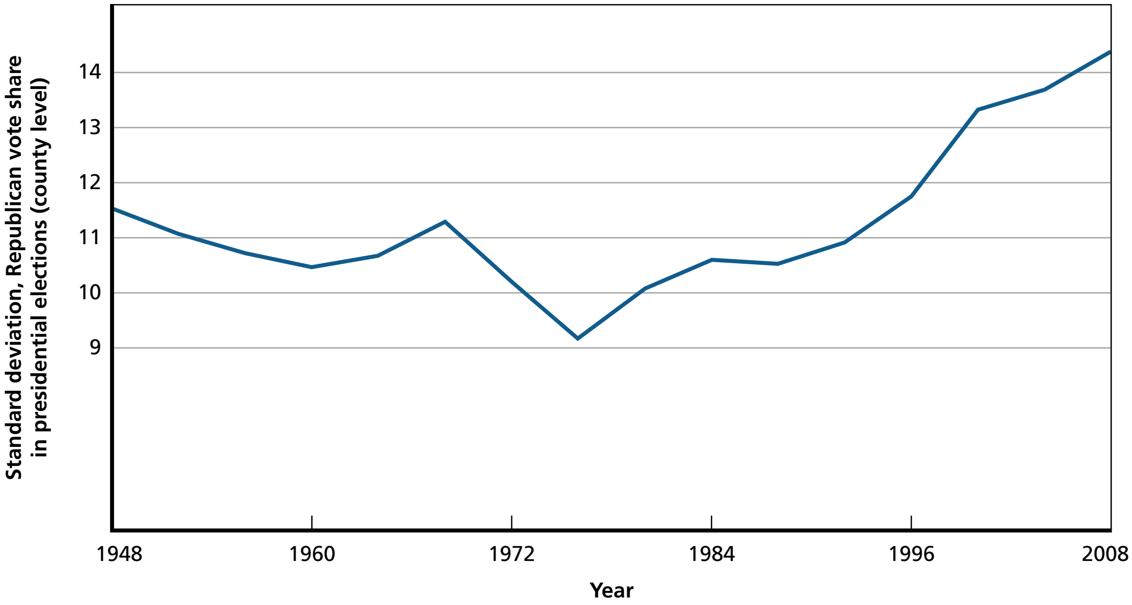
The pattern of Figure 2.2 is identical to that of Figure 2.1: There is a local trough in 1976, and a general upward trend following, especially since 1992.[7]

---

[6] See "Weighting County Data by Population" in the Appendix for details.

[7] As an alternative to PREP, Abrams and Fiorina examined party voter registration statistics in the 21 states for which voter registration data were available for both 1975 and 2008, and found no evidence of increase in the number of landslide counties in the bulk of the states. They conclude that when party registration statistics are analyzed, counties do not appear to be becoming more Democratic and more Republican, but rather more Republican and, in particular, more independent. This

Figure 2.1
Weighted Standard Deviation of Average PREP Across Counties, 1948–2008



RAND *RR896-2.1*

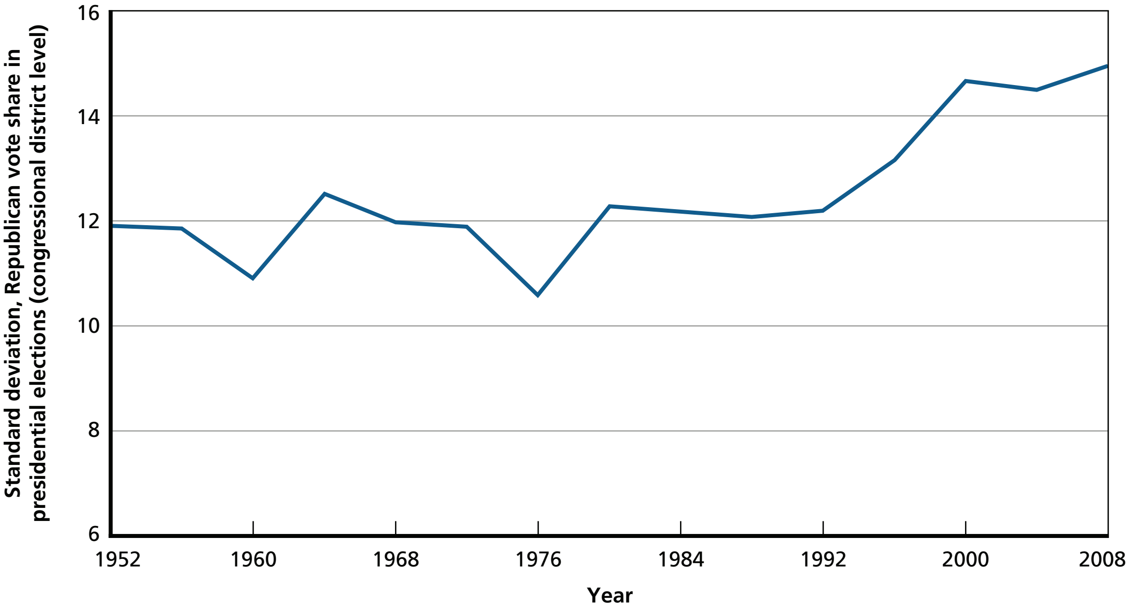
Bishop describes clustering as a neighborhood-level phenomenon. As Abrams and Fiorina point out, an additional (and important) critique of his original analysis is that as a geographical unit, counties are a relatively poor proxy for the concept of neighborhoods. In an alternative effort to look for political clustering at subcounty levels of geography, one of us also examined block group, tract, and county levels of both PREP and party registrations in California, using two statistical indices developed for studies of racial segregation.[8] This study provides strong evidence for the clustering hypothesis, albeit only in California. There is evidence of clustering across both indices, both political measures and all three geographic levels. The evidence is especially strong at the county level, and stronger at that level for registrations than for PREP, the opposite of Abrams' and Fiorina's conclusion. Other authors have used a variety of methods

---

increase in independent voters (technically those who list their party registration status as "independent," "decline to state," or "other") could easily produce a decline in the number of landslide counties according to registrations. However, this does not necessarily suggest a reduction in geographic clustering, as Abrams and Fiorina imply. As Abramowitz and other scholars have shown, registered independents are frequently partisans, and the terms "moderate" and "independent" should not be used interchangeably. (John R. Petrocik, "Measuring Party Support: Leaners Are not Independents," *Electoral Studies*, December 2009; Abramowitz, 2010.) Most independents are so-called "leaners" who vote fairly consistently for one party or the other in presidential elections. Admittedly, both PREP and party registration statistics have weaknesses as measures of local geographic political preferences. Despite its problems, PREP is the only one that classifies "leaners" based on revealed preference.

[8] Jesse Sussell, "New Support for the Big Sort Hypothesis: An Assessment of Partisan Geographic Sorting in California, 1992–2010," *PS: Political Science and Politics*, October 2013.

Figure 2.2
Standard Deviation of PREP Across Congressional Districts, 1952–2010



NOTE: Data for this figure begin in 1952, the first year the best data was available, instead of in 1948 as in Figure 2.1.
RAND *RR896-2.2*

to demonstrate that microgeographic partisan clustering has increased in Texas,[9] Cincinnati,[10] and Minneapolis.[11]

We feel that the trend in standard deviation shown in Figures 2.1 and 2.2, in combination with these other works, provides reasonably strong support for the hypothesis that partisan clustering is increasing, even in the face of the criticisms raised by Abrams and Fiorina.[12]

Bishop believed that clustering might be occurring because political ideologues share similar lifestyle preferences, and argued that an unintended consequence of clustering related to lifestyle choices (e.g., preferences for fair trade coffee or strong religious communities), aggregated over decades, might be an increase in internal political homogeneity at the local level. We do not have good measurements for many of the dimensions along which Bishop argued clustering was occurring: Even in 2015, things like preferences for hunting or vegetarian food are not reliably measured at meaningfully small levels of geography, and there is much more data available today compared with 40 years ago. Denominational religious affiliation is similarly not well measured at smaller geographic levels, particularly when looking backward in time. However, a number of variables that may serve as very general proxies for Bishop's concept of cultural and lifestyle differences are reliably measured in the decennial U.S. Census, with accurate data available at the congressional district level across several decades. We are particularly interested in trends over time in average educational attainment, median family income, and marriage prevalence.[13] Observed clustering along these lines (i.e., more and more places with high and low values over time) would be consistent with Bishop's hypothesis. Our ultimate goals are to examine whether geographic clustering along these attributes has increased over time—and if so, to examine how those changes have affected polarization in the House of Representatives.

---

[9] Adam S. Meyers, "Secular Geographical Polarization in the American South: The Case of Texas, 1996–2010," *Electoral Studies*, Vol. 32, No. 1, March 2013.

[10] Chad J. Kinsella, *The Little Sort: A Spatial Analysis of Polarization and the Sorting of Politically Like-Minded People,* dissertation, University of Cincinnati, Arts and Sciences: Political Science, May 2011.

[11] Kyle E. Walker, "Political Segregation of the Metropolis: Spatial Sorting by Partisan Voting in Metropolitan Minneapolis-St Paul," *City and Community*, Vol. 12, No. 1, March 2013.

[12] The question of *why* this clustering is occurring is an important one, but beyond the scope of this paper. However, we offer the following brief comment: One possible explanation is migration, which was in fact one of the central elements of Bishop's hypothesis. Two other factors besides migration are probably contributing to the observed patterns in the data. The first (*replacement*) would occur if the ideologies of new generations of voters differ within a region from those of older cohorts. The second (*realignment*) would occur if individuals changed their voting patterns or party registration status over time; for example, switching from voting for Democrats to Republicans in the presidential election. Realignment could occur if an individual's ideology changed over time or among individuals whose policy preferences are fixed over time. If the national parties and their candidates are systematically changing *which* constituencies they target (by altering the policy positions they espouse), then some individuals might switch parties without experiencing any personal ideological shift. The Reagan Democrat is a canonical example of this type of realignment.

We begin with an analysis of the geographic distribution of educational attainment across time. It is well known that the last 40 years have brought significant increases in educational attainment in the United States: According to data from the 1970 U.S. Census, 10.7 percent of the total population age 25 and older had earned a bachelor's degree or higher; by 2010, that value had risen to 27.4 percent, an increase of 16.7 percentage points.[14] What is interesting to us about these changes is that the gains were not distributed evenly across the country, as Figure 2.3 demonstrates.

The first frame of Figure 2.3 shows the actual distribution of college attainment across congressional districts in 1970.[15] In the average district, about 10 percent of adults had attained a college degree, but there was also a fair degree of dispersion—in many places, the value was lower or higher, and in a few places it was much higher, above 25 percent.

The second frame shows the actual distribution of educational attainment across congressional districts as observed in 2010—40 years after the data in the first frame. Two changes are evident: First, the mean is equal to 27.7 percent (equal to the true mean value in 2010), a significant increase over the 1970 mean. Second, the overall level of dispersion in the distribution has also increased dramatically.

The third frame isolates the first change from the second by rendering a hypothetical scenario in which the national gains in educational attainment between 1970 and 2010 were distributed evenly across all districts—in other words, a scenario in which every district saw its educational attainment rise by 17.1 percentage points. The mean of this distribution is 27.7 percent, but the shape and overall level of dispersion are identical to those of 1970. What is

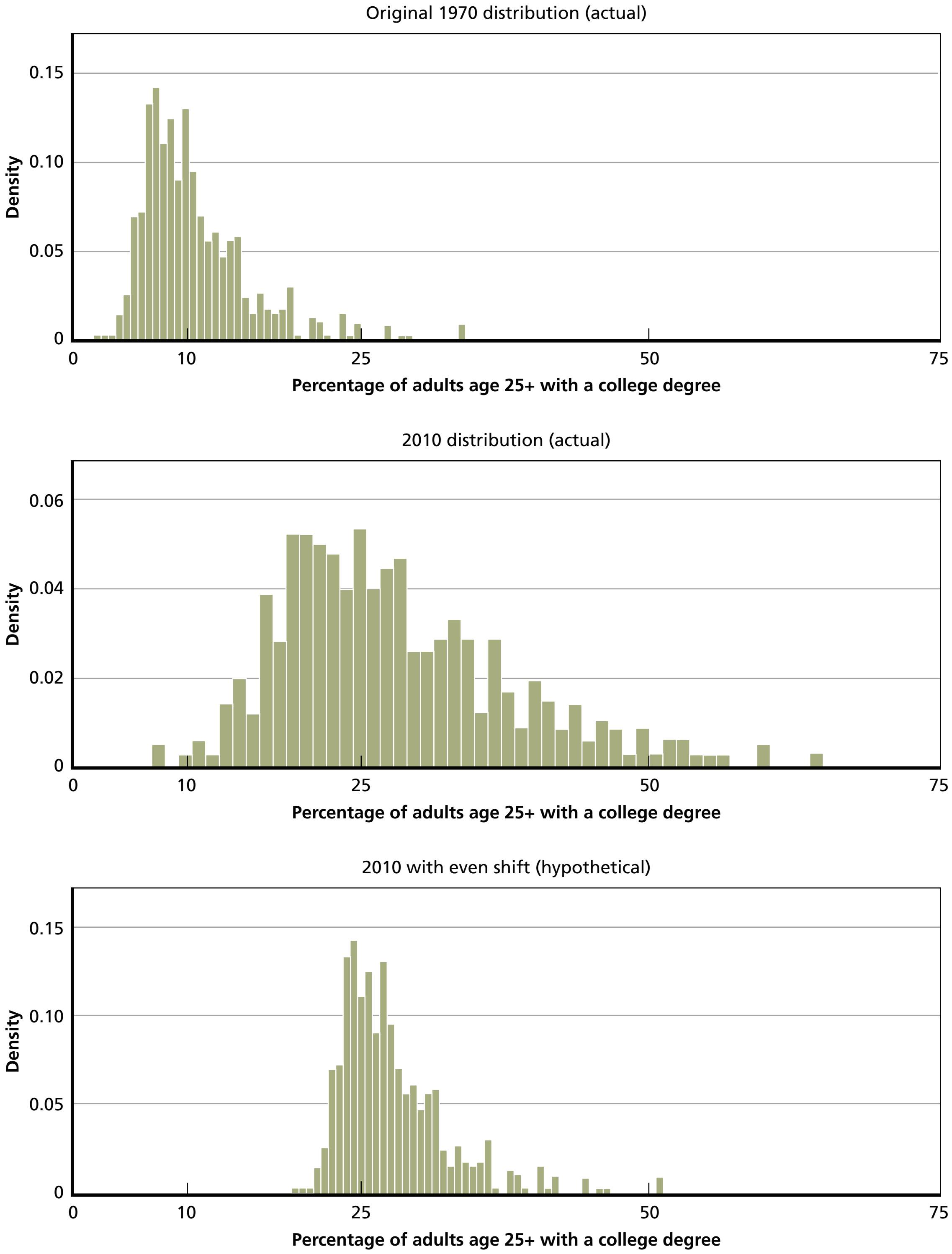
---

[13] We present results for these particular demographic variables for two reasons: First, it is reasonable to think that these things are associated with both lifestyle choices and with political preferences; second, during our initial exploratory analyses we examined all demographic variables that were available from the U.S. Census Bureau across the time frame of this study. Other factors (such as racial composition and average age) showed no increase over time in distributional dispersion.

