Beyond the Median: Voter Preferences, District Heterogeneity, and Representation

Elisabeth R. Gerber
Gerald R. Ford School of Public Policy
University of Michigan

Jeffrey B. Lewis
Department of Political Science
University of California, Los Angeles

August 27, 2002

We thank Mo Fiorina, Scott Page, Keith Poole, Tom Romer, and Jim Snyder for helpful comments. Gerber acknowledges the Center for Advanced Study in the Behavioral Science, Stanford, CA. Please direct correspondence to Elisabeth Gerber, Ford School of Public Policy, 440 Lorch Hall, 611 Tappan St., University of Michigan, Ann Arbor, MI 48109-1220, 734-647-4004, ergerber@umich.edu.
Abstract

Empirical studies of the connection between citizen preferences and legislator actions have largely focused on the preferences of a district’s average or median voter. While many theories suggest that preference heterogeneity should also affect this connection, few empirical analyses have rigorously tested claims about the distribution of individual voter preferences. In this paper, we use a unique set of individual-level proposition voting data from Los Angeles County that allows us to estimate the distribution of voter preferences, including the mean, median, and variance (heterogeneity), for various subsets of voters in each of 55 State Senate, State Assembly, and U.S. House districts in the county. We then develop measures of legislator ideology for each district based on Poole and Rosenthal’s NOMINATE scores. We use these data to test hypotheses about the relationship between district preferences, district heterogeneity, and legislator behavior. Regression analyses reveal that legislators’ roll call votes are more closely related to district mean preferences in homogeneous districts than they are in heterogeneous districts. Roll call votes of legislators from heterogeneous districts are better predicted by the mean preference of members of the legislator’s own party. Thus, in the presence of significant preference heterogeneity, we find that the convergence predictions of the median voter result fail to hold.
1 Introduction

Political representation concerns the relationship between government behavior and citizen preferences. Studies of representation seek to compare what legislators do with what the people in their districts want. The typical theoretical framework underlying such analyses is a simple spatial voting model in which citizens’ preferences are indexed by their ideal points (i.e., their most preferred outcomes), usually along a single policy dimension. A mapping between this set of ideal points and each legislator’s roll call voting or other behavior is then either derived from assumptions about legislators’ motivations, citizens’ information, and so forth, or is estimated empirically. In either case, the object is essentially to identify weights to apply to each citizen’s ideal point in the determination of each legislator’s votes. Perhaps the most widely employed theoretical mapping derives from the median voter theorem, which assigns all of the weight to the voter with the median ideal point. Other models suggest that weight be applied to voters at non-median positions. Still others suggest that substantial weight be applied to legislators’ own ideal points, the ideal points of contributors.

---

1 Legislative scholars refer to the relationship between an individual legislator and his/her district as “dyadic representation.” Weissberg (1978), Hurley (1982), and others have also considered the relationship between the behavior of the legislative body as a whole, and the preferences of the citizenry as a whole (i.e., collective representation).

2 Extending the policy space to multiple dimensions significantly complicates the basic underlying spatial model. See Schofield (1978), McKelvey (1979), and many others on majority voting in multidimensional space.

3 In the tradition of Downs (1957), Mayhew (1974), and others, many studies assume that legislators seek to maximize their election/ re-election prospects. Others assume that they seek to maximize rents. Still others acknowledge the possibility of mixed motives (e.g., Fenno 1978).


5 See Bender and Lott 1993, Goff and Grier 1996 for recent reviews.

6 This mapping follows from the Hotelling/ Downsian model of party competition (Hotelling 1929, Downs 1957, Black 1958, Hinich and Enelow 1984). In this simple spatial representation of electoral competition, election-oriented candidates maximize their election prospects by taking positions that converge to their district’s median voter’s position.

7 For example, entry models such as Palfrey (1989) note that some divergence may deter potential entrants. Setter models such as Romer and Rosenthal (1979) predict that extremist (e.g., budget maximizing) proposers will implement policies which depend on the location of the status quo (reversion point), the location of the median voter, and the uncertainty in the median’s location. Hinich (1977) predicts candidate convergence to the district mean in two-party elections. In multiple dimensions, predictions depend on the joint distribution of ideal points across dimensions (see, for example, McKelvey 1986 and Goff and Grier 1993 on the uncovered set). See Fiorina (1999) for a review of theories of candidate divergence.

tors to their campaigns,\textsuperscript{9} members of their parties and so-called reelection constituencies,\textsuperscript{10} legislative leaders,\textsuperscript{11} or voters in other districts.\textsuperscript{12}

While the theoretical literature suggests a large number of potential weightings, relatively few of these have been subjected to extensive empirical scrutiny. Limited by the lack of data necessary to construct reliable measures of all of the potentially relevant preferences, the empirical literature has typically focused on measuring district average preferences. One recent series of papers,\textsuperscript{13} building on early works by Kau and Rubin (1979), Peltzman (1984, 1985), Kalt and Zupan (1984), and others, seeks to estimate the relative importance or weighting of two sets of factors on legislator behavior – average citizen preferences and legislator ideology – in some cases also accounting for the effects of campaign contributions,\textsuperscript{14} national party preferences,\textsuperscript{15} or electoral competition.\textsuperscript{16} Two main approaches to estimating citizen preferences appear in this literature: either preferences are written as a linear combination of district-level economic and demographic variables, or they are measured by election returns or survey data, aggregated to the district level.\textsuperscript{17} For the most part, these studies concur that both citizen preferences and legislator ideology shape legislative behavior; they disagree about the relative importance and significance of these effects.\textsuperscript{18}

Given their approaches to estimating citizen preferences described above, the extant empirical studies imply that the influence of “citizen preferences” can be fully captured by

\textsuperscript{9}See Stratmann (1995).
\textsuperscript{13}See Bender and Lott 1993, Goff and Grier 1996 for recent reviews of this literature.
\textsuperscript{15}Levitt (1996)
\textsuperscript{17}To estimate legislator ideology, these studies follow two basic approaches. One approach employs a two-stage procedure, which first estimates legislator ideology as the residual from a regression of interest group ratings on measures of citizen preferences and other factors; this “ideology residual” is then included in a regression model of roll call voting behavior (Kau and Rubin 1979, Kalt and Zupan 1984, Carson and Oppenheimer 1984, Wittman 1983). The second approach includes both measures of citizen preferences and direct proxies of ideology such as the interest group ratings (Peltzman 1984, 1985) or legislator characteristics (Coates and Munger 1995, Levitt 1996) as explanatory variables in a roll call voting regression.
\textsuperscript{18}Kalt and Zupan (1984) and others decry the independent impact of legislator ideology as agency loss or shirking. Peltzman (1985) and others argue that some degree of legislator discretion is acceptable to the functioning of democratic governments.
a single statistic measuring “median citizen preferences.”\(^{19}\) We acknowledge that there are certain conditions under which a single statistic could, in fact, be sufficient to represent the influence of the full distribution of voter preferences. For example, to the extent that the predictions of the median voter theorem hold, then knowledge of the median preference is sufficient to predict a legislator’s voting behavior. Moreover, if variation in the distributions of citizen preferences across districts can be captured by a single district-level parameter, then measures of this parameter would fully describe the influence of citizen preferences on legislative voting, even if the median voter theorem did not hold. On the other hand, if voter preference distributions vary in more complex ways across districts, then even if a mean or median voter theorem prediction does not hold, a single measure of average constituency preference is still not sufficient to describe legislator behavior. The question, then, is how valid is the assumption that citizen preferences in each district can be summarized by a single (perhaps composite) indicator.

