Political Impact of US Trade Unions

Matan Kolerman

November 2024

Click here for latest version

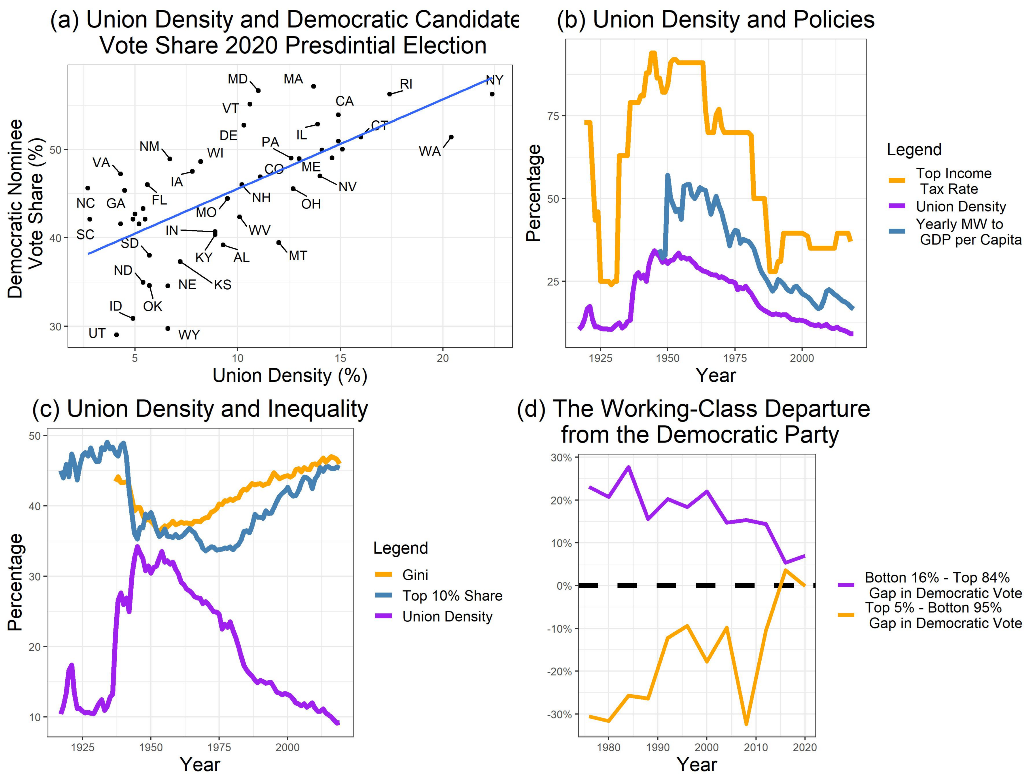
Abstract

I apply the novel RD Aggregation method to asses the political impact of US trade unions. The method allows the aggregation of several discontinuity events–close unionization elections–into a commuting zone level shock that measures the unions' "Luck" in each zone in each period. Using this methodology, I find that, on average, a newly unionized worker is worth 1.5 new votes for the Democratic party candidate in the following presidential elections and that unions shift local congressmen to the left. Further analysis suggests that the significant effects partly stem from increased campaign contributions, strategic political resource allocation by unions in areas with new members, and direct impact on union members.

1 Introduction

In the 2020 US presidential election, a very high correlation was observed between states' trade union density and states' vote share for Joe Biden. This high correlation is presented in panel a of fig. 1. The slope of the linear graph is 1.02 with a high R^2 (0.44). Appendix fig. A.1 shows that a similar correlation appeared in each presidential election since 1980. Panel b of fig. 1 presents a second correlation regarding the relationship between unions and the political field. It shows union density and two measures of policies promoting equality - minimum wage and top income tax rate. The correlation between union density and those measures is very high, 0.9 and 0.93, respectively. A possible explanation for this correlation is that unions have many channels to influence political results- lobbying, contributions, and mobilizing. They can use these channels to promote candidates who promote pro-worker policies and

Figure 1: Motivations



to motivate representatives endorsed by them to support such policies. In this paper, I will try to estimate the impact of unions on electoral outcomes. A substantial impact can explain both correlations.

A significant political effect of unions on election results may be one of the factors behind the long-established negative correlation between national union density and national inequality in the US. Panel c of fig. 1 is taken from Farber et al. (2021) and illustrates this negative relationship using the Gini index and the share of income held by the top 10%. Most of the literature that tries to explain this relationship causally emphasizes unions' effects on compensation and wage gaps in the labor market (DiNardo et al., 1995; Fortin et al., 2021). A complementary explanation can arise from the high impact of unions on the political field that materialized through pro-worker policies promoting equality.