[14] U.S. Dept. of Commerce, Bureau of the Census, *United States Congressional District Data Book for the Ninety-Third Congress, 1973*. ICPSR ed. Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research [producer and distributor], 1999; Minnesota Population Center. *National Historical Geographic Information System*, Version 2.0. Minneapolis, Minn.: University of Minnesota 2011. . http://doi.org/10.3886/ICPSR00011.v1

[15] These estimates are unweighted because the populations of congressional districts are roughly equal by design.

Figure 2.3
Distribution of College Attainment Across U.S. Congressional Districts, 1970 (Actual), 2010 (Actual and Hypothetical)



RAND *RR896-2.3*

clear in this comparison is that the overall gains in education in the United States have not been distributed uniformly with respect to place.

We can conduct a similar analysis for the overall rate of marriage prevalence—the percentage of persons age 15 and older who are married.[16] As with educational attainment, the national-level trend is well known: Marriage prevalence has been falling. In 1970, 61.5 percent of persons over 15 were married; by 2010, the value had fallen to 50.2 percent. And again, as with educational attainment, this shift over time did not occur equally in all areas of the country, as shown in Figure 2.4.

As with the distribution of educational attainment, the actual distribution of marriage prevalence in 2010 (second frame) has a greater level of dispersion than the hypothetical distribution (third frame)—there is more density in the tails and less in the center of the distribution.

We can make our analysis more precise by computing numerical measures of the dispersion we are describing (the standard deviation of the population-weighted distribution), and comparing values across time, as we did with PREP. Figure 2.5 shows the value of this statistic, generated at the congressional-district level for college attainment, adult marriage prevalence, and inflation-adjusted median family income. The values are normalized to 0 in 1970, so that percentage changes since 1970 are pictured.

Each of the normalized dispersion statistics grows significantly between 1970 and 2010: The largest growth is in the standard deviation of college attainment, which more than doubles over the time frame, but the indices derived from income and marriage prevalence also grow significantly. This figure suggests that the degree to which different parts of the country are dissimilar with respect to average education, income, and rates of marriage has increased over the last four decades. These trends are consistent with the hypothesis that a widespread, gradual clustering of the electorate is occurring.

The patterns we describe here also occur at the county level.[17] Because county boundaries are static over time, this is strong evidence that gerrymandering is not the primary explanatory factor for changes observed at the congressional district level. We also note that in separate analyses subset to the Southern states alone and to the non-Southern states alone, the patterns of rising dispersion evident in Figure 2.5 were present in both areas. Growth patterns were qualitatively similar in nature, although overall dispersion in the South was somewhat lower than in the rest of the country.

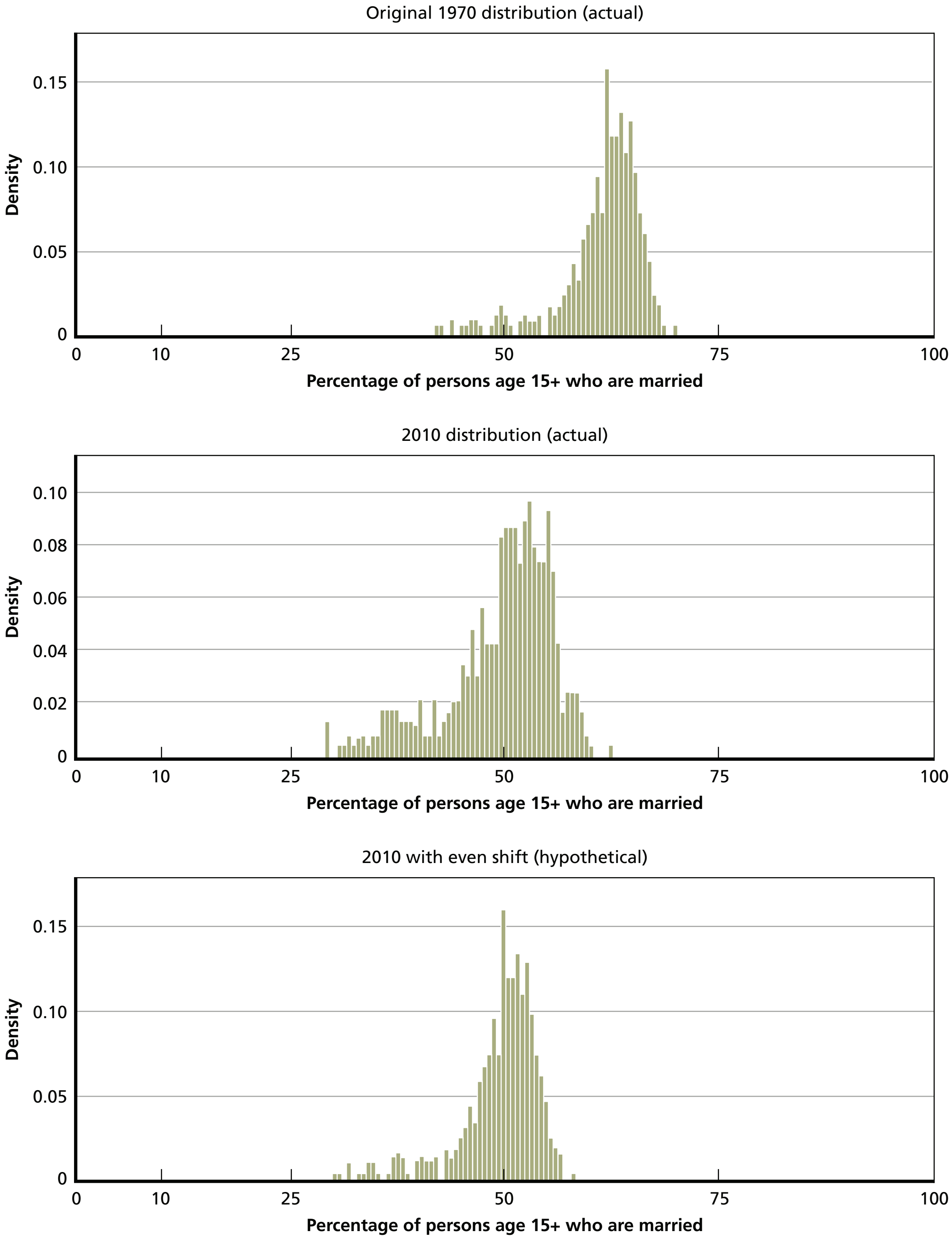
We observe the same patterns of a fairly dramatic increase in the relative dispersion of income, education, marriage, and party voting over time. We conclude this section by stating that these findings are entirely consistent with Bishop's hypothesis: If Americans were not gradually becoming clustered along political, educational, income, and marriage lines, we would expect the indices portrayed in Figures 2.1, 2.2, and 2.5 to be relatively flat over time. Instead, they are rising dramatically.

---

[16] Throughout this document, *marriage prevalence* is defined this way—as the percentage of persons age 15 and older who are currently married. Data from later censuses (2000, 2010) allow for differentiation between those who are married and cohabitating and those who are married but living apart, but earlier periods do not. We use this broader definition because it allows for consistency across time periods.

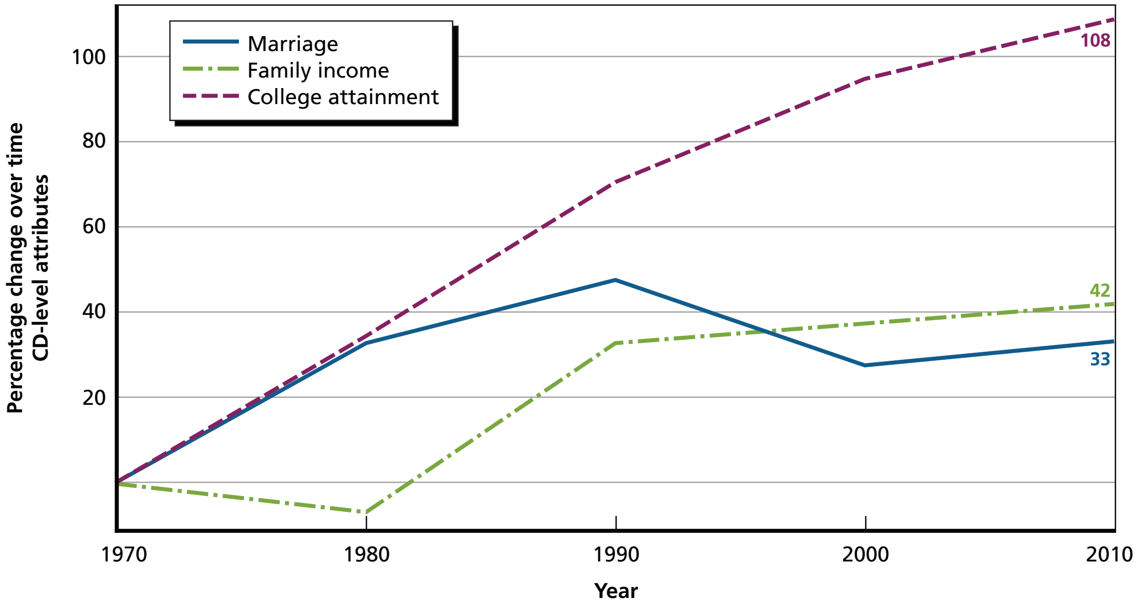
[17] See "Clustering over Time in Demographic Attributes, County-Level Data" in theAppendix for details.

Figure 2.4
Distribution of Marriage Prevalence Across U.S. Congressional Districts, 1970 (Actual), 2010 (Actual and Hypothetical)



RAND *RR896-2.4*

Figure 2.5
Percentage Change in Congressional District–Level Standard Deviation of College Attainment, Marriage Prevalence, and Median Family Income, 1970–2010



RAND *RR896-2.5*

CHAPTER THREE

Is Geographic Clustering of Voters Driving Rising Polarization in Congress?

Before beginning our discussion of the relationship between clustering of people (and therefore of voters) and polarization in Congress, we must first decide how to measure the latter construct. In the abstract, one way to think about congressional polarization is as the ideological "distance" between the average Democratic member of Congress and the average Republican. If we are interested in estimating polarization in this way, we must decide on a quantitative measure of individual legislators' ideologies. Imagine a score in which legislators are assigned a value according to how liberal or conservative they are, with 0 being perfectly liberal and 100 being perfectly conservative. If the average value for the Democratic caucus is 20, and the average value for the Republican caucus is 80, then we can operationalize the concept of congressional polarization as 60, the difference between the two party means.

There are several competing measures for the construct of individual legislator ideology, including scores produced by interest groups such as the American Conservative Union and Americans for Democratic Action. There are also several measures created by the academic community, including those estimated from roll-call votes and from the sources of individual members' fundraising.[1] In this paper we use the Dynamic Weighted Nominal Three-Step Estimation (DW-NOMINATE) scores developed by Nolan McCarty, Keith Poole, and Howard Rosenthal (MPR).[2]

These DW-NOMINATE estimates of ideological positions are derived from analyses of almost all recorded roll-call votes throughout U.S. history (except for unanimous and near unanimous votes). In the DW-NOMINATE framework, each legislator receives a multidimensional score; this permits any number of political dimensions (or ideological tendencies) to influence votes in Congress. A single, unidimensional score would classify legislators along a single liberal-conservative axis; a two-dimensional score would allow for distinctions such as "socially liberal, but fiscally conservative." For most of U.S. history, two dimensions have been sufficient to correctly classify the vast majority of roll-call votes; since the end of the civil rights struggle, one has sufficed.[3] That first dimension originally captured the traditional left-right differences over the role of the government in the economy.[4] We used the first-dimension DW-NOMINATE score as the measure of legislator ideology in this report for two reasons: It is the most widely used in the academic literature, and the scores are constructed in such a way as to permit valid comparisons across time.

---

[1] Joshua Clinton, Simon Jackman, and Douglas Rivers, "The Statistical Analysis of Roll Call Data," *American Political Science Review*, May 2004; Adam Bonica, *Ideology and Interests in the Political Marketplace*, Social Science Research Network, February 2012.

[2] For a more detailed discussion of the family of DW-NOMINATE measures, see Keith Poole and Howard Rosenthal, "D-NOMINATE After 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting," *Legislative Studies Quarterly*, February 2001.

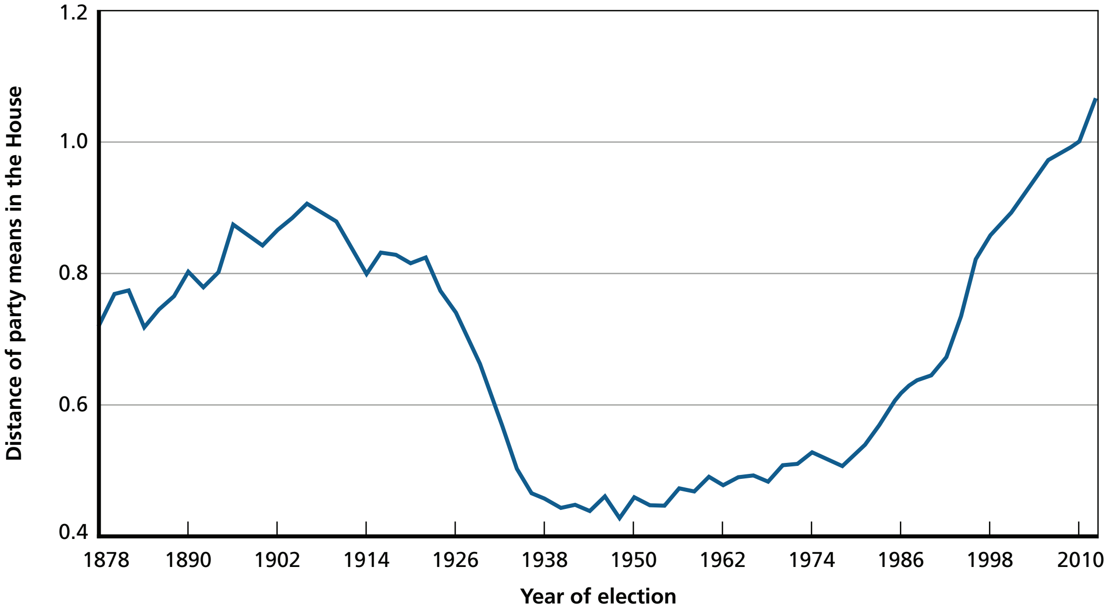
[3] Keith Poole, "The Roots of the Polarization of Modern U.S. Politics," *Revista Ciencia Politica*, October 2008.