We address this question by constructing measures of district preference that go beyond central tendencies. Using individual-level voting data on statewide ballot propositions, we estimate the distribution of voter preferences within each of 55 legislative districts, as described below. Our estimation procedures enable us to estimate the medians, means, and variances of various subsets of voter preferences. If the median voter theorem prediction holds, then information about district heterogeneity (variance) would contribute little to our estimation of legislative behavior. On the other hand, if more a complex relationship between citizen preferences and legislative behavior exists, then we expect preference heterogeneity to affect the relationship between representatives and the represented. In relatively homogeneous (low variance) districts, the distances between the average voter’s position and positions of other potentially relevant members of the electorate will be relatively small and proxying those positions with the location of the median is relatively unproblematic. In heterogeneous (high variance) districts, the average voter location may be quite far from other relevant locations in the voter distribution; in this case, the median cannot effectively

\(^{19}\)As mentioned above and described in more detail below, some of the empirical literature does includes measures of the average policy preference of legislators’ support coalitions in addition to measures of the district-wide average preference.
proxy the locations of other relevant members of the constituency. Thus, only if district heterogeneity does not intervene in the relationship between “average” voter preference and legislator behavior, can the distribution of voter preferences be summarized by the location of the average voter. In our analyses reported in this paper, we find that district heterogeneity does intervene; we find a significantly stronger relationship between average voter preference and legislator behavior in homogenous districts than in heterogeneous districts. This conclusion supports theoretical models predicting non-median electoral outcomes.

Our interest in district heterogeneity is not unique. Indeed, a number of existing empirical studies that attempt to estimate the impact of district heterogeneity on legislator behavior find that composition matters in important ways. For example, Peltzman (1984) finds that much of the apparent impact of ideology on roll call votes disappears once one accounts for deviations between the average characteristics in a Senator’s state and the characteristics of his/her supporters. Goff and Grier (1993) find that differences in same-state Senators’ voting records can be largely explained by heterogeneity in the state’s income distribution, ethnic makeup, and workforce composition. Similarly, Bailey and Brady (1998) find that state population heterogeneity, as measured by an index of state socioeconomic and cultural diversity, significantly impacts Senate voting on trade legislation. All three of these studies conclude that ignoring the effect of heterogeneity produces misleading inferences. They imply that the effect of citizen preferences on legislator behavior is fundamentally different in heterogeneous states.

We believe that by recognizing the potential importance of population composition on representation, and the consequences of preference heterogeneity for empirical applications of the spatial model, these three papers are a major conceptual step in the right direction. However, due to data limitations, their operationalizations of composition/heterogeneity are quite simplistic and may or may not be related to heterogeneity of preferences towards the policies of interest. We build upon these studies by applying new data, and largely confirm their results. We further find, consistent with Peltzman (1985) and others, that when members’ principles are identified as partisans within their districts, much of the apparent ideological or shirking behavior disappears. While this conclusion is not novel, it has important implications for how we evaluate the quality of American political representation.
The paper is organized as follows. In the next section, we construct measures of district preferences that capture both the central ideological tendency in the district as well as the degree of heterogeneity and the preferences of important subgroups. We then use these measures as independent variables in analyses of legislator roll call voting.\textsuperscript{20} In our analyses, we first establish that a district’s estimated median preference is a poor predictor of legislator roll call voting behavior in heterogeneous districts. We then test a number of alternative hypotheses about the effects of heterogeneity. We find strong support for one of these hypotheses (our core/partisan constituency hypothesis) and little evidence for the other two alternatives (our multidimensionality and avoidance hypotheses). We conclude by discussing the implications of our analyses for our understanding of legislative representation.

Empirical studies of political representation require two things to be measured – what constituents want and what legislators do. The next two sections describe how we measure these two things in this study.

2 Measuring District Preferences

We study the linkages between district composition and legislator behavior in Los Angeles County’s US House, State Assembly, and State Senate districts in the early 1990s. In most cases, a study of political representation in state and federal legislatures based on observations made in a single county would not be very informative. The average county in the United States has about 1/7th of a member of Congress, 1 3/4ths state house members, and 6/10ths of a state senator. In total, the average county would yield about 2 1/2 districts in which to explore the constituency-representative relationship.

Fortunately, Los Angeles is no mean county. Los Angeles’ population of 9.3 million makes it larger than all but 8 states and as large the smallest 11 states combined.\textsuperscript{21} With population, of course, comes legislative districts. Angelenos elect 24 members of the California Assembly.

\textsuperscript{20}In our analyses, we focus on legislator policy positions. District composition may also affect, for example, how candidates raise money and from whom, the committee assignments they seek, the amount of time they spend in the district, etc. Computational models by Kollman et al. (1992, 1998) and de Marchi (1999) show that district heterogeneity may also be associated with increased incumbency advantage, as incumbents in heterogeneous districts are better able than challengers to find optimally policy positions in rugged electoral landscapes.

\textsuperscript{21}Based on the 1999 population estimates from the U.S. Census Bureau.
14 members of the California Senate, and 17 members of the US House of Representatives. All but four of the Congressional districts and three of the Senate districts are entirely contained within the county borders. While an \( N \) of 55 districts is by no means huge, it is certainly large enough to yield interesting and informative results.

Size is not everything. In addition to being very large, Los Angeles is also very diverse. The county includes both inner-city and suburban districts, as well as ethnically diverse and ethnically homogeneous districts. The legislators that LA County sent to the House of Representatives in 1992 include three of that House’s 15 most liberal (Waters, Tucker, and Roybal-Allard) and two of its 15 most conservative (Royce and Drier). Because of its size and diversity, Los Angeles County presents an ideal case in which to study the linkage between citizens and their representatives.

To estimate the preferences of Los Angeles County voters, we follow the work of Kukliniski (1979) and Snyder (1996) and infer citizens’ preferences from their votes on ballot propositions. Unlike those authors, however, our estimates are not based on aggregate voting returns. Rather, our analysis employs a canvas of all 2.8 million ballots cast in the county in the 1992 General Election. The computer file that contains these data was generated as a by-product of the punch card ballot system used in the election. Each record contains a complete enumeration of all the vote choices made by each voter in the county. By observing the joint distribution of vote choices of all voters at the individual level, we are able to recover much more detailed information about the distribution voter preferences in a large number of legislative districts than has previously been available.