The dramatic decline in US unionization rates over recent decades makes the political impact of US unions a particularly relevant research topic. Union density has fallen from 35% in the 1960s to just 9% (as shown in panel b of fig. 1), with a marked decrease in new unionization efforts since the Reagan era as demonstrated in fig. A.2. This notable decline aligns with large unions' political impact may help explain several key trends: the shift of the working class away from the Democratic Party depicted in panel d of fig. 1 and described more comprehensively and for many more western countries in Gethin et al. (2021), and the rise of the 'New Democrat' faction within the Democratic Party, known for appealing to more educated voters and their skepticism towards redistribution (Kuziemko et al., 2023).

Two primary challenges may account for the limited research on the impact of unions in the political realm. The first challenge is the significant shortage of data. There are almost no administrative datasets that include union membership information at the sub-national level, and the available survey data is notably limited.

Secondly, left-leaning parties and unions are often described as "Siamese Twins".[1]. In the context of empirical research, this close relationship translates into a high correlation between shocks to unions and shocks to left-leaning parties, making it difficult to isolate the causal effect of unions on voting. For example, the passage of states' right-to-work laws has been a massive negative shock to union membership in the US. Legislating these laws has been central to the Republican party's agenda. Thus, the legislation is negatively correlated with the share of Democratic Party votes (Feigenbaum et al., 2019). Moreover, when the Republican Party is in power, it pushes for additional legislation that limits voting accessibility. This practice is known to hurt the Democratic Party in future election cycles.

This paper proposes a novel approach to address the aforementioned challenges. In the US, the process of establishing a union in a workplace involves holding elections under the oversight of the National Labor Relations Board (NLRB), and data on these elections have been available since the 1960s. Close NLRB elections have been used as a source of exogenous variation in regression discontinuity (RD) studies that examine the impact of unions on economic indicators in the workplace, such as wages, employment, and workplace survival (DiNardo and Lee, 2004; Sojourner et al., 2015; Knepper, 2020; Frandsen, 2021; Matzat and Schmeißer, 2022). However, the challenge arises as data on national election voting are unavailable at the individual workplace level and are instead aggregated over larger geographic areas. This aggregation prevents the standard RD design from effectively estimating the impact of unionization on voting behaviors. In other words, within any given geographic area, there can be several close unionization elections per election cycle. This over-

[1]The term "Siamese Twins" to describe left parties and trade unions was first used by Viktor Adler (1852-1918), the founder and leader of the Austrian Social-Democratic Workers Party (1889). Many scholars (Padgett and Paterson, 1991; Ebbinghaus, 1995; Allern and Bale, 2017) borrowed the term to describe these two social institutions' shared history, culture, ideology, and interests

lap complicates the application of the RD method for assessing how unions influence national election outcomes.

To overcome this limitation, a new method will be offered, which is called RDA - "Regression Discontinuity Aggregation." RDA allows for the aggregation of several close unionization elections into a single shock to the share of newly unionized workers in a specific commuting zone during a specific election cycle. I will show that the shock exogeneity can arise from both frameworks for RDD - Local Polynomial Regression Discontinuity (Lee, 2008; Cattaneo and Titiunik, 2021) and Local Randomization Regression Discontinuity (Cattaneo et al., 2015). Several simple tests will be offered to show their validity in the context of this specific paper.

The intuition behind the method is to estimate unions' "Luck" in each CZ and each period. Luck is defined as the difference between the observed share of newly unionized workers through close elections and the expected share.

Using this method, I find that each newly unionized worker contributed, on average, an additional 1.4 to 1.6 votes to the Democratic party's presidential nominee. These results remain robust across various specifications and have been validated through a placebo test. When analyzing congressional election returns, similar patterns emerge, albeit with larger standard errors. A linear projection analysis indicates that this effect persists over multiple election cycles. Further investigation into the mechanisms reveals that the observed impact cannot be solely explained by increased voter turnout among Democratic-leaning voters or shifts in political alignment following successful unionization. Instead, the effect likely operates through indirect channels. Evidence points to two primary pathways: a rise in political donations and unions directing more resources to areas with newly unionized members. Another notable finding is that unions tend to shift local congressmen's positions to the left, particularly on union-related issues, suggesting an additional possible mechanism at play.

