Labor Unions and Unequal Representation*

Michael Becher†
Daniel Stegmueller‡

This version: January 2019

Abstract

Recent research has documented that lawmakers are more responsive to the views of the affluent than to the less well-off. This raises the important question of whether there are institutions that can limit unequal representation. We argue that labor unions play this role and we provide evidence from the contemporary U.S. House of Representatives. Our novel dataset combines income-specific estimates of constituency preferences based on 223,000 survey respondents matched to 27 roll-call votes with measure of district-level union strength, drawn from 350,000 administrative records. Exploiting within-district variation in preference polarization, within-state variation in union strength and rich data on confounds, our analysis rules out a host of alternative explanations. In contrast to the view that unions have become too weak or fragmented to matter, they significantly dampen unequal responsiveness: a standard deviation increase in union membership increases legislative responsiveness towards the poor by about 6 to 8 percentage points.

∗We thank John Ahlquist, Lucio Baccaro, Alexander Hertel-Fernandez, Patricia A. Kirkland, Clara Park, Jonas Pontusson and Elizabeth Rigby for very helpful comments on earlier drafts of the paper. We are also grateful for feedback from conference/seminar participants at the annual meetings of APSA (2017) and MPSA (2018), IAST/TSE, and the Geneva Workshop on Unions and the Politics of Inequality. We thank Konstantin Käppner for contributing to an earlier version and Spencer Dorsey and Marco Morucci for excellent research assistance. Becher gratefully acknowledges financial support from the Agence Nationale de la Recherche (ANR)-Labex IAST. Stegmueller’s research was supported by the National Research Foundation of Korea (NRF-2017S1A3A2066657).
†Institute for Advanced Study in Toulouse, University of Toulouse 1 Capitole, michael.becher@iast.fr
‡Duke University, daniel.stegmueller@duke.edu
I. Introduction

Over the last 15 years or so, political scientists have paid increasing attention to the link between economic inequality and political representation. In contrast to the principle of political equality that is central to the ideal of democratic governance, this vibrant strand of research has repeatedly found disparities in political representation by income. Specifically, elected officials and policy outcomes are more responsive to the views of affluent citizens than to middle-income and low-income citizens, and sometimes they are not responsive to low-income citizens at all.1 As summarized by Bartels (2016: 235), evidence of unequal representation has been found for legislators, party platforms, national policy and state policy. While scholarship has initially focused mostly on the United States, recent comparative work has revealed similar patterns of unequal representation across a larger range of political systems (Bartels 2017; Elsässer et al. 2017; Lupu and Warner 2017), including democracies with proportional electoral systems and multi-party governments that had been previously associated with kinder and gentler (and presumably more equal) representation (Lijphart 1999). Given these results, it is germane to ask whether there are institutions or organizations that dampen unequal responsiveness in the democratic process.

In this paper, we argue that stronger labor unions systematically decrease the extent of unequal representation by elected representatives in the US. They do so even in a context of high income inequality, expensive electoral campaigns, and comparatively low union membership. Existing social science research has documented that union membership is associated with lower income differentials in political participation (Leighley and Nagler 2007; Rosenfeld 2014). Moreover, unions tend to take positions favored by less affluent citizens (Gilens 2012), and they are one of the few organizations in national politics that advocate on the behalf of non-managerial workers, spending a substantial amount of resources in the process (Schlozman et al. 2012). Some also take costly strike action in the interest of others (Ahlquist and Levy 2013).2

However, the literature provides little evidence on whether unions actually cause a meaningful reduction in the pro-affluent bias of national politicians. Several scholars of representation suggest that unions have become too weak, too narrow, or too fragmented to have a significant egalitarian political impact in national policymaking (Gilens 2012: 175; Hacker and Pierson 2010: 143). Moreover, a key issue is that the relationship between union strength and more equal responsiveness by politicians may be spuriously driven by the same underlying determinants. For example, due to differences in social capital (Putnam 1993, 2000) workers in some electoral districts may be better at solving collective action problems than others. As a result, they would be more likely to unionize their

---

1For instance, see Bartels (2008: ch. 9); Bartels (2016: ch. 8); Bhatti and Erikson (2011); Flavin (2012); Ellis (2013); Gilens (2012); Gilens and Page (2014); Rhodes and Schaffner (2017); Rigby and Wright (2013). For examples of different findings or interpretations, see Brunner et al. (2013); Enns (2015); Erikson (2015).

2See Ahlquist (2017) for a review of the large, interdisciplinary literature on union effects.
workplace in the first place and, independently, politicians would be more responsive to them. Another concern is that the activity of unions may influence “parties and policy, but policy and institutions also affect unionization rates” (Ahlquist 2017: 427). While collective action problems dilute incentives of politicians to make politics using policies, in some circumstances they are overcome (Anzia and Moe 2016; Hacker and Pierson 2010). Thus, unequal representation may produce policies that make it more difficult to organize unions in the first place. In particular, ‘right-to-work’ and collective bargaining laws hamper unionization efforts, and recent research demonstrates that these laws can have profound political effects (Feigenbaum et al. 2018; Flavin and Hartney 2015).

Our empirical strategy addresses these problems based on a combination of fine-grained data, a within-district research design, and robust inferential models. We assess our argument using the contemporary Congress, where unequal responsiveness by elected representatives and their policy choices has been well documented (Bartels 2008, 2016; Ellis 2013; Gilens 2012; Rhodes and Schaffner 2017) and the playing field for organized interest is skewed against the less affluent (Schlozman et al. 2012). We focus on members of the House of Representatives during the 109–112th Congress (2005-2012) since this setting enables us to capture within-state variation in union strength as well as within-district variation in preference polarization by income across a large number of policy issues. Our design provides leverage to rule out alternative explanations using state and district fixed effects and allows us to measure theoretically important confounders not accounted for in previous work.

At its core, our dataset combines estimated income-specific measures of constituency preferences based on 223,000 survey respondents matched to 27 roll-call votes with information on local unions extracted from more than 350,000 administrative records. To measure district-level policy preferences, we use multiple waves of the Cooperative Congressional Election Study (CCES) and calculate preferences on 27 concrete policy issues for each income group in each congressional district. We employ small area estimation as the CCES is not designed to be representative at the district level (we also show that our findings are robust to using alternative approaches, such as multilevel regression and poststratification [MRP]). To measure the district-level strength of unions, we use mandatory reports filed by local unions to the Department of Labor. Following recent work by Becher et al. (2018), this largely neglected administrative data source is used to construct measures of union membership at the district level. This measurement strategy overcomes major limitations of standard survey data used to measure union strength.3

Our empirical analysis traces the legislative responsiveness of House members to the preferences of different income groups in their constituency conditional on district-level

---

3Prior research is almost exclusively based on survey data that are not suited for a district-level analysis due to missing identifiers or sampling design. In contrast to surveys, moreover, filing LM forms is mandatory for most unions, non-submission and incorrect submissions are penalized, reports are audited and contain precise geographic information.
union strength. We find that district-level union membership dampens unequal responsiveness by national legislators. In line with previous research, on average House members are significantly less responsive to the policy preferences of low-income constituents. However, this gap in responsiveness is smaller where unions are stronger, and it decreases significantly where union members are numerous. This moderating effect of unions is not an artifact of existing state-level union policies or largely time-invariant state-level or district-level unobservables (such as institutions, history, or culture). Extended specifications allow other district-level characteristics to also moderate legislative responsiveness to different income groups. They demonstrate that the union effect is not driven by district-specific levels of socio-economic factors such as education, race, gender, median household income, urbanization, or a district’s employment structure. We also rule out the possibility that our finding simply represents the general capacity of workers or people to organize (or be organized) by accounting for explicit measures of district-level organization capacity based on new data on unionization attempts from the National Labor Relations Board, the predominance of religious organizations, and behavioral measures of social capital. We also go beyond standard regression models, and employ estimates from a Double Selection Estimator (Belloni et al. 2014) and Kernel Regularized Least Squares (Hainmueller and Hazlett 2014) to show that the moderating effect of unions is robust to relaxing potentially important modeling assumptions.

An exploration of possible mechanisms points to campaign contributions and partisan selection as two relevant channels through which local unions enhance the representation of the less well-off. Relatedly, the equality-enhancing effect of unions is stronger for bills on which the largest union confederation AFL-CIO has staked out a clear position.

We are aware of only two previous investigations of the effect of organized labor on unequal responsiveness and they differ considerably in their approach from ours. Focusing on a recent cross-section of 47 US states, Flavin (2018) shows that states with stronger unions exhibit less unequal representation, as estimated from regressions of income-weighted voter preferences on state-level policy liberalism. Studying the 110th House of Representatives, Ellis (2013) finds mixed results. District-level unionization is related to a smaller rich-poor gap for key legislative votes, but there is no such effect for overall ideological representation. Our analysis addresses the problem that survey samples are not representative for congressional districts by design, which can lead to biased estimates of income biases in representation. It confirms the finding that unions are linked to more equal representation covering four Congresses and three times as many roll call votes as studied by Ellis (2013). However, our main empirical contribution is that we can go much further in ruling out alternative explanations. Our research design leverages within-district variation in preference polarization, within-state variation in union strength, as well as extensive district-level data on alternative moderating factors that may be bundled with union strength. In contrast to the pure cross-sectional designs of these two previous studies, our analysis can thus account for state and district fixed effects that capture important sources of unobserved
heterogeneity and it directly measures many important confounders. As a result, we can state with more confidence that the impact of unions is not spurious.

Against the backdrop of current scientific and public debates about labor unions and political representation, these findings may come as somewhat of a surprise. While some strands of research and political discourse portray unions as an egalitarian force in politics, others see them fatally weakened, effects as much as causes of unequal representation, or simply as just another organized group fighting for special interests (that do not generally overlap with those of lower income individuals). The latter view is held by a large strand of scholarship in economics (cf. Freeman and Medoff 1984) and by researchers studying the role of teachers’ unions in political science (Anzia 2011; Moe 2011). Most extant research on Congress simply does not have the required data to directly assess the effect of unions on representational equality. Numerous studies of union strength and congressional roll-call voting do not measure voter preferences, which makes it difficult to interpret who is being represented (Becher et al. 2018; Box-Steffensmeier et al. 1997; Freeman and Medoff 1984).

Altogether, our results suggest that unequal responsiveness is not an unavoidable feature of democratic capitalism. The results are especially striking given that recent cross-national studies have found consistent patterns of unequal representation across different political institutions (Bartels 2017; Lupu and Warner 2017). In contrast, we find considerable heterogeneity in differential responsiveness across districts affected by local labor unions—a fundamental economic institution. The moderating effect of unions uncovered in our analysis is large enough to swing key votes in Congress. That said, our results support the view that political efforts to (further) weaken unions, as evidenced in recent reforms in states like Michigan and Wisconsin, are, if anything, likely to exacerbate unequal responsiveness in representation. They may also explain why unions are (still) under attack.

II. Moderating biased responsiveness in Congress?

While few studies have directly assessed the impact of labor unions on unequal responsiveness in Congress or elsewhere, various strands of scholarship in political science and related fields suggest that labor unions are one of the few mass-membership organization that provide collective voice to lower income individuals in the political arena, with potentially important consequences for political representation (Ahlquist and Levy 2013; Bartels 2016; Freeman and Medoff 1984; Schlozman et al. 2012). Consistent with a central premise of the collective voice perspective, unions tend to take positions favored by less affluent citizens. Gilens (2012: 154-161) compares public positions of national unions with mass policy preferences across several hundred policy issues and finds that unions’ positions are most strongly correlated with the preferences of the less well-off (see also Hacker and
Similarly, Schlozman et al. (2012: 87) conclude that unions are one of the few organizations in national politics “that advocate on behalf of the economic interest of workers who are not professionals or managers.”

However, shared preferences between the less well-off and organized labor are by no means sufficient to alter inequalities in political representation in national politics. This requires an effective political transmission mechanism. To guide the empirical analysis, we sketch key elements of a framework of union organization and political responsiveness.

Labor unions are organizations formed to bargain collectively, on behalf of their members, with employers over wages and conditions. Unions are thus created at the local (i.e., establishment) level (Freeman and Medoff 1984). Once formed, unions may (and often do) enter the political arena. The ability of unions to increase the rate of political participation—including voting, contacting officials, attending rallies, or making donations—of low- and middle-income citizens is often considered to be their key channel of political influence. Importantly, unions may also increase participation among non-members with similar policy preferences through get-out-the-vote campaigns and social networks (Leighley and Nagler 2007; Rosenfeld 2014; Schlozman et al. 2012). Making contributions to favored candidates and campaigns complements the ability of unions to communicate with and mobilize members or to provide campaign volunteers. Indeed, unions are among the leading contributors to political action committees (PAC), accounting for a quarter of total PAC spending in 2009 (Schlozman et al. 2012: ch. 14). In contrast to corporations and business organizations, union contributions “represent the aggregation of a large number of small individual donations” (Schlozman et al. 2012: 428).

The credible threat of political mobilization can affect policy decisions by representatives in two general ways. First, it may shape who is elected in a given electoral district. If politicians are not exchangeable (because they differ in their preferences and beliefs), political selection is important. In an age of elite polarization (McCarty et al. 2006), the partisan identity of a representative is often crucial for determining legislative voting (Bartels 2016; Lee et al. 2004). Since the New Deal era, unions and union members have largely allied with the Democratic Party, given its stronger support for many of their broader policy demands (Lichtenstein 2013; Schlozman 2015). Political selection might also shape other political characteristics of representatives, such as their class background or race (Butler 2014; Carnes 2013).