The ballot measures contained on the 1992 ballot tap numerous important policy questions. The statewide propositions on the 1992 ballot are shown in table 1. These propo-

---

22 California’s Assembly has 80 members and its Senate has 40 members.
23 As measured by the NOMINATE scores described and presented below.
24 Due to changes in the computers and software used to tabulate ballots in Los Angeles, these data are not available for elections after 1992.
25 One limitation of these data is that they measure the preferences of only those citizens who participate in the election, i.e., of actual voters. However, given the logic of the electoral competition underlying our basic spatial framework, we expect legislators to respond to precisely these individuals.
26 Because it posed a very similar question to Proposition 158, we dropped Proposition 159 from the analysis. Votes on Propositions 158 and 159 are highly correlated (much more highly than voting on any other proposition pair) and including both of them results in the recovery of a policy dimension that is dominated by these two items.
# Propositions on the November 1992 ballot

<table>
<thead>
<tr>
<th>Proposition</th>
<th>Topic</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>155</td>
<td>$900m School facilities bond</td>
<td>(Referendum passed, 52—48)</td>
</tr>
<tr>
<td>156</td>
<td>$1b Passenger rail and clean air bond</td>
<td>(Referendum failed, 49—51)</td>
</tr>
<tr>
<td>157</td>
<td>Toll Roads and highways</td>
<td>(Referendum failed, 29—71)</td>
</tr>
<tr>
<td>158</td>
<td>Office of the California Analyst</td>
<td>(Referendum failed, 41—59)</td>
</tr>
<tr>
<td>160</td>
<td>Property tax exemption</td>
<td>(Referendum passed, 52—49)</td>
</tr>
<tr>
<td>161</td>
<td>Physician-assisted death</td>
<td>(Initiative failed, 46—54)</td>
</tr>
<tr>
<td>162</td>
<td>Public employee’s retirement system</td>
<td>(Initiative passed, 51—49)</td>
</tr>
<tr>
<td>163</td>
<td>Taxation of food products</td>
<td>(Initiative passed, 66—34)</td>
</tr>
<tr>
<td>164</td>
<td>Congressional term limits</td>
<td>(Initiative passed, 62—38)</td>
</tr>
<tr>
<td>165</td>
<td>Budget process and reduction in AFDC</td>
<td>(Initiative failed, 46—54)</td>
</tr>
<tr>
<td>166</td>
<td>Mandate employer provided healthcare</td>
<td>(Initiative failed, 32—68)</td>
</tr>
<tr>
<td>167</td>
<td>Corporate, Income and sales tax rates</td>
<td>(Initiative failed, 42—58)</td>
</tr>
</tbody>
</table>

Table 1: *Outcomes of propositions on the 1992 ballot*

Propositions include issues that tap some of the main dimension of political conflict in the US (bonds, taxes, and healthcare reform), as well as some technical government operations questions and social issues (for example, physician assisted suicide). Overall, they provide a range of questions from which we can estimate voters’ underlying political preferences. Further, compared to respondents on most surveys, voters had the opportunity and the incentive to carefully consider the ”questions” posed to them on these propositions before “answering” by casting their ballots.

The voting data contained on these ballots is analogous to the roll call voting data commonly used to construct measures of legislators’ policy positions. For each voter, we have a voting record on a series of binary legislative proposals. Building on the models of Bock and Aitkin (1981), Lewis (2001) develops a technique for estimating the mean and variance of the distribution of voter preferences within each of a set of predetermined groups.

---

27Given the structure of the data, it would seem logical to use a technique such as Poole and Rosenthal’s NOMEAT to extract each voter’s ideal point from their voting record (see Poole and Rosenthal 1985, 1997). However, with a maximum of only 12 votes observed for each voter, NOMEAT estimates are not consistent. The exact conditions under which NOMEAT scores are consistent is unclear. However, it is clear that a necessary though not sufficient condition is that the number of votes taken goes to infinity (see Londregan 2000 or Lewis 2001).
(e.g., electoral districts) from data on a small number of observed binary choices. We use a similar estimation procedure in this analysis.

Each voter’s decisions on each of a set of ballot propositions, \( j = 1, 2, \ldots, J \), is described by a simple spatial voting model. In particular, the utility that voter \( i = 1, 2, \ldots, N \) receives as a result of the outcome of the proposition election \( j \) is

\[
U(A_j, \theta_i) = -(A_j - \theta_i)^2 + \epsilon_{ijA}
\]

where \( A_j \in \{P_j, S_j\} \), \( P_j \) is the policy that is implemented if proposition \( j \) passes, \( S_j \) is the status quo policy that continues if proposition \( j \) fails, \( \theta_i \) is voter \( i \)’s ideal point, and \( \epsilon_{ijA} \) is a random utility shock. Assuming sincere voting in the sense that voter \( i \) votes for proposition \( j \) if \( U(P_j, \theta_i) > U(S_j, \theta_i) \) and against it otherwise and letting

\[
V_{ij} = \begin{cases} 
1 & \text{if voter } i \text{ votes for proposition } j \\
0 & \text{if voter } i \text{ votes against proposition } j 
\end{cases}
\]

we can write for each \( i \) and \( j \):

\[
\text{Prob}(V_{ij} = 1|\theta_i) = \text{Prob} \left( U(P_j, \theta_i) - U(S_j, \theta_i) > 0 \right) = \text{Prob} \left( -P_j^2 + 2P_j \theta_i + \theta_i^2 + \epsilon_{ijP} + S_j^2 - 2S_j \theta_i + \theta_i^2 - \epsilon_{ijS} > 0 \right) = \text{Prob} \left( S_j^2 - P_j^2 + 2(P_{ij} - S_{ij})\theta_i > \epsilon_{ijS} - \epsilon_{ijP} \right).
\]

Assuming each \( \epsilon_{ijA} \) is i.i.d. \( N(0, 1) \), \( \epsilon_{ijS} - \epsilon_{ijP} \sim N(0, \sqrt{2}) \). Now we can write,

\[
\text{Prob}(V_{ij} = 1) = \Phi \left( \frac{S_j^2 - P_j^2 + 2(P_{ij} - S_{ij})\theta_i}{\sqrt{2}} \right)
\]

where \( \Phi \) is the standard normal cumulative distribution function. Letting,

\[
\alpha_j = \frac{S_j^2 - P_j^2}{\sqrt{2}}
\]

and

\[
\beta_j = \frac{2(P_{ij} - S_{ij})}{\sqrt{2}},
\]
we find that $V_{ij}$ is described by a simple Probit regression,\footnote{This sort of result is well-known in the literature. See, for recent examples, Heckman & Snyder (1997) or Londregan (2000).}

$$\text{Prob}(V_{ij} = 1) = \Phi (\alpha_j + \beta_j \theta_i).$$

Assuming the $V_j$'s are (conditionally) independent, we write the likelihood that voter $i$ casts a pattern of votes, $V_i = (V_{i1}, V_{i2}, \ldots, V_{iJ})$ as

$$L(V_i | \alpha, \beta, \theta_i) = \prod_j \Phi (\alpha_j + \beta_j \theta_i)^{v_{ij}} (1 - \Phi (\alpha_j + \beta_j \theta_i))^{1 - v_{ij}}$$

where $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_J$ and $\beta = \beta_1, \beta_2, \ldots, \beta_J$.

