Measuring Constituent Policy Preferences in Congress, State Legislatures and Cities[1]

Chris Tausanovitch
Department of Political Science
UCLA
ctausanovitch@ucla.edu

Christopher Warshaw
Department of Political Science
Massachusetts Institute of Technology
cwarshaw@mit.edu

August 16, 2012

Abstract:

Little is known about the American public's policy preferences at the level of Congressional districts, state legislative districts, and local municipalities. In this paper, we overcome the limited sample sizes that have hindered previous research by jointly scaling the policy preferences of 275,000 Americans based on their responses to policy questions. We combine this large dataset of Americans' policy preferences with recent advances in opinion estimation to estimate the preferences of every state, congressional district, state legislative district, and large city. We show that our estimates outperform previous measures of citizens' policy preferences. These new estimates enable scholars to examine representation at a variety of geographic levels. We demonstrate the utility of these estimates through applications of our measures to examine representation in state legislatures and city governments.

Keywords: representation, public opinion, state politics, urban politics, MRP modeling

A well-functioning democracy requires legislators to represent the will of their constituents. Despite this fact, political scientists still have only a limited understanding of the extent of constituency influence in Congress (Clinton 2006). Moreover, we know even less about the extent of constituency influence at other levels of government in the United States. Previous empirical work has been hindered by the fact that the sample size in national surveys is generally too small to make inferences about the preferences of individual geographic units. Even the largest national surveys have only about one hundred people in each congressional district. Making inferences below the congressional district level is usually even more difficult. Most surveys have just a handful of respondents in each state legislative district and municipality. As a result, scholars have been severely limited in their ability to evaluate state or city-level institutional factors that might mediate the link between citizens' preferences and political outcomes.

Scholars have adopted various techniques to cope with the sparse availability of data on citizens' policy preferences at lower levels of aggregation. Some scholars have disaggregated survey data to the district level (Miller and Stokes 1963; Clinton 2006). Other scholars have used district-level presidential vote as a proxy for district public opinion (e.g. Canes-Wrone, Cogan, and Brady 2002). Still others have employed demographics (Peltzman 1984) or simulation techniques (Ardoin and Garand 2003). All of these methods, however, have clear drawbacks. In particular, they are ill-suited to estimate the policy preferences of geographic sub-constituencies or the preferences of non-standard geographic units such as cities.

In this paper, we provide a new method to estimate the policy preferences of small geographic units. First, we use an original survey that allows us to jointly scale the

policy preferences of respondents to seven recent, large-scale national surveys using an item-response theory (IRT) model (Clinton, Jackman, and Rivers 2004; Shor and McCarty 2011). Our original survey serves as a mechanism to pool the other datasets, allowing a much larger dataset than was previously possible. This approach enables us to develop a continuous measure of the policy preferences of 275,000 citizens in all fifty states.

Next, we use this large national sample to estimate the average policy preferences of citizens in every state, congressional district, state legislative district, and large city in the country. We generate estimates of mean policy preferences using both simple disaggregation and multi-level regression with post-stratification (MRP). In general, we find that both approaches yield accurate estimates, even in small geographic units. Despite our very large sample size, however, we find that MRP almost always produces more accurate estimates of the mean policy preferences at each geographic level than disaggregation (Warshaw and Rodden 2012).

We also show how our large sample of citizens' policy preferences can be used to estimate the preferences of geographic sub-constituencies, such as partisan sub-constituencies in each congressional district. Finally, we move beyond estimates of mean district preferences to examine other quantiles of the distribution of preferences. For instance, it is straightforward to use our approach to estimate the ideological heterogeneity of citizens' policy preferences in each geographic unit.

These new estimates of citizens' policy preferences can be used to address a variety of substantive questions on representation. In this paper, we demonstrate two applications to the study of representation at lower levels of government that are under-

studied in political science: state politics and urban politics. These literatures are important in their own right, but studying lower levels of government is also vital for understanding representation more broadly, because it allows us to examine the role of institutional moderating factors on representation (Lax and Phillips 2009).

The paper proceeds as follows. First, we discuss our approach to conceptualizing policy preferences. Next, we discuss the datasets we use for our analysis. Then we describe our methodology for jointly scaling respondents from multiple contexts, and using this data to estimate the mean policy preferences of citizens at a variety of geographic levels. Next, we validate and describe our estimates. Then, we present applications of our measures to examine representation in state legislatures and city governments. Finally, we briefly conclude with some suggestions for future research.

HOW HAVE PAST SCHOLARS MEASURED SUB-STATE LEVEL POLICY PREFERENCES?

Previous scholars have used a variety of approaches to measure citizens' policy preferences at the state level. The most straightforward approach is to use data from a representative survey that asks respondents for their preferences on individual issues (Erikson, Wright, and Mclver 1993). But national surveys generally do not have enough respondents to develop accurate estimates at the sub-state level (Erikson 1978).

Other scholars have used election returns to estimate district preferences (e.g., Canes-Wrone, Cogan and Brady 2002; Erikson and Wright 1980). The advantage of this approach is that it is explicitly based on electoral behavior, it is available across all states and districts, and it is updated frequently (Kernell 2009). However, presidential vote shares in any given election may be largely the product of short-term forces (Levendusky,

Pope, and Jackman 2008). In addition, even if short-term forces could be removed, the medians of district preferences can only be ranked ordinally based on presidential vote share if researchers are willing to assume equal variance across districts (Kernell 2009). Finally, it is impossible to measure the preferences of district sub-constituencies (e.g., the preferences of Democrats or Latinos) using presidential vote shares.

The most recent development is the use of Bayesian approaches to measure district-level policy preferences (e.g., Levendusky, Pope, and Jackman 2008). Several scholars have used multi-level regression and post-stratification (MRP) to estimate state and district-level public opinion on individual issues using survey data (Park, Gelman, and Bafumi 2004; Lax and Phillips 2009). This approach builds on earlier simulation approaches (e.g., Ardoin and Garand 2003). It employs Bayesian statistics and multi-level modeling to incorporate information about respondents' demographics and geography to estimate the preferences of each geographic sub-unit even if survey samples are small. Warshaw and Rodden (2012) show that MRP produces more accurate estimates of district-level public opinion on individual issues than either disaggregation of national surveys or presidential vote shares.

In this paper, we use item response theory estimation and a dataset of surveys linked by our "super survey" to estimate of citizens' policy preferences at a variety of geographic levels. We further improve our estimates by applying MRP to incorporate information about respondents' demographics and geography into our model

IDEAL POINTS AND PUBLIC OPINION

We assume that both citizens and legislators have a unique set of policies that they "prefer" to all others.[2] This point in the policy space is called an "ideal point." Scholars studying the United States Congress have long recognized the utility of thinking about legislators' preferences in terms of ideal points derived from a spatial model of choice (e.g., Poole and Rosenthal 2000; Clinton, Jackman, and Rivers 2004). An ideal point is a convenient summary of how far to the "left" or the "right" a person's policy preferences are on each policy dimension. We assume that on any given dimension, people prefer policies that are closer to their ideal point over policies that are farther away. Ideal points are latent traits because we cannot observe them; we can only estimate them based on the observed policy choices of each person.