Second, unions’ mobilization potential shapes the incentives of elected representatives, beyond their partisan affiliation and personal traits. Policymakers’ rational anticipation of

---

4This is consistent with the argument that organized labor fosters norms of solidarity and support for the less well-off, through leadership (Ahlquist and Levy 2013; Kim and Margalit 2017) or social interactions (Berelson et al. 1954).

5While evidence on the direct effect of contributions on legislative behavior is mixed, recent field-experimental results indicate that contributions help to provide access (Kalla and Broockman 2016) or sway congressional staffers (Hertel-Fernandez et al. 2018).
public reactions plays a central role in theories of accountability and dynamic responsiveness (Arnold 1990; Stimson et al. 1995). While many individual legislative votes do not affect the reelection prospects of representatives, on potentially salient votes they can face hard choices between party ideology and competing constituency preferences. On international trade agreements, for instance, Democratic representatives have faced cross-pressures between a more skeptical stance taken by unions and low-income constituents versus that of their own party (Box-Steffensmeier et al. 1997). On the other side of the aisle, in the wake of the financial crisis, Republican legislators found themselves torn between their own partisan views on stimulus spending and the pressure from less well-off constituents (Mian et al. 2010).

Politicians’ incentives are also linked to information. Theories of representation emphasize that members of Congress, and especially the House, face numerous voting decisions in each term, and it would be unrealistic to assume that they have access to reliable, unbiased polling data on constituency preferences on all the issues they face (Arnold 1990; Miller and Stokes 1963). Instead, representatives—with the help of their staffers—rely on alternative methods to assess public opinion, including constituent correspondence, town halls, contacts with community leaders, or local interest groups (Miller 2007). In this limited information context, the strength of local unions may enhance the visibility and perception of constituent preferences (Hertel-Fernandez et al. 2018).

Following seminal theories of congressional action (Arnold 1990; Miller and Stokes 1963), our argument emphasizes that the strength of local unions underpins a credible mobilization threat that impacts the action of candidates and legislators. Anticipating mobilizing efforts by unions, a potential candidate may not even enter into the race; an elected, career-oriented, politician might be pressured to alter his or her vote even without a full mobilization effort as long as unions’ mobilization capacity is visible. Thus, both campaign contributions and candidate selection should matter as a channel linking local union strength and representation, since they are linked to credible threats of mobilization.

Our argument implies that the district-level strength of labor unions increases the responsiveness by members of Congress to the less affluent. While we know from previous work that politicians are considerably more responsive to the preferences of the affluent than those of the less well-off, this bias should be reduced in districts with relatively higher union membership. Substantively, it is crucial to assess how far the presence of unions can move responsiveness toward the ideal of political equality.

Butler and Nickerson (2011) find that politicians respond when provided with more accurate opinion data. However, behavioral biases may lead politicians to discount constituent preferences they disagree with (Butler and Dynes 2016).

In line with a large literature, we focus on union membership as a key component of union strength. In a study of the effect of unions on legislative ideology rather than income-biased responsiveness, Becher et al. (2018) argue that structure of local unions (i.e., the concentration of unions in a given locality) matters as well. However, they also show empirically that union density and concentration are separable dimensions.
III. DATA AND EMPIRICAL STRATEGY

Any effort to test the relevance of unions for unequal representation confronts major challenges of measurement and causal interpretation. The dataset we have compiled allows us to address these issues to an extent previously impossible. We have created a panel of legislators’ roll call votes matched to income-specific policy preferences at the district level, and district-level measures of union membership. Our main empirical strategy to examine the influence of unions on unequal representation is built on two basic pillars: district fixed effects and interactive controls. The fact that we observe several roll calls within a given congressional district allows us to specify a model with district fixed effects, which capture unobservable characteristics of districts (and states) that are constant over roll-calls, such as historical legacies or the strength of partisan organization. To provide for a stricter test of the moderating effect of unions, we also allow a rich set of other district characteristics to moderate the link between income groups and legislators’ voting behavior. This amounts to estimating models including interactions between observed district characteristics and group preferences. In our most flexible specification we allow these to be non-linear (we describe our models in more detail below).

The data required to implement these models were constructed in three steps. First, we match information on roll call items for 223,000 CCES respondents to actual roll call votes cast in the House of Representatives in the 109th to the 112th Congress.\(^8\) Second, we estimate policy preferences for low and high income constituents in each district for 27 roll calls. To deal with the fact that the CCES is not a representative sample of district populations we use a small area estimation strategy combining the CCES sample with unit record Census data matching the full distribution of age, education, gender, race, and income using a chained Random Forests algorithm (more below and in Appendix B). Third, we measure district-level union membership based on digitized administrative records from the Department of Labor.

III.A. CCES data and Congressional roll calls

The CCES is an ideal starting point for our analysis, since it is a nationally representative study, includes a considerable number of roll call questions, and provides us with a large enough sample size to decompose income-group preferences by district. It addresses several data concerns that plagued initial research on unequal responsiveness in Congress (Bhatti and Erikson 2011). The roll calls included in the CCES concern key votes as identified by Congressional Quarterly and the Washington Post and cover a broad range of issues.

---

\(^8\)Our analysis focuses on one apportionment period, which generally holds district boundaries constant (we show that the results are robust to cases of mid-period redistricting).
(Ansolabehere and Jones 2010). Respondents are presented with the key wording of the bill (as used on the floor and in media reports) and are then asked to cast their own vote: “What about you? If you were faced with this decision would you vote for, against, or not sure?” Contrary to widely usual agree–disagree survey measures of issue preferences, matched roll call votes provide us with unequivocal evidence of policy congruence between respondent and legislator (Jessee 2009, Ansolabehere and Jones 2010: 585). We match 27 roll call items in the CCES to roll call votes cast in the House of the 109th to 112th Congress. These cover important legislative decisions, such as Dodd-Frank, the Affordable Care Act (and attempts to repeal it), the minimum wage increase, the ratification of the Central America Free Trade Agreement, or the Lilly Ledbetter Fair Pay Act. Table A.1 in the Appendix lists all matched CCES items and House bills included in our estimation sample.

III.B. Measuring constituency preferences by income group

The CCES provides us with a comparatively large sample size per district. However, an important potential issue is that it is not designed to be representative for congressional district populations. Thus, individuals with certain characteristics, such as particular combinations of income, race and education, may be underrepresented in the CCES sample for a given district. If this is the case, unadjusted policy preferences from the CCES will not reflect the target population and using them can lead to biased estimates of unequal representation in Congress, as politicians are held to the wrong benchmark. The solution to this issue is to employ some form of small area estimation to rebalance the survey sample to represent the district population. The machine-learning solution we propose is relatively new to the representation literature in political science, but it has some attractive features that merit its application to this topic. It does not require distributional and functional form assumptions, it allows for arbitrary higher-order interactions of covariates, and it can fully leverage fine-grained census data to construct representative samples of congressional districts. However, we stress that our findings do not depend on this particular approach. As shown in Online Appendix B, our approach leads to somewhat more conservative estimates of the impact of unions on the representation of different income groups compared to the MRP approach widely used by political scientists (Lax and Phillips 2009). Qualitatively both approaches yield the same conclusions.

Our approach, small area estimation using chained random forests, matches CCES survey respondents to corresponding cases from unit record Census data. The design of the Census ensures an accurate representation of the distribution of population characteristics in a given district (Torrieri et al. 2014: Ch.4). Matching these two data sources is essentially a prediction problem, which we address using a flexible, non-parametric machine learning approach based on random forests (Stekhoven and Bühlmann 2011). Put simply, the idea is

Honaker and Plutzer (2016) use a similar approach (but relying on multivariate normal imputations) and further discuss its empirical performance in estimating small area attitudes and preferences.
that rich census data exist for every district, whereas survey data on preference are scarce in some districts and may not be fully representative. Using general machine learning tools, we can attach preferences to the Census by matching it to CCES respondents based on common demographic characteristics. The resulting data set of public preferences is representative of congressional districts.

Concretely, we use about 3 million individual-level records from a synthetic sample of the Census Bureau’s American Community Survey from 2006 to 2011. We stack both datasets, creating a structure where we have common district identifiers and individual covariates while responses to policy preference questions are missing in the Census portion of the data. As common covariates bridging CCES and Census we use the following demographic characteristics: gender, race (3 categories), education (5 categories), age (continuous) and family income (continuous). The latter is of particular relevance as we are interested in producing district–income group specific preferences.

In the next step we fill missing roll call preferences in the Census with matching data from CCES respondents. Since this is essentially a prediction problem, we can use powerful tools developed in the machine learning literature to achieve this task. We use an algorithm proposed by Stekhoven and Bühlmann (2011), which uses chained random forests (Breiman 2001) to impute missing cells. Compared to commonly used multivariate normal or regression imputation techniques, this strategy has the advantage that it is fully nonparametric, allowing for complex interactions between covariates, and deals with both continuous and categorical data (Tang and Ishwaran 2017). Our completed data-set now contains preferences for 27 roll call items of synthetic ‘Census individuals’, which are a representative sample of each House district.

With these data in hand, we assign individuals to income groups and calculate group-specific preferences for each roll call in each district. Following previous work in the representation literature (Bartels 2008, 2016), we delineate low- and high-income respondents using the 33rd and 67th percentile of the distribution of family incomes. Note that in line with theories of constituency representation in Congress we specify these income thresholds separately by congressional district. This accounts for the substantial differences in both average income and income inequality between US districts. It also ensures that within each district, income groups are of comparable size. Online Appendix Table A.2 shows the distribution of income-group cutoffs. On average, our chosen cutoffs are close to those used in the established literature. The mean of our district-specific low-income cutoffs is around $39,000, while Bartels uses $40,000 (Bartels 2016: 240); our mean high-income cutoff is around $81,000, where Bartels employs a threshold of $80,000. However, beyond these averages lies considerable variation. In some districts, the 33rd percentile cutoff is as low as $16,500, while the 67th percentile reaches almost $160,000 in others.

See Appendix B for more details on the construction of our Census sample and our matching/imputation procedure.

Results are relatively invariant to using alternative income thresholds (see Table C.1).
District-level income gap in public support for 6 selected policies

*Note:* Each histogram plots the difference in support for a matched roll-call vote question between people in lower third and people in upper third of their district’s income distribution for all House districts.

For each roll call, we then estimate district-level preferences of low- and high-income constituents, which we denote by \((\theta_l, \theta_h)\), as the proportion of individuals voting ‘yea’. Since preference estimates are in \([0, 1]\) they can be directly related to legislators’ probability of voting ‘yea’ on a given roll call. Our data shows considerable variation in the distance of the policy preferences of those at the top and those at the bottom as illustrated in Figure I. It plots histograms of the difference between low-income and high-income preferences \((\theta_h - \theta_l)\) in congressional districts for six selected roll calls. For salient bills, such as increasing the minimum wage (the Fair Minimum Wage Act), housing crisis assistance (the Housing and Economic Recovery Act), or Affordable Care Act, the vast majority of low-income constituents are more supportive than their high-income counterparts in each and every district. On other issues, such as the ratification of the Central America Free Trade Agreement, high income constituents are clearly in favor. In all examples, we find considerable across-district variation in the preference gap between low- and high-income constituents.\(^\text{12}\) We will employ this variation over both roll calls and districts to estimate legislators’ differential

\(^{12}\text{Averaged over all districts and roll calls, there is a statistically significant gap between the preferences of the bottom third and the top. The mean of the (absolute) preference difference is 17 percentage points; the 10th percentile is 3 points while the 90th percentile is 32 percentage points.}\)
responsiveness to changes in policy preferences of different income groups, and how it might be moderated by union strength.

III.C. District-level union membership

To measure district-level union membership we draw on fine-grained administrative data. Based on the Labor-Management Reporting and Disclosure Act (LMRDA) of 1959, unions have to file mandatory yearly reports (called LM forms) with Office of Labor-Management Standards (OLMS). The Civil Service Reform Act of 1978 introduced a similarly comprehensive system of reporting for federal employees (see Budd 2018). A mandatory part of each report is the number of members a union has. Failure to report, or reporting falsified information, is made a criminal offense under the LMRDA, and reports filed by unions are audited by the OLMS. This makes LM forms a reliable source of information on unions and their members.

Using LM forms provides important advantages over using measures derived from surveys. First, mandatory administrative filings are likely more reliable than population surveys, which often suffer from over-reporting and unit-nonresponse (Southworth and Stepan-Norris 2009: 311; Card 1996). Second, they allow us to estimate union membership numbers for smaller geographical units, which are usually unavailable in population surveys (to protect respondents’ confidentiality) or only covered with insufficient sample sizes. Another advantage for the study of politics is that the presence of union locales is observable to politicians on the ground even in the absence of survey data.

The resulting database contains almost 30,000 local union. It is based on 358,051 digitized individual reports that were cleaned, validated, geocoded, and matched to congressional districts. The number of union members in each congressional district can then be readily obtained as the sum of all reported union members. Figure II shows the distribution of union membership in House districts averaged for the 109th to 112th Congress. It demonstrates that there is substantial variation in unionization between electoral districts even within states, which would be ignored by a state-level analysis.

A potential drawback of using LM forms is that some unions are exempt from filing requirements. Each and every private sector union is required to submit a report, but under some specific conditions public sector unions are exempt. Thus, while unions representing postal or federal employees are covered, unions that exclusively represent state, county, or municipal government employees are exempt. However, even these have to file if at least one of their members is a private sector employee. In practice, this leads to almost

---

13 Even the primary source for union data, the Current Population Survey (CPS) suffers from these issues, partly as a result of its rather broad question wording.