MPR have used the DW-NOMINATE measure and its variants to document the extent to which the Republican and Democratic congressional delegations have become polarized. In their 2006 book, *Polarized America*, they established that the trend toward increase polarization in the House of Representatives began around the mid-1970s.[5] In updates to their estimates published on Poole's website, they demonstrate that the trends have continued up to today and that—at the end of the 113th Congress—polarization in the House is now at an all-time (post-Civil War) high.[6]

Figure 3.1 displays indices of political polarization for the House of Representatives for the period 1878–2010. There is strong evidence that polarization has risen dramatically since the 1970s.

As we have previously described, the fact that congressional polarization has been rising is clear both to the informed observer and the political scientist, but the causes of this trend are less so. One key factor, of course, was the "Southern Realignment," a shift that began when

Figure 3.1
House Polarization Has Been Rising Since the 1970s



NOTE: The DW-NOMINATE scores for this figure were taken on February 22, 2013, from Keith Poole, voteview.com website, undated.
RAND *RR896-3.1*

---

[4] For details, see Keith Poole and Howard Rosenthal, *Ideology and Congress*, New Brunswick, New Jersey: Transactional Press, 2007.

[5] Nolan McCarty, Keith T. Poole, and Howard Rosenthal, *Polarized America: The Dance of Ideology and Unequal Riches*, Cambridge, Massachusetts: MIT Press, 2006.

[6] Keith Poole, voteview.com, website, undated.

the Voting Rights Act (VRA) of 1965 disrupted the status quo that had permitted Southern Democrats (or "Dixiecrats") to maintain an iron grip on politics.[7] To the degree that it existed, much of the ideological overlap between Republican and Democratic members prior to the passage of the VRA was due to the Dixiecrats being relatively more conservative than their non-Southern Democratic brethren. The post-VRA era marked the beginning of a process in which the South gradually shifted from being a Democratic stronghold to a Republican one. In Congress this process was largely gradual as retiring Dixiecrats were replaced by Republicans; however, in a few instances sitting members changed parties or were defeated. As conservative Democrats switched parties or were replaced by Republicans, the average ideology of the Democratic distribution shifted left. But over the past 40 years, the Republican shift to the right has been considerably larger than the Democratic shift to the left—about three times larger.[8] More has been going on than the Southern Realignment.

In addition to their examination of the Southern Realignment, MPR explored several alternative explanations for rising polarization, including institutional changes within Congress, House redistricting ("gerrymandering"), and closed primaries. They did not find evidence that any of these had measurable effects on the trend. They subsequently conducted a more detailed study of the gerrymandering hypothesis, probably the most popular explanation for polarization among pundits. They concluded that gerrymandering was responsible for, at most, 15 percent of the increase in post-1970s polarization in the House.[9] In any case, polarization in the Senate—which has grown at a level comparable to that of the House—cannot be explained by gerrymandering.[10]

Figure 3.2 displays the trend in House polarization alongside the trend in district-level political dispersion: The dashed red line is our calculation of the standard deviation of the average of PREP across congressional districts; [11] the solid blue line is a subset of the same index of House polarization in Figure 3.1.

Clearly, the lines are correlated (correlation coefficient=0.94). We emphasize that this correlation is not sufficient to establish causality in either direction. However it is useful as a way of *framing the question of causality* originally posed by Bishop: Are a set of underlying social, economic, demographic, or way-of-life trends combining with geographically based elections to promote polarization in Congress? In the next section, we describe three different analytical techniques for answering this question, and we present evidence on the possible relationship between voter clustering and rising polarization in the House of Representatives.

---

[7] Richard Pildes, *Why the Center Does Not Hold: The Causes of Hyperpolarized Democracy in America*, New York University Public Law and Legal Theory Working Papers, Paper 207, 2010.

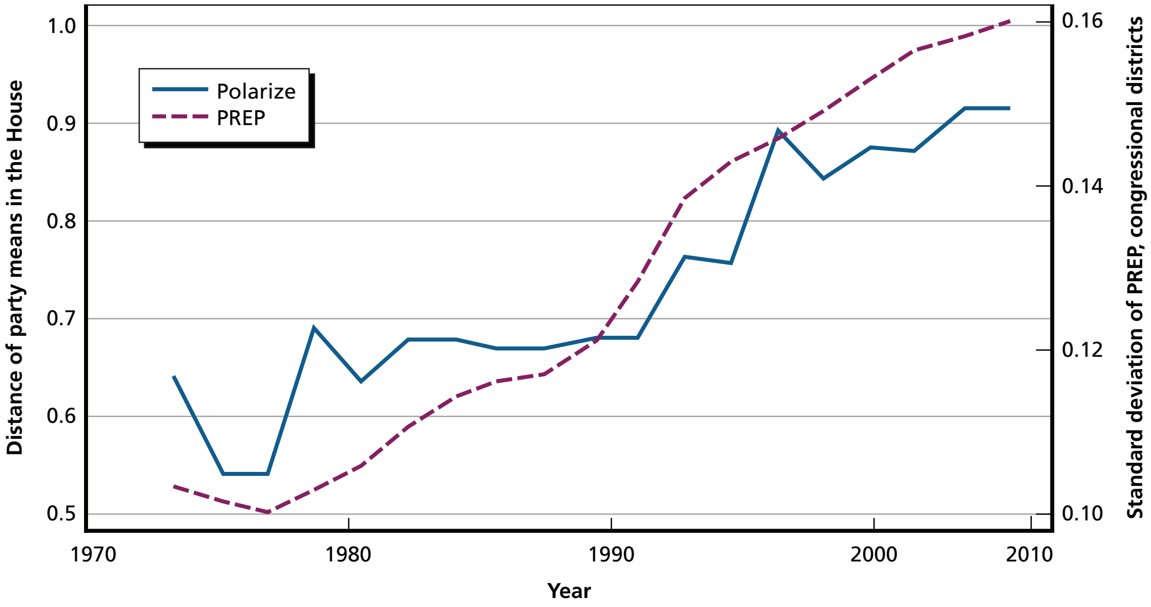
[8] See "Relative Growth in Extremism, Democratic and Republican House Caucuses" in the Appendix for details.

[9] MPR, "Does Gerrymandering Cause Polarization?" *American Journal of Political Science*, Vol. 53, No. 2, July 2009; Thomas E. Mann, "Polarizing the House of Representatives: How Much Does Gerrymandering Matter?" in Pietro S. Nivola and David W. Brady, eds., *Red and Blue Nation? Characteristics and Causes of America's Polarized Politics*, Vol. 1, Washington, D.C.: Brookings Institution Press, 2006.

[10] Our analysis focuses on the House of Representatives both because of the larger number of observations per Congress (435 vs. 100) and because House districts, particularly in urban areas, conform more closely to the idea of local place. The question of a link between shifting constituencies and rising polarization in the Senate is a worthy one, but beyond the scope of this paper.

[11] We did not weight the PREP values by population since congressional districts contain roughly equivalent populations by design.

Figure 3.2
Geographic Dispersion in Presidential Voting and Polarization in Roll-Call Votes in the House of Representatives



RAND *RR896-3.2*

This is a difficult question—no perfect method exists for answering it. Our approach is to use three different models and compare each set of results. Each model suffers from significant limitations, but each also has a relatively distinct methodological approach. If the three models yield results that are consistent with one another—that is, if our results are robust to the choice of model—then our confidence in any findings of a link between clustering in the electorate and polarization in Congress will be bolstered.

In the sections that follow, we provide brief methodological details for each of these three approaches and emphasize the empirical findings; in the Appendix we provide complete specifications for the design and implementation of each model.

Method 1: The Regression Discontinuity Model

One way to estimate the effect of changes in constituency attributes on polarization in Congress is through the use of simple regression analysis, and in fact this approach has been taken by two previous studies.[12] The basic idea is to regress the first-dimension DW-NOMINATE score (the measure of legislator ideology) on an indicator for political party plus a vector of district-level covariates—things like race, education, income, or the within-district presidential vote share. If the model is properly specified, an estimate of the upper bound of the contribution of constituency variables can be calculated as the difference between total polarization (the difference in party means of the DW-NOMINATE score) and the coefficient on the political party indicator, which can be thought of as measuring the effect of political party on legislator ideology after district-level attributes have been controlled for.

---

[12] MPR, 2006; Abramowitz, 2010.

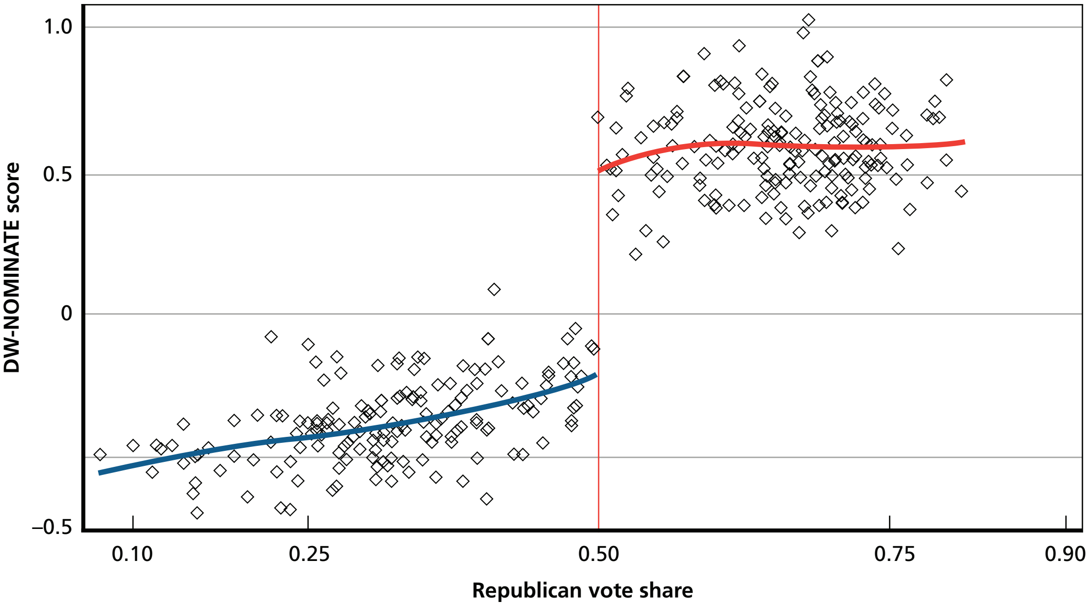
The trouble with this approach is that it is generally not possible to know if the model is properly specified. In cross-sectional regression models such as these, there is simply no way to guarantee this because it is not possible to control for all of the other relevant factors that might influence legislator ideology; this problem is known as "omitted variable bias" (OVB).

Regression discontinuity presents a potential solution to this problem. Regression discontinuity designs work when the values of the variable whose effect is to be measured are determined by a cutoff point in some secondary assignment variable. In this case, we are interested in the effect of political party identification on legislator ideology, and the political party of legislators is completely determined by their share of the vote in the district election. Districts that vote 50 percent +1 (or greater) for the Democratic candidate are assigned Democrats; districts that vote the other way are assigned Republicans.

Figure 3.3 allows us to visualize this assignment process, and how it allows for an unbiased estimate of the effect of political party on legislator ideology.

The regression discontinuity approach allows us to generate a quasi-random estimate of the effect of political party on legislator ideology—conceptually, this estimate is the vertical distance between the smoothed lines on either side of the discontinuity. With this in hand, we can create and compare estimates across time of the upper bound of the effect of voter clustering on House polarization by employing the same algebra used by both MPR and Abramowitz.[13]

Figure 3.3
DW-NOMINATE Scores and District Election Republican Vote Share, 108th Congress (2003)



RAND *RR896-3.3*

---

[13] See "Regression Discontinuity Approach" in the Appendix for a more detailed explanation of the regression discontinuity approach.

Method 2: The Rescaling Model

Our second strategy for estimating the contribution of clustering to polarization in Congress centers on the question of what the geographic distribution of socioeconomic variables might have looked like *if clustering had not occurred*. The first step in this model is to generate "rescaled" or "simulated" distributions of the constituency variables that we described earlier: race, educational attainment, income, etc. The goal of this procedure is to create new distributions of covariate values that allow the major demographic changes of the past several decades—growing Hispanic populations, rising average education as pictured in Figure 2.3 in Chapter Two, falling marriage prevalence as pictured in Figure 2.4 in Chapter Two, and so on—to occur, while fixing the *relative dispersion* of observations at the 1970 level.[14]

The second step is to use regression analysis to model legislator ideology as a function of district-level covariates, fitting one model for each of the 93rd, 98th, 103rd, 108th, 111th, and 112th Congresses (those immediately following reapportionment after the last four censuses plus the two most recent Congresses at the time of our analysis.) The coefficients from these models theoretically capture the relationships between legislator ideology and district-level attributes.

The final step is to estimate what polarization would have been in the absence of clustering. This is accomplished by combining the rescaled distributions of constituency attributes from the first step with the regression coefficients from the second step, and then generating predicted values for legislator ideology for the scenario in which clustering does not occur.

By comparing these predicted values with predicted values generated from the original regression models, we can estimate the overall contribution of demographic clustering to polarization in the House of Representatives.

Method 3: The Multistage Model

Our third approach for assessing the relationship between voter clustering and polarization in Congress begins by acknowledging that DW-NOMINATE, PREP, and the political party of House legislators are all correlated, both with each other and with within-district constituency attributes.

We initially regress PREP and an indicator for whether the legislator was a Republican ("R") separately on a vector of constituency variables, using ordinary least squares (OLS) regression for PREP and logistic regression for R. This is similar to the approach taken by MPR;[15] however, we were able to expand the set of variables used in their original analysis: In addition to the race, education, and income variables, as well as an indicator for district presence in the South, we included (from the census) variables associated with marital status (percentage married, divorced, etc.), the percentage of each district that is urban, and the population density of each district. We fit one such model for each of the 93rd, 98th, 103rd, 108th, 111th, and 112th Congresses.

---

[14] More details on how this is accomplished are available in "Processes for Step One of the Rescaling Model" in the Appendix.