In the estimation, voters are partitioned into groups based on geography and partisan voting patterns.\footnote{Note that this is the only other information about the voter present in the ballot data.} Each voter is classified as a Democrat, Independent, or Republican based on the their votes for President, US Senate, US Congress, and State Assembly, and is further classified by their US House, State Senate, and State Assembly district.\footnote{We include the four partisan races that were common to all voters in the county; only a subset lived in districts in which the State Senator faced election in 1992.} Thus each voter belongs to a unique group defined by their partisan preference and the set of legislative districts in which they live. Let $g = 1, 2, \ldots, G$ enumerate the non-empty district-partisan combinations and $\Psi : i \rightarrow g$ be a function that associates each voter with her district-partisan combination. The ideal points of voters from the same district-partisan combination, $g$, are assumed to be independently drawn from the same normal distribution with mean $\mu_g$ and variance $\sigma_g^2$. For any voter $i$ in a given group $g$, the likelihood of casting a pattern of proposition votes $V_i$, unconditional on $\theta$, is

$$L(V_i | \alpha, \beta, \mu_g, \sigma_g) = \int L(V_i | \alpha, \beta, \mu) \phi(\mu | \mu_g, \sigma_g) d\mu$$

where $\phi$ is the normal density function with mean $\mu_g$ and variance $\sigma_g^2$. Assuming that the voters cast their ballots independently, we can write the full log likelihood function for the
set of $N$ votes, $V$ as

$$
\ln L(V|\alpha, \beta, \mu, \sigma) = \sum_i \ln \left( \int L(V_i|\alpha, \beta, \theta) \phi(\theta|\mu_{\psi(i)}, \sigma_{\psi(i)}) d\theta \right).
$$

(1)

where $\mu = \mu_1, \mu_2, \ldots, \mu_D$ and $\sigma = \sigma_1, \sigma_2, \ldots, \sigma_G$. We maximize this log likelihood function to obtain estimates of $\alpha, \beta, \mu,$ and $\sigma$. In order to identify the location and scale of $\theta$, the values of $\mu_1$ and $\sigma_1$ are fixed at 0 and 1 respectively.\footnote{Identification in factor analytic models such as the one presented here generally require either the fixing of location and scale of the latent factor (as done here) or the fixing of one $(\alpha_j, \beta_j)$ pair.} The likelihood function is maximized using an EM algorithm. Standard errors are obtained in the usual way using numerical integration to evaluate the full log likelihood.\footnote{The estimating similar likelihood functions was first introduced in the psychometric literature on test taking (see Bock & Aitken (1981)). Similar models have been considered in the context of roll-call voting by Londregan (2000) and Bailey (2001). A complete treatment of the exact statistical model presented here including estimation and robustness issues and extensions can be found in Lewis (2001).}

The distribution of preferences within each legislative district is formed by mixing the estimated preferences distributions of all of the groups that have a particular district in common. For example, the distribution of voters in Assembly district $d$ can written as

$$
f_d(\theta) = \sum_{g=1}^{G} P(g, d) \phi(\theta|\mu_g, \sigma_g).
$$

where $P$ is a function returning the share of voters in Assembly district $d$ who belong to group $g$. The distribution of voters in each legislative district is the discrete mixture of at least three normal distributions (one for each partisan group). In the case of Congressional and State Senate districts the mixture is generally comprised of many more groups because Senate districts overlap more than one Assembly district and Congressional districts typically overlap multiple Assembly and Senate districts. While the assumption that voter preferences are normally distributed within groups is restrictive, the possible distributions of voter preferences at the full district level are far more general. Thus, while group preferences must be unimodal and symmetric, estimated preference distributions at the district level can be skewed and even multimodal.
3 Measuring Legislator Behavior

We now turn to the estimation of legislator ideal points. Several methods exist that allow researchers to infer legislators’ policy positions from a large set of roll call votes. Most notable and commonly used are interest group ratings such as ADA scores, NOMINATE scores (Poole and Rosenthal, 1985, 1997), and Heckman-Snyder scores (Heckman and Snyder, 1997). A general finding of this literature is that a very large amount of the variation in legislators’ roll call voting records can be accounted for by a single left-right dimension. For this study, we use a variant of the Poole and Rosenthal NOMINATE procedure to estimate legislators’ left-right positions.

As mentioned above, our dataset contains roll call voting data for members from three different legislative bodies (the California State Assembly, the California State Senate, and the US House of Representatives). In order to make comparisons across these three chambers, we need a way of placing members from each of the chambers in the same policy space. The problem is that no roll call scaling technique (including NOMINATE) produces a scale that has any natural metric. In particular, incomparability can enter in three ways. First, the underlying dimension or dimensions may not be the same in each chamber. Second, the units of the NOMINATE scores may not be comparable. Third, the zero point of the scales may not be comparable. Without a way to address these problems, comparisons of NOMINATE scores generated from separate sets of roll call votes (for example, votes from different chambers) are invalid.

In order to construct NOMINATE scores that are comparable across legislatures, it is necessary to have either some members who vote in all of the chambers, or some roll call votes that are over the same bill (and status quo) in all chambers, or both. Each of these approaches have been used to recover comparable estimates of the location of US House and Senate members and to get comparable estimates of the location of US House and Senate members at different points in time. In each case, the common space is recovered because the members or votes that are common to all chambers serve to anchor the rotation, scale, and location of the space. However, to assume that members’ ideal points remain the same

---

33 The most common approach to this problem is to assume that members’ positions are fixed over time
when their constituencies change (as they do when a member moves from one chamber to another) is to assume away the very possibility of responsiveness to constituency that we are interested in exploring. Moreover, there are few (if any) roll call votes that involve that the same alternative and status quo across national and state legislatures. Thus, neither of these approaches is appropriate to the current study.