14 The most prominent data set on union membership, compiled by Hirsch et al. (2001), provides CPS-based estimates for states and metropolitan statistical areas; district identifiers are not available.
IV.3. Statistical specifications

For each roll call vote $j$ ($j = 1, \ldots, J$), we have measured preferences of low and high income citizens in a given congressional district $d$ ($d = 1, \ldots, D$) denoted by $(\theta_{jd}^l, \theta_{jd}^h)$. For each district, the level of (logged) union membership is denoted by $U_d$. Given that population size is approximately identical in districts within states, we sometimes simply refer to this as union density. We specify relevant confounders in $X_d$. Depending on the particular specification (discussed in the next section) these will include (i) socio-economic district characteristics, (ii) measures of historical state union policies and state fixed effects, (iii) measures for the capability of districts’ workers to organize collective action, (iv) as well as non-linear transformations of these. For ease of interpretation, we have scaled all inputs to have mean zero and unit standard deviation. Our model for the voting behavior of House

---

While there is no "gold standard" of accurate union membership numbers, we can compare aggregate membership based on our LM form data with widely used survey-based measure from the CPS (Hirsch et al. 2001). This confirms that LM forms provide a rather comprehensive accounting of unions. At the national level, the average number of union members in our dataset is 13.21 million (excluding Washington, D.C., which is not represented in Congress). The CPS figure for the same period is 15.22 million. This modest difference is consistent with some degree of over-reporting in the CPS given its broad question wording (Southworth and Stepan-Norris 2009: 311). It can also be interpreted as an upper bound for the non-coverage of some public sector unions in our data. A more detailed analysis by Becher et al. (2018) shows that state-level aggregates from LM forms and the CPS are strongly correlated ($r = 0.86$).
members is the following linear probability specification:

\[
y_{ijd} = \mu_l \theta_{jd} + \mu_h \theta_{jd} + \eta_l(U_d \times \theta_{jd}) + \eta^h(U_d \times \theta_{jd}) + \beta_l(X_d \times \theta_{jd}) + \beta^h(X_d \times \theta_{jd}) + \alpha_d + \epsilon_{ijd}
\]

The key terms here are the interactions between union membership and the respective preferences of the affluent and the poor, \(U_d \theta^h_{jd}\) and \(U_d \theta^l_{jd}\). Thus, when \(\eta_l\) and \(\eta^h\) are zero the group-specific preference coefficients \(\mu_l\) and \(\mu^h\) indicate the change in the probability of legislators casting a supportive vote induced by a standard deviation change in the respective preferences of the poor and the affluent. The coefficient \(\eta_l\) indicates the marginal effect of a standard deviation change in logged union membership on the responsiveness of legislators’ votes to the preferences of the poor. The corresponding marginal effect for the affluent is given by \(\eta^h\). Our theoretical expectation is that \(\eta_l > 0\) and \(\eta^h \leq 0\).

In order to mitigate the influence of unobserved confounders affecting legislators’ voting behavior, we account for time-constant unobservables on the district-level by including district fixed effects, \(\alpha_d\).\(^{16}\) Despite this, one may be worried that changes in responsiveness attributed to unions are spurious. To provide a stricter test of the moderating effect of unions, we include the interactions between controls (both on the district- and state-level) and group preferences \(X_d \theta^l_{jd}\) and \(X_d \theta^h_{jd}\). They use within-district variation over roll-calls and preferences to estimate the conditional marginal effect of group preferences, making it less likely that our estimated effect of union membership is simply due to omitted confounders. In more sophisticated analyses, detailed below, we allow these confounds to be strongly non-linear as well. Finally, \(\epsilon_{ijd}\) are white-noise errors assumed independent of covariates.

We account for heteroscedasticity and arbitrary within-district correlations when calculating standard errors (Abadie et al. 2017; Cameron and Miller 2015: 324).

**IV. Results**

Before presenting evidence on the moderating effect of unions, we want to give a sense of the overall picture of legislators’ responsiveness emerging from our data. Estimating a model as described above with district fixed effects but without accounting for local union organization (setting \(\beta^l, \beta^h\) and \(\eta_l, \eta^h\) to zero) or any other moderators, we find a clear gap in the responsiveness of legislators to the preferences of low- versus high-income individuals. A standard deviation increase in the preferences of the affluent is linked to an increase in the probability of legislators to cast a corresponding vote of 13.6 (±1.2) percentage points. In contrast, a standard deviation increase in the preferences of the less well-off induces a much smaller change in legislators’ behavior of 1.6 (±1.4) percentage points. With a

\(^{16}\)Note that non-interacted effects of district-level union membership and covariates (which vary between districts, but are constant over roll calls) are absorbed in \(\alpha_d\).
confidence interval ranging from $-1.1$ to $4.4$ points, we cannot reject the null hypothesis that legislators do not respond to the preferences of low-income constituents in the average electoral district. The responsiveness gap between the two groups is sizable (at 11.9 ($\pm2.5$) percentage points) and significantly different from zero. We show below that the extent of legislators’ non-responsiveness depends crucially on the strength of local unions.

**IV.A. Unions and unequal legislative responsiveness**

We start by summarizing our key finding graphically and then discuss more extensive model specifications. Figure III plots marginal effects of low- and high-income constituency preferences on representatives’ roll-call votes at varying levels of union membership with 95% confidence intervals.\(^{17}\) It shows that legislators’ responsiveness to the policy preferences and low-income and high-income constituents depends on district-level union membership: as unionization increases, legislators’ responsiveness to low-income constituents increases, while their responsiveness to high-income constituents declines by a similar amount. For example, moving from a district with median levels of union density to one at the 75th percentile increases the responsiveness of legislators to low-income preferences by 8 percentage points, while it decreases responsiveness to high-income preferences by about 5 points. Given the initial responsiveness gap, this change is substantial enough to substantially level the playing field between affluent and poor.

Are these findings robust to confounding factors? Table I presents parameter estimates from a number of increasingly rich specifications designed to capture potential confounds. In specification (1), we begin with a baseline model (also plotted in Figure III) that includes district fixed effects but no further preferences-confounder interactions (setting $\beta^l$ and $\beta^h$ to zero). We find that a standard deviation increase in district union membership increases legislators’ responsiveness to the poor by about 11 ($\pm1$) percentage points, while at the same time decreasing the advantage in responsiveness enjoyed by the affluent by about 6 ($\pm1$) points.

Even after accounting for district fixed effects, however, our results are still vulnerable to omitted variables that interact with group preferences. Following accounts of winner-take-all politics (Hacker and Pierson 2010), one alternative interpretation is that the moderating effect we have ascribed to unions mostly reflects the fact that state governments have chosen policies that strengthen or weaken the ability of unions to organize (also see Ahlquist 2017; Anzia and Moe 2016). If the likelihood of adapting pro- or anti union policies is correlated with biased representation, our estimated effect of unions might be spurious. In line with this concern, recent studies have demonstrated that right-to-work and collective bargaining laws regulating the formation and management of unions in the private or public sector have clear political effects on turnout and partisan vote shares (Feigenbaum et al. 2018; Feigenbaum et al. 2018; Feigenbaum et al. 2018; Feigenbaum et al. 2018).

---

\(^{17}\)Calculated from a LPM of vote choice on preferences and union membership. It includes district fixed effects and clusters standard errors on the district level. See also specification (1) in Table I below.
Figure III

District-level union membership as moderator of unequal representation.

Note: This figure plots changes in marginal effects of low- and high-income constituency preferences on representatives’ roll-call votes conditional on district-level union membership. Shaded areas are 95% confidence intervals based on district-clustered standard errors. The sample distribution of (z-standardized) union membership is indicated above the x-axis.

Flavin and Hartney 2015). In specification (2), we therefore add two measures of historical state union policy: the share of years with right-to-work legislation and the share of years with mandatory collective bargaining laws for teachers since 1955, taken from Flavin and Hartney (2015). These enter $X_d$ and are interacted with income group preferences $\theta^l$ and $\theta^h$. In specification (3) we go one step further and allow for any state-level characteristic (such as institutions or historically-rooted popular anti-union sentiments) to moderate the marginal effect of income group preferences on legislators vote choice by including state-specific constants in $X_d$ which are interacted with group preferences. The results from both extended specifications show that accounting for state-level policies and institutions as potential moderators does not change our core picture of the role of local union organization: where local unions are stronger the responsiveness gap between the affluent and the poor is reduced.

A more subtle problem concerns a form of simultaneity bias at the district level. There may be district-level factors shaping both the propensity to be a union member and to be politically active. If less affluent individuals with a higher capacity to organize and solve collective action problems cluster in specific districts, our estimates of the marginal impact of district union membership on responsiveness will be overly optimistic. Such a propensity may reflect critical historical junctures in labor organizations (Ahlquist and Levy 2013) or
Table I
Union density and representation. Marginal effect of standard deviation increase in union membership on marginal effect of income group preferences on legislator vote.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income preferences</td>
<td>0.106</td>
<td>0.082</td>
<td>0.098</td>
<td>0.084</td>
<td>0.068</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>−0.063</td>
<td>−0.036</td>
<td>−0.053</td>
<td>−0.051</td>
<td>−0.050</td>
<td>−0.040</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>District fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Group preferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× union policy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>× state constants</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>× organizational capacity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>× district covariates</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: N=15,780. $N_d = 534$. 27 roll call votes, 109th to 112th Congress. Linear probability models with standard errors robust to arbitrary within-district correlation and heteroscedasticity. All models include district fixed effects. Entries are marginal effects of union membership, $\eta^l$ and $\eta^h$. Specifications (2) to (5) include coefficients for interaction ($\beta^l$, $\beta^h$) of income group preferences with state- or district-level confounders. Specification (2) includes two measures of historical state union policymaking, the share of years with right-to-work legislation and collective bargaining agreements. (3) interacts preferences with state fixed effects. (4) includes a measure of district-level capacity to organize collective action captured by the number of churches per inhabitant and the number of NLRB union certification elections. (5) includes a large set of district-level characteristics (population size, degree of urbanization, shares of female, Black, Hispanic, BA degrees, employed in manufacturing, as well as median household income). Specification (6) includes all of the previously described measured variables.

Social capital (Putnam 1993, 2000). Consistent with the latter, for instance, Nannicini et al. (2013) find that political accountability in Italy is higher in districts with higher social capital.

To tackle this problem, we gathered additional data capturing the organizational capacity of a district: (i) the capability of workers to organize collective action measured via the average number of union certification elections in a district; (ii) the stock of social capital captured by the number of congregations per 1000 inhabitants (as well as two alternative measures of social capital, a behavioral index and the number of bowling alleys used in robustness tests).

Union certification elections, conducted by the National Labor Relations Board (NLRB), are a useful proxy, since holding such an election requires overcoming a costly organizational hurdle: at least 30 percent of employees have to sign authorization cards stating that they want to be represented by a union. Union organizers also face a non-trivial probability of being (illegally) fired by her employer (Budd 2018: ch. 6). We use the NLRB’s database to

\[18\text{Certification elections are not a foregone conclusion: during the 112th Congress, unions won 59%.}\]
extract all attempts to certify (or de-certify) a local union.\textsuperscript{19} We geocode each individual case report and locate it in a district. We then use the (logged) average number of cases in a district over the last seven years to proxy organizational potential. To count the number of congregations in a district, we use county-level data from the 2000 Religious Congregations and Membership Study and spatially interpolate it to districts. Appendix D provides more details. Both measures (interacted with group preferences) proxy a district’s organizational capacity in specification (4).

Perhaps surprisingly, we find that accounting for organizational capacity only dampens the union effect by a modest amount. The estimated impact of unions on responsiveness is reduced by about 1 percentage point. Note that this may also reflect the fact that existing union strength shapes attempts to organize new firms or establishments. However, specification (4) in Table I makes clear that even after accounting for organizational capacity we find that local union membership shapes responsiveness: a standard deviation increase in union membership still increases legislators’ responsiveness to the preferences of the poor by 9 ($\pm$1) percentage points, and lowers their responsiveness to the preferences of the affluent. This rules out the interpretation that the moderating effect of unions is merely an artifact of a broader propensity to overcome collective action problems.

In specification (5) we measure a large number of districts’ socio-economic characteristics and allow them to interact with constituency preferences: population size, race (share of African Americans and Hispanics), education (share with BA or higher), the share of the working population employed in manufacturing, median household income, and the degree of urbanization (for descriptive statistics, see Table A.3). This set of covariates excludes “bad controls” (Samii 2016) such as partisanship that are a mechanism through which unions influence representation.\textsuperscript{20} Again, our results point towards the existence of a clear moderating effect of unions, albeit at a somewhat smaller magnitude of about 7 percentage points. Our final specification, column (6) of Table I, includes all previous covariates and, again, confirms our core finding.

\textsuperscript{19}There are about 2200 elections each year. Not included is voluntary card check recognition by employers. Despite several high-profile voluntary recognition campaigns in recent years, Budd (2018: 199) notes that this is “the exception rather than the norm because employers typically refuse to recognize unions voluntarily.”