[15] MPR, 2006.

We next fit a series of regression models—one for each of the aforementioned Congresses— with DW-NOMINATE as the dependent variable and PREP and R as the explanatory variables. Both coefficients were positive and highly significant in all years: Unsurprisingly, the within-district Republican vote share and membership in the Republican caucus are highly predictive of conservative ideology. We then used a model selection procedure (outlined in "Multi-Stage Approach" in the Appendix) to winnow the list of included district-level attributes.

The question that remains is how to isolate the effect of the constituency attributes on legislator ideology from the effect of PREP and R (to examine how much of the growth in House polarization has been driven by changes over time in constituency attributes).

We attempt to do this through the use of a multistage regression model, in which the predicted values from first-stage regressions (of PREP and R on constituency attributes) were substituted into a second-stage regression (of DW-NOMINATE on PREP and R). The effect of this procedure is to generate estimates of the share of House polarization attributable to constituency variables, with the confounding effect of PREP and R controlled for.[16] By comparing the growth in this index over time with the growth in the index of actual polarization in the House, we generate our third estimate of the contribution of constituency changes to House polarization over the last four decades.

Findings

The results of our three models are consistent with one another (see Table 3.1). The models estimate that between 23.4 percent and 31.1 percent of the growth in House polarization between 1972 and 2010 can be attributed to changes in the distributions of constituencies over time.

It is natural to ask *which* constituency attributes matter most. In other words, of the dimensions included in our models, which appear to contribute the most to rising polarization in the House? The regression discontinuity model cannot answer this question—it does not distinguish between individual attributes, but rather estimates a theoretical upper bound

Table 3.1
Summary of Model Estimates: Contribution of Voter Clustering to Rising Polarization in the House of Representatives, 1972–2010

Method	Growth in House Polarization Attributable to Voter Clustering	Percentage of Growth in House Polarization Attributable to Voter Clustering	Percentage of Total Polarization in the 112th Congress Attributable to Constituency Attributes
Regression discontinuity (RD)[a]	0.151	31.1	18.0
Rescaling of district covariates	0.123	23.4	NA
Multistage	0.156	29.5	45.7

[a] RD estimates are for the change in polarization between the pooled 93rd/94th Congresses and the pooled 111th/112th Congresses.

---

[16] See "Multi-Stage Approach" in the Appendix for more details on the multistage model.

for the joint contribution of *all* district-level characteristics. However, the other two models do allow a means for exploring it.

We fit a version of the multistage model in which only a single covariate (the marriage prevalence rate) was used on the right-hand side of both first-stage equations. The increase in polarization predicted from this restricted model is similar in magnitude to the increase predicted from the full model. This suggests that within the component of House polarization growth estimated to be attributable to voter clustering (roughly 30 percent of total polarization), a substantial portion may be related to clustering along marriage lines. Estimating the specific contribution of marriage clustering to rising polarization in the House is not possible in the context of the multistage model, but by revisiting the second model (rescaling) we can make a somewhat crude attempt at isolating the individual effect of marriage clustering.

The rescaling model functions by first modeling legislator ideology as a function of constituency attributes, and then by predicting new ideology scores derived from hypothetical distributions of covariates in which clustering (operationalized as an increase in the dispersion of each distribution) is reversed. Polarization in these predicted ideology scores is compared to polarization in the predicted values derived from the actual covariate distributions. In this framework, we can estimate the incremental effect of marriage clustering on congressional polarization by generating a third set of predicted ideology scores in which the *actual* values of marriage prevalence are used alongside the *rescaled* values of all other covariates. This is analogous to allowing clustering to occur with respect to marriage but not to other factors, and allows us to estimate the contribution of the marriage dimension as the difference between the two indices. Of the increase in polarization attributable to changes in constituencies between the 93rd and 112th Congresses (1972–2012), we estimate that about 85 percent is associated with increased clustering along marriage lines.[17]

The increasingly strong relationship between marriage prevalence and politics can also be demonstrated in a more intuitive way. Table 3.2 shows the proportion of districts within the top and bottom quartiles of marriage prevalence represented by Republican and Democratic members, as well as the average DW-NOMINATE score within those districts.

Table 3.2
Republican and Democratic Membership of the House of Representatives, by Year of Election and Quartile of Within-District Marriage Prevalence

	1972	1982	1992	2002	2008	2010
Percentage of seats and average member ideology for districts in the top quartile of marriage prevalence						
Democrats	41.3%	45.9%	33.0%	15.6%	33.0%	15.6%
Republicans	58.7%	54.1%	67.0%	84.4%	67.0%	84.4%
Average ideology	0.076	0.109	0.223	0.438	0.349	0.551
Percentage of seats and average member ideology for districts in the bottom quartile of marriage prevalence						
Democrats	85.3%	80.7%	89.9%	91.7%	94.5%	91.7%
Republicans	14.7%	19.3%	10.1%	8.3%	5.5%	8.3%
Average ideology	–0.262	–0.283	–0.367	–0.359	–0.389	–0.346

---

[17] See “Marriage Prevalence and Polarization Estimates” in the Appendix for details on how these estimates were derived. In particular, the estimate of the contribution of marriage clustering to House polarization is subject to significant limitations.

The results are striking. In every time period, Republicans held the majority of districts in the top quartile of marriage prevalence, while Democrats held the majority of districts in the bottom quartile. More importantly, the extent of these advantages grew dramatically over time—particularly the Republican advantage in the top quartile districts. In the 93rd Congress (1973), Republicans held just under 60 percent of the seats in the top quartile; by 2011, that advantage had grown to almost 85 percent. Similarly, Democrats held 85 percent of the seats in the bottom quartile of marriage prevalence in 1973; in 2011 they held all but nine of the 109 seats in that quartile (92 percent).

Intensity or Clustering?

We have observed that within-district marriage prevalence is an increasingly important positive predictor of Republican legislators.[18] Two factors could theoretically explain this pattern:

- *Intensity*: Individuals who are married might increasingly (or more intensely) favor Republicans, while unmarried people could increasingly favor Democrats (or both could be occurring in combination). Thus, districts that contain a higher percentage of married people would increasingly trend Republican and vice versa.
- *Clustering*, or sorting. Without a change in intensity, married people might increasingly cluster in some districts and unmarried people in others. Thus, districts with an increasing level of married people would trend Republican.[19]

Of course, both intensity and clustering could be happening in combination.

We know from our analysis in the first section that clustering is definitely occurring—at the congressional-district level of geography, we observe higher dispersion over time in the distributions of marriage prevalence, income, and education. Because we observe very similar trends at the county level, we feel confident that gerrymandering is not the cause of this clustering. Some of the growth in the correlation between R and PREP and the constituency variables is due to clustering, but how big a part also depends on the change in intensity.[20]

We can examine trends in intensity at the national level using the American National Election Survey (ANES), a nationally representative, individual-level sample of American political beliefs and behavior. Using this study, we can examine the effect of such attributes as race, education, and marital status on party voting in presidential elections *at the individual level, rather than the congressional-district level*. It is reasonable to assume that national trends are sufficient to indicate any meaningful changes in intensity as we have defined it. As shown in Table A.4 of the Appendix, the effects of socioeconomic factors on individual party voting do not exhibit any significant trends over time, except for a growth in Hispanic voters' preferences for Democratic candidates in recent elections and a growth in black voter preferences for the Democratic candidate (Barack Obama) in 2008. On average, married voters appear to slightly favor the Republican candidate (and unmarried voters the Democratic candidate), but there is no statistically significant evidence that the intensity of this relationship is increasing over time.[21]

---

[18] See Figure A.7 in the Appendix.

[19] This would be related to the cultural clustering hypothesized by Bishop. An example at the individual level: Susan and Anne are sisters who share constituency characteristics that trend Republican (they are married, white, and finished high school but not college). They live in different districts and thus contribute equally to the two constituencies' characteristics. Intensity is determined by the fact they both increasingly support Republicans over time. Clustering occurs if Anne moves to Susan's district.

[20] See Tables A.2 and A.3 in the Appendix.

Thus, we conclude that the growth in correlation between constituency variables (particularly marriage) and district partisanship appears to be largely a result of clustering, as opposed to intensity. Since marital status dominates the analyses, it appears that Bishop's hypothesis of cultural clustering contributing to congressional polarization is correct.

Clearly, there is a growing correlation between the clustering of married people into some districts and unmarried people into others, on the one hand, and growing polarization in the House, on the other hand. Moreover, as we show in the Appendix, the inclusion of the district-level marriage prevalence into Models 2 and 3 swamps the predictive ability of other district-level variables that we would expect to be correlated with polarization in the House. (These include median family income, educational attainment, and race.) In addition, as we show in Figure A.3 in the Appendix, marriage prevalence by district is also strongly correlated to the percentage of the vote received by Republican candidates in the 2010 House election. Only race is similarly (although negatively) correlated.

What Is the Political Significance of District-Level Marriage Prevalence?

It is possible that marriage prevalence is standing in for some other variable that truly explains the geographic variance in House roll-call voting patterns. We examined district-level average age, urbanicity, and home ownership rates using Method 3. None of these reduced the dominance of marriage prevalence as a correlate with polarization in the House. We also obtained a data set of the number of various types of businesses organized by congressional districts in 2010. As we will discuss in a future paper, we chose business types that we thought would have local cultural significance, such as the number of gun shops or Whole Foods markets. Using these data with Method 3, we found that some variables, such as the number of religious bookstores, improved the resulting model. But these variables did not reduce the significance of marriage prevalence much. So, while marriage prevalence may be standing in for some other variable, we have been unable to empirically identify it.

Our conjecture is that marriage prevalence, more than any other available variable, serves as a proxy for values-oriented local cultural differences as described by Cahn and Carbone.[22] They suggested that Red states "emphasize the special status of marriage as the institution that unites sex, procreation, and childrearing." By contrast, Blue states, in their view, are moving away from this emphasis. This same argument could be extended to congressional districts.

We do not suggest that Americans (as they choose whether and where to migrate) or members of Congress (as they vote) are consciously considering local marriage prevalence as part of their decisionmaking process. But we do suggest that they consider, both consciously and unconsciously, local cultural attitudes toward traditional family values—and thus, that marriage prevalence is the best available measure of this construct at the local geographic level.

A Delayed Effect

Observant readers may notice that the most significant increases in geographic dispersion of marriage prevalence and average income occur between 1970 and 1990 (Figure 2.5 in Chapter Two), and may wonder how this reconciles with the fact that the most significant increases in congressional polarization have occurred in the subsequent decades, i.e., 1990 to the present (Figure 3.1). Although we do not model this explicitly, we suspect that there may be a lag effect at work. One possibility is that legislators do not instantly perceive compositional changes within their districts, and thus may take time to respond to those changes. A second possibility is related to incumbency: The most obvious pathway through which compositional change in the constituency base leads to altered voting patterns in Congress is replacement of old legislators with new ones, perhaps belonging to the opposite party. However if that compositional change is insufficient to overcome the electoral advantage enjoyed by sitting legislators (the incumbency advantage), the effect on polarization in Congress could be delayed, either until the changes become large enough to unseat individual incumbents or until they decide to retire. For example, this is arguably how Congressmen Dan Glickman and John Spratt lost their seats in Kansas (1996) and South Carolina (2010), respectively.

---

[21] Details are available in "Influence of Socioeconomic Factors on Individual Party Voting over Time" in the Appendix.

[22] Naomi Cahn and June Carbone, *Red Families v. Blue Families: Legal Polarization and the Creation of Culture*, New York: Oxford University Press, 2010.

CHAPTER FOUR

Discussion and Conclusion

It is well known that place and politics are correlated in the United States. Large cities in general and coastal metropolises in particular are generally liberal bastions, for example, while the South as a whole is more conservative than the rest of the country. Because this is true of our electorate, and because voters elect representatives who share their ideologies, it is also true of our legislators. Members from liberal places (like California Democrat Nancy Pelosi) are themselves liberal, and members from conservative places (like Idaho Republican Raul Labrador) are themselves conservative. In this report, we attempt to address two issues: first, whether clustering of the population, and therefore of the electorate, has occurred over the last 40 years; and second, whether this clustering has been a contributor to rising polarization in the U.S. House of Representatives. We find evidence to support both claims. With respect to the first question, we observe that the distributions of congressional-district measures of income, education, marriage, and voting have become markedly more dispersed since 1970. With respect to the second question, each of three different technical approaches generated results that are consistent with the hypothesis that clustering and congressional polarization are linked.

Regarding the related question of the overall share of polarization that is attributable to constituency differences in the most recent House (the 112th), results are mixed. The regression discontinuity estimate of this factor is 18.0 percent, less than half of the estimate generated by the multistage model (45.7 percent). The rescaling model, of course, is designed only to estimate changes since the baseline period (the 93rd Congress) and cannot address this question.

Each of the models in the second section suffers from important limitations, and these limitations preclude a claim of demonstrated causality. The RD estimate is an estimate of an "upper bound" of the contribution of voter clustering—in the pooled model from which it is generated, it represents the total remaining polarization after the RD estimate of the effect of legislator political party has been accounted for. If there are other, unmeasured factors besides political party and constituency preferences that independently influence legislator ideology, then the RD estimate will be biased upward. Additionally, the RD estimate of the effect of political party on legislator ideology is a Local Average Treatment Effect (LATE) estimate—it measures the effect on ideology of a quasi-random assignment of a Republican legislator to districts *in the vicinity of the discontinuity*—that is, swing districts. The RD model makes the strong assumption that the Average Treatment Effect (ATE) is equal to the LATE estimate: In other words, the RD model assumes that, on average, the true effect on legislator ideology of a party switch in ALL districts (whether they are liberal, conservative, or moderate) will be equal to the estimated effect of a party switch in swing districts; i.e., districts in the neighborhood of the discontinuity.

The rescaling model predicts congressional polarization from a hypothetical set of district-level covariates. These new values were our best guess as to what covariate values would have been in the absence of clustering, and, as previously discussed, they have a number of properties that suggest they are reasonable. Nevertheless, they are still guesses, and to the (unknowable) extent that they are biased, the model results will be biased as well. An additional source of bias for the rescaling model is the possibility of omitted variables: Fundamentally, the rescaling model is derived from a series of OLS regressions and if these models omit variables that are correlated both with the outcome (legislator ideology) and with the existing regressors, then the model results will be biased. As a result, the best we can do is say our results are consistent with the hypothesis of causality, and our estimates of the effect are valid only under the assumptions discussed above. The multistage model is also derived from a series of cross-sectional OLS and logistic regressions; as such, it is also vulnerable to the threat of omitted variable bias.