Instead, we rely on interest groups that rated members in the California Assembly, the California State Senate, and the US House. These interest group ratings are calculated as the percentage of times each legislator shared the position of the group on a roll call vote of salience to the group. Groups report the votes they use to construct their scores and the positions they took on those votes. As such, while these groups did not actually “vote” in the three chambers, they took positions on the issues considered by them. Thus, we can construct a “voting record” for each group in each chamber. The group is said to have abstained on all roll calls that it did not factor into its rating score and to have voted “nay” or “yea” according to its stated preference on all roll calls that were used in the score. These “voting records” are then added to the records of the “real” legislators and the groups’ ideal points are estimated along with those of the legislators. If we assume that groups’ idea points do not vary across chambers, then we can use these groups to anchor the issue space. The more groups’ “voting” patterns we incorporate into our estimations, the better anchored the issue space is and the more confident we are that the scales are indeed comparable.

We have constructed roll call voting records in the three chambers for three interest groups: the League of Conservation Votes (LCV), the Chamber of Commerce (CoC), and the AFL-CIO. The number of positions taken by these groups during 1993-94 is shown in table 2. The number of observed votes for each group is small relative to the total number or between chambers (see Poole and Rosenthal 1997; Groseclose, Snyder, and Levitt 1999). This approach may be generalized to allow members’ positions to change over time in circumscribed ways (see Poole and Rosenthal, 1997). An alternative approach is to assume that certain votes that take place in the each chamber involve the same proposed policy and status quo (see Bailey 1998, Adams, Bailey, and Fastnow 2000), and hence divide the policy space in the same way across chambers. These votes can then be used to anchor the distribution of scores across chambers.

Poole and Rosenthal (1997) estimate interest groups’ NOMINATE scores in the context of the US Congress, finding generally similar placements to those presented below.

It is possible that the positions of state and national chapters of a given interest group might differ, and hence the assumption that the ideal points of the groups are constant across legislatures would be violated.
Legislative positions taken by three interest groups, 1993-94

<table>
<thead>
<tr>
<th></th>
<th>State Assembly</th>
<th>State Senate</th>
<th>U.S. House</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yea</td>
<td>Nay</td>
<td>Yea</td>
</tr>
<tr>
<td>AFL-CIO</td>
<td>59</td>
<td>6</td>
<td>55</td>
</tr>
<tr>
<td>Chamber of Commerce</td>
<td>20</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>League of Conservation Voters</td>
<td>19</td>
<td>15</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 2: Shows the number of roll call votes on which the LCV, CoC, and AFL-CIO took the yea and nay position in 1993-94 by chamber.

of about 9,000 roll call votes used to construct the NOMINATE scores.\(^\text{36}\) However, in absolute numbers, 90 or so votes from each interest group should be enough to yield reasonable ideal point estimates for the groups.\(^\text{37}\)

Given the voting records for these three groups, we then form a large dataset comprised of all non-unanimous votes cast in each of the three chambers in 1993 and 1994. The dataset includes not only those members from Los Angeles County, but all members serving in each chamber. Including these other members improves the efficiency of the estimation and allows us to compare Los Angeles’ delegation to that of California or the nation. The NOMINATE procedure is then applied to this large set of data and comparable estimates of the spatial locations of each member in each chamber are recovered.\(^\text{38}\)

Table 3 shows some summary statistics for the first dimension NOMINATE scores by chamber. The means, medians, and standard deviations are quite similar across chambers, and none of the differences in these statistics are significant. Since the NOMINATE scores are constructed such that -1 represents the most liberal position and +1 the most conservative, we see that the mean and median member of the California House delegation are more liberal

\(^{36}\)The NOMINATE scores are based on all non-unanimous votes taken during 1993 and 1994 in the Assembly, State Senate, and US House. There were 6166 such votes in the State Assembly, 1758 such votes in the State Senate, and 1093 such votes in the US House.

\(^{37}\)The accuracy of these estimates will depend on the accuracy with which the legislators’ ideal points and the vote-specific parameters are estimated.

\(^{38}\)We recover scores in three (orthogonal) dimensions.
Summary Statistics of First Dimension NOMINATE Scores by Chamber

<table>
<thead>
<tr>
<th>Chamber</th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>US House:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All members</td>
<td>-0.09</td>
<td>-0.27</td>
<td>0.61</td>
</tr>
<tr>
<td>CA members</td>
<td>-0.12</td>
<td>-0.47</td>
<td>0.75</td>
</tr>
<tr>
<td>CA Assembly</td>
<td>-0.11</td>
<td>-0.37</td>
<td>0.59</td>
</tr>
<tr>
<td>CA Senate</td>
<td>-0.02</td>
<td>-0.20</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Table 3: Shows summary statistics for first dimension NOMINATE scores by chamber. Note: The variation in means and medians across chambers is not statistically significant at the α = 0.05 level.

than their chamber-wide counterparts, although the standard deviation of the California delegation is higher.\(^{39}\) The mean and median Assembly members’ positions are nearly identical to the House mean and median, and the Senate mean and median are more conservative.

Consistent with Poole and Rosenthal’s findings for the US Congress, the first NOMINATE dimension has great explanatory power. Conditioning voting decisions on the first dimension NOMINATE score results in an overall classification success across all of the roll calls of 89.7 percent. Compared to a naive model, 89.7 percent correct classification represents a 67 percent reduction in error (PRE). Adding two more dimensions to the model increases the overall classification success to nearly 91.3 percent and increases the PRE to 0.72.

4 District Composition and Legislator Behavior

We now use our estimates of district composition and legislator policy positions to test hypotheses about the effects of heterogeneity on political representation. We begin by analyzing the relationship between district mean preference and legislator policy positions in districts with varying degrees of heterogeneity.\(^{40}\) We hypothesize that the logic of spatial

\(^{39}\)Some of this is likely a result of the smaller size of the delegation.

\(^{40}\)In the analyses that follow, we employ estimates of district means rather than medians. We note that substituting means for medians could be problematic, particularly if differences between the medians and means were systematically larger in high variance districts (we thank an anonymous referee for bringing this possibility to our attention). Under the assumptions described in section 2, we can calculate each district’s
competition will lead legislators to take positions that are closer to their district’s mean when heterogeneity is low. We expect the district’s mean preference to be a less powerful predictor of legislator behavior when heterogeneity is high. Using several specifications within each of two general models, we assess the extent to which heterogeneity mediates or conditions the relationship between district preferences and legislator behavior.

Table 4 presents our first set of analyses. Here, we simply compare the relationship between legislator policy positions and district mean preferences in homogenous and heterogeneous districts. Each column reports OLS regression coefficients and standard errors, where the unit of analysis is a legislative district and the dependent variable is the legislator’s first dimension NOMINATE score.

In the first column of table 4, we provide a baseline model in which we regress the legislator’s NOMINATE score on our estimate of the district’s mean preference. We see that, as predicted by the spatial models, the coefficient on district preference is positive and significant. In other words, on average, legislators in conservative districts (high mean) take more conservative policy positions (high NOMINATE score), and vice versa.