Ideal points are only defined relative to the particular choices that are used to estimate the policy space (Bafumi and Herron 2010). For example, ideal point estimates for members of the Senate are not generally comparable with ideal point estimates for members of the U.S. House since each chamber votes on a different set of roll calls. This limitation prevents us from directly comparing ideal points between groups responding to disjoint sets of choices. In order to pool data over multiple surveys, we need ideal point estimates for respondents to each survey that reside in a common policy space. We address this problem by using common questions asked to different sets of people to bridge the ideal points of survey respondents into a common space. In the future, we could jointly-scale the spatial positions of legislators and other institutions into this common scale (Bafumi and Herron 2010; Shor and McCarty 2011).

DATA

Ideal point estimation typically draws on responses to individual-level, binary choices. We use seven recent large-scale surveys of the American public (the 2006, 2007, 2008, 2010, and 2011 Cooperative Congressional Election Surveys (CCES) and the 2000 and 2004 Annenberg National Election Surveys (NAES)). Each of these surveys asked between 14 and 32 policy questions to 30,000-80,000 Americans.[3] Combined, these surveys include more than 275,000 respondents (Table 1). We gain additional information about respondents' policy preferences using modules we placed on the 2010 and 2011 Cooperative Congressional Election Surveys. Most of the survey questions used here are binary, but when they are not, we dichotomize them by separating responses into two ordered categories.

The key to our research design is bridging respondents in a way that allows us to generate common space ideal point estimates (see Bailey 2007 and Bafumi and Herron 2010).[4] As mentioned earlier, we cannot directly compare ideal points if they are estimated using disjoint sets of people answering disjoint sets of choices. However, we can estimate comparable ideal points if there is a sufficient overlapping set of choices and/or people to bridge individuals into a common policy space.

We link together survey respondents using the module we placed on the 2010 Cooperative Congressional Election Survey. In this module, we asked 1,300 survey respondents a large number of questions with wording identical to questions asked on previous CCES and NAES surveys. These common questions allow us to place respondents from all seven large-sample surveys on a common scale. Our module is a superset of all of the questions on the other surveys, hence the name "super survey." We

also asked a large set of additional policy questions, which enables us to estimate more precise ideal points than was possible with the smaller sets of questions on earlier surveys.

STATISTICAL MODEL FOR CITIZENS' POLICY PREFERENCES

We combine observed survey responses from all of our surveys. This yields a set of millions of unique choices. The number of rows of our so-called "roll call matrix" corresponds to the number of respondents, and the number of columns corresponds to the number of survey questions.

We assume that all survey respondents have a quadratic utility function with normal errors (Clinton, Jackman Rivers 2004; Treier and Hillygus 2009). Each item j presents individuals i with a choice between a "Yes" position and a "No" position. Let $y_{i,j}$ =1 if individual i votes yes on the jth roll call and $y_{i,j}$ =0 otherwise. We assume a 1-dimensional policy space, where $x_i \in R$ is the ideal point of respondent *i*. We choose a one-dimensional model because a two-dimensional model shows little improvement in terms of model fit.[5]

We estimate the ideal points using a Bayesian Item-Response (IRT) model (Clinton, Jackman Rivers 2004; Jessee 2009). Let $i = 1, \ldots, n$ index individuals and $j = 1, \ldots, m$ index items. Then our model is

(1) $\Pr(y_{ij} = 1) = \Phi(u_{ij} - \alpha_j)$, where

(2) $u_{ij} = x_i \beta_j$,

$y_{i,j}$ is the *i*-th respondent's answer to question *j*, x_i is the ideal point for respondent *i*, β_j is the "discrimination" parameter for item *j*, α_j is the "difficulty" parameter for item *j*, and $\Phi(\bullet)$ denotes the standard normal cumulative distribution function.

There are three parameters in Equations (1) and (2). The ideal point x_i for individual i signifies the "liberalness" or "conservativeness" of that individual. We orient our x_i values so that lower values are associated with politically left preferences and higher values with politically right preferences. Ideal point estimates lack an absolute alignment. We resolve this problem by normalizing them. The discrimination parameter β_j reveals how well an item discriminates between liberals and conservatives. The difficulty parameter on issue j, α_j, is related to how liberal or conservative a person must be in order to be indifferent toward agreeing or disagreeing with the item.

We assume that a question asked on our CCES module is no different than a question asked on the source surveys. At first this may not seem like an assumption at all. After all, the questions are exactly the same. However, the context of the questions may be different (e.g., the status quo may have changed on particular items). In order to be conservative we only "bridge" questions that have similar margins across surveys and time. Although the composition of districts changes over time, we see little difference over the period in which these surveys are pooled. However, our estimates should be interpreted as an average of the positions of the geographic areas in question between 2000 and 2011.

We approximate the joint posterior density of the model parameters using a Markov chain Monte Carlo (MCMC) method (Clinton, Jackman, and Rivers 2004). We use diffuse normal priors for the discrimination parameters, β_j, and the ability parameters, α_j, with mean 0 and variance 25. We specify normal priors with mean 0 and variance 25 for each x_i. To make the estimation manageable, we use software that does parallel

draws of the Gibbs sampler using graphics processing units (Lewis, Lo, and Tausanovitch 2011). This reduces our computing time by a factor of 15-20.

MEASURING THE POLICY PREFERENCES OF STATES, CONGRESSIONAL DISTRICTS, STATE LEGISLATIVE DISTRICTS, AND CITIES

In this section, we describe how we estimate the mean policy preferences of citizens in each constituency, the preferences of sub-constituencies such as Democrats and Republicans, and the average heterogeneity of each constituency.

Estimating the Mean Preferences of Each Geographic Constituency

The most straightforward way to use our large sample of citizens' policy preferences to estimate citizens' preferences at a variety of geographic levels is to estimate the simple "disaggregated" mean of each state, city, and legislative district (Erikson, Wright, and Mclver 1993). Our large sample size lends itself well to applying disaggregation since we have an average of over 5,000 respondents in each state, 500 respondents in each congressional district, and 100 respondents in each city with more than 25,000 people.[6]

An alternative strategy introduced by Park, Gelman, and Bafumi (2004) and Lax and Phillips (2009b) is to estimate district-level public opinion using multilevel regression and poststratification (MRP). Pairing this technique with our very large dataset of citizens' policy preferences may yield even greater accuracy. MRP models incorporate information about respondents' demographics and geography in order to estimate the public opinion of each geographic subunit (see Gelman and Hill 2007 and Jackman 2009 for more about multilevel modeling). Specifically, each individual's survey responses are modeled as a function of demographic and geographic predictors, partially pooling

respondents across districts to an extent determined by the data. Thus, all individuals in the survey yield information about demographic and geographic patterns, which can be applied to all district estimates. Several recent studies have found that MRP models yield accurate estimates of public opinion in states and congressional districts using national samples of just a few thousand respondents (Park, Gelman, and Bafumi 2004; Lax and Phillips 2009b; Warshaw and Rodden 2012).