This paper contributes to 3 main strands of the literature. First, it enriches the literature regarding the political impact of US unions. Most of this literature is based on surveys (Juravich and Shergold, 1988; Freeman, 2003; Silver, 2011; Kim and Margalit, 2017) and restricted by design to be able to identify only the effect of one's unionization status on one's opinions, attitude, voting, or political involvement (such effects will be defined as direct effects of unionization). [2]. None of them have a natural experiment structure to establish causality.

Few recent papers use geographic units as the basic observation unit and exploit variation in unionization or union regulation (Feigenbaum et al., 2019; Becher and Stegmueller, 2021). Using geographic units allows the estimations of effects that

[2] Freeman (2003) is exceptional for this rule and includes estimations on union members' families

include unions' indirect effects, i.e., the effects of one unionization on other individuals. Building on this work, this paper introduces a robust and reliable identification strategy to examine a clear and intuitive object - the number of votes that each newly unionized worker is worth to the Democratic party. This paper is closely related to Matzat and Schmeißer (2022) that combines unionization election data with campaign contributions data at the workplace establishment level and uses a Diff-in-Diff and IV identification strategies to find that unionization results in a leftward shift of campaign contributions at the workplace. Matzat and Schmeißer use the same source of variation used here and offer strong support for the indirect mechanism of campaign contributions that, to my claim, is one of the main drives of the effect of unionization on voting. While Matzat and Schmeißer used establishment-level data and the DID method, this paper's approach allows for merging unionization data with higher geographic-level outcomes, enabling estimation of spillover effects and using national election outcomes.

Secondly, this paper contributes to the literature that estimates unions' effects on the macro-level (Collins and Niemesh, 2019; Farber et al., 2021; Fortin et al., 2022).[3] It offers a high and robust exogenous shock to shares of unionized workers. In a complementary paper (Borusyak and Kolerman, 2024), we exploit a similar variation to confirm causally the common perception of the large impacts of unions on inequality.

Lastly, this paper contains important methodological contributions. It presents a novel application of the RDA method that allows aggregating several low-level discontinuity events into high-level shock. The method is based on recent theoretical econometric advancements in Regression Discontinuity Designs, re-centered IVs, and Shift Share Instruments (Borusyak and Hull, 2020; Borusyak et al., 2022; Cattaneo and Titiunik, 2022). The method is developed formally and comprehensively in the complementary paper (Borusyak and Kolerman, 2024), with an application in a similar context.

The structure of the paper is as follows: Chapter 2 provides a detailed institutional background. Chapter 3 outlines the identification strategy and introduces the novel RDA method. Chapter 4 is dedicated to presenting the data used in this study. Chapter 5 details various balance tests. Chapter 6 presents the main results, while Chapter 7 focuses on the results for Congressional elections. Chapter 8 delves into the mechanisms underlying the main effect. Finally, Chapter 9 offers concluding remarks.

[3]Macro refers to geographical units containing a mass of workplaces; some are unionized.

2 Institutional Background

The variation exploited in the paper is from close unionization elections. The first subsection of this section details the election procedure and specific issues documented regarding those elections. In the second subsection, I briefly review the historical relationship between labor unions and the Democratic Party in the U.S.

2.1 Unionization Process

The National Labor Relations Board (NLRB) secret ballot election is the most common way for workers to unionize.[4] The process of gaining representation through election consists of four steps:

1. **Petition:** Workers and organizers gather signatures from those seeking union representation. They need at least 30% of the workforce to sign the petition. Once they meet this threshold, they submit it to the NLRB to request an election.

2. **Before Election:**An NLRB agent works to secure an election agreement between the employer and the union. This agreement sets the time and place for voting and defines the bargaining unit—the group of workers eligible to vote and be represented by the union once formed.

3. **Election:** The election will take place, usually in the workplace.

4. **Certification:** The votes are counted. Before certification, parties can challenge certain votes. If the challenges could change the outcome, an NLRB regional director reviews the objections and may order a hearing. Once all challenges are settled, it will be certified if the union has a majority (50%+1). The employer must then bargain "in good faith" with the union. After certification, the union gains legal protection for strikes

The Discontinuity around the 50% cutoff and the sizable amount of NLRB elections ($n = 145,265$ in the period this research deals with- 1976-2020) are attractive to researchers who use the regression discontinuity design and NLRB election data matched with external data sets to learn about the effects of unions (DiNardo and Lee, 2004; Sojourner et al., 2015; Knepper, 2020; Frandsen, 2021).