We then split the sample into low variance and high variance districts. We see that most of the relationship in the baseline model is being driven by the low variance (homogenous) districts. For homogenous districts in column 2, the coefficient on district mean preference is double the magnitude of that in the baseline model, positive, and significant. For heterogeneous districts in column 3, the coefficient on district mean preference is indistinguishable from zero. In other words, legislators in more conservative homogenous districts are themselves more conservative (and vice versa). However, legislators in conservative heterogeneous districts are no more conservative, on average, than legislators in liberal heterogeneous districts. Thus, district mean preference is a much less powerful predictor of legislator behavior.

\[
\text{median by solving } \sum_j p_j F(m|\mu_j, \sigma_j) = 1/2 \text{ for } m \text{ where } F \text{ is the normal CDF, } j \text{ indexes the party subdistrict pairs found in the district and } p_j, \mu_j, \text{ and } \sigma_j \text{ are the fraction of the district’s voters from subgroup } j \text{ and the mean position and standard deviation of subgroup } j, \text{ respectively. The district medians correlate with the estimated means at over } r = 0.98. \text{ Not surprisingly, all the subsequent analysis is unchanged when the medians are used in place of the means.}
\]

\footnote{The cutoff value is slightly above the mean district variance. Our results are robust to our choice of cutoff value in the sense that we obtain qualitatively similar, although slightly less significant results when we use a slightly lower cutoff value.}

15
## Representation in Homogenous and Heterogeneous Districts

<table>
<thead>
<tr>
<th>Variable</th>
<th>All</th>
<th>Low Var</th>
<th>High Var</th>
<th>Low Var</th>
<th>High Var</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.86</td>
<td>1.86</td>
<td>.26</td>
<td>.43</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.35)</td>
<td>(.19)</td>
<td>(.22)</td>
<td>(.07)</td>
</tr>
<tr>
<td>MeanPref</td>
<td>-.13</td>
<td>-.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.08)</td>
<td>(.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>House</td>
<td>-.19</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.08)</td>
<td>(.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senate</td>
<td>-1.11</td>
<td>-1.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party</td>
<td>1.78</td>
<td>-.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.59)</td>
<td>(.72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Ed. BA or more</td>
<td>-2.38</td>
<td>-.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.06)</td>
<td>(1.29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Inc $100k+</td>
<td>-0.48</td>
<td>-1.04</td>
<td>.44</td>
<td>.35</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.45)</td>
<td>(.14)</td>
<td>(.17)</td>
<td>(.10)</td>
</tr>
<tr>
<td>Constant</td>
<td>.36</td>
<td>.44</td>
<td>.04</td>
<td>.91</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>55</td>
<td>34</td>
<td>21</td>
<td>34</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 4: *Regression coefficients, dependent variable is legislator’s first dimension NOMINATE score. Standard errors in parentheses.*
in heterogeneous districts.

In columns 4 and 5, we expand the model to control for other factors that may affect the relationship between district preferences and legislator behavior. These include chamber specific effects (with Assembly as the excluded category), the legislator’s party (scored one for Democrats), the percent of households in which the head of household has at least a bachelor’s degree, and the percent of households with an annual income over $100,000 (as additional measures of heterogeneity). We see that the effects of district mean preference are robust to these controls; the coefficient on mean preference is positive and significant in homogenous districts and is insignificant in heterogeneous districts. We take the results in table 4 as preliminary evidence that legislators are less responsive to their district’s mean preference in heterogeneous districts.

The results in table 4 suggest that district heterogeneity mediates the relationship between district preferences and legislator behavior. In table 5, we report the results from an alternative specification of this relationship. Instead of separating the sample into low- and high-variance districts, we add an interaction term between district mean preference and variance. This allows us to better utilize the information from the full sample, and to analyze the full range of values on the variance variable.

For comparison, column 1 reports the baseline model, omitting the interaction term, as in table 4. In column 2, we add the interaction term. We see that the estimated effect of district mean preference remains positive and significant, while the interaction term is negative and significant. In addition, the magnitude of the preference coefficient increases dramatically. We interpret these estimates as indicating that once we control for the confounding effect of preference heterogeneity, the marginal effect of the district’s mean preference becomes a

---

Another possible source of variation in the degree of political representation across districts might be electoral competitiveness. Competitiveness is known to affect many aspects of political campaigns (see McArthur and Marks 1988, Bender 1991, Coates and Munger 1995, Kahn and Kenney 1999). It seems logical to suppose that representatives from competitive districts more closely reflect the preferences of their districts. While not the primary focus of our study, we investigated this possibility. Defining competitive districts to be those that saw winning margins of less than 10 percentage points in half or more of the elections contested between 1994 and 2000, we found limited support for the notion that competitiveness affects political representation. In some specifications, interactions between competitiveness and mean preference were statistically significant while in others they were not. Of particular importance to the current study, including controls for district competitiveness did nothing to upset the results presented.
Interactive Specification

<table>
<thead>
<tr>
<th>Variable</th>
<th>All</th>
<th>With Interaction</th>
<th>Full Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeanPref</td>
<td>.86</td>
<td>2.88</td>
<td>.83 (.36)</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.40)</td>
<td>(.36)</td>
</tr>
<tr>
<td>Mean*Var</td>
<td>-.94</td>
<td>-.32</td>
<td>(.15)</td>
</tr>
<tr>
<td></td>
<td>(.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>House</td>
<td>-.18</td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>Senate</td>
<td>-.16</td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>Party</td>
<td>-1.15</td>
<td>(.07)</td>
<td></td>
</tr>
<tr>
<td>% Ed-BA+</td>
<td></td>
<td>.62 (.94)</td>
<td></td>
</tr>
<tr>
<td>% Inc-100k+</td>
<td></td>
<td>-.113 (1.64)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-.48</td>
<td>.66 (.31)</td>
<td>.54 (.12)</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.36</td>
<td>.44 (.31)</td>
<td>.91 (.12)</td>
</tr>
<tr>
<td>N</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 5: Regression coefficients, dependent variable is legislator’s first dimension NOMINATE score. Standard errors in parentheses.
much better predictor of legislator behavior. The marginal effect of district preferences is much smaller when variance is high. In the third column, we add the same control variables as in table 4. We again see that the effects of district mean preference and the interaction term are robust to the addition of these other factors, although the magnitudes of both the preference and interaction coefficients are substantially reduced.