Despite these limitations, the models discussed in this paper, when viewed jointly, provide evidence that a nontrivial component of rising polarization in the House of Representatives is due to a widespread, gradual clustering of the electorate; we estimate that proportion to be roughly 30 percent. Our confidence in the robustness of this estimate is increased because it is derived not from a single analysis but from three different models, each of which uses a fairly distinct approach.

We note that the measures used in our analysis—income, race, education, and marriage prevalence—are at best crude proxies for the actual phenomenon that Bishop described—the voluntary clustering of Americans into regions based on shared cultural and lifestyle factors.

While these findings are significant, our analyses leave unexplained the majority (70 percent) of the rise in polarization in the House of Representatives over the past 40 years. The regression discontinuity model describes this component as the "pure" effect of political party, and we admit this is not much of an explanation at all. Our understanding of the factors behind this unexplained majority remains incomplete: reduced turnout in primary elections, the increasingly dominant role of money in politics, the relatively new 24-hour news cycle, and the rise of the Tea Party movement have all been suggested as culprits. Analysis elsewhere has suggested that polarization in the House has increased in part because the last several decades have seen dramatic changes regarding *which* constituencies the national political parties pursue during campaigns and are therefore beholden to.[1]

Still, that the polarization of the U.S. House is so dependent on underlying geographical demographic shifts suggests that solutions to the problem—if, in fact, one thinks it is a problem—will be hard to develop in the context of the American approach to elections; i.e., one House member elected per district with elections decided by the first-past-the post method. As others have shown and we reinforce here, fixing gerrymandering, as desirable as it might otherwise be, will not make much of a dent in House polarization.[2] Alternative approaches to elections, such as the Louisiana and California open primaries, appear to have promise, but it will take time for this to become clear. The key point is that policymakers looking for "solutions" to polarization may need to turn their attention to election laws.

---

[1] Jeffrey M. Stonecash, *Party Pursuits and the Presidential-House Election Connection, 1900-2008,* New York: Cambridge University Press, 2013.

[2] MPR, 2006.

APPENDIX

Notes and Technical Methods

Legislator Ideology as a Function of Constituency Attributes (Chapter One, Footnote 5)

Let D_i denote some measure of the ideology of the legislator representing district *i,* X_i denote a vector of demographic attributes of that district, and ε_i denote other factors that influence legislator ideology. Let I_i denote the ideology of the median voter of district *i.* Without specifying a functional form, we can consider both *D* and *I* as being function of *X*, as in

$$D_i = f(X_i, I_i[X_i]\ \varepsilon_i)$$

So, for example, if X_i is average income and it is true that there is a positive relationship on average between wealth and politically conservative views (via, for example, an increased preference for low levels of taxation and government spending), then we might expect that—everything else being equal—a district with a high level of income would tend to have a more conservative median voter. Under the reasonable assumption that legislators' voting behavior is partially responsive to constituency preferences, we would also expect that district to have a more conservative representative in Congress. By extension, we would expect (again, all else being equal) that the effect of an increase in the average income within a given district over time would be to shift both the median voter and the representative of that district in a more conservative direction.

Weighting County Data by Population (Chapter Two, Footnote 6)

A simple average of county-level data will accord an identical weight to Loving County, Texas (population 82), and Los Angeles County, California (population 9.8 million). For sociological phenomena, such as voting data, this approach is generally incorrect. To account for differences in population, we weighted by the total population in each county. Thus, while the unweighted average is calculated as

$$\overline{x_u} = \frac{\sum_1^N x_c}{N}$$

The weighted average is calculated as

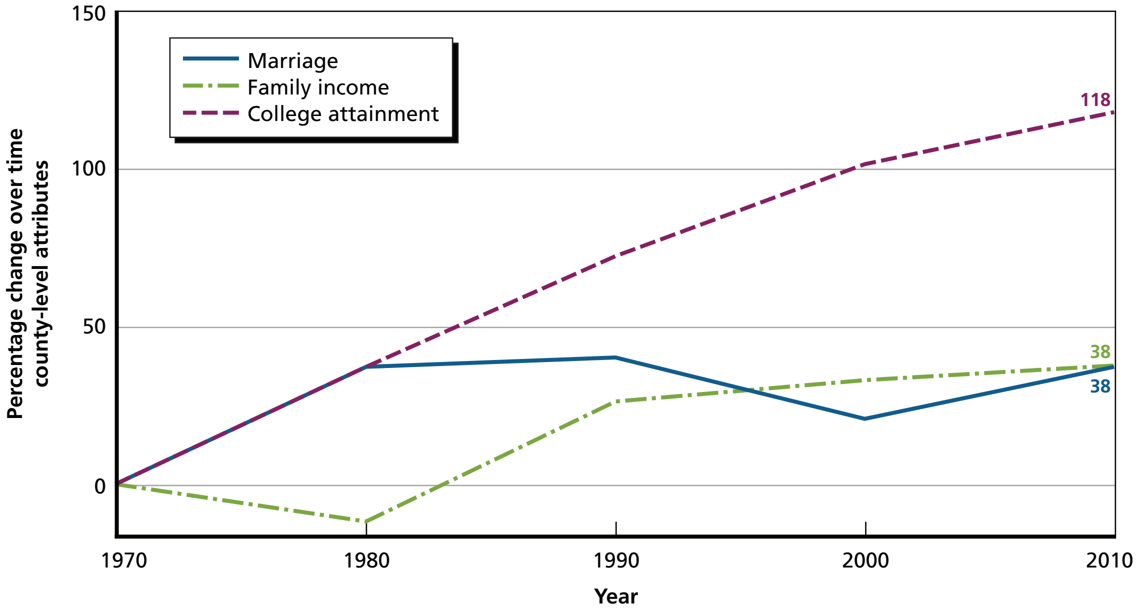
$$\overline{x_w} = \frac{\sum_1^N x_c P_c}{\sum_1^N P_c}$$

where x_c is the value for county c, P_c is the total population of that county, and N is the total number of counties.

Clustering over Time in Demographic Attributes, County-Level Data (Chapter Two, Footnote 17)

In the main body of the text, Figure 2.5 revealed rising geographic dispersion in marriage prevalence, educational attainment, and income at the congressional-district level. Figure A.1 below replicates Figure 2.5 at the county level. The results are remarkably similar. Each of the measurements indicates a significant increase in overall dispersion over time, with the largest increase occurring in college attainment. The consistency of these trends across two levels of geography is important because it refutes the claim that these changes are being driven by the boundary drawing process: Gerrymandering is not the culprit here.

Figure A.1
Percentage Change in Population-Weighted County-Level Standard Deviation of College Attainment, Marriage Prevalence, and Average Income, 1970–2010



RAND *RR896-A.1*

Relative Growth in Extremism, Democratic and Republican House Caucuses (Chapter Three, Footnote 8)

Table A.1 displays the mean DW-NOMINATE scores of the Democratic and Republican caucuses in the House for each Congress since the 93rd. Also displayed are the absolute and percentage changes for both parties over time, relative to the 93rd Congress. The rightward shift by Republicans is about three times greater in magnitude than the leftward shift by Democrats.

Regression Discontinuity Approach (Chapter Three, Footnote 13)

One of the primary ways in which OLS regression can produce biased estimates is if variables are endogenous—that is to say, correlated with other things not included in the statistical model. For example, we replicated the OLS models fit by MPR and generated identical results: Their model found that total polarization in the 108th Congress was 0.864, and that after controlling for district-level attributes, the coefficient on the party indicator variable was 0.799;

Table A.1
U.S. House of Representatives' Mean Democratic and Republican First-Dimension DW-NOMINATE Scores, 1972–2010

Year Elected	Mean Democratic Score	Mean Republican Score	Absolute Change from 93rd (Democrats)	Absolute Change from 93rd (Republicans)	Percent Change from 93rd (Democrats)	Percent Change from 93rd (Republicans)
1972	–0.324	0.273	n/a	n/a	n/a	n/a
1974	–0.322	0.275	0.002	0.002	–0.6	0.8
1976	–0.316	0.273	0.009	0.000	–2.7	0.1
1978	–0.315	0.296	0.010	0.023	–3.0	8.2
1980	–0.312	0.313	0.012	0.040	–3.8	14.5
1982	–0.315	0.327	0.009	0.053	–2.9	19.5
1984	–0.323	0.334	0.001	0.061	–0.3	22.2
1986	–0.322	0.335	0.003	0.061	–0.8	22.4
1988	–0.326	0.337	–0.002	0.064	0.6	23.4
1990	–0.328	0.344	–0.004	0.071	1.3	25.9
1992	–0.346	0.368	–0.022	0.094	6.8	34.6
1994	–0.372	0.394	–0.048	0.121	14.8	44.1
1996	–0.389	0.404	–0.065	0.130	19.9	47.6
1998	–0.384	0.404	–0.059	0.131	18.3	47.8
2000	–0.386	0.412	–0.062	0.139	19.1	50.7
2002	–0.385	0.416	–0.060	0.143	18.6	52.2
2004	–0.398	0.424	–0.074	0.151	22.7	55.2
2006	–0.379	0.440	–0.055	0.167	16.9	61.1
2008	–0.364	0.457	–0.040	0.183	12.2	67.1
2010 (112th Cong.)	–0.402	0.482	–0.078	0.208	24.0	76.2

they thus estimated the upper bound of the contribution of constituency variables to congressional polarization to be $1-\frac{0.799}{0.864}=7.5$ percent. Abramowitz conducted a similar analysis, using a normalized measure of PREP in place of MPR's race, education, and income variables; he estimated that 20 percent of polarization in the 108th House was attributable to differences in constituencies.[1]

When we add three other variables to MPR's model to tell us more about the constituency in question (PREP, the percentage of the adult population that is married, and the percentage that lives in urban areas), the estimate of the effect of political party on legislator ideology falls from 0.799 to 0.699; as a result, the estimate of the upper bound of the effect of constituency more than doubles, rising from 7.5 percent to 19.1 percent. It turns out that each of these things (Republican voting, marriage, and low percentages of urbanicity) are all correlated both with Republican representation and with conservative legislator ideology; their exclusion from the original MPR models causes those estimates to be biased.

This is not to say that either Abramowitz's estimate or this new revised estimate (19.1 percent of polarization in Congress is due to constituency differences) do not suffer from the same problem. In the same way that MPR's estimate was biased because of the omission of our three variables, our estimates may also be biased because of the omission of variables we are unaware of, or cannot measure. This endogeneity/bias problem is probably the biggest single shortcoming of regression analysis on cross-sectional data, and the reason why randomized experiments are preferred by researchers. As is the case with much social science research, randomization is simply infeasible here: We cannot simply assign Republicans and Democrats to different congressional districts and then observe their voting behavior.

One technique that can address this problem is regression discontinuity (RD). In Figure 3.3, a discontinuity exists in the ideology data at the 50-percent mark in the local election results: This is the point at which districts switch from being represented by Republicans to Democrats, and vice versa. The RD estimate is equal to the vertical distance between the endpoints of the smoothed curves on either side of the discontinuity.

As it turns out, the RD estimate of the contribution of political party (0.747) lies in between the MPR estimate (0.799) and the estimate generated when the marriage, PREP, and urbanicity are added to the original model (0.699). In the case of the data from Figure 3.3, the RD estimate for the upper bound of the contribution of constituency attributes to congressional polarization in the 108th congress is equal to $1-\frac{0.747}{0.864}=13.5$ percent.

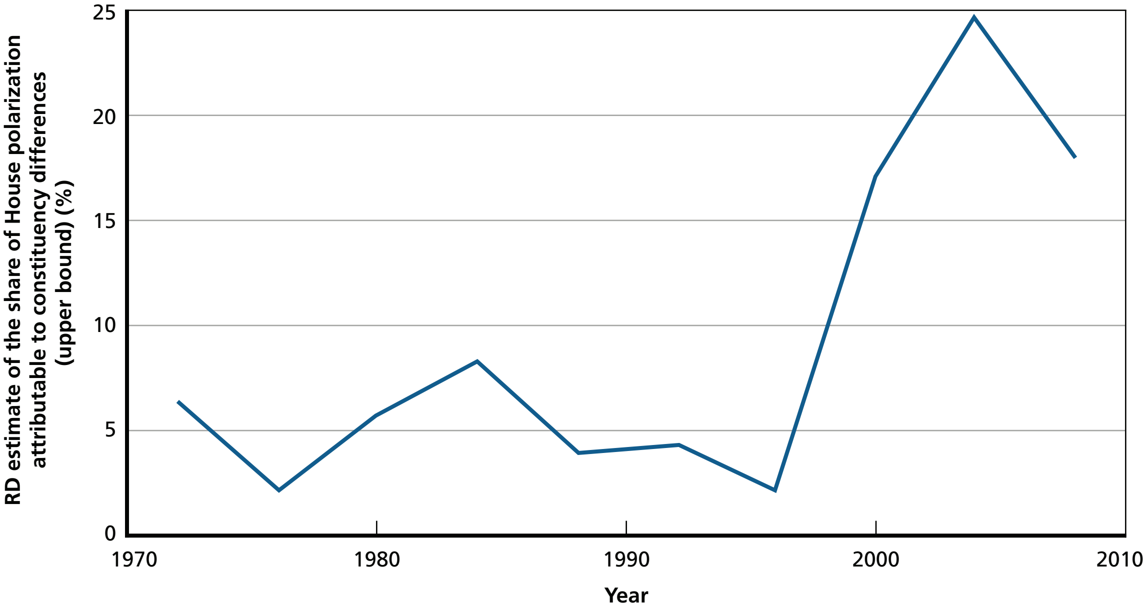
The critical question of course, is whether this contribution is stable over time or is increasing; the latter finding would be consistent with the hypothesis that clustering is a partial driver of polarization in Congress. We can address this question by computing the estimated contribution of constituency differences to congressional polarization not just for the 108th Congress, but for every Congress in the period under consideration. If Bishop's hypothesis is true, and legislators are responding to their increasingly homogenous constituencies by voting more ideologically, then we would expect a plot of this statistic over time to be increasing. This plot is presented in Figure A.2.

Because the RD estimates become more precise as more observations are added, we elected to pool data from individual Congresses into pairs (93rd/94th, etc.) with a single statistic calculated for each pair. Each observation in Figure A.2 is calculated from such a pair.

---

[1] Abramowitz, 2010.

Figure A.2
RD-Derived Estimate of the Upper Bound of the Percentage of House Polarization Attributable to Differences in Constituencies, 93rd–112th Congresses (Pooled)



RAND *RR896-A.2*

The values appear relatively stable over time until the 107th/108th Congresses (2001–2004), at which point an upward trend begins. The trend is not entirely consistent: the value peaks in the 109th/110th at 24.6 percent, then declines in the 111th/112th to 18.0 percent. In a general sense, the RD model suggests that differences in constituencies contribute more to polarization in the House today than they did in previous decades.