\textsuperscript{20}Theoretically and empirically, unions shape voting and election outcomes (see our analysis of possible mechanisms below and the literature cited in the introduction). Union membership is mainly driven by economic considerations and state-level policies that are accounted for in the analysis (Feigenbaum et al. 2018). To the degree that historical district-level partisanship is linked to union organization beyond state-level policies and district socio-economic structure, this should be captured by our measure of certification elections.
IV.B. Further robustness tests

*Alternative measures of social capital*  We consider two additional measures of social capital. Our first measure is the number of bowling alleys in an area popularized in “Bowling Alone” ([Putnam 2000](#)) based on data collected by [Rupasingha and Goetz (2008)](#). Our second measure is a composite social capital index combining information on membership in voluntary associations, voter turnout, the Census response rate, and the number of non-profit organizations ([Rupasingha and Goetz 2008](#)). We aggregate both measures to congressional districts (both refer to 2009 values) using spatial population-based weighting. Our results show that using these alternative measures does not change our core results.

<table>
<thead>
<tr>
<th></th>
<th>Low income</th>
<th>High income</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1a) Social capital: bowling alleys</td>
<td>0.067 (0.014)</td>
<td>−0.051 (0.013)</td>
</tr>
<tr>
<td>(1b) Social capital: index</td>
<td>0.065 (0.014)</td>
<td>−0.048 (0.013)</td>
</tr>
<tr>
<td>(2) Redistricting</td>
<td>0.067 (0.014)</td>
<td>−0.051 (0.013)</td>
</tr>
<tr>
<td>(3) MRP estimated preferences</td>
<td>0.115 (0.022)</td>
<td>−0.091 (0.018)</td>
</tr>
</tbody>
</table>

*Note:* Based on specification (5) in Table I. Entries are parameter estimates for $\eta^l$ and $\eta^h$. Cluster-robust standard errors in parentheses. Specification (1) includes measures of social capital, the number of bowling establishments and the social capital index of [Rupasingha and Goetz (2008)](#), spatially interpolated to congressional districts $N=15,420$. Specification (2) exclude both states (Texas and Georgia) where inter-census redistricting occurred, $N=14,150$. Specification (3) uses preferences estimated using MRP. See appendix B for more details. $N=15,647$.

*Redistricting*  Our analysis is confined to a single apportionment period during which district borders remain constant. The exceptions are several cases of court-ordered redistricting in Georgia and Texas. We exclude these two states in our second robustness test and find that our results are virtually unchanged.

*MRP estimated preferences*  An alternative approach to estimating district preferences is to use multilevel regression followed by poststratification ([Lax and Phillips 2009](#) or [Gelman 2014](#)). We discuss the differences in statistical assumptions made by the two approaches in detail in Appendix B. Here, we show in specification (3) that using estimates based on the MRP methodology yields results that are qualitatively similar to ours. Estimated marginal effects for responsiveness towards low income constituents are somewhat larger at about 12 ($\pm 2$) percentage points while marginal effects for high income constituents are more pronounced as well. In Table B.1 in the online appendix we estimate more specifications and show that responsiveness estimates based on MRP preferences are always somewhat larger than the ones based on matching using chained Random Forests. In
the same table we also show that our core results are also obtained when simply aggregating raw preference data from the CCES.

**Additional robustness tests** In Appendix E we report additional ‘technical’ robustness tests, such as removing extreme district preferences in each district, accounting for measurement error in district preferences, or using the robust trimmed linear probability estimator suggested by Horrace and Oaxaca (2006).

**IV.C. Relaxing modeling assumptions**

So far we have mainly studied the robustness of our results by adding potential confounders. In this subsection we implement two rather different statistical specifications in order deal with issues of omitted variable bias and functional form dependence.

*Post-double-selection estimator* Our first model, using the post-double-selection estimator (Belloni et al. 2014; Chernozhukov et al. 2015), addresses bias arising from omitted variables using two strategies. First, it constructs a high-dimensional vector of controls by allowing functional transforms of observables and their higher order interactions. It thus creates a partially linear model (Robinson 1988) using controls without the functional form restrictions commonly employed in the linear model. Second, it models both the legislative voting equation that we considered so far as well as “treatment” equations that model variation in the interaction of union membership and preferences. Importantly, the high-dimensional control vector enters both outcome and treatment equations. Out of the (possibly large) number of terms one selects confounders that predict both preferences and roll call votes using standard Machine Learning tools, such as the LASSO.\(^{21}\) The selected set of covariates is used in a post-LASSO estimation step to account for relevant confounders. The resulting estimator has low bias and yields accurate confidence intervals even under moderate selection mistakes (Belloni et al. 2014). Appendix F provides more technical details. Responsible for this robustness property is the LASSO step selecting the control set from both treatment and outcome equations. It finds controls whose omission leads to “large” omitted variable bias and includes them in the model. Any variables that are not included are therefore at most mildly associated to the treatment and the outcome, which decidedly limits the scope of omitted variable bias (Chernozhukov et al. 2015).

Table III shows the resulting estimates from three specifications. In the first one we include all district variables, their pairwise interactions and their interactions with district preferences, all in both linear and quadratic form. This leads to a vector of 144 covariate terms. In specification (2) we extend the set of possible controls and additionally include union policy variables and our measures of organizational capacity (as well as all their transforms) leaving us with 312 terms. Specification (3) allows for even more nonlinearity

\(^{21}\)The key is to transform this system of equations into one that represents a predictive relationship (where the application of machine learning tools such as the LASSO make sense).
Table III
Post-double-selection estimator. Marginal effect of unionization on legislative responsiveness to low and high income groups.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income preferences</td>
<td>0.063</td>
<td>0.066</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>High Income preferences</td>
<td>-0.054</td>
<td>-0.036</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Semi-parametric terms</td>
<td>144</td>
<td>312</td>
<td>624</td>
</tr>
<tr>
<td>post-LASSO terms</td>
<td>18</td>
<td>45</td>
<td>112</td>
</tr>
</tbody>
</table>

Note: Double Selection Estimator (Belloni et al. 2014) consists of LASSO selection of confounders in both outcome and union-preferences equations and post-selection least squares estimation of model; see Appendix F for details. Selection performed using root-LASSO (Belloni et al. 2011). We employ sample splitting: LASSO selection performed on 50% sample, parameter estimates performed on remaining 50% (N=7,884). Table entries are estimates for $\eta_L$ and $\eta_H$ with cluster-robust standard errors in parentheses. Specification (1) includes district characteristics in both linear and quadratic form and all their pairwise interactions. Specification (2) adds union policy and organizational capacity terms. Specification (3) additionally includes cubic splines (at four knots) of all terms.

by using cubic splines for all covariate terms leading to a high-dimensional vector of 624 controls. As the last line of Table III shows, the estimator selects a subset of these producing more flexible model specifications with the number of included controls ranging from 18 to 112. Even under these much more demanding specifications, we find that increasing unionization positively affects the representation of low-income constituents. A standard deviation increase in union membership increases legislators’ responsiveness to low-income preferences by about 6 to 7 percentage points, while decreasing the responsiveness to the preferences of the affluent by about 4 points. The magnitude of our estimates is in line with the ones we obtained in the richer specifications of our previous linear model (compare specifications (4) and (5) in Table I).

Kernel Regularized Least Squares (KRLS) While the previous modeling strategy is rather flexible, it did not relax one key assumption: the existence of an interaction between district preferences and union membership (our $\eta$ terms). This interaction is, of course, the center of our analysis and one might ask why its exclusion should be considered at all. The issue here is that we specify this interaction in a restrictive—linear—form, which might not be supported by the data and only found in our model estimates due to functional form misspecification. In a recent replication survey, Hainmueller et al. (2018) warn that “a large portion of published findings based on multiplicative interaction models are artifacts of misspecification or are at best highly model dependent.” It is thus prudent to consider an analysis that “lets the data
speak”. In the model below, estimated using KRLS (Hainmueller and Hazlett 2014), we do not specify any interaction \textit{a priori}, nor do we specify any functional form.

Intuitively, one can think of KRLS as a local regression method, which predicts the outcome at each covariate point by calculating an optimally weighted sum of locally fitted functions. The KRLS algorithm uses Gaussian kernels centered around an observation. The weights are chosen to produce the best fit to the data.\textsuperscript{22} The benefit of this approach is twofold. First, it allows for an approximation of highly nonlinear and non-additive functional forms. Second, it allows us to check if the marginal effects of group preferences changes with levels of unionization \textit{without} explicitly specifying this interaction term. To do the latter, we calculate pointwise partial derivatives of district preferences with respect to levels of union membership (Hainmueller and Hazlett 2014: 156).

Figure IV summarizes results from this approach. It plots a locally smoothed summary of pointwise partial effects for low and high income group preferences (on the y-axis) against levels of union membership (on the x-axis). Perhaps unsurprisingly, we find that the assumption of an exactly linear interaction specification is too restrictive, especially in the case of the preferences of high income constituents.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Nonparametric estimate of interaction between union membership and preferences}
\end{figure}

\textit{Note:} This figure plots partial effects (summarized using thin-plate spline smoothing) of preferences of low and high income constituents on legislative votes at levels of district union membership. Estimates obtained via KRLS.

\textsuperscript{22}See Appendix G for details on the approach and parameter selection.
However, the most noteworthy result clearly is the fact that, using a non-parametric model not including an \textit{a priori} interaction between union membership and preferences, we find clear evidence that union membership moderates the relationship between preferences and legislative voting. For low income constituents, increasing district-level union membership steadily increases the marginal effect of their preferences on legislators’ vote choice. Moving from low levels of union membership (at the 25th percentile) to median levels of union membership increase low-income preference responsiveness by about 5 percentage points. An equally sized increase from the median to the 75th percentile increases responsiveness by almost 8 percentage points. We also find similar (albeit weaker) evidence for an interaction between high income group preferences and union membership.

V. Heterogeneity

\textit{Union type}  Is our finding driven by a particular type of union? A recent strand of research stresses the special characteristics of public unions and their political influence (e.g., Anzia and Moe 2016; Flavin and Hartney 2015). Hence, one may ask whether our findings mainly reflect the influence of private-sector unions since public sector unions are too narrow in their interests to mitigate unequal responsiveness. Panel (A) of Table IV provides some evidence on this question. The administrative forms used to measure union membership do not distinguish between private and public unions, and local unions may contain workers from both the private and the public sector. To calculate an approximate measure of district public union membership, we identify unions with public sector members (based on their name) and create separate union membership counts for “public” and the remaining “non-public” unions (see appendix A for details).

Our findings suggest that the coefficient for the impact of a district’s public union membership on the responsiveness of legislators to the preferences of the poor is sizable (at about 7 percentage points) and clearly statistically different from zero. At the same time, the coefficient for the remaining “non-public” unions is slightly reduced. The difference between the two estimates is not statistically distinguishable from zero. This finding does not support the hypothesis of a null-effect of public sector unions. It also suggests that the changing private-public union composition will not necessarily lead to less collective voice in Congress.

\textit{Bill ideology}  Panel (B) explores whether the effect of unions varies with the ideological direction of the bill that is voted on. Based on the partisan vote margin of the roll call vote, we define an indicator variable for conservative roll calls and estimate separate coefficients for each bill type. We find that union effects are relevant (and significant) for both bill types, they are larger for conservative votes. A standard deviation increase in union membership increases responsiveness to the preferences of low-income constituents by about 9 (±2) percentage points for conservative bills compared to about 5 (±1) points for liberal bills.
The difference is larger for the preferences of high income constituents. In both cases the difference in marginal effects between liberal and conservative bills is statistically significant. Our findings suggest that union influence is more relevant for bills that have (potentially) adverse consequences for low income constituents. We trace this issue further in the next specification.

### Table IV

<table>
<thead>
<tr>
<th>Effect heterogeneity. Marginal effects of unionization on legislative responsiveness to low and high income groups.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Private vs. Public unions</td>
</tr>
<tr>
<td>Public unions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Non-public unions</td>
</tr>
<tr>
<td>(B) Bill ideology</td>
</tr>
<tr>
<td>Conservative bill</td>
</tr>
<tr>
<td>Liberal bill</td>
</tr>
<tr>
<td>(C) AFL-CIO endorsement</td>
</tr>
<tr>
<td>No position</td>
</tr>
<tr>
<td>Endorsement</td>
</tr>
</tbody>
</table>

*Note: Estimates for $\eta^L$ and $\eta^H$ with cluster-robust standard errors in parentheses. N=15,780. Panel (A) shows separate effects for district counts of union members for unions classified as public or non-public (see text). Statistical tests for the difference in union type yield $p = 0.172$ for low income preferences and $p = 0.027$ for high income ones. Panel (B) estimates separate effects for bills classified as conservative or liberal based on their predominant party vote. Tests for significance of difference: $p = 0.009$ for low and $p = 0.000$ for high income preferences. Panel (C) classifies bills with economic content where the AFL-CIO has taken a public stand for or against it (depending on bill content). Tests for significance of difference: $p = 0.003$ for low income, $p = 0.049$ for high income preferences.*

**Union voting recommendations** In panel (C) we consider bills with economic content and that have (or have not) been endorsed explicitly by the largest union confederation, the AFL-CIO. Our definition of endorsement is based on voting recommendations made publicly by the AFL-CIO.\(^{23}\) AFL-CIO recommendations signal the salience of the issue to unions, and they were made for more than half of the votes in the analysis. Panel (C) shows that the impact of union membership on legislators’ responsiveness for bills especially relevant to low-income citizens is about 2 percentage points larger for votes on which the AFL-CIO had taken a prior position. This difference is statistically different from zero ($p = 0.003$).\(^{24}\) The fact that districts with higher union membership see better representation of the less affluent

---


\(^{24}\)For high-income preferences the estimate for $\eta^H$ is smaller for endorsed bills but still significantly different from zero.
more so when issues are salient to unions bolsters the interpretation that our main result is actually driven by unions’ capacity for political action. This finding is also consistent with micro-level studies of the effects of union position-taking (Ahlquist et al. 2014; Kim and Margalit 2017).