To estimate the policy preferences of citizens in each state, congressional district, state legislative district, and city, we use an MRP model similar to the one in Warshaw and Rodden (2012). In the first stage of the model, we estimate each individual's policy preferences as a function of his or her demographics and geographic location. We assume that the "geographic effects" in the model are a function of a vector of demographic factors that previous studies have found to influence constituency preferences. For instance, the congressional district effects are modeled as a function of the state into which the district falls, the district's average income, the percent of the district's residents that live in urban areas, the percentage of the district's residents that are military veterans, and the percentage of couples in each district that are in same-sex couples.[7] The state effects are modeled as a function of the region into which the state falls, the percentage of the state's residents that are union members, and the state's percentage of evangelical or Mormon residents. The second stage is post-stratification. In this stage, we use the multi-level regression to make a prediction of public opinion in each demographic-geographic sub-type. The estimates for each respondent demographic geographic type are then weighted by the percentages of each type in the actual district populations.[8] Finally, these predictions are summed to produce an estimate of public opinion in each district.

Estimating the Policy Preferences of Geographic Sub-Constituencies

Many questions of representation concern the relative weight that elected officials attach to the preferences of various sub-constituencies (Fenno 1978). Our large sample enables us to go beyond estimates of the average preferences in each state or district to estimate the preferences of various types of sub-constituencies in each district. These estimates could enable scholars to better address a variety of substantive questions on representation. For instance, scholars could use our estimates of the preferences of partisan sub-constituencies to examine whether legislators are differentially responsive to citizens in their own party (Clinton 2006). In Online Appendix A, we use simple disaggregation and MRP to generate estimates of the mean Democrat and Republican in each state. We validate our measures against partisan sub-constituencies' voting behavior and self-identified ideology in recent exit polls.

Estimating Quantiles Beyond the Mean

Our large dataset of American' policy preferences also provides sufficient sample size and granularity to move beyond estimates of median district preferences to examine other quantiles of the distribution of preferences. For instance, a frequent hypothesis about the distribution of district preferences is that greater heterogeneity in district preferences should weaken the link between the median voter and representatives (Bailey and Brady 1998, Ensley 2010; Gerber and Lewis 2004). The literature on polarization and electoral constituencies also emphasizes the role that ideological extremists play in candidate reelection. This suggests that there are other quantiles and summary statistics that will be of theoretical interest to future work in representation, and that our method will allow empirical investigation of these quantities. In Online Appendix B, we use our

large sample to estimate the heterogeneity of citizens in each state and congressional district and we validate these estimates by comparing them to other recent estimates of heterogeneity in the electorate.

VALIDATION AND DESCRIPTIVE RESULTS

How well do our measures of citizens' policy preferences perform? Figure 1 compares the correlations of mean disaggregated policy preferences and MRP policy preferences with 2008 presidential vote shares at the level of states, congressional districts, state senate districts, state house districts, and cities. As we noted above, presidential vote shares are not a perfect measure of citizens' policy preferences. But a high correlation with presidential vote shares would suggest our estimates are accurate measures of citizens' policy preferences. Moreover, it is useful to compare the correlations of disaggregated and MRP estimates of policy preferences with presidential vote shares to evaluate which one performs better.

Figure 1 about here

Figure 1 compares three different measures of citizens' policy preferences at a variety of geographic levels:

- Disaggregated estimates of citizens' policy preferences generated using the 2006 CCES, with a sample of approximately 36,000 Americans.
- Disaggregated estimates of citizens' policy preferences generated using our large sample of the policy preferences of 275,000 Americans.
- MRP estimates of citizens' policy preferences generated using our sample of the policy preferences of 275,000 Americans.

Most importantly, Figure 1 demonstrates the value of our large sample of Americans' policy preferences compared to smaller datasets. Both the disaggregated and MRP measures estimated using our super-sample dramatically outperform estimates

generated using the 2006 CCES. The differences are particularly large at lower levels of aggregation. For instance, disaggregated estimates of the preferences of state senate districts from the 2006 CCES are only correlated with presidential vote share at about .46, compared with .77 for disaggregated estimates from our pooled dataset.

For larger geographic units, the MRP and disaggregated estimates of citizens' policy preferences are roughly equivalent. At the state-level, the MRP and disaggregated estimates of citizens' policy preferences are both highly correlated with presidential vote share. At the congressional district-level, the MRP estimates are correlated with presidential vote shares at .92, compared with .90 for our disaggregated estimates.

The MRP estimates substantially outperform disaggregation, however, at lower levels of aggregation.[9] Despite our very large sample size, the disaggregated estimates of the policy preferences of state senate district are correlated with 2008 presidential vote shares at just 0.77, compared with 0.88 for the MRP estimates. For state house districts, the gap is even larger. The disaggregated estimates are correlated with 2008 presidential vote shares at 0.64, compared with 0.85 for the MRP estimates. Finally, disaggregated estimates of the preferences of cities with more than 25,000 people are correlated with presidential vote shares at about 0.66, compared with 0.76 for the MRP estimates.

Even though our disaggregated estimates are based on a very large sample of 275,000 Americans, our MRP estimates outperform disaggregation in all geographic units smaller than states. These results suggest that MRP estimates of citizens' mean policy preferences should almost always be preferred to simpler disaggregated estimates (Warshaw and Rodden 2012). In the remainder of this paper, we use our MRP estimates for all analyses.

Figure 2 about here

Figure 2 shows the policy preferences of voters by state. It shows that our estimates of policy preferences vary sensibly across geographic units. Idaho, Oklahoma, and Utah are the most conservative states; Washington DC, New York, Vermont, and Massachusetts are the most liberal states. The figure also shows that precision of each method is proportional to the size of the state. In large states such as California, both MRP and disaggregation yield very similar estimates. In smaller states, the MRP estimates are partially pooled toward the national distribution. Nonetheless, the MRP estimates are slightly more precise than the disaggregated estimates. Finally, the figure shows that our estimates are precise enough that the preferences of different states can generally be distinguished from one another.

Figure 3 about here

Figures 3 illustrates our estimates for cities. It shows the policy preferences of citizens in 34 cities in Texas with more than 50,000 people. Once again, the estimates vary sensibly across cities. Our estimates suggest that Austin is the most liberal city in Texas, while Amarillo is the most conservative city. These estimates are highly correlated with the 2008 presidential vote share: in Austin, President Obama received 71% of the vote, while he received just 26% of the vote in Amarillo.

On our website, we provide our full results for the mean disaggregated and MRP ideology in every state, congressional district, state legislative district, and large city, as well the standard errors of each estimate. We also provide estimates of the preferences of Democratic and Republican sub-constituencies in each state and congressional district (Online Appendix A), as well as the ideological heterogeneity of each state and

congressional district (Online Appendix B).

APPLICATIONS

In this section, we demonstrate two applications of our estimates of constituent preferences. In each application, we use the super-survey to estimate ideological preferences at a different level of geographic aggregation. The variation in the applications illustrates the range of substantive questions that our new dataset could help scholars answer.

Representation in State Legislatures

In the past, state politics scholars have been hindered by the unavailability of data on policy preferences at the level of state legislative districts. As a result, most studies of representation have focused on the U.S. Congress. This focus on Congress has hindered scholars' ability to study institutional factors that affect representation (Wright, Osborn, Winburn, and Clark 2009). For instance, it remains unclear how factors such as term limits and initiatives affect state legislators' responsiveness to public opinion.