[4] A voluntary recognition of the union by the employer is also possible, but only a marginal part of new unions gain voluntary recognition

However, Frandsen (2017) recently identified a flaw in this approach—imbalances in very close elections. As shown in panel B of fig. B.6, far fewer elections end with a one-vote margin for unions than a tie or a one-vote margin for the employer. Frandsen (2021) suggests a possible explanation: vote counts reflect results after challenges, and employers tend to challenge more votes. In addition, rulings often favor employers, especially when Republicans control the NLRB. This potential bias poses a significant problem for research on unions' political influence. This article will address this issue in detail later.

2.2 Unions and the Democratic Party

Since the Roosevelt era, unions have been key allies of the Democratic Party (Rosenfeld, 2014). Arguably, this relationship is mutually beneficial: Democratic lawmakers promote pro-labor legislation and block anti-union measures, while unions support the party's campaigns at local, state, and national levels. Unions engage in two main types of political activities. The first aims to increase voter turnout and Democratic support among union members—we define these as the direct effects of unionization. The second targets the general population to boost voting for the Democratic Party. These activities include contributions to PACs and door-to-door canvassing for Democratic candidates, which we refer to as the indirect effects of unionization.

Despite their potential significance, indirect effects have received less attention in the literature due to data limitations and identification challenges. Yet, they may be substantial. According to Labor Organization Annual Financial Reports (LM forms), which nearly all unions are required to file,[5] unions' total disbursements in 2020 amounted to over 8 billion. Of this, more than 600 million dollars was spent on various political activities.[6] This translates to average expenditures of 760$ per member overall and 60$ per member on political activities. Appendix fig. A.3 shows that these numbers have remained relatively stable over time, with modest increases in political spending during national election years.

[5]Unions that cover at least one private-sector employee must file a report.

[6]Disbursements associated with, but not limited to, the following: (1) Political disbursements or contributions. (2) Dealing with executive and legislative branches of federal, state, and local governments. (3) Advancing or defeating laws or regulations (including litigation expenses). (4) Influencing the selection, nomination, election, or appointment of individuals to public or political office. (5) Supporting or opposing ballot referenda. (6) Communications with members and their families for voter registration, get-out-the-vote, and voter education campaigns. (7) Establishing, administering, and soliciting contributions to union-segregated political funds or PACs.

3 Identification Strategy

This paper leverages quasi-experimental variation in union election outcomes across U.S. commuting zones (CZs) and counties. My identification rests on close union certification elections, where union victory can be considered as-good-as-random near the 50% vote threshold.[7] The key insight of my approach is that aggregating multiple close-election discontinuities within a region generates plausibly exogenous variation in local unionization rates. I implement this using the Regression Discontinuity Aggregation (RDA) framework developed by Borusyak and Kolerman (2024), which also includes an application in a similar setting, exploiting the same variation to study the effect of unions on inequality.

The basic unit of observation is one of the 762 commuting zones in the mainland United States, denoted by i, observed over 11 four-year Presidential election cycles, denoted by t.[8]

My treatment variable measures new union formation intensity:

$$NewUnions_{it} = \frac{\text{Workers Unionized}_{it}}{\text{Presidential Voters}_{it}}$$

where i indexes CZs and t indexes presidential election cycles. Formally, this can be expressed as:

$$NewUnions_{it} = \sum_{j \in \mathcal{J}_{it}} s_j \cdot w_j$$

where $\mathcal{J}_{it}$ is the set of unionization elections in CZ i during period t, w_j indicates union victory, and $s_j = \frac{\text{Workers}_j}{\text{Presidential Voters}_{it}}$ represents workplace size relative to the local electorate. I use the number of voters in the presidential election as the denominator for the treatment to be comparable to the outcome, which will be the share of Democratic candidate votes in those elections. [9]

A naive OLS approach faces clear endogeneity concerns. Union organizing likely responds to local political conditions – areas with pro-labor politicians may facilitate unionization (Ellwood and Fine, 1987), in addition, grassroots political movements often coincide with labor organizing (Ferguson et al., 2018) and also may affect political outcomes (Madestam et al., 2013).

[7] This research design builds on earlier work using union election RDs, including DiNardo and Lee (2004) and Frandsen (2021).

[8] I use the 1980s commuting zones rather than the more commonly used 1990s commuting zones due to both data availability and their smaller size, which allows for more precise identification.

[9] Note that $NewUnions_{it}$ captures new union formation rather than net changes in union density, as workers may join or leave existing unionized workplaces through other channels.