VI. EXPLORING POSSIBLE MECHANISMS

In this final empirical section, we assess two mechanisms of union influence discussed before: campaign contributions and partisan selection. If contributions are a channel of union influence, we should observe that (i) in districts where unions are stronger, local unions and their members contribute more to sitting members of Congress; and (ii) that these contributions are positively linked to legislative responsiveness. We examine both relationships in Panel (A) of Table V. The first two columns show district-level regressions (with state fixed effects) relating union strength to (logged) contributions. We find that under two specifications (with and without extensive district controls) an increase in union membership systematically increases the amount of contributions from labor in that district. Converted to Dollar amounts (following Duan (1983)), a standard deviation increase in union membership increases contributions from Labor by about $81,000. Our measure of contributions is calculated from raw campaign finance contribution data obtained from the Center for Responsive Politics. We sum contributions reported to the Federal Election Commission to candidates from the “labor” sector (excluding single-issue donations). Our count includes both individuals and PACs (but using either alone does not change our results).

The last two columns of Panel (A) examine how contributions moderate legislators’ responsiveness. Following the specification used in Table I, we estimate linear probability models regressing roll call votes on contributions interacted with constituency preferences, district fixed effects, and, in column (4), district covariates interacted with preferences. We find that in districts where labor contributions are higher, the marginal effect capturing a legislator’s responsiveness to the preferences of low income constituents is significantly higher. This holds when accounting for district characteristics in the second specification, which also hold constant the amount donated by business interests.

Turning to the selection of partisan politicians, if unions rally around Democratic candidates and manage to influence electoral outcomes through contributions and other mobilization efforts, we expect to find that higher union membership is associated with a higher probability of a Democratic candidate being elected. We examine this relationships in Panel (B). The first two columns show LPMs with state fixed effects modeling a Democrat being elected in a given district as a function of union membership (and district-level controls). We find our expectation to be borne out: an increase in union membership is significantly associated with an increase in the election probability of a Democratic candidate. Consistent with previous research (Rhodes and Schaffner 2017), the selection of Democratic legislators
Table V
Labor contributions and selection of Democratic legislators.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Contributions channel</strong></td>
<td>DV: Contrib.</td>
<td>DV: roll call</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Union membership</td>
<td>0.056</td>
<td>0.046</td>
<td>0.946</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.036)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Contributions × low income prefs.</td>
<td>0.946</td>
<td>0.865</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.034)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributions × high income prefs.</td>
<td>−0.735</td>
<td>−0.714</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.031)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B: Selection channel</strong></td>
<td>DV: Democrat</td>
<td>DV: roll call</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Union membership</td>
<td>0.161</td>
<td>0.106</td>
<td>0.576</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.012)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Democrat × low income prefs.</td>
<td>0.576</td>
<td>0.542</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Democrat × high income prefs.</td>
<td>−0.411</td>
<td>−0.423</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

District controls ✓ ✓

Note: Panel (A), column (1) shows district-level regression of (log) labor contributions on (log) union membership with state fixed effects. Column (2) adds district-level controls (population size, degree of urbanization, shares of female, Black, Hispanic, BA degrees, employed in manufacturing, median household income, organizational capacity). N=428 (at-large districts are excluded). Column (3) shows LPMs with district fixed effects for legislators’ vote as function of the interaction between (log) labor contributions and district preferences. Column (4) adds district-level controls interacted with preferences. N=428. Panel (B), columns (1) and (2) show district-level LPM with state fixed effects of presence of Democratic representative on (log) union membership. N=428. Columns (3) and (4) show LPMs with district fixed effects for legislators’ vote as function of the interaction between (log) labor contributions and Democratic representative. N=15776. All specifications employ cluster-robust standard errors.

is then associated with higher responsiveness to the preferences of low income constituents compared to their Republican counterparts as shown in the least two columns of Panel (B).

Local unions are not necessarily the primary actor lobbying Congress relative to state associations or national/international affiliates (Dark 1999). The evidence that district-level union membership nonetheless matters for legislative responsiveness is consistent with the argument that local union strength underpins a credible threat of mobilization that shapes political equality through political selection and post-electoral incentives. The importance of electoral selection visible in our results is in line with a larger body of research on elections and representation (Bartels 2016; Lee et al. 2004; Miller and Stokes 1963). Mobilization efforts by unions remain strongly linked to available human resources on the ground (Rosenfeld 2014; Zullo 2008). As has already been shown by Berelson et al. (1954), local unions provide an
important social basis for electoral mobilization. Furthermore, national associations may also have incentives to target contributions to districts where unions are stronger, to demonstrate that members’ contributions are used in an effective way. Finally, recent evidence also shows that the presence of local unions is linked to the perceptions of constituent preferences by congressional staffers. Hertel-Fernandez et al. (2018) find that congressional staffers’ views are biased toward the preferences of conservative and business interest groups (also see Broockman and Skovron 2018). Strikingly, however, they find that this bias declines as district-level union membership increases. This is consistent with the (old) argument that the visible presence of an organized group in a district makes legislators more alert to its preferences (Arnold 1990; Miller and Stokes 1963).

In sum, we find that the political power of unions rests in part on their ability to mobilize campaign contributions and to help getting Democratic candidates elected. Consistent with arguments based on mobilization threats and rational politicians, these results also help to explain the puzzle documented by previous studies that inequalities in turnout or contacting officials alone do not appear to explain most of the observed income gap in political responsiveness (Bartels 2008; Ellis 2013; Erikson 2015).

VII. Conclusion

As Dahl (1961) famously asked, who governs in a polity where political rights are equally distributed, but where large inequalities in income and wealth (may) bias representation? In the wake of rising income inequality in the United States and other advanced economies, scholars have identified the question of political inequality as one of the central challenges facing democracy in the twenty-first century (see, for example, the report of the task force on Inequality and Democracy of the American Political Science Association (APSA Task Force 2004)). While the scientific debate is ongoing and some results are open to different interpretations (Erikson 2015), a growing number of studies has documented striking patterns of unequal responsiveness by income. When policy preferences diverge across income groups, legislators and public policy are biased toward the affluent at the expense of the middle-class and—especially—the poor. Many recent works conclude by asking what factors may improve political representation of the economically disadvantaged.

We contribute to this body of research by analyzing whether labor unions serve as a collective voice institution that limits unequal representation in the House of Representatives. Against the wide-spread view that unions are either too weak or too narrow to mitigate political inequality in the national arena, we find that the district-level strength of unions is clearly linked to the responsiveness of legislators to different income groups. While legislators are, on average, more responsive to the preferences of the affluent than to the preferences of the poor, this representation gap is highly variable. It is much less pronounced in districts where union membership is relatively higher. This result is in line with evidence on state-level policy responsiveness (Flavin 2018).
Our findings cast a somewhat less pessimistic light on democratic representation in Congress. Despite high income inequality, polarization, expensive campaigns, and a legislature dominated by affluent politicians (Carnes 2013; Gilens 2012; Hacker and Pierson 2010; McCarty et al. 2006), our evidence indicates that unequal representation is not hard-wired into the fabric of American democracy. We also find suggestive evidence that public sector unions, to whom union membership has been shifting over the last decades, do not appear to be less of a collective voice for the less well-off than private sector unions.

Admittedly, the observational nature of our data makes it challenging to draw causal conclusions. However, our within-district research design combined with rich data on possible confounds and flexible statistical specifications allows us to rule out a host of alternative explanations. Going beyond the few existing studies that directly examine the effect of unions on unequal representation, we demonstrate that the moderating effect of unions on legislative responsiveness is not simply a result of state-level policies or institutions, district-level socio-economic structure, workers’ propensity to organize, or broader patterns of associational life, and it is robust to relaxing parametric modeling assumptions. Our empirical strategy was made possible by combining local-level administrative data on unions with extensive public opinion data capturing within-district variation in opinion polarization across numerous issues. As a result, our interpretation of the results is that it is unlikely that the effects of unions are spurious. More broadly, a focus on real-world variation in mass organizations is a necessary complement to field-experimental studies of unequal responsiveness and their ability to isolate biases in response to personal contacts as well as the effectiveness of particular strategies of influence (Butler 2014; Kalla and Broockman 2016).

Our findings have important implications for the direction of future research on representation. First, they encourage research on unequal representation to pay more attention to unions. Beyond Congress, our data on local unions can also be mapped to districts of state legislatures. Similarly, existing work in the nascent comparative literature on the topic has directed its focus on political institutions (Bartels 2017; Lupu and Warner 2017); including the role of labor unions—traditionally a strong force in many European countries—would paint a clearer picture of the drivers of equal versus unequal representation of citizens’ interests in the political arena. Second, a fuller understanding of representation requires going beyond taking citizens’ preferences as given. Unions are a prime target for studying how economic groups may shape mass preferences as well as political responses to those preferences. Unions’ influence on preferences may work through leadership or socialization (Ahlquist et al. 2014; Kim and Margalit 2017) but also through directly through labor markets and economic inequality (Ahlquist 2017).
Appendices

A. Data

In this appendix we present additional details on our dataset including details on the creation of some control variables and descriptive statistics.

Matched roll calls  Table A.1 displays Congressional roll calls matched to CCES items. We selected congressional roll calls based on content and, when several choices were available, based on their proximity to CCES fieldwork periods.

Income thresholds  Table A.2 presents an overview of the income thresholds we use to classify CCES respondents into income groups. We use two thresholds separating the lowest and highest income terciles. We calculate them from yearly American Community Survey files excluding individuals living in group quarters. For each congress, Table A.2 shows the average of all district-specific thresholds as well as the smallest and largest ones.

Descriptive statistics  Table A.3 shows descriptive statistics for all variables used in our analysis. Note that these are for the untransformed variables. In our empirical models, we standardize all inputs to have mean zero and unit standard deviation.

Public unions  Public unions captured (by name) in our data include the American Federation of State, County & Municipal Employees, National Education Association, American Federation of Teachers, American Federation of Government Employees, National Association of Government Employees, United Public Service Employees Union, National Treasury Employees Union, American Postal Workers Union, National Association of Letter Carriers, Rural Letter Carriers Association, National Postal Mail Handlers Union, National Alliance of Postal and Federal Employees, Patent Office Professional Association, National Labor Relations Board Union, International Association of Fire Fighters, Fraternal Order of Police, National Association of Police Organizations, various local police associations, and various local public school unions.
**Table A.1**
Matched CCES–House roll calls included in our analysis.