The relatively few studies that have focused on representation in state legislatures have used various proxies for the preferences of state legislative constituents. Some studies have used data on demographics to estimate the preferences of district constituencies (Hogan 2008). But the relationship between demographics and ideology is generally weak and heterogeneous across states (Erikson, Wright, and McIver 1993). Other studies have used the distribution of presidential vote shares as a proxy for state legislative districts' ideological preferences (Shor and McCarty 2011; Wright, Osborn, Winburn, and Clark 2009; Shor 2010). However, this measure is vulnerable to home-

state effects, regional biases, and heterogeneity in the relationship between policy preferences and presidential vote shares across districts. In addition, presidential vote data is difficult to collect at the state legislative level.

Figure 4 about here

Our approach enables us to provide a direct estimate of citizens' policy preferences in each state legislative district. Figure 4 illustrates an application of our estimates to representation in state legislatures. This figure shows the relationship between district ideologies and state house members' ideal points in Pennsylvania, California, Wisconsin, and Texas.[10] We find a large and statistically significant relationship between district policy preferences and roll call voting in all four states: legislators in more liberal districts tend to have more liberal ideal points. However, even conditional on the policy preferences of a district, Democrats tend to have much more liberal voting patterns than Republicans. State legislatures appear to resemble Congress, where scholars have found significant splits between Democrats and Republicans after accounting for the policy preferences of their constituencies. These results also illustrate that within-party representation varies across states. In some states, such as Wisconsin and Texas, there is little relationship between the preferences of a district and legislators' ideal points within parties. In other states, the ideal points of Democrats and Republican appear to be vary significantly within parties due to the preferences of their districts.

Representation in City Governments

One of the most important questions in the study of local politics is whether city governments respond to the will of their citizens. For instance, do more liberal cities have more progressive tax regimes, or higher per capita spending rates? There is significant

evidence that policy outcomes at the state (Lax and Phillips 2009; Erikson, Wright, and McIver 1993) and national levels (Stimson, MacKuen, and Erikson 1995) are highly responsive to citizens' preferences. At the city-level, however, the lack of data on public opinion has stymied research on the link between citizens' preferences and salient policy outcomes. As Trounstine (2010) puts it, "in order to explain how well and under what conditions city policy reflects constituent preferences, we need … some knowledge of different constituents' preferences." But, "[b]ecause we lack survey data on local public opinion, we lack a sense of the underlying distribution of interests at the local level..."

Scholars have used two approaches to overcome the unavailability of public opinion data. As in the state politics and Congress literatures, some scholars have used demographic information as a proxy for preferences (Trounstine 2010). But the weak link between demographics and public opinion applies as much in the city context as it does in other contexts (Erikson, Wright, and McIver 1993). Other scholars have focused on a small number of urban areas with large survey samples (e.g., Palus 2010). However, there is no reason to believe that the link between public opinion and policy outputs in these large cities is similar to other types of cities.

To address these problems, we estimate the policy preferences of citizens in 1,502 cities with more than 25,000 people. We find significant variation in the policy preferences of cities. Not surprisingly, we find that San Francisco, Berkeley, and Cambridge are three of the most liberal cities in the country. Mesa AZ, Provo UT and Waco TX are three of the most conservative cities.

Figure 5 about here

These new estimates enable scholars to re-examine the link between public opinion and city policy outcomes. For instance, one important policy decision made by city governments is the choice regarding whether to institute progressive or regressive tax regimes. Sales taxes are one of the most regressive sources of tax revenues. Thus, one way to measure the progressivity of a city's tax revenues is to examine the percentage of its revenue that come from sales taxes (Gerber and Hopkins 2011). In Figure 5, we present a simple scatterplot of the relationship between city policy preferences and the percentage revenues that come from sales taxes (in the states that allow municipalities to collect sales taxes). In general, we find that conservative cities obtain significantly more revenues from sales taxes than liberal cities.[11] Thus, the linkage between public opinion and policy outputs at the city level appears to mirror the link between public opinion and policy outputs at the state and federal levels. This analysis could easily be extended to other policy areas, and it could incorporate the effect of elections (Gerber and Hopkins 2011) and other institutional and economic factors (Hajnal and Trounstine 2010).

CONCLUSION

This article addresses a crucial question in the study of Congress, state politics, public opinion, and political geography: How should we measure policy preferences at the sub-national level? Even the largest national surveys lack sufficient statistical power to estimate citizens' preferences at the level of congressional districts, let alone cities, state legislative districts, and other small geographic units. As a result, scholars have relied on a variety of proxies for public opinion, all of which have serious flaws. In this paper, we have developed a new survey-based estimate of the public opinion of 275,000

people. This new measure enables scholars to estimate citizens' policy preferences at a variety of levels of geographic aggregation where public opinion estimates were previously unavailable.[12]

In this paper, we have described two illustrative applications. First, we show how our data will enable scholars to examine the link between public opinion and state legislative representation. We estimate the policy preferences of every state legislative district in the country, and use these estimates to examine the link between public opinion and roll call voting in four state houses. We find a moderate link between public opinion and roll call voting patterns in these states, even after controlling for legislators' partisanship. Our estimates of the policy preferences of state senate districts could be combined with recent estimates of legislators' ideal points in all 50 states (McCarty and Shor 2010) to examine how institutional factors affects the link between voters preferences and legislators' voting behavior.

Second, we show how our data will enable scholars to examine the link between public opinion and city policy outcomes. In the past, scholars of urban politics have estimated public opinion using city demographics or they have limited their analysis to a small number of cities with large survey samples. However, our estimates enable scholars to examine representation in over 1,500 cities with a population of more than 25,000 people. We find a strong link between public opinion and city taxation regimes. Conservative cities have much more regressive tax regimes than liberal cities.

Beyond our substantive findings on representation we have shown the vast potential our new estimates of constituent preferences provide for testing different hypotheses about the mapping of public preferences into legislative action. For instance,

our sample can be jointly scaled to include legislators or candidates who responded to the National Political Awareness Test survey. We can also extend our analysis to develop multidimensional models to examine whether preferences that don't fit into the most prominent political cleavage are important in legislative actions and electoral contexts (Bailey and Brady 1999).

The study of elections, representation and policy preferences has historically been driven by large data sets. Even as great progress has been made, there is a clear need for this empirical project to get even bigger, that even in a time when surveys of 50,000 people and more are available, even more data is needed. We solve this need for more data by using a tried and true workaround, which has advanced science and social science many times before: the pooling of data. Our innovation is to create a new data set "merely" by pooling others. One of the great virtues of this innovation is that it is scalable: we now have the ability to pool even more data as it becomes available.

Although the advance introduced here is one of data and measurement, it is as much substantive as methodological. Better measurement enables us to ask new questions, while giving better answers to old ones. We are better equipped to answer the question Miller and Stokes asked in 1963: what is the extent of constituency influence in Congress? At the same time, we can ask this question about every level of government, furthering our ability to understand the factors that improve democratic representation.

References

Ansolabehere, Stephen, Jonathan Rodden & James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2): 215–232.

Ansolabehere, Stephen and Jonathan Rodden. 2012. Harvard Election Data Archive.