More precisely, the RD model estimates that 68.9 percent of the rise in polarization is directly attributable to political party, with the remainder (31.1 percent) being the theoretical upper bound for the contribution of voter clustering to rising polarization.

The underlying premise of the RD approach is that in the neighborhood of the discontinuity, the *only* way in which the districts on either side differ is with respect to the assignment variable (the within-district vote) and thus the outcome variable (legislator ideology), which depends on the assignment variable. This cannot be proven, but if we demonstrate that districts on either side of the discontinuity are not systematically different with respect to available covariates, then our confidence in the results of the RD approach is bolstered.

Following the practice of other scholars, we make the argument that districts in the neighborhood of the discontinuity are statistically indistinguishable from one another graphically.[2] This is accomplished by examining plots of different covariates (race, income, education) in the neighborhood of the discontinuity in a manner identical to the actual RD analysis for legislator ideology, the outcome variable of interest. If the assumption of indistinguishability holds on either side of the cut point, we should not expect to see a discontinuous jump at the

---

[2] David Lee, Enrico Moretti, and Matthew Butler, "Do Voters Affect or Elect Policies," *Quarterly Journal of Economics*, August 2004.

50-percent vote threshold for any observable covariate. Any such jump would cast into question the validity of the RD approach in this context.

Figure A.3 displays these scatterplots for six covariates: the percentage of the population that is black; the percentage of the adult population with a college degree; the percentage with "some college" education; median income within the district; adult marriage prevalence; and PREP, the Republican vote share in the presidential election. The most significant finding in this figure is that there is no discontinuity at the 50-percent vote threshold for any of the measured covariates—this is consistent with the underlying assumption of RD. In particular, if we look at the fifth frame, there is a strong positive relationship between marriage prevalence within the district and the vote share of the Republican candidate. However, in the neighborhood of the discontinuity there is virtually no difference between districts that were close victories for Democrats and districts that were close victories for Republicans. This finding holds across all measured covariates; these results are for 2010, but similar results were found for all prior years.

Processes for Step One of the Rescaling Model (Chapter Three, Footnote 14)

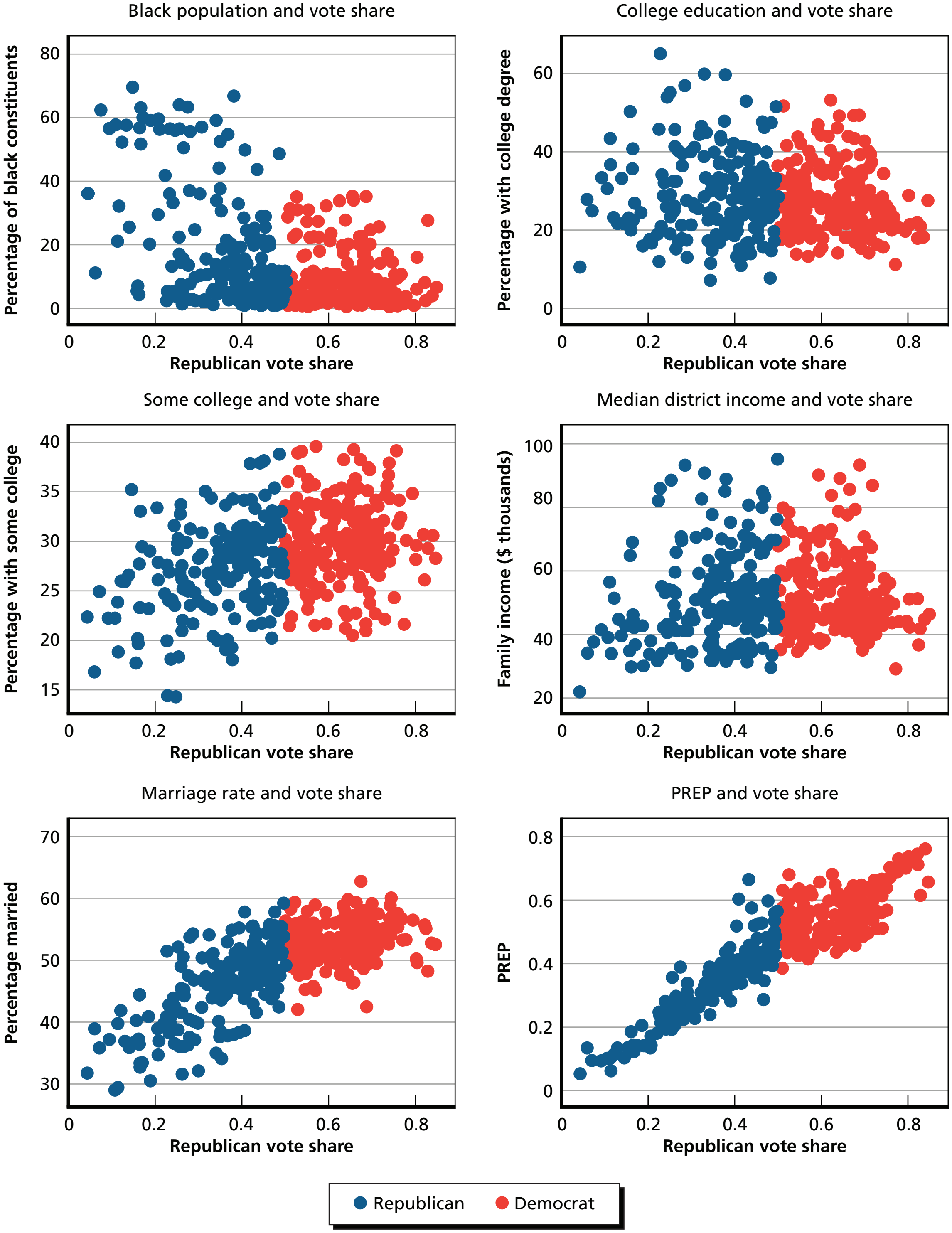
The "rescaling" procedure is best explained using an example for a single variable, and we elect to use the proportion of adults with a college degree. Figure A.4 shows the distribution of this statistic at the congressional district level, derived from the 1970 and 2010 censuses. Two obvious changes are apparent from the data. First, between 1970 and 2010, the mean of the distribution shifts upward by about 16 percentage points, from 10.6 percent to 27.7 percent, consistent with the well-known phenomenon of rising rates of educational attainment in the United States. Second, the relative shape of the distribution changes—it becomes much more dispersed. This is indicative of a greater number of congressional districts with low or high rates of college education, relative to population averages. We can quantify this increased dispersion by observing that the standard deviation more than doubles between 1970 and 2010; this change is consistent with the hypothesis that clustering is occurring.

We implement a rescaling procedure that seeks to isolate the first change (mean shifting) from the second (clustering). This is done by creating a distribution of covariates in which all of the major demographic trends of the past 40 years (e.g., falling marriage prevalence, rising Hispanic population) are retained, but with the added constraint that the overall level of dispersion in the data be identical to the level of dispersion present in the 1970 distribution. We created one such rescaled data set for each census period between 1980 and 2010. The result of this exercise is visualized in Figure A.5, in which the actual and rescaled 2010 distributions for college education rates are displayed.

The actual values of this statistic—the congressional district–level college education rate in 2010—are pictured in a solid density; the rescaled values are pictured in a hollow density with red borders. For reference, the actual distribution from 1970 is displayed in a smaller frame to the right.

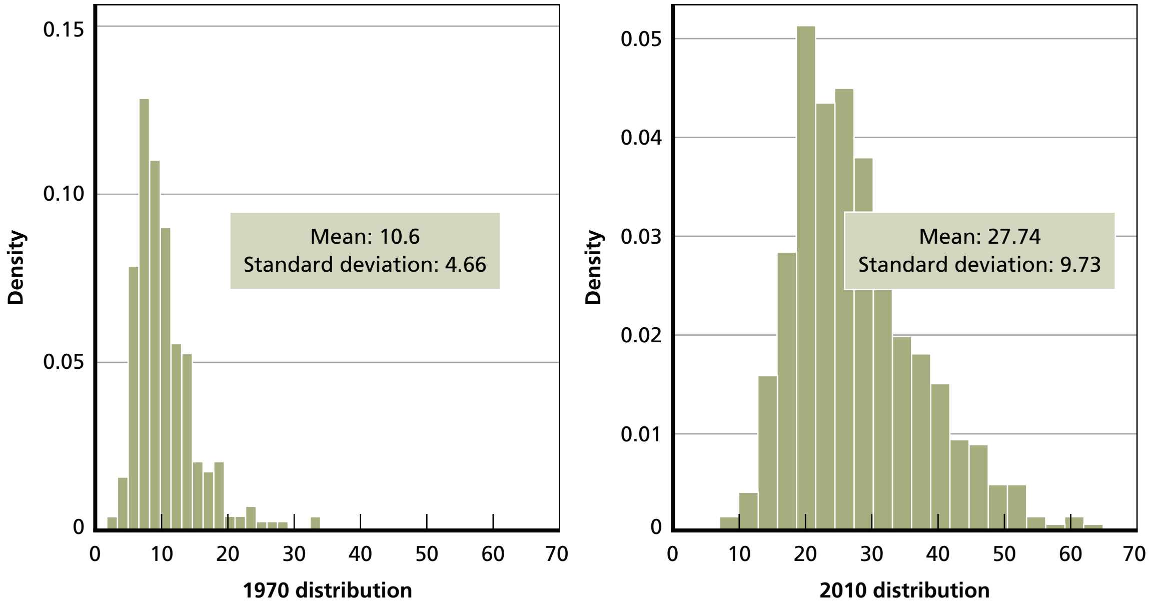
The rescaled distribution has a number of important properties. First, its mean is identical to the mean of the actual 2010 distribution. Second, its shape is identical to the actual distribution from 1970. Third, it preserves rank order: In both the actual and rescaled 2010 distribu-

Figure A.3
Absence of Discontinuity at 0.5 Vote Threshold for District-Level Covariates, 2010 Midterm Elections



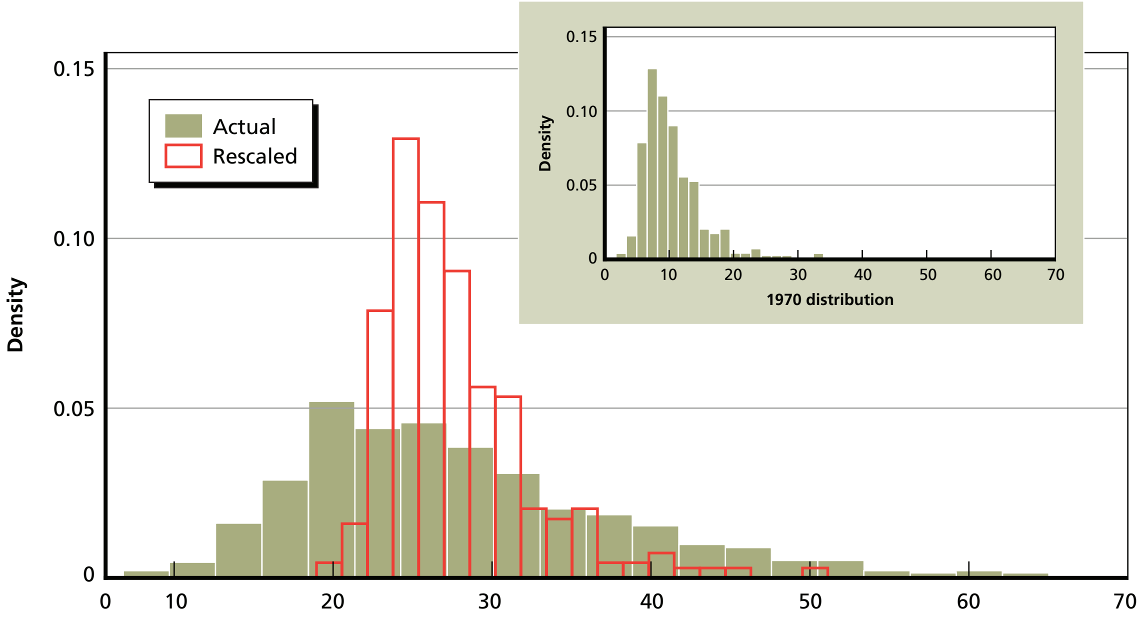
RAND *RR896-A.3*

Figure A.4
Distribution of College Education Rates by U.S. Congressional District, 1970 and 2010



RAND *RR896-A.4*

Figure A.5
Actual and Rescaled Distributions for 2010 College Education Rates, U.S. Congressional Districts



RAND *RR896-A.5*

tions, the congressional district with the highest college education rate is NY-14 (Manhattan) and the lowest is TX-29 (east of Houston).

Formally, the rescaling procedure is accomplished as follows:

Let X_{ijtk} denote the observation of variable *i* in congressional district *j*, at time period *t*, with rank *k*. Let D_{ijtk} denote the deviation of X_{ijtk} from its mean, i.e., $X_{ijtk} - \overline{X}_{it}$. For example, in 1972, the congressional district with the lowest marriage prevalence was NY-19 (Harlem), with a marriage prevalence of 41.9 percent. In our notation, this is $X_{married,NY(19),1972,435}$ and $D_{married,NY(19),1972,435}$ (the deviation) is equal to (41.9%–61.2%=–19.3%): this observation is 19.3 units below the mean of the data.