My proposed instrumental variable exploits the quasi-random nature of close union elections using the RDA method. The instrument Z_{it} captures the share of workers unionized through close elections:

$$Z_{it} = \sum_{j \in \mathcal{C}_{it}} s_j \cdot w_j$$

where $\mathcal{C}_{it}$ is the set of close elections within bandwidth δ in CZ i during period t.[10]

The RDA framework requires three controls to ensure identification under RD continuity assumptions (Cattaneo and Titiunik, 2021). First, I control for the intensity of union election activity:

$$S_{it} = \sum_{j \in \mathcal{C}_{it}} s_j$$

This accounts for the underlying propensity for union organizing in a region and would be sufficient for the instrument's independence under the assumption of constant union win probability in close races, similar to the local randomization approach to RD developed in Cattaneo et al. (2015). In such a case, the expected value of the instrument would simply be S_{it} multiplied by a constant, making this control sufficient to address omitted variable bias (Borusyak et al., 2022). This approach has precedent in other settings requiring discontinuity aggregation (Clots-Figueras, 2011; Nellis and Siddiqui, 2018).

However, with non-negligible bandwidths, the probability of union victory likely varies with local conditions, making the S_{it} control insufficient for identification. In standard RD designs, controlling for the running variable addresses this varying treatment probability. The challenge in my setting is that each region may contain multiple close elections, each with its own running variable. Following Borusyak and Kolerman (2024), I address this by controlling for aggregate running variables:

$$SR_{it} = \sum_{j \in \mathcal{C}_{it}} s_j \cdot r_j; \quad SR_{it}^{+} = \sum_{j \in \mathcal{C}_{it}} s_j \cdot r_j \cdot w_j$$

Where r_j is the union vote share margin (centered at 0.5), this control strategy effectively localizes my comparison to regions with similar underlying propensities for union organizing and similar voting patterns but different realized election outcomes.

My main instrumental variables specification is thus:

$$NewUnions_{it} = \mu Z_{it} + \gamma_1 S_{it} + \gamma_2 SR_{it} + \gamma_3 SR_{it}^{+} + X_{it}\gamma_4' + \eta_{it} \tag{1}$$

[10] Currently, the RDA method does not offer a technique for choosing the optimal bandwidth. Therefore, several bandwidths will be used. The results are robust to all of them.

$$\Delta Y_{it} = \tau NewUnions_{it} + \beta_1 S_{it} + \beta_2 SR_{it} + \beta_3 SR^{+}_{it} + X_{it}\beta'_4 + \epsilon_{it} \tag{2}$$

The coefficient τ captures the causal effect of union formation on political outcomes, identified from regions where close election outcomes created quasi-random variation in unionization rates. Here, ΔY_{it} represents changes in political outcomes (e.g., Democratic Presidential vote share) in CZ i in year t. I use differenced outcome variables to increase estimation power. Another advantage of using differenced outcomes is that it mitigates potential bias arising from manipulated unionization election results Frandsen (2017), as will be further discussed later. X_{it} includes additional region-cycle controls to increase precision, though the RDA controls alone are sufficient for identification.

Note that the instrument (Z_{it}) and the three RDA controls are weighted sums of canonical RD variables, with s_j, the number of workers in each establishment divided by the number of Presidential voters, used as weights. Specifically, the region-level instrument is the weighted sum of the (close election) establishments' treatments. The first control (S_{it}) is the sum of the weights (equivalent to the weighted sum of the intercept included in an RD regression specification). The second RDA control (SR_{it}) is the weighted sum of the running variables, and the third (SR^{+}_{it}) is the weighted sum of the running variables interacted with being to the right of the cutoff. As shown in (Borusyak and Kolerman, 2024), an important attribute of this aggregated IV specification is that it leads the τ estimator from Equation 2 to be algebraically equivalent to a fuzzy RD estimator at the level of a single unionization election, thereby inheriting its properties. Thus, the assumptions required for identification are thus the canonical RD continuity assumptions (Cattaneo and Titiunik, 2021).

4 Data

My empirical analysis uses data at two geographic levels: establishment-level unionization elections and county-level electoral and other political outcomes. The unit of observation is a county-presidential election cycle pair between 1976 and 2020. The treatment, the instrument, and the main RDA controls are aggregations of establishment-level unionization election data. A detailed description of data sources and construction is provided in appendix B.