Ardoin, Phillip and James C. Garand. 2003. "Measuring Constituency Policy preferences in U.S. House Districts: A Top-Down Simulation Approach." *Journal of Politics*. 65(4): 1165-1189.

Bafumi, Joseph & Michael Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and their Members in Congress." *American Political Science Review* 104(3): 519-542.

Bailey, Michael A. "Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency." 2007. *American Journal of Political Science* 51(3): 433-448.

Bailey, Michael A. & David W. Brady. 1998. "Heterogeneity and Representation: The Senate and Free Trade." *American Journal of Political Science* 42(2): 524-544.

Canes-Wrone, Brandice, David W. Brady & John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1): 127–140.

Clinton, Joshua D. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *Journal of Politics* 68(2): 397-409.

Clinton, Joshua D., Simon Jackman & Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355–370.

Cooperative Congressional Election Study. 2006, 2008, 2010.

Ensley, Michael. 2010. Incumbent positioning, ideological heterogeneity and mobilization in U.S. House elections. *Public Choice*.

Erikson, Robert S. 1978. "Constituency Opinion and Congressional Behavior: A Reexamination of the Miller-Stokes Representation Data." *American Journal of Political Science* 22(3): 511-535.

Erikson, Robert and Gerald C. Wright. 1980. "Policy Representation of Constituency Interests." *Political Behavior* 2(1): 91–106.

Erikson, Robert S., Gerald C. Wright, and John P. McIver. 1993. *Statehouse Democracy: Public Opinion in the American States*. Cambridge University Press.

Fenno, Richard F. 1978. Home *Style: House Members in Their Districts*. Boston: Little, Brown and Company.

Gerber, Elisabeth R. & Jeffrey B. Lewis. 2004. "Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation." *Journal of Political Economy* 112(6): 1364-1383.

Gerber, Elizabeth and Daniel Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55(2): 326–339.

Hajnal, Zoltan L. and Jessica Trounstine. 2010. "Who or What Governs?: The Effects of Economics, Politics, Institutions, and Needs on Local Spending." *American Politics Research* 38(6): 1130-1163.

Hillygus, Sunshine D. & Shawn Treier. 2009. "The Nature of Political Policy preferences in the Contemporary Electorate." *Public Opinion Quarterly* 73(4): 679-703.

Hogan, Robert E. 2008. “Policy Responsiveness and Incumbent Reelection in State Legislatures.” *American Journal of Political Science* 52(4): 858-873.

Jackman, Simon. 2009. Bayesian Analysis for the Social Sciences. West Sussex, UK: John Wiley & Sons Ltd.

Jessee, Stephen A. 2009. “Spatial Voting in the 2004 Presidential Election.” *American Political Science Review* 103(1): 59-81.

Kernell,Georgia. 2009. “Giving Order to Districts: Estimating Voter Distributions with National Election Returns.” *Political Analysis* 17(3): 215-235.

Lax, Jeffrey, and Justin Phillips. 2009. “Gay Rights in the States: Public Opinion and Policy Responsiveness.” *American Political Science Review* 103(3): 367-386.

Lax, Jeffrey, and Justin Phillips. 2009b. “How Should We Estimate Public Opinion in the States?” *American Journal of Political Science* 53 (1): 197–21.

Levendusky, Matthew S., & Jeremy C. Pope. 2010. “Measuring Aggregate-Level Ideological Heterogeneity.” *Legislative Studies Quarterly* 35(2): 259-282.

Levendusky, Matthew S., Jeremy C. Pope and Simon Jackman. 2008. “Measuring District Level Preferences with Implications for the Analysis of U.S. Elections.” *Journal of Politics* 70(3): 736-53.

Lewis, Jeffrey, James Lo and Chris Tausanovitch. 2011. “Parallelizing Item Response Theory Estimation with CUDA.” Unpublished manuscript..

Miller, Warren E. & Donald E. Stokes. 1963. “Constituency Influence in Congress.” *The American Political Science Review* 57(1): 45–56.

National Annenberg Election Survey. 2000, 2004.

Palus, Christine Kelleher. 2010. “Responsiveness in American Local Governments.”

State and Local Government Review 42(2): 133-150.

Park, David, Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–85.

Peltzman, Sam. 1984. ''Constituent Interest and Congressional Voting.'' *Journal of Law and Economics* 27: 181--210.

Poole, Keith T. & Howard Rosenthal. 2000. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press US.

Shor, Boris & Nolan McCarty. 2011."The Ideological Mapping of American Legislatures." *The American Political Science Review* 105(3): 530-551.

Shor, Boris. 2010. "All Together Now: Putting Congress, State Legislatures, and Individuals in a Common Ideological Space to Assess Representation at the Macro and Micro Levels." Working Paper.

Stimson, James A., Michael B. Mackuen & Robert S. Erikson. 1995. "Dynamic Representation." *The American Political Science Review* 89(3): 543–565.

Trounstine, Jessica. 2010. "Representation and Accountability in Cities." *Annual Review of Political Science* 13: 407-23

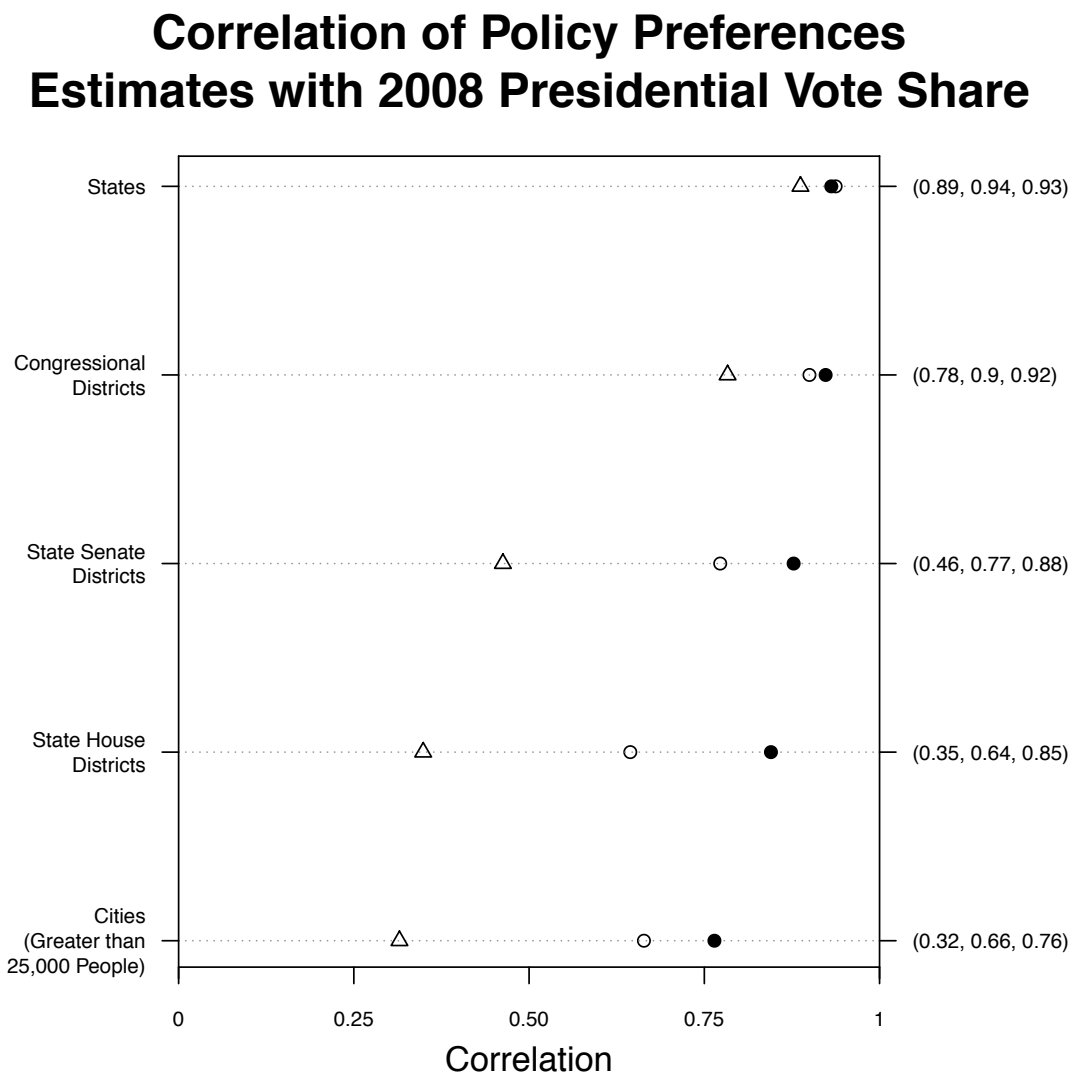
Warshaw, Christopher & Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74(1): 203-219.