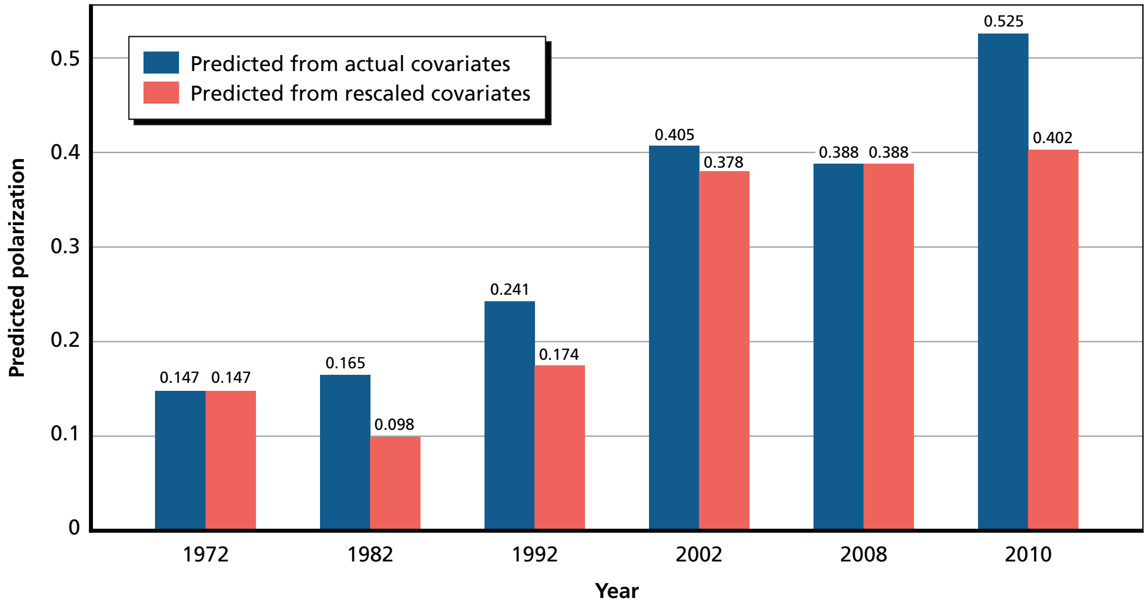
Under this setup, the simulated distributions of covariates are defined by the following condition:

$$X_{ijtk}^{sim} = \bar{X}_{it} + \left(D_{ijk,1972}\right) \forall k = k_{1972}, t \neq 1972$$

In other words, the simulated value of each covariate and each year is obtained by adding the mean of the data in that year to the number of standard deviation units away from the mean of the current-year distribution that the 1970 observation of identical rank was away from the 1970 mean.

Figure A.6 demonstrates the predicted polarization from the original and rescaled models. The model regresses the first dimension DW-NOMINATE score on median family income, the percentage black and Hispanic, the percentage of adults with “some college” and college degree attainment, the percentage of the population living in urban areas, and marriage prevalence. Right-hand side variables are demeaned (within periods, the mean of each variable is

Figure A.6
Difference in Inter-Party Means of DW-NOMINATE Scores as Predicted by Actual and Rescaled District-Level Covariate Values, 1972–2010



RAND *RR896-A.6*

subtracted from each observation of that variable), and a squared term is added in the marriage prevalence variable. The first series (blue, on the left) shows polarization in the original model's predicted values; the second series (red, on the right) shows polarization in the predicted values from the rescaled model.

By the end of the 112th Congress (2011–2012), the level of polarization in the DW-NOMINATE scores predicted from a regression on actual district level covariates was 0.525. The level of polarization predicted following the rescaling procedure was 0.402. Because the rescaling procedure essentially estimates what district-level attributes would have been in the absence of voter clustering, the rescaling model estimates that (0.525–0.402)/(0.525)=23.4 percent of the growth in House polarization since the 93rd Congress is attributable to voter clustering.

Multi-Stage Approach (Chapter Three, Footnote 16)

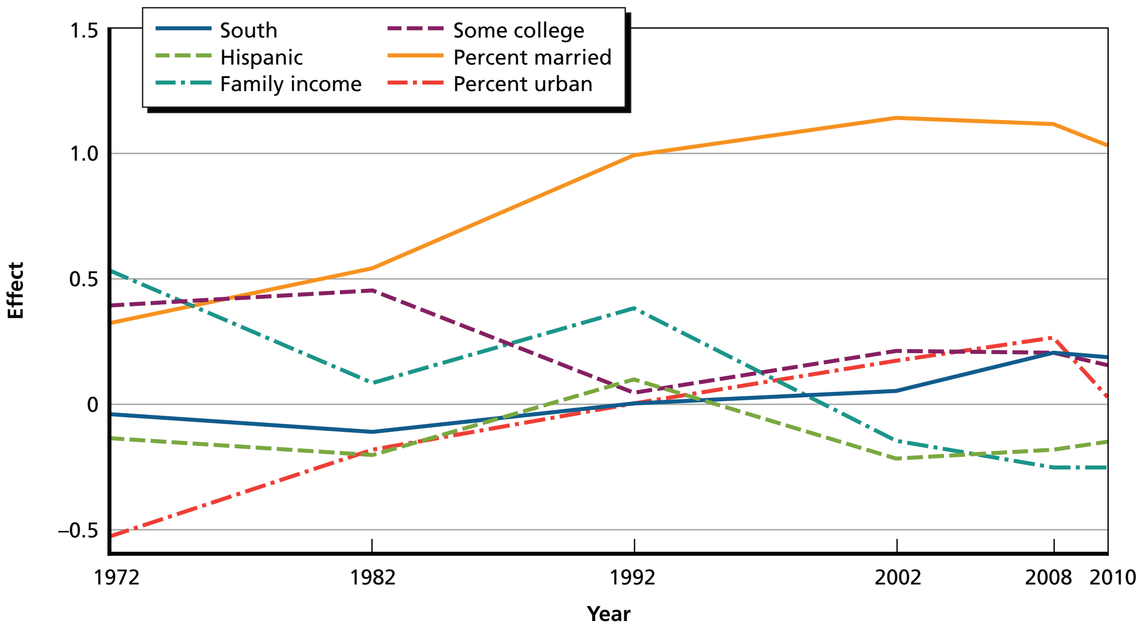
We began with a data set consisting of pooled observations from the 111th and 112th Congresses (2009 and 2011), and demeaned all continuous variables. We then fit two initial models: a logit model for R, the indicator for a Republican member, and an OLS model for PREP, the proportion of the vote in the 2008 presidential election going to McCain. In both of these models, the initial set of predictors consisted of the previously described race, education, and income variables, as well as the percentage of the population residing in an urban area, the percentage of the population that is married, and the indicator for South. These models were then winnowed using a backward stepwise approach with $p<0.05$ being the threshold for retention in the model. In fact, almost all variables were significant at a level of $p<0.001$. The exceptions were the percentage Hispanic and family income in the logit model for R ($p=0.008$ and $p=0.002$, respectively). These stepwise procedures resulted in two sets of "final" predictor variables: South, some college, Hispanic, married, percentage urban, and family income for the logit model; and black, Hispanic, college, some college, South, and married for the OLS model on PREP.

These variables were then used in analysis of the previous Congresses for R and national presidential elections for PREP, again across congressional districts. The overall importance of each variable was estimated by computing the product of the marginal effect in the regression model with the range of the variable from the fifth to 95th percentiles; we denote this product the "effect" of the variable.

We considered and rejected the approach of pooling data across all Congresses and introducing a variable to account for the different time periods. We did such analyses, but realized that this obscured important changes in the distributions of key variables over time, as discussed in Chapter Two. Thus, we chose to analyze each Congress separately.

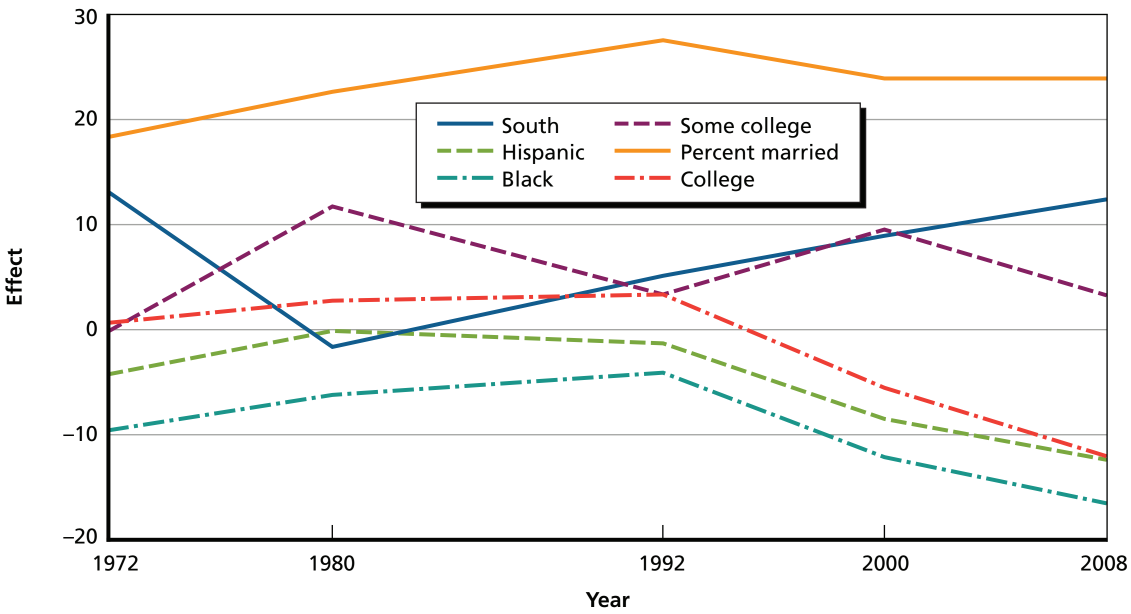
The trends in the importance of these variables are displayed in Figures A.7 and A.8, which summarize the results of these multivariate analyses. Each figure displays, for each variable, the product of the incremental or marginal effect from a cross-sectional regression with the fifth/95th percentile spread of that variable. In both the R and PREP analyses, only a few variables suffice in the sense that almost all of the variation in the outcome is explained; this is especially true for R. The variable effects shown on the graphs are important relative to each other, not in absolute terms.

Figure A.7
Variable Effects from Logit Model for "R"



RAND *RR896-A.7*

Figure A.8
Variable Effects from OLS Model of PREP



RAND *RR896-A.8*

Figure A.7 measures the relative importance of different factors in predicting the political party of district representatives for each of the Congresses in our study. Looking at the 93rd Congress (elected in 1972), we see that the value for the percent urban variable is negative: This means that in that Congress, urbanicity was predictive of Democratic representation. The value for the marriage variable is positive, indicating a positive relationship with Republican representation. Two things are noteworthy about this chart: First, even when multiple factors are accounted for, the marriage variable appears far more important than any of the other variables in terms of predictive ability. Second, that importance appears to be increasing over time.

The effect of the marriage prevalence rate on PREP also dominates, in terms of magnitude, in all years; however, the upward trend is much less pronounced relative to the model that predicts a Republican legislator.

The multistage model was constructed to address the following issue: DW-NOMINATE scores are correlated with constituency variables, but also with the political party of the representative, and with PREP, the Republican share of the two-party vote in the district. These factors—the party of the elected representative R and PREP—are themselves also correlated with the district-level covariates. In a causal framework, this means the district-level covariates influence DW-NOMINATE via two pathways: directly and indirectly through both R and PREP. Our model addresses this problem.

For each Congress, three regression models were fit (Tables A.2 and A.3). In the first, each representative's DW-NOMINATE score was regressed on PREP and the indicator for political party (using OLS). In the second, also using OLS, the PREP within district was regressed on the following district-level covariates: South, the proportion of adults with "some college" educational attainment, the proportion of adults with a college degree, the black and Hispanic population proportions, and marriage prevalence. In the third model, the indicator for representative political party was modeled logistically as a function of South, the "some college" variable, the Hispanic population proportion, marriage prevalence, median family income in the district, and the percentage of the population living in an urban area. As previously discussed, these predictor variables were chosen using a model-selection procedure that discarded variables failing to meet a significance criterion of $p<0.05$.

Table A.2
OLS Regression of PREP on Constituency Variables, 1972–2008

Variable	1972	1980	1992	2000	2008
South	13.055***	–1.622	5.080***	8.946***	12.574***
College	0.045	0.146	0.131**	–0.184***	–0.378***
Some college	–0.007	0.550***	0.196**	0.608***	0.225**
Black	–0.236***	–0.146***	–0.073	–0.222***	–0.312***
Hispanic	–0.211***	–0.004	–0.03	–0.155***	–0.205***
Married	1.094***	1.099***	1.220***	1.179***	1.141***
Constant	–6.002	–19.591**	–28.640***	–26.425***	–3.93
N	434	434	435	435	435
R^2	0.646	0.655	0.676	0.72	0.775

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table A.3
Marginal Effects from Logistic Regression of R (Indicator for Republican Legislator) on Constituency Variables, 1972–2010

Variable	1972	1982	1992	2002	2008	2010
South	−0.150**	−0.122**	−0.068	0.062	0.224***	0.246***
Some college	0.035***	0.020***	0.010*	0.015***	0.017***	0.016***
Hispanic	−0.016***	−0.010**	0.001	−0.002	0.000	−0.002*
Married	0.035***	0.032***	0.042***	0.050***	0.045***	0.044***
N	434	434	435	435	435	434
Pseudo R^2	0.167	0.166	0.187	0.331	0.283	0.395

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Formally, the following set of equations was fit for each of the 93rd, 98th, 103rd, 108th, 111th, and 112th Congresses:

$$Nom_i=\beta_0+\beta_1 PREP_i+\beta_2 REP_i+\varepsilon_i \tag{1}$$

$$PREP_i=\alpha_0+\alpha_1 South_i+\alpha_2 somcollege_i+\alpha_3 college_i+\alpha_4 black_i+\alpha_5 hispanic_i+\alpha_6 married_i+\upsilon_i \tag{2}$$

$$REP_i=\Lambda\left(\gamma_0+\gamma_1 South_i+\gamma_2 somcollege_i+\gamma_3 hispanic_i+\gamma_4 married_i\right) \tag{3}$$

where the variables are as described and $\Lambda(\cdot)$ is the logistic function.

Next, the predicted values from equations (2) and (3) were combined with the regression coefficients from equation (1) to generate a new predicted DW-NOMINATE score:

$$\widehat{Nom_i} = \beta_0 + \beta_1\widehat{PREP_i} + \beta_2\widehat{REP_i}$$

Finally, growth in the polarization statistic (the difference in interparty means) was generated from these predicted values and compared across Congresses to the polarization statistic generated from actual DW-NOMINATE scores.

This score vector is an estimate of what the ideology of representatives is, predicted from the observed constituency variables and after removing the confounding effects of R and PREP.