4.1 Union Election Data

The backbone of my analysis is the universe of National Labor Relations Board (NLRB) certification elections from 1976 to 2020, compiled from databases maintained by Henry Farber and J. P. Ferguson. Each observation represents a bargaining unit election and contains information on the number of eligible workers, vote counts, employer name, and location.[11] For elections between 2009 and 2020, I geocoded missing county identifiers using establishment addresses.

Following standard practice in the union election literature, I restrict the sample to elections with at least 20 voters. A key concern with NLRB election data is the potential manipulation of close election results through post-election challenges (Frandsen, 2017, 2021). The tendency for these ex-post changes to favor unions or employers could be influenced by existing conditions prior to the election. [12] This manipulation appears particularly pronounced when Republican appointees hold a majority on the five-member NLRB board, which prominently favors employers in those periods. Evidence of such manipulation appears in appendix fig. B.6, which shows an unusually sharp drop in elections decided by a single vote for the union.

I address this threat to identification in three ways. First, I use first-differenced outcome variables, which should absorb bias from time-invariant manipulation patterns. Second, I exclude tied elections and elections with a 1-vote margin. Lastly, I will demonstrate robustness in excluding various ranges of close elections through "donut hole" RD specifications. These approaches follow similar strategies in recent work on union elections (Knepper, 2020; Frandsen, 2021).

Table 1 presents the NLRB election data summary statistics. The average union vote share is 52.5%, with unions winning 48.5% of elections. The data exhibit the expected features of NLRB certification elections, which were documented in previous research.

4.2 High Level Data

I define the treatment and instrument at the commuting zone (CZ) level for each 4-year presidential election cycle. Commuting zones, developed by Tolbert and Sizer (1996) and popularized in economics through work by David and Dorn (2013); David et al. (2013), offer two key advantages for my analysis. First, CZs better capture local labor markets than counties, reducing measurement error from workers who live and vote in jurisdictions different from where they work. Second, CZs' larger

[11] A bargaining unit represents workers of one establishment or a subset thereof.

[12] While pre-challenge vote counts could address this issue in a fuzzy RD design, such data is unavailable for most of my sample period.

Table 1: NLRB elections summary statistics

Variable	All	Treatment- Vote share>50%	Control- Vote share≤50%
Won(%)	48.66	98.80	0.45
Won(at least 20 votes)(%)	44.68	98.97	0.43
Vote share(%)	52.97	77.72	29.17
Vote share(at least 20 votes)(%)	48.67	70.81	33.85
Number of votes (average)	53.42	44.80	61.72
Number of votes (total)	7860472	3230676	4629796
Number of elections	147132	72115	75017
Number of elections(at least 20 votes)	77152	34643	42509
Democrat President(%)	48.35	49.08	47.64
Democrat Governer(%)	51.47	51.45	51.50

Notes: Summary statistics about Low-Level dataset- US NLRB elections between 1976 to 2020.
Source: Source Source Source Source Source Source Source Source Source

size helps address a methodological challenge in the RDA framework. In small geographic units, a few close union elections might represent a large share of the local workforce, creating extreme variation in the instrument that the RDA controls cannot adequately address. The larger CZ geography ensures that most areas contain multiple elections, allowing the law of large numbers to stabilize the ratio of union wins to total close elections. This stabilization strengthens the correlation between my instrument and the first RDA control (the share of workers in close elections), leading it to absorb much of the variation in the instrument.

Figure 2 shows the $LuckShock$ distribution in the 1980 presidential election cycle. For constructing the shock, close elections were defined based on a 5% bandwidth. In this figure, no solid geographical patterns in the distribution of the $LuckShock$ are observed. Another point that emerges from the figure is the small magnitude of the $LuckShock$. In half CZs, the shock is precisely equal to 0, meaning there was no close unionization election. In most cases with non-zero shock, the shock is very small, in the range of -0.1%-0.1% of the voters.

Table 2 summarises information about the $LuckShock$ over the whole high-level dataset. Column 2 of the table shows the number of observations in each range of $LuckShock$ size. Column 3 displays the average shock size without dividing the shock size in each CZ by the number of voters in presidential elections. The table is another indication that the $luckShock$ is small in magnitude, with only a small number of observations with shocks outside the range of $[-1\%, 1\%]$. The sum of absolute values of the shock exhibits the statistical power of the identification strategy offered in this paper. This statistic is the exogenous variation used in the paper. For a bandwidth of 5%, it equals 425,701 voters or 0.03% of the total votes in the 11 presidential elections between 1980 and 2020.