Wright, Gerald, Tracy Osborn, Jon Winburn, and Jennifer Hayes Clark. 2009. "Patterns of Representation in the American States and Congress." Working Paper.

Table 1: Data Sources for Super-Survey

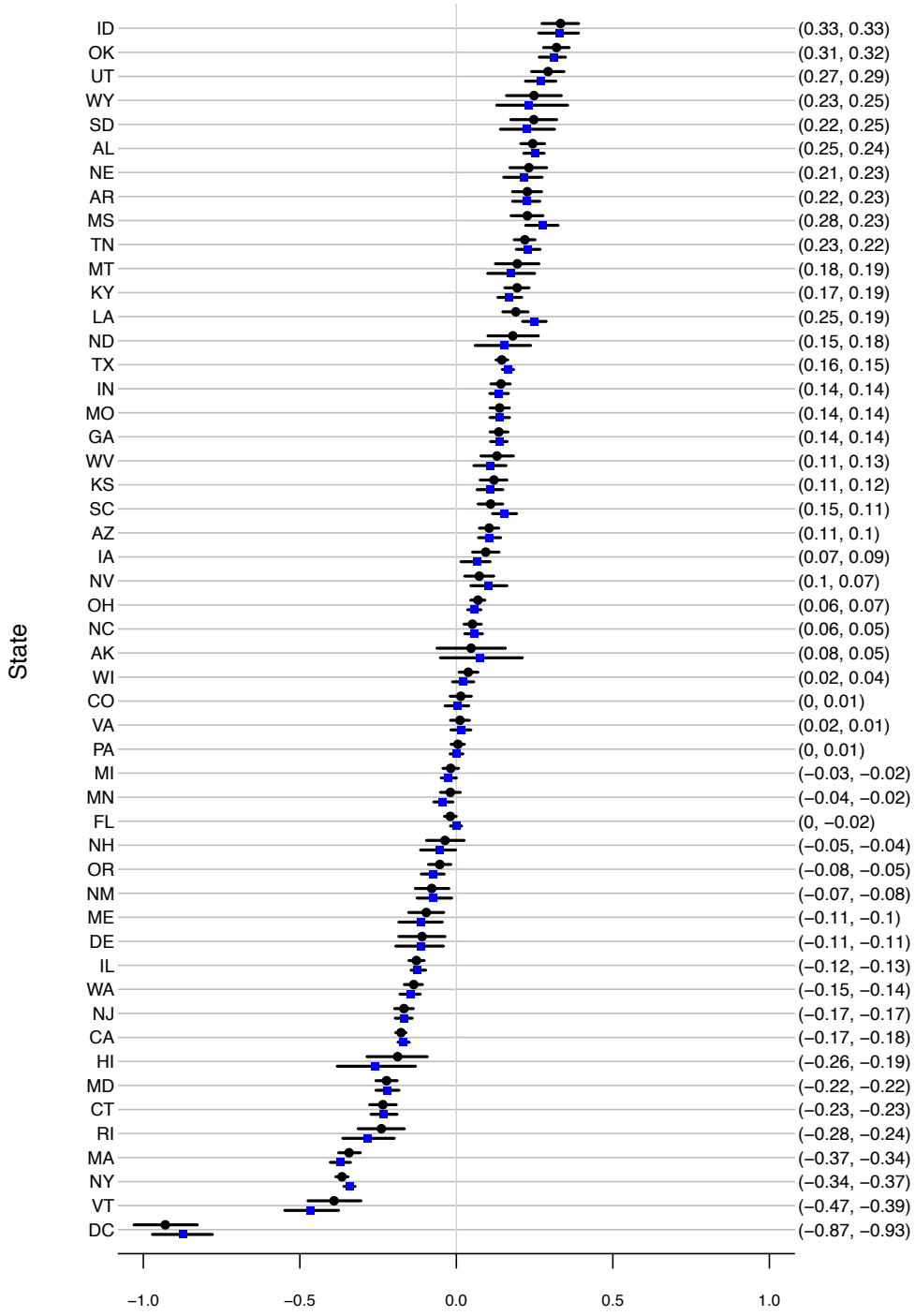
Survey	Respondents	Policy Items
2010 CCES module	1,300	136
2011 CCES module	2,500	41
2006 CCES	30,000	16
2007 CCES	10,150	14
2008 CCES	32,800	15
2010 CCES (Common Content)	55,000	22
2011 CCES (Common Content)	20,000	14
2000 NAES	58,400	28
2004 NAES	81,400	25