5 Representing Heterogeneous Districts

If legislators from heterogeneous districts do not represent their district’s median or mean position, who or what do they represent? In this section, we explore several possibilities. One possibility is that in our operationalization, what we refer to as heterogeneity is really something else. Specifically, it might be the case that legislators from districts that are heterogeneous on one dimension better represent the interests of their constituents on a different dimension. For example, it is possible that the first dimension NOMINATE score we use to operationalize legislator preferences in fact taps into a different set of issues or dimensions than the dimension or scale measured by the ballot proposition votes. Legislators may still represent their districts on issues that matter to them, but operationally these issues show up on a different dimension. Thus, we explore the possibility that we might be picking up different dimensions between the legislature and the electorate. We refer to this as the multidimensionality hypothesis.43

To test the multidimensionality hypothesis, we compare the relationship between a legislator’s second and third dimension NOMINATE scores and his or her district’s mean preference. Table 6 reports the results of this analysis. In the first and second columns, we report OLS regression estimates for the relationship between district mean preference and a legislator’s second dimension NOMINATE score in low- and high- variance districts, respectively. In columns three and four, we report comparable analyses where the dependent variable is the legislator’s third dimension NOMINATE score. Since it is unclear what issue dimensions these scores are tapping, we do not have specific hypotheses about the direction of these

\footnote{It is possible that this problem of multidimensionality would also show up in districts that are homogeneous on the first dimension. We allow for this possibility in our estimations, below.}
Test of Multidimensionality Hypothesis

<table>
<thead>
<tr>
<th>Variable</th>
<th>DV=NOM2 (Low Var)</th>
<th>DV=NOM2 (High Var)</th>
<th>DV=NOM3 (Low Var)</th>
<th>DV=NOM3 (High Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeanPref</td>
<td>-.12 (.19)</td>
<td>.07 (.10)</td>
<td>-.07 (.18)</td>
<td>-.12 (.09)</td>
</tr>
<tr>
<td>House</td>
<td>-.07 (.07)</td>
<td>.10 (.09)</td>
<td>.10 (.07)</td>
<td>.14 (.08)</td>
</tr>
<tr>
<td>Senate</td>
<td>.14 (.07)</td>
<td>-.02 (.09)</td>
<td>-.03 (.07)</td>
<td>.17 (.08)</td>
</tr>
<tr>
<td>Party</td>
<td>-.19 (.08)</td>
<td>-.15 (.15)</td>
<td>-.09 (.07)</td>
<td>-.13 (.13)</td>
</tr>
<tr>
<td>% Ed-BA+</td>
<td>-.67 (1.41)</td>
<td>-.17 (1.12)</td>
<td>-.94 (1.31)</td>
<td>.30 (1.00)</td>
</tr>
<tr>
<td>% Inc-100k+</td>
<td>1.75 (2.71)</td>
<td>-.56 (2.01)</td>
<td>.51 (2.53)</td>
<td>-2.13 (1.79)</td>
</tr>
<tr>
<td>Constant</td>
<td>.15 (.15)</td>
<td>-.06 (.16)</td>
<td>.14 (.14)</td>
<td>.20 (.14)</td>
</tr>
</tbody>
</table>

$R^2$ | .27 | -.06 | .05 | .40 |
N    | 34  | 21  | 34  | 21  |

Table 6: Regression coefficients, dependent variable in columns 1-2 is legislator’s second dimension NOMINATE score. Dependent variable in columns 3-4 is legislator’s third dimension NOMINATE score. Standard errors in parentheses.

effects. Therefore, we simply test whether the effect of mean preference is different in low- and high-variance districts.

Table 6 shows little support for the multidimensionality hypothesis. The relationship between district mean preference and a legislator’s second dimension NOMINATE score in high-variance districts is positive, but the effect is not statistically significant. In low-variance districts, it is negative and insignificant. The relationship between district mean preference and a legislator’s third dimension NOMINATE score in both high- and low-variance districts is negative but insignificant.

Given that our results do not appear to be an artifact of our operationalization, we next consider a number of behavioral explanations. We refer to the first as the avoidance/
abstention hypothesis. We hypothesize that to the extent that legislators from heterogeneous districts face more conflicted or ambiguous constraints from their districts, and seek to avoid issues that alienate important constituencies, they may also abstain more often.\textsuperscript{44}

Table 7 presents a test of the avoidance/abstention hypothesis with our data. In these analyses, we test whether legislators from heterogeneous districts abstain more often than legislators from homogenous districts. The dependent variable is the percent of votes on which the legislator abstained. The independent variables are measures of district heterogeneity. In the first column, we regress abstention rate on the district’s estimated variance. In the second column, we regress abstention rate on a dummy variable scored one if the district’s variance is above the mean variance across all districts. In columns three and four, we add control variables to the baseline specifications to account for other factors that might affect legislators’ abstention rates.

There is little support for the hypothesis that legislators from heterogeneous districts abstain rather than try to represent their diverse constituents in their votes. The effect of overall district variance is significant in the baseline specification, but insignificant once we control for other factors. The effect of the high variance dummy is significant in the baseline model and insignificant in the full model.\textsuperscript{45}

A second alternative behavioral hypothesis is that legislators represent not their overall district mean or median, but rather their “core” or re-election constituencies. Variations of this argument are advanced by Fiorina (1974), Fenno (1978), Peltzman (1984), Goff and Grier (1993), Swain (1993), Jung et al. (1994), Brady and Schwartz (1995), Levitt (1996), Stratmann (1996) and others. In the context of our analysis, we interpret this core constituency hypothesis as implying that legislators from both homogenous and heterogeneous districts are most responsive to their core constituencies. In homogenous districts, the mean and median preferences of this subgroup resemble the overall district mean and median quite closely, but in heterogeneous districts, the preferences of the core mean and median and the overall district mean and median may diverge substantially.

\textsuperscript{44}See Rothenberg and Sanders (2000) for a similar analysis.

\textsuperscript{45}Since our research design is quite different, we do not view this analysis as a critical test of Rothenberg and Sanders’ theory, but rather note the limits of their hypotheses to our specific application.
## Test of Avoidance Hypothesis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td>.04</td>
<td>.02</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>High Variance</td>
<td>.05</td>
<td>.03</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>House</td>
<td>-.03</td>
<td>-.03</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>Senate</td>
<td>.09</td>
<td>.09</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>Party</td>
<td>.04</td>
<td>.03</td>
<td>(.02)</td>
<td>(.02)</td>
</tr>
<tr>
<td>% Ed-BA+</td>
<td>.41</td>
<td>.36</td>
<td>(.29)</td>
<td>(.29)</td>
</tr>
<tr>
<td>% Inc-100k+</td>
<td>-.47</td>
<td>-.40</td>
<td>(.52)</td>
<td>(.52)</td>
</tr>
<tr>
<td>Constant</td>
<td>.002</td>
<td>-.02</td>
<td>-.02</td>
<td>.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.05</td>
<td>.08</td>
<td>.39</td>
<td>.40</td>
</tr>
<tr>
<td>N</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 7: Regression coefficients, dependent variable is percent of roll call votes on which legislator abstained. Standard errors in parentheses.
Because our voter preference data is based on direct individual-level measurements, we are able to estimate to the preferences of subgroups within a districts much more directly than has previously been possible. Where previous studies have relied on auxiliary aggregate-level regressions (Peltzman 1984) or proxies based on the roll-call voting of legislators in lower chamber (Levitt 1996), we are able to identify voters who belong to a member’s core constituency and use their proposition votes to estimate the distribution citizen preference within and among these more narrowly defined constituencies.