<table>
<thead>
<tr>
<th>Match</th>
<th>Bill</th>
<th>Date</th>
<th>Name</th>
<th>House Vote (Yea-Nay)</th>
<th>Bill Ideology†</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>HR 810</td>
<td>07/19/2006</td>
<td>Stem Cell Research Enhancement Act (Presidential Veto override)</td>
<td>235-193</td>
<td>L</td>
</tr>
<tr>
<td>(1)</td>
<td>HR 3</td>
<td>01/11/2007</td>
<td>Stem Cell Research Enhancement Act of 2007 (House)</td>
<td>253-174</td>
<td>L</td>
</tr>
<tr>
<td>(2)</td>
<td>HR 2956</td>
<td>07/12/2007</td>
<td>Responsible Redeployment from Iraq Act</td>
<td>223-201</td>
<td>L</td>
</tr>
<tr>
<td>(3)</td>
<td>HR 2</td>
<td>01/10/2007</td>
<td>Fair Minimum Wage Act</td>
<td>315-116</td>
<td>L</td>
</tr>
<tr>
<td>(4)</td>
<td>HR 4297</td>
<td>12/08/2005</td>
<td>Tax Relief Extension Reconciliation Act (Passage)</td>
<td>234-197</td>
<td>C</td>
</tr>
<tr>
<td>(4)</td>
<td>HR 4297</td>
<td>05/10/2006</td>
<td>Tax Relief Extension Reconciliation Act (Agreeing to Conference Report)</td>
<td>244-185</td>
<td>C</td>
</tr>
<tr>
<td>(6)</td>
<td>S 1927</td>
<td>08/04/2007</td>
<td>Protect America Act</td>
<td>227-183</td>
<td>C</td>
</tr>
<tr>
<td>(6)</td>
<td>HR 6304</td>
<td>06/20/2008</td>
<td>FISA Amendments Act of 2008</td>
<td>293-129</td>
<td>C</td>
</tr>
<tr>
<td>(7)</td>
<td>HR 3162</td>
<td>08/01/2007</td>
<td>Children’s Health and Medicare Protection Act</td>
<td>225-204</td>
<td>L</td>
</tr>
<tr>
<td>(7)</td>
<td>HR 976</td>
<td>10/18/2007</td>
<td>Children’s Health Insurance Program Reauthorization Act (Presidential Veto Override)</td>
<td>273-156</td>
<td>L</td>
</tr>
<tr>
<td>(7)</td>
<td>HR 3963</td>
<td>01/23/2008</td>
<td>Children’s Health Insurance Program Reauthorization Act (Presidential Veto Override)</td>
<td>260-152</td>
<td>L</td>
</tr>
<tr>
<td>(7)</td>
<td>HR 2</td>
<td>02/04/2009</td>
<td>Children’s Health Insurance Program Reauthorization Act</td>
<td>290-135</td>
<td>L</td>
</tr>
<tr>
<td>(8)</td>
<td>HR 3221</td>
<td>07/23/2008</td>
<td>Foreclosure Prevention Act of 2008</td>
<td>272-152</td>
<td>L</td>
</tr>
<tr>
<td>(9)</td>
<td>HR 3688</td>
<td>11/08/2007</td>
<td>United States-Peru Trade Promotion Agreement</td>
<td>285-132</td>
<td>C</td>
</tr>
<tr>
<td>(10)</td>
<td>HR 1424</td>
<td>10/03/2008</td>
<td>Emergency Economic Stabilization Act of 2008</td>
<td>263-171</td>
<td>L</td>
</tr>
<tr>
<td>(11)</td>
<td>HR 3080</td>
<td>10/12/2011</td>
<td>To implement the United States-Korea Trade Agreement</td>
<td>278-151</td>
<td>C</td>
</tr>
<tr>
<td>(12)</td>
<td>HR 3078</td>
<td>10/12/2011</td>
<td>To implement the United States-Colombia Trade Promotion Agreement</td>
<td>262-167</td>
<td>C</td>
</tr>
<tr>
<td>(13)</td>
<td>HR 2346</td>
<td>06/16/2009</td>
<td>Supplemental Appropriations, Fiscal Year 2009 (Agreeing to conference report)</td>
<td>226-202</td>
<td>L</td>
</tr>
<tr>
<td>(14)</td>
<td>HR 2831</td>
<td>07/31/2007</td>
<td>Lilly Ledbetter Fair Pay Act</td>
<td>225-199</td>
<td>L</td>
</tr>
<tr>
<td>(14)</td>
<td>HR 11</td>
<td>01/09/2009</td>
<td>Lilly Ledbetter Fair Pay Act of 2009</td>
<td>247-171</td>
<td>L</td>
</tr>
<tr>
<td>(14)</td>
<td>S 181</td>
<td>01/27/2009</td>
<td>Lilly Ledbetter Fair Pay Act of 2009</td>
<td>250-177</td>
<td>L</td>
</tr>
<tr>
<td>(15)</td>
<td>HR 1913</td>
<td>04/29/2009</td>
<td>Local Law Enforcement Hate Crimes Prevention Act</td>
<td>249-175</td>
<td>L</td>
</tr>
<tr>
<td>(16)</td>
<td>HR 1</td>
<td>02/13/2009</td>
<td>American Recovery and Reinvestment Act of 2009 (Agreeing to Conference Report)</td>
<td>246-183</td>
<td>L</td>
</tr>
<tr>
<td>(17)</td>
<td>HR 2454</td>
<td>06/26/2009</td>
<td>American Clean Energy and Security Act</td>
<td>219-212</td>
<td>L</td>
</tr>
<tr>
<td>(18)</td>
<td>HR 3590</td>
<td>03/21/2010</td>
<td>Patient Protection and Affordable Care Act</td>
<td>220-212</td>
<td>L</td>
</tr>
<tr>
<td>(19)</td>
<td>HR 3962</td>
<td>11/07/2009</td>
<td>Affordable Health Care for America Act</td>
<td>221-215</td>
<td>L</td>
</tr>
<tr>
<td>(20)</td>
<td>HR 4173</td>
<td>06/30/2010</td>
<td>Wall Street Reform and Consumer Protection Act of 2009</td>
<td>237-192</td>
<td>L</td>
</tr>
<tr>
<td>(21)</td>
<td>HR 2965</td>
<td>12/15/2010</td>
<td>Don’t Ask, Don’t Tell Repeal Act of 2010</td>
<td>250-175</td>
<td>L</td>
</tr>
<tr>
<td>(22)</td>
<td>S 365</td>
<td>08/01/2011</td>
<td>Budget Control Act of 2011</td>
<td>269-161</td>
<td>C</td>
</tr>
<tr>
<td>(23)</td>
<td>H CR 34</td>
<td>04/15/2011</td>
<td>House Budget Plan of 2011</td>
<td>235-193</td>
<td>C</td>
</tr>
<tr>
<td>(24)</td>
<td>H CR 112</td>
<td>05/28/2012</td>
<td>Simpson-Bowles/Copper Amendment to House Budget Plan</td>
<td>38-382</td>
<td>C</td>
</tr>
<tr>
<td>(25)</td>
<td>HR 8</td>
<td>08/01/2012</td>
<td>American Taxpayer Relief Act of 2012 (Levin Amendment)</td>
<td>170-257</td>
<td>L</td>
</tr>
<tr>
<td>(26)</td>
<td>HR 2</td>
<td>01/19/2011</td>
<td>Repealing the Job-Killing Health Care Law Act</td>
<td>245-189</td>
<td>C</td>
</tr>
<tr>
<td>(26)</td>
<td>HR 6079</td>
<td>07/11/2012</td>
<td>Repeal the Patient Protection and Affordable Care Act and [...]</td>
<td>244-185</td>
<td>C</td>
</tr>
</tbody>
</table>

*Note: The matching of roll calls to CCES items can be many-to-one.*

† Coding of a bill’s ideological character as (L)iberal or (C)onservative based on predominant support of bill by Democratic or Republican representatives, respectively.
### Table A.2
Distribution of district income-group reference points. Average threshold over all districts, smallest and largest value.

<table>
<thead>
<tr>
<th>Congress</th>
<th>33th percentile</th>
<th>67th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>109</td>
<td>38123</td>
<td>16800</td>
</tr>
<tr>
<td>110</td>
<td>40127</td>
<td>18000</td>
</tr>
<tr>
<td>111</td>
<td>39021</td>
<td>17500</td>
</tr>
<tr>
<td>112</td>
<td>37381</td>
<td>16500</td>
</tr>
</tbody>
</table>

Note: Calculated from American Community Survey 1-year files. Household sample excluding group quarters. Missing income information imputed using Chained Random Forests.

### Table A.3
Descriptive statistics of analysis sample

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roll-call vote: yea</td>
<td>0.568</td>
<td>0.495</td>
<td>0.000</td>
<td>1.000</td>
<td>15780</td>
</tr>
<tr>
<td>Constituent preferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>0.593</td>
<td>0.220</td>
<td>0.047</td>
<td>0.979</td>
<td>15934</td>
</tr>
<tr>
<td>High income</td>
<td>0.555</td>
<td>0.198</td>
<td>0.037</td>
<td>0.967</td>
<td>15934</td>
</tr>
<tr>
<td>Low-High Gap</td>
<td>0.172</td>
<td>0.121</td>
<td>0.000</td>
<td>0.588</td>
<td>15934</td>
</tr>
<tr>
<td>Population</td>
<td>7.022</td>
<td>0.723</td>
<td>4.697</td>
<td>9.980</td>
<td>15934</td>
</tr>
<tr>
<td>Share African American</td>
<td>0.124</td>
<td>0.146</td>
<td>0.004</td>
<td>0.680</td>
<td>15934</td>
</tr>
<tr>
<td>Share Hispanic</td>
<td>0.156</td>
<td>0.174</td>
<td>0.005</td>
<td>0.812</td>
<td>15934</td>
</tr>
<tr>
<td>Share BA or higher</td>
<td>0.275</td>
<td>0.097</td>
<td>0.073</td>
<td>0.645</td>
<td>15934</td>
</tr>
<tr>
<td>Median income [$10,000]</td>
<td>5.177</td>
<td>1.356</td>
<td>2.282</td>
<td>10.439</td>
<td>15934</td>
</tr>
<tr>
<td>Share female</td>
<td>0.508</td>
<td>0.010</td>
<td>0.462</td>
<td>0.543</td>
<td>15934</td>
</tr>
<tr>
<td>Manufacturing share</td>
<td>0.110</td>
<td>0.047</td>
<td>0.025</td>
<td>0.281</td>
<td>15934</td>
</tr>
<tr>
<td>Urbanization</td>
<td>0.790</td>
<td>0.199</td>
<td>0.213</td>
<td>1.000</td>
<td>15934</td>
</tr>
<tr>
<td>Certification elections [log]</td>
<td>3.347</td>
<td>0.861</td>
<td>0.000</td>
<td>5.100</td>
<td>15934</td>
</tr>
<tr>
<td>Congregations [per 1000 persons]</td>
<td>0.765</td>
<td>1.147</td>
<td>0.062</td>
<td>6.453</td>
<td>15934</td>
</tr>
</tbody>
</table>

Note: Calculated from American Community Survey, 2006-2013. Note that when entered in models, variables are scaled to mean zero and unit SD. Preference gap is absolute difference in preferences between low and high income constituents in sample. Urbanization is calculated as the share of the district population living in an urban area based on the Census’ definition of urban Census blocks (matched to congressional districts using the MABLE database). Congregations per 1000 inhabitants calculated from RCMS 2000 (spatially interpolated).
B. Estimation of District Preferences

In this section we describe how we estimate district-level preferences using three different strategies: (i) small area estimation using a matching approach based on random forests (which we use in the main text of our paper), (ii) estimation using multilevel regression and post-stratification (MRP), and (iii) unadjusted cell means. Each approach invokes different statistical and substantive assumptions. In the spirit of consilience, our aim here is to show that our substantive results do not depend on any particular choice.

B.1. Small Area Estimation via Chained Random Forests

The core idea of our small area estimation strategy is based on the fact that we have access to two samples: one that is likely not representative of the population of all Congressional districts (the CCES), while the second one is representative of district populations by virtue of its sampling design (the Census or American Community Survey). By matching or imputing preferences from the former to the latter based on a common vector of observable individual characteristics, we can use the district-representative sample to estimate the preferences of individuals in a given district.\(^{25}\)

Combining CCES and Census data using Random Forests Figure B.1 illustrates this approach in more detail. We have data from \(m\) individuals in the CCES and \(n\) individuals in the Census (with \(n \gg m\)). Both sets of individuals share \(K\) common characteristics \(Z_k\), such as age, race, or education. The first task at hand is then to match \(P\) roll call preferences \(Y_p\) that are only observed in the CCES to the census sample. This is a purely predictive task and it is thus well suited for machine learning approaches. We use random forests (Breiman 2001) to learn about \(Y_p = f(Z_1, \ldots , Z_K)\) for \(p = 1, \ldots , P\) using the algorithm proposed by Stekhoven and Bühlmann (2011). This approach has two key advantages. First, as is typical for approaches based on regression trees, it deals with both categorical and continuous data, allows for arbitrary functional forms, and can include higher order interactions between covariates (such as age \(\times\) race \(\times\) education). Second, we can assess the quality of the predictions based on our model before we deploy it to predict preferences in the Census. With the trained model in hand we can use \(\hat{f}(Z_1, \ldots , Z_K)\) in combination with observed \(Z\) in the Census sample to fill in preferences (i.e., completing the square in the lower right of Figure B.1). Using the completed Census data, we can estimate constituent district preferences as simple averages by district and income group since the Census sample is representative for each Congressional district’s population.

Data details Due to data confidentiality constraints the Census Bureau does not provide district identifiers in its micro-data records. Instead, it identifies 630 Public Use Microdata

\(^{25}\)See Honaker and Plutzer (2016) for a more explicit exposition of this idea, evidence for its empirical reliability, and a comparison to MRP estimates.
Illustration of Small Area Estimation of District Preferences.

We use a sample of \( m \) individuals from the CCES that is not necessarily representative on the district-level, while a sample of \( n \) individuals from the Census is representative of district populations by design (Torrieri et al. 2014: Ch.4). We have access to bridging covariates \( Z_k \) that are common to both samples, while roll call preferences \( Y_p \) are only observed in the CCES. We train a flexible non-parametric model relating \( Y_p \) to \( Z \) and use it to predict preferences \( Y_{p}^{*} \) for Census individuals with characteristics \( Z \). With preference values filled in, a district’s income-group specific roll call preference can be estimated as the average of all units in that district.

Areas. We create a synthetic Census sample for Congressional districts by sampling individuals from the full Census PUMA regions proportional to their relative share in a given districts. This information is based on a crosswalk from PUMA regions to Congressional districts created by recreating one from the other based on Census tract level population data in the MABLE Geocorr2K database. The ‘donor pool’ for this synthetic sample are the 1% extracts for the American Community Survey 2006-2011. We limit the sample to non-group quarter households and to individuals aged 17 and older providing us with data on 14 million (13,711,248) Americans. From this we create the synthetic district file which is comprised of 3,040,265 cases. This provides us with a Census sample including Congressional district identifiers. The sample for each district is representative of the district population (save for errors induced by the crosswalk). We thus use the distribution of important population characteristics (age, gender, education, race, income) to match data on policy preferences from the CCES.

We harmonize all covariates to be comparable between CCES and Census. For family income this entails an adjustment to the measure provided in the CCES. It asks respondents to place their family’s total household income into 14 income bins.\(^{26}\) We transform this discretized measure of income into a continuous one using a nonparametric midpoint

\(^{26}\)The exact question wording is: “Thinking back over the last year, what was your family’s annual income?” The obvious issue here is that it is not clear which income concept this refers to (or, rather, which on the
Pareto estimator (Henson 1967). It replaces each bin with its midpoint (e.g., the third category $20,000$ to $29,999$ gets assigned $25,000$), while the value for the final, open-ended, bin is imputed from a Pareto distribution (e.g., Kopczuk et al. 2010). Using midpoints has been recognized for some time as an appropriate way to create scores for income categories (without making explicit distributional modeling assumptions). They have been used extensively, for example, in the American politics literature analyzing General Social Survey (GSS) data (Hout 2004).