The dataset consists of basic yearly demographics statistics based on the SEER

Figure 2: Luck Shock Map- 1980 Presidential Election

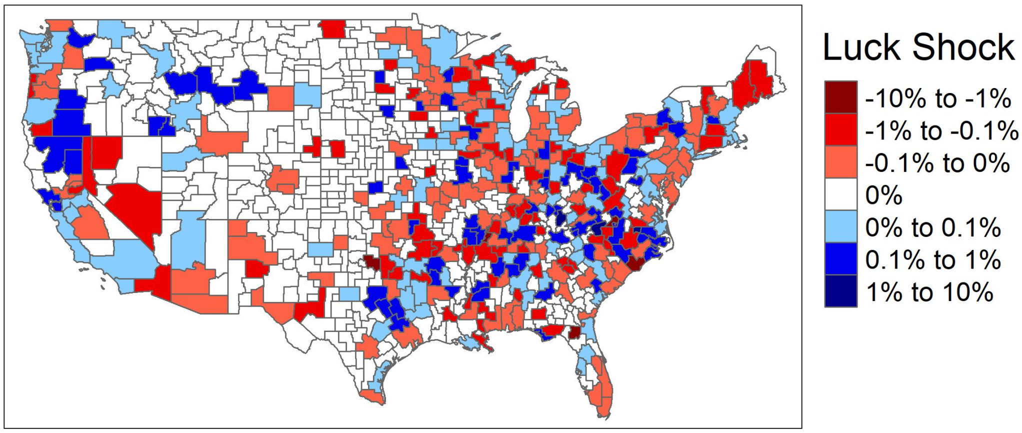


Table 2: Luck Shock Summary Table

Range	Observations	meanLuck	Luck(Not Normalized)
$[-10\%, -1\%)$	20	-473.00	-9460.00
$[-1\%, -0.1\%)$	479	-224.89	-107722.00
$[-0.1\%, 0\%)$	1091	-87.64	-95616.00
$[0\%]$	5316	0.00	0.00
$(0\%, 0.1\%]$	1017	107.46	109284.00
$(0.1\%, 1\%]$	436	214.52	93533.00
$(1\%, 10\%]$	23	438.53	10086.00
Sum of Absulute Values	8382	0.00	425701.00

Notes: This table summarises information about the *LuckShock* from 1976 to 2020. The luck is defined based on a 7.5% bandwidth. Each row of the table represents a range of the shock. Column 2 of the table shows the density of the *LuckShock* based on those ranges. Column 3 displays the average numerator shock size - the number of workers unionized due to luck. Column 4 displays the aggregates of the numerators.

Source: Source Source Source Source Source Source Source Source Source

County Population dataset (National Cancer Institute, 2022). Basic industry composition information was taken from the County Business Patterns (CBP) dataset. The CBP dataset, including imputed missing values, was obtained from Eckert et al. (2021). Measures of the share of workers in high and medium unionization industries were created based on classifications from (Fortin et al., 2022).

4.3 Congress Data

For analysis of the effects on Congress elections, In line with my primary analysis, I will estimate the effect of unionization over 4 year period. For this, I will employ as an outcome variable the first difference between the vote share of the Democratic candidate in the congressional election during period t and that of the Democratic candidate from the election occurring two cycles before (t-2 election cycles or elections taking place 4 years before period t). [13]

A first complication with the data construction required for such analyses is that while for presidential elections, the same candidates are on the ballot in all polling stations in each county, there are many counties that cross several congressional districts. I broadly followed Autor et al. (2020) in dealing with this complication; shortly, I used a county-by-congressional-district cell as the geographic unit of analysis. That is, I used a congressional district-level outcome and county-level controls and shock (in the main analysis, a commuting level shock is assigned to each county). In the regression analysis, I weighted each county-district cell by its total population to maintain equal representation across congressional districts. In the case of several districts crossing the same county, I duplicated the county observation and assigned the same county to each district; the weights in those cases will be the population living in the intersection of the counties and districts (that can be calculated from census-tract population data).

Another complication arises from redistricting. In instances of a redistricted district, defining the pre-period outcome necessary for constructing a first-difference outcome measure becomes ambiguous. In the main analysis, only districts with a matched district in the pre-period, characterized by overlapping boundaries, are kept. [14] This approach deviates from Autor et al. (2020), which, in the case of redistricting, creates a synthetic district in the pre-period. Given the emphasis on estimating effects over short periods, my simpler approach is possible without dropping a substantial portion of the sample.[15] In addition, I excluded districts with two leading candidates from the same party or districts with effectively only one candidate.[16]

[13] An additional benefit of this approach over the one involving the first difference between two consecutive congressional elections is the substantial divergence between elections held in presidential and non-presidential years. Utilizing a 4-year difference conceals noise in the outcome variable arising from these disparities.