Figure 1: Validation of Our Policy Preferences Estimates



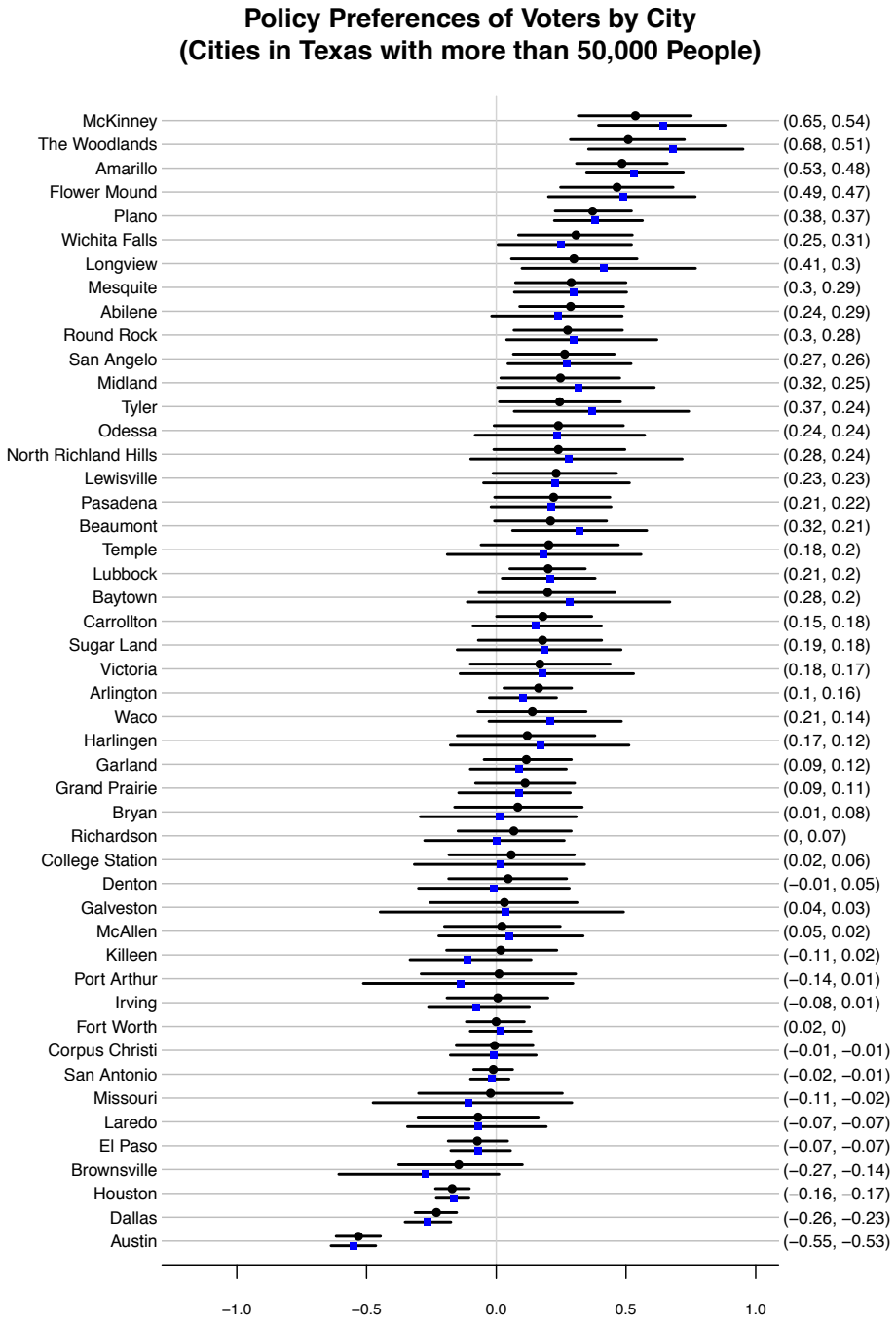
This figure shows the tight relationship between our estimates of citizens' policy preferences at each geographic level and 2008 presidential vote shares. The hollow triangles show the correlation with disaggregated estimates of citizens' policy preferences using the 2006 CCES, the hollow dots show the correlation with disaggregated estimates of citizens' policy preferences using our super-survey, and the black dots show the correlation with MRP estimates of citizens' policy preferences using our super-survey.

Figure 2



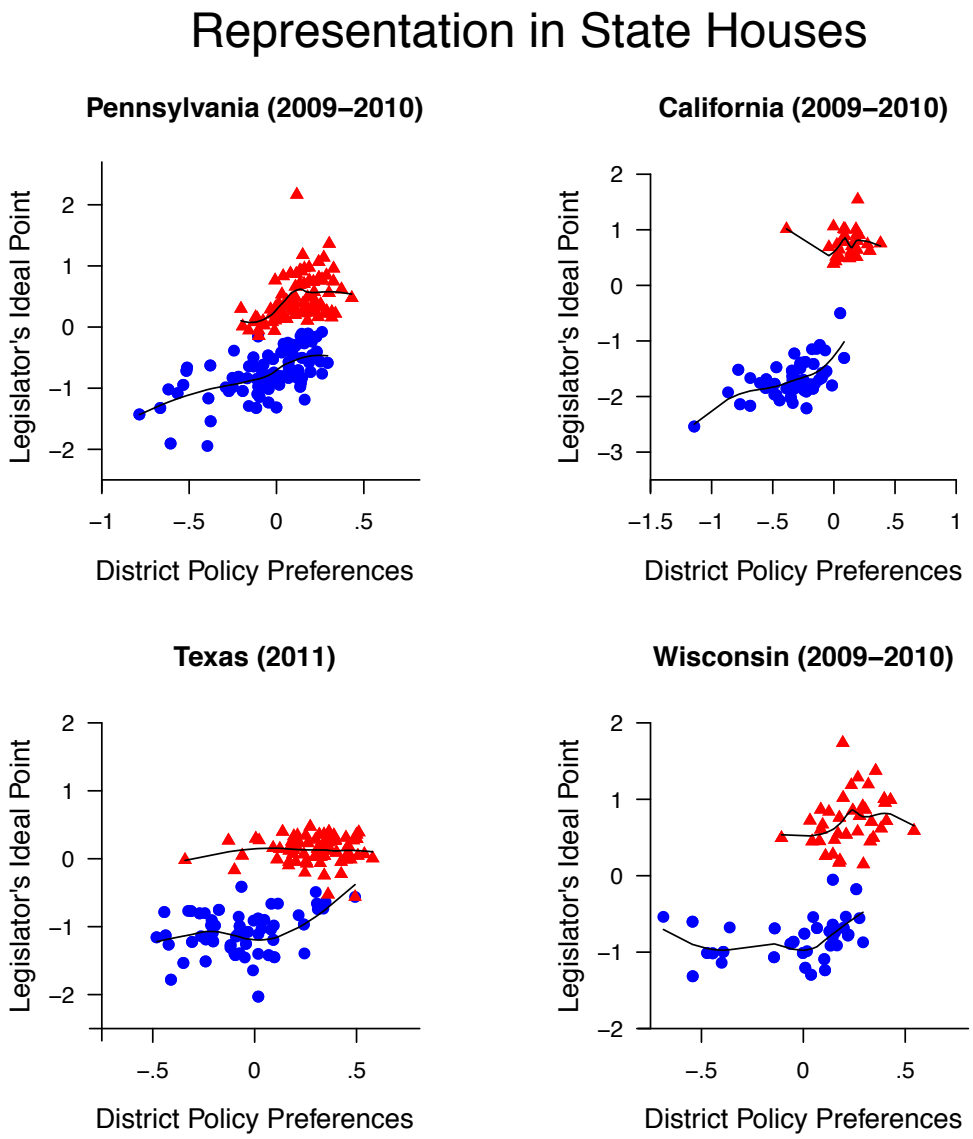
This figure shows the disaggregated and MRP estimates of the policy preferences of the mean citizen in each state. The circular dots are MRP estimates and the squares are disaggregated estimates. The graph also shows confidence intervals for each estimate.