**Algorithm details**  For easier exposition define a matrix $D$ that contains both individual characteristics and roll call preferences. Let $N$ be the number of rows of $D$. For any given variable $v$ of $D$, $D_v$, with missing entries at locations $i_{mis}^{(v)} \subseteq \{1, \ldots, N\}$ we can separate out four parts: 

- Observed values of $D_v$: denoted as $y^{(v)}_{obs}$
- Missing values of $D_v$: $y^{(v)}_{mis}$
- Variables other than $D_v$ with available observations $i^{(v)}_{obs}$, $x^{(v)}_{obs}$
- Variables other than $D_v$ with observations $i^{(v)}_{mis}$, $x^{(v)}_{mis}$

We now cycle through variables iteratively fitting random forest and filling in unobserved values until a stopping criterion $c$ (indicating no further change in filled-in values) is met. Algorithmically, we proceed as follows:

---

**Algorithm 1 Chained Random Forests**

1. Start with initial guesses of missing values in $D$
2. $w \leftarrow$ vector of column indices sorted by increasing fraction of NA
3. **while** not $c$ **do**
4. $D_{old}^{imp} \leftarrow$ previously imputed $D$
5. **for** $v$ in $w$ **do**
6. Fit Random Forest: $y^{(v)}_{obs} \sim x^{(v)}_{obs}$
7. Predict $y^{(v)}_{mis}$ using $x^{(v)}_{mis}$
8. $D_{new}^{imp} \leftarrow$ updated imputed matrix using predicted $y^{(v)}_{mis}$
9. Updated stopping criterion $c$
10. Return completed $D^{imp}$
---

To assess the quality of this scheme, we inspect the prediction error of the random forests using the out-of-bag (OOB) estimate (which can be obtaining during the bootstrap for each respondent employs). In line with the wording used in many other US surveys, we interpret it as referring to market income.

---

27 Note that this setup deals transparently with missing values in individual characteristics (such as missing education).
We find it to be rather small in our application: most normalized root mean squared errors are around 0.11. This result is in line with simulations by Stekhoven and Bühlmann (2011) who compare it to other prediction schemes based on K nearest neighbors, EM-type LASSO algorithms, or multivariate normal schemes and find it to perform comparatively well with both continuous and categorical variables.\footnote{See Tang and Ishwaran (2017) for further empirical validation of this strategy. See also Honaker and Plutzer (2016), who compare a similar matching strategy (but based on a multivariate normal model) with MRP estimated preferences using the CCES.}

\section*{B.2. Multilevel Regression and Poststratification}

The approach described in the last section is closely related to MRP (Gelman and Little 1997; Park et al. 2006; Lax and Phillips 2013), which has become quite popular in political science. Both strategies involve fitting a model that is predictive of preferences given observed characteristics followed by a weighting step that re-balances observed characteristics to their distribution in the Census. What differentiates MRP from the previous approach is that it imposes more structure in the modeling step both in terms of functional form and distributional assumptions. By utilizing the advantages of hierarchical models with normally distributed random coefficients it produces preference estimates that are shrunken towards group means (Gelman et al. 2013: 116f.).\footnote{This might be especially appropriate when some groups are small. The median number of respondents per district in the CCES is 506 and no district has fewer than 192 sampled respondents. But since we slice preferences further by income sub-groups, one may be worried that the sample size in some districts is small. MRP deals with this potential issue at the cost of making distributional assumptions.} No such structural assumptions are made when matching preferences to the Census using Random Forests. It will thus be instructive to compare how much our results depend on such modeling choices, which we do in the next section.

\textit{MRP implementation} For each roll call item in the CCES we estimate a separate model expressing the probability of supporting a proposal as a function of demographic characteristics. The demographic attributes included in our model broadly follow Lax and Phillips (2009, 2013) and are race, gender, education, age, and income.\footnote{We also estimated a version of the model including a macro-level predictor, which has been found to improve the quality of the model. We use the demographically purged state predictor of Lax and Phillips (2013: 15), that is, the average liberal–conservative variation in state-level public opinion that is not due to variation in demographic predictors. In our case this produces rather similar MRP estimates.} Race is captured in three categories (white, black, other), education in five (high school or less, some college, 2-year college degree, 4-year college degree, graduate degree). Age is comprised of 6 categories (18-29, 30-39, 40-49, 50-59, 60-69, 70+) while income is comprised of 13 categories (with thresholds 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 100, 120, 150 [in $1,000]). Our model also includes district-specific intercepts. For each roll-call, we estimate the following hierarchical
model using penalized maximum likelihood \citep{Chung2013}:

$$Pr(Y_i = 1) = \logit^{-1} \left( \beta_0 + \alpha_{race}^{i[j]} + \alpha_{gender}^{k[i]} + \alpha_{age}^{l[i]} + \alpha_{educ}^{m[i]} + \alpha_{income}^{n[i]} + \alpha_{district}^{d[i]} \right)$$ \hfill (B.1)

We employ the notation of \citet{Gelman2007} and denote by \(j[i]\) the category \(j\) to which individual \(i\) belongs. Here, \(\beta_0\) is an intercept and the \(\alpha_s\) are hierarchically modeled effects for the various demographic groups. Each is drawn from a common normal distribution with mean zero and estimated variance \(\sigma^2\):

\[
\begin{align*}
\alpha_{race}^{j} & \sim N \left( 0, \sigma^2_{race} \right), \quad j = 1, \ldots, 3 \\
\alpha_{gender}^{k} & \sim N \left( 0, \sigma^2_{gender} \right), \quad k = 1, \ldots, 2 \\
\alpha_{age}^{l} & \sim N \left( 0, \sigma^2_{age} \right), \quad l = 1, \ldots, 6 \\
\alpha_{educ}^{m} & \sim N \left( 0, \sigma^2_{educ} \right), \quad m = 1, \ldots, 5 \\
\alpha_{income}^{n} & \sim N \left( 0, \sigma^2_{income} \right), \quad n = 1, \ldots, 13 
\end{align*}
\hfill (B.2-6)
\]

This setup induces shrinkage estimates for the same demographic categories in different districts. Note that using fixed effects for characteristics with few categories (Specifically, gender) does not impact our results. The district intercepts are drawn from a normal distribution with state-specific means \(\alpha_{s[d]}\) and freely estimated variance:

$$\alpha_{d} \sim N \left( \alpha_{state}^{s[d]}, \sigma^2_{state} \right).$$ \hfill (B.7)

Our final preferences estimates for each income group on each roll call are obtained by using cell-specific predictions from the above hierarchical model, weighted by the population frequencies (obtained from our Census file) for each cell in each congressional district.

### B.3. Model results under various preference estimation strategies

The estimates of district-level preferences obtained via our SAE approach and MRP are in broad agreement: The median difference in district preferences between SAE and MRP is 2.5 percentage points for low income and −0.1 percentage points for high income constituents. A large part of this difference is due to the heavier tails of the distribution of district preferences for each roll call estimated by our approach—perhaps not surprising given the shrinkage characteristics of MRP. To what extent do these differences in the distribution of preferences affect our estimated union effects?

Table B.1 shows estimates for our six main specifications using three different measurement strategies for district preferences. Panel (A) shows our approach contrasted to MRP-based preferences in panel (B). The results are unequivocal: using MRP estimated preferences leads to more pronounced estimates in all specifications. Using specification (6),
which includes state policies, measures of district organizational capacity, district covariates interacted with preferences, as well as district fixed effects, we find that a unit increase in union membership increased responsiveness of legislators towards the preferences of low income constituents by about 12 (±2) percentage points (compared to only 6 points using our measurement strategy). Responsiveness estimated for high income preferences are similarly larger. Note that while larger, all estimates also carry increased confidence intervals.

Table B.1
Model results using different strategies to estimate district-level preferences. Entries are marginal effects of standard deviation increase in union membership on marginal effect of income group preferences on legislator vote.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: Small Area Estimation via Chained Random Forests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.106</td>
<td>0.082</td>
<td>0.098</td>
<td>0.084</td>
<td>0.068</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.063</td>
<td>-0.036</td>
<td>-0.053</td>
<td>-0.051</td>
<td>-0.050</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td><strong>B: Multilevel Regression &amp; Poststratification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.182</td>
<td>0.158</td>
<td>0.181</td>
<td>0.162</td>
<td>0.115</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.022)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.136</td>
<td>-0.119</td>
<td>-0.139</td>
<td>-0.122</td>
<td>-0.091</td>
<td>-0.091</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.021)</td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td><strong>C: Raw CCES means</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.080</td>
<td>0.061</td>
<td>0.063</td>
<td>0.072</td>
<td>0.043</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.027</td>
<td>-0.013</td>
<td>-0.010</td>
<td>-0.027</td>
<td>-0.018</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.009)</td>
</tr>
</tbody>
</table>

District fixed effects ✓ ✓ ✓ ✓ ✓ ✓
Group preferences
× union policy ✓
× state constants ✓
× organizational capacity ✓ ✓
× district covariates ✓ ✓

Note: Replicates Table I in the main text using different strategies to estimate district-level preferences of three income groups.

As a further point of comparison, panel (C) shows preferences estimated via raw cell means in the CCES. Due to the the issues discussed above, the raw data should not be taken
as a yardstick, but it is nonetheless informative to see how much the results vary. Our core results even obtain when we simply use raw cell means without any statistical modeling to counter non-representative distributions of individual characteristics and small cell sizes. We find that in our strictest specification, a unit increase in union membership still increases responsiveness towards low income constituents by about 5 (±1) percentage points.

In sum, all three approaches lead to the same qualitative conclusions about the moderating effect of unions on unequal representation in Congress. The two alternative approaches to deal with the problem that CCS surveys are not representative of congressional districts by design suggest that a larger effect of unions than the naive approach using the unadjusted survey data. Quantitatively, our preferred estimates are based on small area estimation via random forests as they are less reliant on normality assumptions and are systematically more conservative than those based on MRP.

C. Alternative Income Thresholds

This section discusses the impact of different income thresholds on our results. Panel (A) of Table C.1 replicates Table I in the main text. Here, preferences of income groups are based on a district-specific income thresholds splitting the population into three groups (at the 33rd and 66th percentile). Thus, in our model voters are classified as ‘low income’ relative to other voters in their congressional district. For example, during the 111th Congress a voter with an income of $40,000 would be part of the low income group in most of Massachusetts’ districts (where low income thresholds vary from about $40,000 to $50,000), but not in the 8th (where the threshold is about $30,000). If income threshold were state-specific instead, he or she would be considered low income everywhere in the state (as the state-specific low income threshold is now ≈$47,000). Not all states display as much variation in income-group thresholds. Thus, using state- instead of district-specific thresholds does not alter our core results in an appreciable way. As Panel (B) shows, the resulting marginal effects estimates for all six model specifications are remarkably similar when using preferences of income groups defined by state-specific thresholds. In panel (C) we no longer divide the population into three equally sized income groups. Instead, we restrict the low-income group to only those below the 20th percentile of the (district-specific) income distribution. Similarly, we classified as high income only those above the 80th percentile. Our resulting estimates for the union-responsiveness marginal effects are slightly smaller, but still of a substantively relevant magnitude and statistically different from zero.
Model results using different definitions of income groups. Marginal effect of standard deviation increase in union membership on marginal effect of income group preferences on legislator vote.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A: District-specific income thresholds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.106</td>
<td>0.082</td>
<td>0.098</td>
<td>0.084</td>
<td>0.068</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.063</td>
<td>-0.036</td>
<td>-0.053</td>
<td>-0.051</td>
<td>-0.050</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td><strong>B: State-specific income thresholds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.105</td>
<td>0.082</td>
<td>0.097</td>
<td>0.083</td>
<td>0.067</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.062</td>
<td>-0.036</td>
<td>-0.052</td>
<td>-0.050</td>
<td>-0.049</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td><strong>C: Shifted income thresholds: p20 - p80</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income preferences</td>
<td>0.098</td>
<td>0.077</td>
<td>0.09</td>
<td>0.078</td>
<td>0.063</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>High income preferences</td>
<td>-0.054</td>
<td>-0.031</td>
<td>-0.046</td>
<td>-0.044</td>
<td>-0.044</td>
<td>-0.034</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>District fixed effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Group preferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× union policy</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× state constants</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× organizational capacity</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>× district covariates</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: Replicates Table I in the main text using income groups defined via different income thresholds.
D. Measures of District Organizational Capacity

In the empirical analysis reported in the main text, we use two proxies for the organizational capacity of workers, union certification elections and the number of religious congregations. Here we provide some background and explain in more detail how we calculate both variables.

*NLRB certification elections* The formation of unions is regulated by the National Labor Relations Act (NLRB) enacted in 1935 (see Budd 2018: ch. 6). A successful union organization process usually requires an absolute majority of employees voting for the proposed union in a certification election held under the guidelines of the NLRB. Getting the NLRB to conduct an election requires that there is sufficient interest among employees in an appropriate bargaining unit to be represented by a union. For proof of sufficient interest, the NLRB requires that at least 30% of employees sign an authorization card stating they authorize a particular union to represent them for the purpose of collective bargaining. Building support and collecting the required signatures takes organizational effort. For workers, unionization has features of a public good. Everybody may gain through better conditions from collective bargaining, but contributing to the organizational drive is costly for each individual. Beyond mere opportunity costs, there also is a non-zero risk of being (illegally) fired by the employer for those especially active. If more than 50% of employees sign authorization cards, then the union can request voluntary recognition without a certification election. However, the employer has the right to deny this, in which case a certification election is held. In his labor relations textbook, Budd (2018: 199) notes that voluntary card check recognition is “the exception rather than the norm because employers typically refuse to recognize unions voluntarily.”