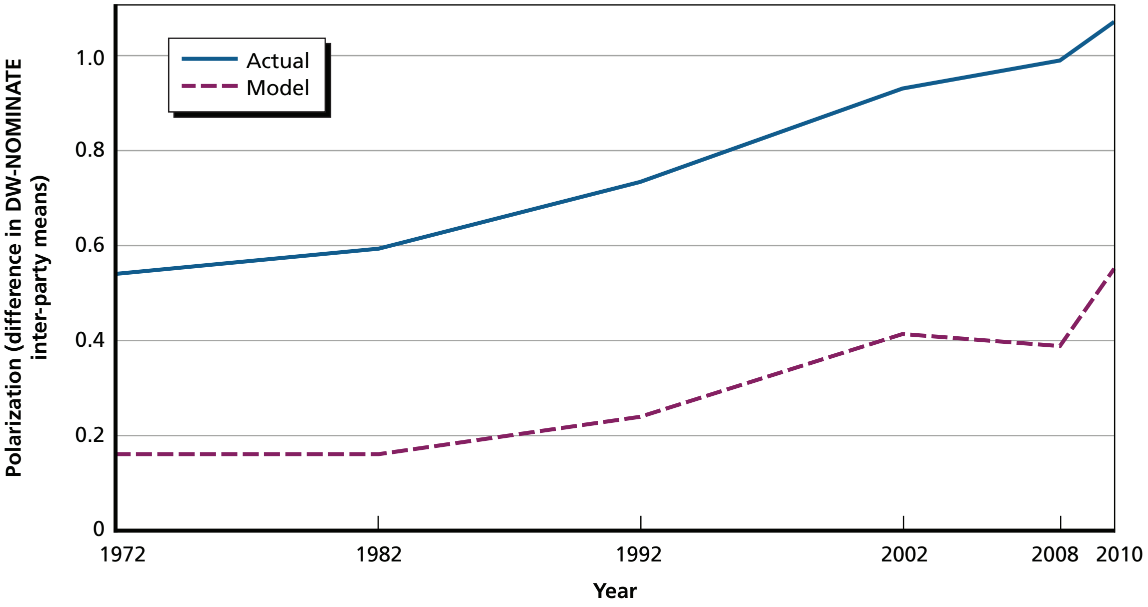
Marriage Prevalence and Polarization Estimates (Chapter Three, Footnote 17)

Figure A.9 displays the actual growth in House polarization alongside the index of predicted polarization generated from the multistage model.

In Figure A.10, we present a second version of the multistage model in which only a single covariate—the marriage prevalence—is used in both of the first-stage regression models.

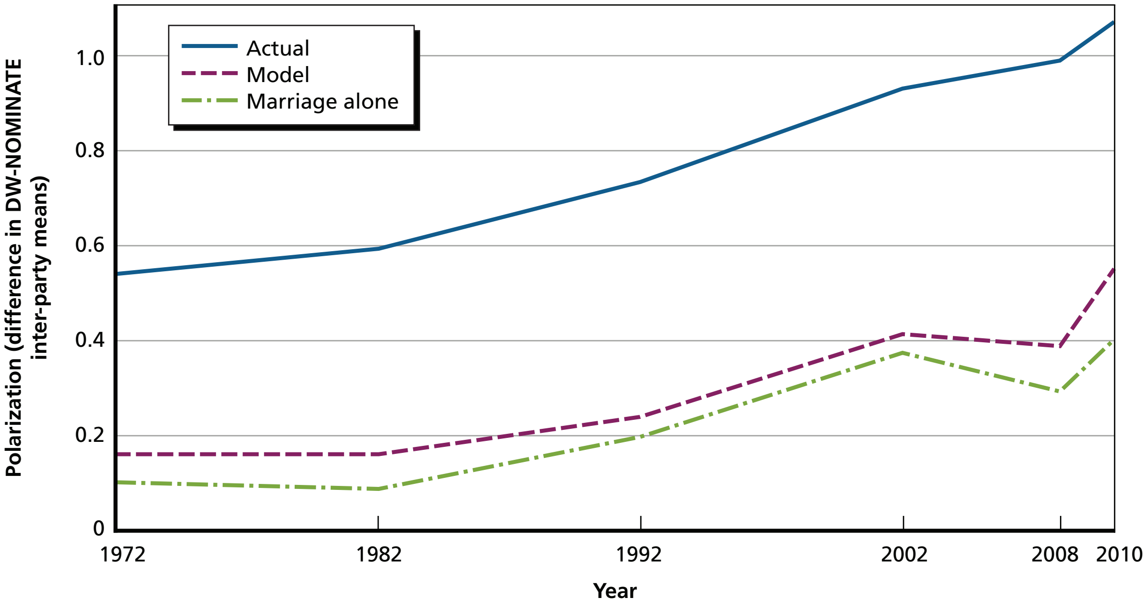
Because the polarization index derived from the reduced model (the model using marriage alone) tracks so closely with the index from the full model, we sought to investigate the overall contribution of marriage clustering to House polarization within the context of the

Figure A.9
Actual Polarization and Polarization Predicted from Model of Constituency Characteristics



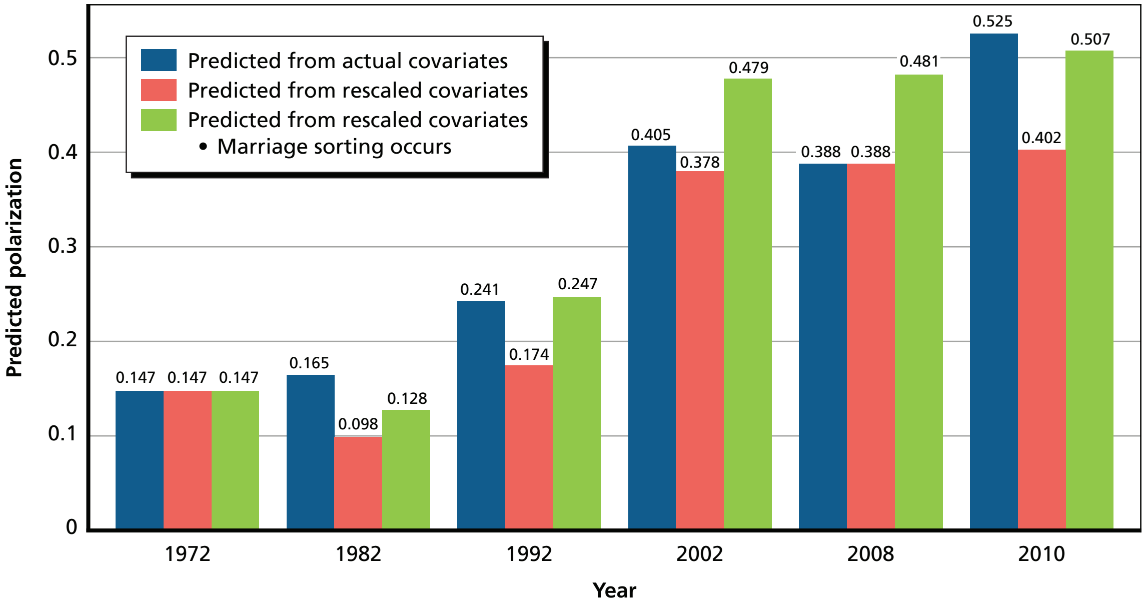
RAND *RR896-A.9*

Figure A.10
Actual Polarization and Polarization Predicted from Model of Constituency Characteristics (Full Model and Model Restricted to Marriage Covariate Alone)



RAND *RR896-A.10*

Figure A.11
Difference in Interparty Means of DW-NOMINATE Scores as Predicted by Actual, Rescaled, and Rescaled (Excluding Marriage) District-Level Covariate Values, 1972–2010



RAND *RR896-A.11*

rescaling model. As discussed in the main body of this text, this involves estimating polarization from predicted ideology scores using rescaled values for all covariates *except* the marriage prevalence rate, and using actual values for the marriage prevalence rate. The results of this analysis are displayed in Figure A.11.

The individual contribution of voter clustering along marriage lines may be significant. In the most recent Congress (the 112th), the gap between polarization as predicted by the actual and fully rescaled covariates is 0.525–0.402=0.123; of this, 0.105 (0.507–0.402) or 85.4 percent appears to be attributable to marriage clustering. This is estimated to be 19.9 percent of the overall increase in House polarization during the study period. However, we note that there is considerable variance in this estimate across years: In the 111th Congress (2009–2010), polarization predicted by the model in which only marriage clustering occurs (0.481) is actually greater than polarization predicted by the model in which clustering across all dimensions occurs (0.388). So while there is some evidence that clustering across marriage lines is contributing to rising House polarization, this evidence depends heavily on which year is the terminal year of the analysis and is therefore not decisive. One possible explanation for this high level of year-to-year variance is that the models are being affected by significant residual variation; i.e., noise.

Influence of Socioeconomic Factors on Individual Party Voting over Time (Chapter Three, Footnote 21)

We can formally test whether the effect of marriage is increasing over time by switching from a cross-sectional approach to a pooled approach; instead of fitting distinct models for each presidential election year, we fit a single model and include a time index and a term that interacts with the indicator for marriage with the time index (Tables A.4 and A.5).

In this formulation, the coefficient on the interaction term is positive but not statistically significant; this is consistent with visual inspection of the cross-sectional models and leads us to conclude that at the individual level, marriage is not an increasingly strong predictor of Republican partisanship.

Although the coefficient on the noninteracted marriage variable in the ANES study is not statistically significant at $p<0.05$, it is close to being so ($p=0.07$), suggesting that married voters marginally favor Republican presidential candidates (and nonmarried voters, Democratic ones). But the fact that the time-interaction term is not statistically significant ($p=0.41$), suggests that this marginally positive effect has not been growing over time when we control for other demographic variables. The large marriage effects that we see in the multivariate analysis (also controlling for other demographic variables) of congressional voting behavior is then due to clustering and not intensity.

Table A.4
Marginal and Incremental Effects of Individual Characteristics on the Probability of Voting for the Republican Candidate (Logistic Regression), 1972–2008 Presidential Elections (ANES)

Variable	1972	1976	1980	1984	1988	1992	1996	2000	2004	2008
Black	–0.55***	–0.64***	–0.63***	–0.52***	–0.62***	–0.53***	–0.92***	–0.60***	–0.55***	–0.80***
Hispanic	–0.21*	–0.16	–0.16	–0.19**	–0.26***	–0.12	–0.24**	–0.08	–0.20**	–0.25***
College	–0.09**	0.09*	0.06	–0.06	0.03	0.06	0.04	–0.08	–0.09*	–0.04
Some college	–0.02	0.04	0.05	0.02	0.08*	0.04	0.07	–0.01	0.02	0.02
South	0.10***	0.03	0.00	0.06*	0.07*	0.03	0.07	0.16***	0.12	0.12***
Income	0.02*	0.06***	0.03	0.08***	0.03	0.03	0.04*	0.04**	0.04*	0.04***
Male	0.05*	–0.04	0.06	0.03	0.04	0.01	0.09**	0.08**	0.07*	0.04
Married	0.04	0.00	0.05	0.01	0.00	0.11***	0.04	0.07*	0.05	0.06**

* significant at .05; ** significant at 0.01, *** significant at 0.001

Table A.5
Logistic Regression Results: Outcome Is Voting for the Republican Candidate in the Presidential Election (as Opposed to the Democratic Candidate), 1972–2008

Trait	Result
Black	–2.887***
Hispanic	–0.986***
College	0.005
Some College	0.204***
South	0.358***
Income	0.193***
Male	0.167***
Married	0.163
Time	–0.141***
tXmarried	0.013
Constant	0.031
N=49,127 F-statistic=131.46	

* p<0.05; ** p<0.01; *** p<0.001

References

Abramowitz, Alan I., *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*, New Haven: Yale University Press, 2010.

Abrams, Samuel J., and Morris P. Fiorina, "The Big Sort That Wasn't: A Skeptical Reexamination," *PS: Political Science and Politics*, April 2012.

Bishop, Bill, with Robert Cushing, *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*, New York: Houghton Mifflin, 2008.

Bonica, Adam, *Ideology and Interests in the Political Marketplace,* Social Science Research Network, February 2012.

Cahn, Naomi, and June Carbone, *Red Families v. Blue Families: Legal Polarization and the Creation of Culture*, New York: Oxford University Press, 2010.

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

Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope, *Culture War? The Myth of a Polarized America*, 3rd. ed., New York: Pearson Education Inc., 2010.

Fleisher, Richard, and John R. Bond, "The Shrinking Middle in Congress," *British Journal of Political Science*, 2004.

Kinsella, Chad J., *The Little Sort: A Spatial Analysis of Polarization and the Sorting of Politically Like-Minded People*" dissertation, University of Cincinnati, Arts and Sciences: Political Science, May 2011. As of December 11, 2014:
http://rave.ohiolink.edu/etdc/view?acc_num=ucin1318607836

Lee, David, Enrico Moretti, and Matthew Butler, "Do Voters Affect or Elect Policies," *Quarterly Journal of Economics*, August 2004.

Levitt, Steven D., "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology," *American Economic Review*, June 1996.

Mann, Thomas E., "Polarizing the House of Representatives: How Much Does Gerrymandering Matter?" in Pietro S. Nivola and David W. Brady, eds., *Red and Blue Nation? Characteristics and Causes of America's Polarized Politics*, Vol. 1, Washington, D.C.: Brookings Institution Press, 2006.

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

———, "Does Gerrymandering Cause Polarization?" *American Journal of Political Science*, Vol. 53, No. 2, July 2009.

Meyers, Adam S., "Secular Geographical Polarization in the American South: The Case of Texas, 1996–2010," *Electoral Studies*, Vol. 32, No. 1, March 2013.

Miller, Kristina C., "The Limitations of Heuristics for Political Elites," *Political Psychology*, December 2009.

MPR—*See* McCarty, Nolan, Keith T. Poole, and Howard Rosenthal.

Pildes, Richard, *Why the Center Does Not Hold: The Causes of Hyperpolarized Democracy in America*, New York University Public Law and Legal Theory Working Papers, Paper 207, 2010. As of February 2013:
http://lsr.nellco.org/nyu_plltwp/207

Petrocik, John R., "Measuring Party Support: Leaners Are not Independents," *Electoral Studies*, December 2009.

Poole, Keith, "The Roots of the Polarization of Modern U.S. Politics," *Revista Ciencia Politica*, October 2008.

———, voteview.com, website, undated. As of February 2013:
http://www.voteview.com

Poole, Keith, and Howard Rosenthal, "D-NOMINATE After 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting," *Legislative Studies Quarterly*, February 2001.

———, *Ideology and Congress*, New Brunswick, New Jersey: Transactional Press, 2007.

Stonecash, Jeffrey M., *Party Pursuits and the Presidential-House Election Connection, 1900–2008*, New York: Cambridge University Press, 2013.

Stonecash, Jeffrey M., Mark D. Brewer, and Mack D. Mariani, *Diverging Parties: Social Change, Realignment, and Party Polarization*, Boulder, Colo.: Westview Press, 2003.

Sussell, Jesse, "New Support for the Big Sort Hypothesis: An Assessment of Partisan Geographic Sorting in California, 1992–2010," *PS: Political Science and Politics*, October 2013.

Walker, Kyle E., "Political Segregation of the Metropolis: Spatial Sorting by Partisan Voting in Metropolitan Minneapolis-St Paul," *City and Community*, Vol. 12, No. 1, March 2013.