We operationalize a legislator’s core constituency as voters in his or her district from the same party. We construct a new variable, \emph{Partisan Mean}, as the mean preference of a district’s Democratic voters (as previously defined) for Democratic legislators and the mean preference of Republican voters for Republican legislators.46

Table 8 reports the results of several analyses that test the partisan constituency hypothesis. In the first column, we regress a legislator’s first dimension NOMINATE score on his or her partisan mean preference. In the second and third columns, we separate low- and high-variance districts, respectively. In the fourth and fifth columns, we control for the marginal effect of district overall mean. The partisan constituency hypothesis predicts that legislators in high-variance districts will more closely represent their district’s partisan mean than their district’s overall mean.

Table 8 shows strong support for the partisan constituency hypothesis. The bivariate relationship between a legislator’s policy position and his or her district’s partisan mean is strong, positive, and significant, as expected. In contrast to the results presented in table 4, where legislators from high variance districts did not represent their district’s overall mean, we see in column 3 that legislators from high-variance districts do much more closely represent their district’s partisan means. This relationship is also strong, positive, and significant, and is robust to the other standard controls. We note that legislators from low-

\footnote{46In our data, the relationship hypothesized above holds for Republican districts - the correlation between overall district mean and partisan mean is .37 in high variance districts and is .63 in low variance districts. In Democratic districts, however, the correlation between overall and partisan mean is extremely high for both high and low variance districts, and is, in fact, slightly higher in the high variance districts. We attribute this anomaly to the presence of several extremely liberal but high variance districts in which there are very few Republican voters, and hence in which the partisan mean and the overall district mean are virtually identical.}
## Test of the Partisan Constituency Hypothesis

<table>
<thead>
<tr>
<th>Variable</th>
<th>All</th>
<th>Low Var</th>
<th>High Var</th>
<th>Low Var</th>
<th>High Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>PartisanMean</td>
<td>.61</td>
<td>.78</td>
<td>.44</td>
<td>.77</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.06)</td>
<td>(.06)</td>
<td>(.07)</td>
<td>(.06)</td>
</tr>
<tr>
<td>MeanPref</td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td>-.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.26)</td>
<td>(.10)</td>
</tr>
<tr>
<td>House</td>
<td>-.13</td>
<td>-.31</td>
<td>-.13</td>
<td>-.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.09)</td>
<td>(.09)</td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>Senate</td>
<td>-.17</td>
<td>-.05</td>
<td>-.17</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.09)</td>
<td>(.09)</td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>% Ed-BA+</td>
<td>3.59</td>
<td>.22</td>
<td>3.77</td>
<td>-.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.63)</td>
<td>(1.14)</td>
<td>(1.80)</td>
<td>(.75)</td>
<td></td>
</tr>
<tr>
<td>% Inc-100k+</td>
<td>-4.62</td>
<td>-1.92</td>
<td>-5.04</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.05)</td>
<td>(2.00)</td>
<td>(3.46)</td>
<td>(1.36)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-.44</td>
<td>-.55</td>
<td>-.12</td>
<td>-.58</td>
<td>-.07</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td>(.09)</td>
<td>(.10)</td>
<td>(.15)</td>
<td>(.06)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.82</td>
<td>.89</td>
<td>.80</td>
<td>.89</td>
<td>.92</td>
</tr>
<tr>
<td>N</td>
<td>55</td>
<td>34</td>
<td>21</td>
<td>34</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 8: Regression coefficients, dependent variable is legislator’s first dimension NOMINATE score. Standard errors in parentheses.
variance districts also closely represent their partisan means. Finally, in column four, we see that even in low variance districts, most of the apparent effect of the district overall mean is picked up by the partisan mean variable, indicating that on the margin, legislators from low-variance districts more closely represent their partisan means than their overall means. This effect is more dramatic in high variance districts, where the inclusion of the partisan mean variable, whose effect is again positive, strong, and significant, not only picks up much of the apparent effect of the overall district mean, but even produces a reversal in the overall mean variable’s sign. The negative sign on district mean suggests that holding fixed the partisan mean locations, members become more extreme as their district mean locations move away from their party mean locations. Because, holding the partisan mean constant, the changes in the district mean must be the result of changes in the opposite party and independent means, the negative sign suggests that members choose more extreme positions when the two partisan constituencies are more polarized. This finding is consistent with models of electoral competition such as Aldrich (1983) that argue that electoral or activist polarization in the electorate leads to partisan polarization in the legislature.

6 Conclusion

To summarize, this paper considers the question of how district composition affects the relationship between citizen preferences and legislator behavior. This question is fundamental to our understanding of representation in modern American legislatures, since many districts are characterized by a high degree of preference heterogeneity. We employ a unique dataset to estimate the distribution of voter preferences within legislative districts. We find that district composition affects the relationship between citizen preferences and legislator behavior in a number of ways. District mean preference is a much better predictor of legislator behavior in homogenous districts than in heterogeneous districts. In other words, legislators take policy positions that are close to their district’s median when many constituents share these preferences. In heterogeneous districts, mean preference is a much less powerful predictor of legislative behavior. Legislators from heterogeneous districts often take policy positions that diverge substantially from their district’s mean or median. This result is robust to a
variety of model specifications and measures of preference heterogeneity. It suggests that our standard spatial voting models, with their predictions of convergence to the median, do not apply well to situations where legislators represent heterogeneous districts.

We then test several alternative hypotheses about how legislators behave in heterogeneous districts. We find very little support for the hypotheses that legislators represent their districts on other issues or dimensions, or that they engage in participatory shirking or abstention. However, we do find strong support for the hypothesis that legislators from all types of districts - homogenous and heterogeneous - closely represent their partisan constituencies. In homogenous districts, the overall district mean and the partisan mean are similar, by definition. In heterogeneous districts, by contrast, the overall district mean and the partisan mean may diverge substantially. Our data show that legislators pay close attention to like-partisans.

These results have important implications for the study of political representation. If legislators from heterogeneous districts represent the interests of only a subset of their constituents, then the people who live in those districts but who are not part of the legislator’s core constituency really lose out. In terms of having their preferences expressed in policy, they would be better off in different districts along with like-minded citizens and representatives. Some research has explored this question, both theoretically and empirically, in the context of racial redistricting (Lublin 1997, Epstein and O’Halloran 1999). We see our research as a generalization of the insights from that literature. In addition, our research has implications for the future study of legislative representation. To the extent that legislators fail to converge to their district’s mean or median position, standard spatial explanations offer little insight. Numerous studies, including many discussed in our Introduction, have considered some of the factors that lead legislators to diverge from their district’s median preference (or fail to converge), however most of these studies rely on non-spatial explanations. An important direction for future research involves specifying the conditions under which the standard spatial explanations of legislative behavior do and do not apply.
References