We use the NLRB’s database on election reports to extract all attempts to certify (or de-certify) a local union. They are available from www.nlrb.gov. Each database entry is a vote concerning a bargaining unit; the average unit size is 25 employees. There are about 2200 elections each year. Each individual case file usually provides address information on the employer and the site where the election was held. Using this information, we geocode each individual case report and locate it in a congressional district. Figure D.1 shows the resulting variation in certification elections over districts.

*Congregations* As a proxy for district level social capital we use the number of congregations per inhabitant. The number of congregations in a given district is not readily available for the years covered in our study. Therefore, we spatially aggregate county-level measures from the 2010 Religious Congregations and Membership Study to the congressional district level using areal interpolation techniques that take into account the population distribution between counties and districts. We use a geographic country-to-district equivalence file calculated from Census shapefiles. This is combined with population weights for each country-district intersection derived using the Master Area Block Level Equivalency
Figure D.1
Total number of union certification elections in House districts (109th-112th Congress).

database v1.3.3 (available from the Missouri Census Data Center), which calculates them based on about 5.3 million Census blocks. With these weights in hand we can interpolate county-level to district-level congregation counts using weighted means (for states with at-large districts, this reduces to a simple summation, as counties are perfectly nested within districts).
E. ADDITIONAL ROBUSTNESS TEST

In this section we describe several additional robustness tests.

1:1 mapping of CCES preferences to roll calls

We begin by limiting our sample by creating a unique mapping between preferences and roll call votes. Some of our CCEs preferences estimates are linked to more than one Congressional roll call. To investigate if this affects our results, specification (1) uses a 1:1 map dropping additionally available roll calls after the first match. This reduces the sample size to 11,104 respondents. We find that our results are not influenced by this change.

Table E.1

<table>
<thead>
<tr>
<th></th>
<th>Low income preferences</th>
<th>High income preferences</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Injective preference roll call map</td>
<td>0.063 (0.013)</td>
<td>−0.041 (0.013)</td>
<td>11,104</td>
</tr>
<tr>
<td>(2) Extreme preferences excl.</td>
<td>0.074 (0.016)</td>
<td>−0.048 (0.015)</td>
<td>13,308</td>
</tr>
<tr>
<td>(3) New York excluded</td>
<td>0.070 (0.015)</td>
<td>−0.048 (0.014)</td>
<td>14,730</td>
</tr>
<tr>
<td>(4) Local Union Concentration</td>
<td>0.065 (0.014)</td>
<td>−0.047 (0.014)</td>
<td>15,780</td>
</tr>
<tr>
<td>(5) Trimmed LPM estimator</td>
<td>0.074 (0.015)</td>
<td>−0.055 (0.014)</td>
<td>15,426</td>
</tr>
<tr>
<td>(6) Errors-in-variables</td>
<td>0.062 (0.004)</td>
<td>−0.054 (0.004)</td>
<td>15,345</td>
</tr>
</tbody>
</table>

Note: Based on specification (5) of Table I. (4) used trimmed estimator of Horrace and Oaxaca (2006). Specification (5) shows results from an errors-in-variables model implemented in a Bayesian framework. See text for details. Table entries are posterior means and standard deviations.

Extreme preferences excluded

In specification (2) we investigate if extreme district preferences on some roll calls drive our results. To do so, we trim the distribution of preferences at the bottom and the top. For each roll call we exclude districts with preference estimates below the 5th and above the 95th percentile. Using only trimmed preferences has no appreciable impact on our estimates.

New York excluded

Another test estimates our model with the state of New York excluded from the sample. In earlier work we found that our estimates of union strength correlate highly with aggregated state-level estimates derived from the Current Population survey. One state where this correlation is lower is New York (cf. Becher et al. 2018). In specification (3) we show that our results are not affected by its exclusion.

Union Concentration

Our data on local unions are from Becher et al. (2018), who also find that the local concentration of unions is an important dimension. While Becher et al. (2018) show that both dimensions (membership and concentration) vary independently, it is prudent to check if our results on the impact of union membership on representation
still obtain when accounting for the structure of union organization. In specification (4) we show this to be the case.

*Trimmed LPM estimator*  A fifth, more technical, specification implements the trimmed estimator suggested by Horrace and Oaxaca (2006). It accounts for the fact that we estimate a linear probability model to a binary dependent variable, which entails the possibility that the model-implied linear predictor lies outside the unit interval. Our results in Table E.1 indicate that this change does not materially affect our core results (if anything, they become slightly larger).

*Errors-in-variables*  Our final test accounts for the errors-in-variables problem caused by the fact that our district preference measures are based on estimates. While, in general, standard errors for our district-level estimates are quite small relative to the quantity being measured and one expects a downward bias in parameter estimates in a linear model with errors-in-variables, we estimate this specification to get a sense of the quantitative magnitude of the change in parameter estimates. We find that adjusting for measurement error produces very little quantitative change; both estimates are within the confidence bounds of our non-corrected estimates.

**F. Post-Double-Selection Estimator**

The post-double-selection models in the main text provide a relaxation of the linearity and exogeneity assumptions made in our main model. To do so we use the double-post-selection estimator proposed by Belloni et al. (Belloni et al. 2013, 2017). Specifically, this model setup aims to reduce the possible impact of omitted variable bias by accounting for a large number of confounders in the most flexible way possible. This can be achieved by moving beyond restricting confounders to be linear and additive, and instead considering a flexible, unrestricted (non-parametric) function. This leads to the formulation of the following partially linear model (Robinson 1988) equation (for ease of exposition we omit

---

31 We implement this model in a Bayesian framework, where we incorporate the measurement error model directly into the posterior distribution. To specify the variance of the measurement error for low and high income group preferences, we average the standard errors of the district-group means from the raw CCES data (pre-Census matching). Measurement error variance is slightly larger for low income preferences (0.029) than for high income preferences (0.025). We use the setup proposed in Richardson and Gilks (1993), implemented in Stan (v.2.17.0) and estimated (due to the size of our data set) using mean field variational inference. We use normal priors with mean zero and standard deviation (SD) of 100 for all regression coefficients, and inverse Gamma priors with shape and scale 0.01 for residuals. In the measurement error equation, we use normal priors with mean zero and SD of 10 for the mean of the measurement error and a student-t prior with 3 degrees of freedom and mean 1, SD 10 for the standard deviation of the measurement. The reported entries are posterior means and standard deviations.
district fixed effects in the notation and ignore $i$ subscripts):

$$y_{jd} = \mu^l\theta^l_{jd} + \mu^h\theta^h_{jd} + \eta^l U_d\theta^l_{jd} + \eta^h U_d\theta^h_{jd} + g(Z_d) + \epsilon_{jd} \quad (F.1)$$

with $E(\epsilon_{jd}|Z_s, U_d, \theta_{jd}) = 0$. Here, $y$ is the vote of a representative in a given district, $U_d$ is the level of union density. The function $g(Z_d)$ captures the possibly high-dimensional and nonlinear influence of confounders (interacted with income group preferences). The utility of this specification as a robustness tests stems from the fact that it imposes no a priori restriction on the functional form of confounding variables. A second key ingredient in a model capturing biases due to omitted variables is the relationship between the treatment (union density) and confounders. Therefore, we consider the following auxiliary treatment equation

$$U_d = m(Z_d) + \nu_i, \quad E(\nu_i|Z_d = 0), \quad (F.2)$$

which relates treatment to covariates $Z_d$. The function $m(Z_d)$ summarizes the confounding effect that potentially create omitted variable bias if $m \neq 0$, which is to be expected in an observational study such as ours.

The next step is to create approximations to both $g(\cdot)$ and $m(\cdot)$ by including a large number ($p$) of control terms $w_d = P(Z_d) \in \mathbb{R}^p$. These control terms can be spline transforms of covariates, higher order interaction terms, etc. Even with an initially limited set of variables, the number of control terms can grow large, say $p > 200$. To limit the number of estimated coefficients, we assume that $g$ and $m$ are approximately sparse (Belloni et al. 2013) and can be modeled using $s$ non-zero coefficients (with $s \ll p$) selected using regularization techniques, such as the LASSO (see Tibshirani 1996; see Ratkovic and Tingley 2017 for a recent exposition in a political science context):

$$y_{jd} = \mu^l\theta^l_{jd} + \mu^h\theta^h_{jd} + \eta^l U_d\theta^l_{jd} + \eta^h U_d\theta^h_{jd} + w_d' \beta_0 + r_{gd} + \zeta_{jd} \quad (F.3)$$

$$U_d = w_d' m_0 + r_{mi} + \nu_d \quad (F.4)$$

Here, $r_{gi}$ and $r_{mi}$ are approximation errors.

However, before proceeding we need to consider the problem that variable selection techniques, such as the LASSO, are intended for prediction, not inference. In fact, a “naive” application of variable selection, where one keeps only the significant $w$ variables in equation (F.3) fails. It relies on perfect model selection and can lead to biased inferences and misleading confidence intervals (see Leeb and Pötscher 2008). Thus, one can re-express the problem as one of prediction by substituting the auxiliary treatment equation (F.4) for $D_d$ in (F.3) yielding a reduced form equation with a composite approximation error (cf. Belloni et al. 2013). Now both equations in the system represent predictive relationships and are thus amenable to high-dimensional selection techniques.

Note that using this dual equation setup is also necessary to guard against variable selection errors. To see this, consider the consequence of applying variable selection tech-
niques to the outcome equation only. In trying to predict $y$ with $w$, an algorithm (such as LASSO) will favor variables with large coefficients in $\hat{\beta}_0$ but will ignore those of intermediate impact. However, omitted variables that are strongly related to the treatment, i.e., with large coefficients in $\hat{\beta}_{m0}$, can lead to large omitted variable bias in the estimate of $\eta$ even when the size of their coefficient in $\hat{\beta}_0$ is moderate. The Post-double selection estimator suggested by Belloni et al. (2013) addresses this problem, by basing selection on both reduced form equations. Let $\hat{I}_1$ be the control set selected by LASSO of $y_{jd}$ on $w_d$ in the first predictive equation, and let $\hat{I}_2$ be the control set selected by LASSO of $U_d$ on $w_d$ in the second equation. Then, parameter estimates for the effects of union density and the regularized control set are obtained by OLS estimation of equation (F.1) with the set $\hat{I} = \hat{I}_1 \cup \hat{I}_2$ included as controls (replacing $g(\cdot)$). In our implementation we employ the root-LASSO (Belloni et al. 2011) in each selection step.

This estimator has low bias and yields accurate confidence intervals even under moderate selection mistakes (Belloni and Chernozhukov 2009; Belloni et al. 2014). Responsible for this robustness is the indirect LASSO step selecting the $U_d$-control set. It finds controls whose omission leads to “large” omitted variable bias and includes them in the model. Any variables that are not included (“omitted”) are therefore at most mildly associated to $U_d$ and $y_{jd}$, which decidedly limits the scope of omitted variable bias (Chernozhukov et al. 2015).

G. Nonparametric Evidence for Union-Preferences Interaction

As discussed in the main text, we want to estimate a specification that makes as little a priori assumptions about functional form relationships between variables (including their interactions). Thus, we non-parametrically model $y_{ijd} = f(z)$ with $z = [\theta_{jd}^l, \theta_{jd}^h, U_d, X_d]$ by approximating it via Kernel Regularized Least Squares (Hainmueller and Hazlett 2014),

$$y = Kc.$$  \hspace{1cm} (G.1)

Here, $K$ is an $N \times N$ Gaussian Kernel matrix

$$K = \exp \left( \frac{-\|Z_d - z_j\|^2}{\sigma^2} \right) \hspace{1cm} (G.2)$$

with an associated vector of weights $c$. Intuitively, one can think of KRLS as a local regression method, which predicts the outcome at each covariate point by calculating an optimally weighted sum of locally fitted functions. The KRLS algorithm uses Gaussian kernels centered around an observation. The weights $c$ are chosen to produce the best fit to the data. Since a possibly large number of $c$ values provide (approximately) optimal weights it makes sense to prefer values of $c$ that produce “smoother” function surfaces. This is achieved via

\footnote{For a very general discussion see Belloni et al. (2017).}
regularization by adding a squared L2 penalty to the least squares criterion:

\[ c^* = \arg\min_{c \in \mathbb{R}^D} [(y - Kc)'(y - Kc) + \lambda c'Kc], \quad (G.3) \]

which yields an estimator for \( c \) as \( c^* = (K + \lambda I)^{-1}y \) (see Hainmueller and Hazlett 2014, appendix). This leaves two parameters to be set, \( \sigma^2 \) and \( \lambda \). Following Hainmueller and Hazlett (2014), we set \( \sigma^2 = D \) the number of columns in \( z \) and let \( \lambda \) be chosen by minimizing leave-one-out loss.

The benefit of this approach is twofold. First, it allows for an approximation of highly nonlinear and non-additive functional forms (without having to construct non-linear terms as we do in the post-double selection LASSO). Second, it allows us to check if the marginal effects of group preferences changes with levels of union density without explicitly specifying this interaction term (and instead learning it from the data). To do the latter one can calculate pointwise partial derivatives of \( y \) with respect to a chosen covariate \( z^{(d)} \) (Hainmueller and Hazlett 2014: 156). For any given observation \( j \) we calculate

\[ \frac{\partial y}{\partial z_{j}^{(d)}} = \frac{-2}{\sigma^2} \sum_{i} c_i \exp\left(\frac{-||Z_d - z_j||^2}{\sigma^2}\right) \left(Z_{ Ud} - z_{ Ud}^j\right). \quad (G.4) \]

These yields as many partial derivatives as there are cases. We apply a thin plate smoother (with parameters chosen via cross-validation) to plot these against district-level union membership in Figure IV.

References