Figure 3



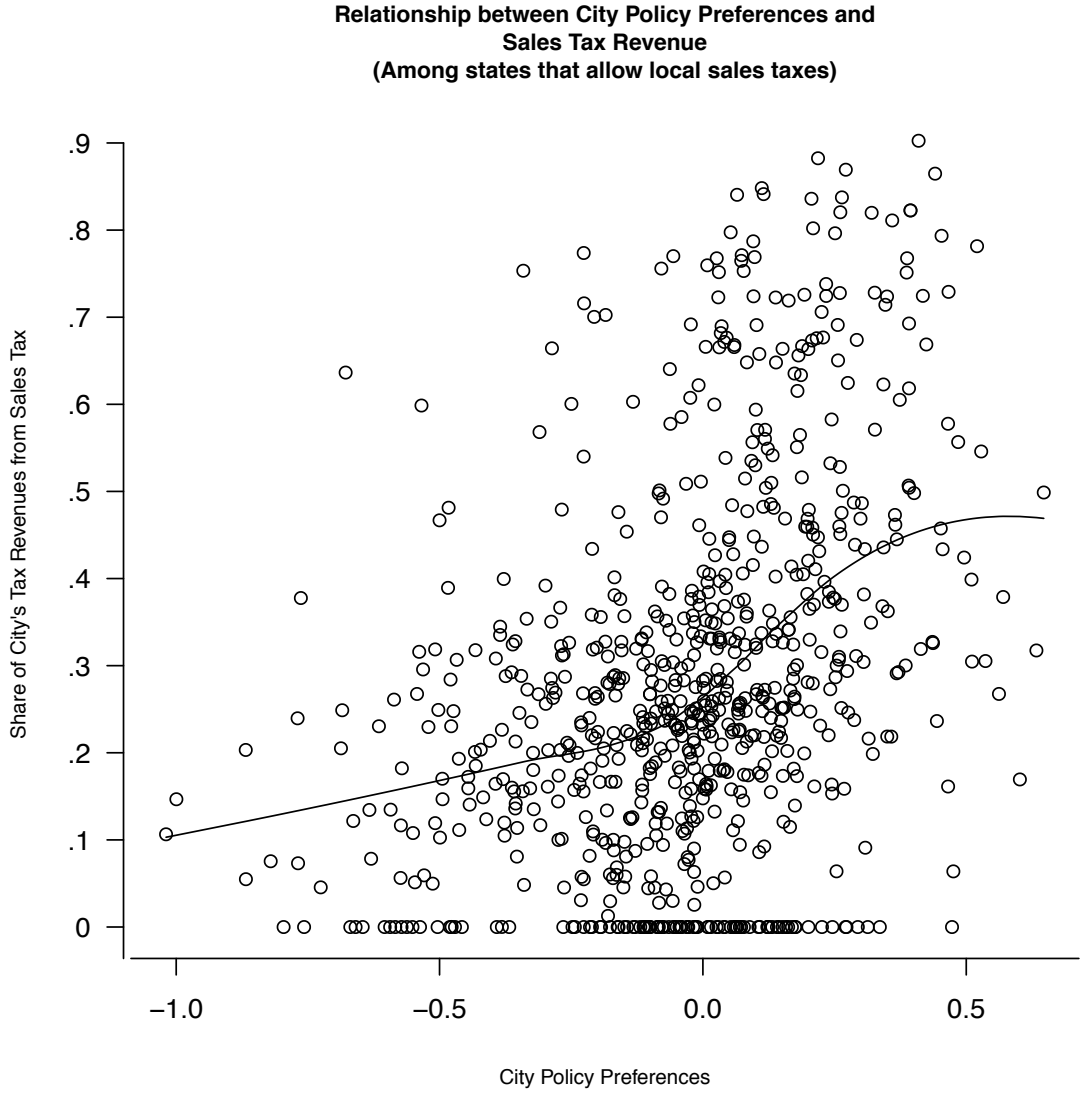
This figure shows disaggregated and MRP estimates of the policy preferences of the mean citizen in each city in Texas with more than 50,000 people. The circular dots are MRP estimates and the squares are disaggregated estimates. The graph also shows confidence intervals for each estimate.

Figure 4



This figure shows the relationship between district policy preferences and legislators' ideal points in the Pennsylvania, California, Texas, and Wisconsin state houses. The lines are loess plots of the relationship between district policy preferences and legislators' ideal points in each party.

Figure 5



This figure shows the relationship between city policy preferences and the share of city revenues collected from sales taxes in states that allow municipalities to collect sales taxes.

[1] We are grateful for feedback on previous drafts of this paper from four anonymous reviewers, Adam Bonica, Joshua Clinton, Morris Fiorina, Robert Gulotty, Simon Jackman, Howard Rosenthal, Jed Stiglitz, and participants at the Stanford American Politics Workshop and 2011 Cooperative Congressional Election Study Conference. We are especially indebted to David Brady, Jeff Lewis, Stephen Ansolabehere, and Jonathan Rodden for their generous advice and support. Data and supporting materials will be made available at www.americanideologyproject.com/data upon publication.

[2] Individuals do not need to be able to identify this policy bundle. It merely must be true that they would choose this policy bundle over any other one in a pairwise comparison.

[3] The large number of questions in these surveys reduces the measurement error in our estimates of citizens' policy preferences (Ansolabehere, Rodden, and Snyder 2008).

[4] Bailey (2007) compares ideal points of the president, senators, representatives, and Supreme Court justices. Bafumi and Herron (2010) compare the ideal points of president, senators, representatives, and voters.

[5] Scaling our super survey alone, we find that a one-dimensional model correctly classifies 78.8% of all responses. A two dimensional model increases the percent correctly predicted to 80.2%, an increase of only 1.4 percentage points. This is less than the increase in fit that is used in the Congressional literature as a barometer of whether roll call voting in Congress has a one-dimensional structure.

[6] The survey data does not include identifiers for state legislative districts and cities. As a result, we use respondents' zipcodes to match respondents to these geographic units. Specifically, we estimate the proportion of people in each zipcode that live in each state legislative district or city using GIS software. Then, we probabilistically assign survey

respondents to state legislative districts and cities based on the proportion of people in their zipcode that live in each geographic area. Overall, this process introduces a small amount of noise into our estimates, but it does not introduce any systematic bias.

[7] These data were obtained from Census factfinder.

[8] Previous work using MRP at the state level has used the "Public Use Microdata Sample" (PUMS) from the Census (e.g., Lax and Phillips 2009). However, the PUMS data does not include information about respondents' congressional, state-legislative districts, or cities. Fortunately, the Census Factfinder includes demographic breakdowns for each city, state legislative district, congressional district, and state for the population 25 and over, which we use to calculate the necessary population frequencies for our analysis. This approach introduces some error into our analysis. But this error is likely minimal since only about 10% of the voting population is under 25 and the demographic breakdown of the 25 and over population is generally similar to the voting-age population.

[9] To estimate the presidential vote share in state legislative districts and cities, we aggregated precinct-level 2008 presidential election data collected by Ansolabehere and Rodden (2012) for 39 states.

[10] We estimate state legislators' ideal points using roll call data from the 2009-2010 and 2011 session collected by the Sunlight Foundation.

[11] In cities in states that allow municipalities to collect a sales tax, there is a correlation of .34 between a city's policy preferences and the share of its revenues from sales taxes.

[12] We will make all of our new estimates available on our website soon after publication.