[14] Overlapping districts are defined as those with an overlapping area covering at least 99% of the area of each district.

[15] As a consequence of this approach, a considerable number of observations will be discarded in the years following redistricting (Most of the time, years ending with 2 or 4).

[16] I define such cases as districts where a candidate received more than 95% of the votes.

In addition to congressional election results obtained from Dave Leip's Atlas (Leip, 2022), I will estimate the unions' effects on congressmen's ideology alignment with unions' positions. The main measure for alignment is derived from the AFL-CIO scoreboard. Annually since 1957, the Committee on Political Education of the AFL-CIO, the largest federation of unions in the US, has released a scoreboard evaluating all serving federal legislators based on their voting record in 10-30 issues broadly related to unions interests and goals. The score, ranging from 0 to 100, reflects the percentage of votes aligning with the union position. To represent the measure for each congressman in each two-year term, I computed a simple average of their scores across both years. [17].

As complementary measures, I will use two additional measures. The first is the first dimension of the widely used Nominate score (Lewis et al., 2023), which quantifies a lawmaker's political stance in the Liberal-Conservative range relative to their peers based on the universe of roll calls. The second is their votes on roll calls directly related to union issues. To identify those roll calls, I rely on the Comparative Agendas Project (Baumgartner et al., 2019).

To complement my analysis, I employ two additional metrics. Firstly, I use the widely used Nominate score's first dimension(Lewis et al., 2023). This score quantifies a lawmaker's position on the Liberal-Conservative spectrum relative to their peers based on an analysis of all roll call votes. Secondly, I will examine their voting behavior on roll calls directly related to unions. For identifying these particular roll calls, I rely on the Comparative Agendas Project (Baumgartner et al., 2019).

4.4 Auxiliary Data

An additional dataset used in the analysis is the Database on Ideology, Money in Politics and Elections (DIME) created by Bonica (Bonica, 2016) based mainly on the Federal Election Commission public information. The database contains over 500 million political contributions primarily made by individuals from 1979 to 2022. A comprehensive description of the dataset is in data appendix B. The primary metrics I use are the share of political contributors of the total voter population and the share of contributors to the democratic party. Additional measures will be based on the contributors' ideology.

As a source for individual-level data with unionization information, I will use the "Cooperative Election Study" dataset that contains 372,242 observations for the six Presidential Elections between 2004 and 2020.

[17] In cases where two congressmen served in the same district during the same term (e.g., due to the resignation of the elected member), I calculated a simple average of their scores.

5 Balance

To lay the ground for my identification utilizing the RDA design, I will proceed by presenting evidence supporting the instrument, *LuckShock*, as being exogenous. To do so, I regress the instruments on a series of CZ-period pre-determined covariates. The total number of instruments is 24 from 4 categories: Demographics measures, industry composition, lagged unionization measures and lagged political controls. The controls are detailed in the notes of table 3. Those controls will be used throughout my analysis. To reduce my degrees of freedom, I broadly followed Autor et al. (2020) in covariates selection (mainly in the demographic and political controls).

In Table 3, I show summary statistics of regressing the instrument and the independent variable on the list of pre-determined observables described above. I am using different close elections bandwidths to calculate the instrument. For the larger bandwidths (10%, 15%), I residual the running variable controls from the instrument as those controls are included in the identification model. From column (1), it is evident that the pre-determined variables are strongly statistically associated with the share of newly unionized workers with an adjusted R^2 of 0.16. This aligns with the argument that this variable is endogenous. In all columns where the instrument is put on the left-hand side, the adjusted R^2 yields values of a maximum of 0.0004. This is three orders of magnitude smaller than the equivalent test with the independent variable in the LHS, providing strong evidence that the instrument conceals the correlation between unionization rates and pre-determined conditions. The F-test values present a similar picture with a very high and significant value for the independent variable and essentially zero for the instruments.

Several additional balance tests are conducted in appendix C. In the first one, the independent variable and the shocks are placed in the RHS, and a subset of the pre-determined covariates included in table 3 are in the RHS; a simple OLS regression is conducted. Results are presented in table C.8. While the independent variable (NUW) coefficients significantly differ from zero for most pre-determined covariates, only two of the instrument ($LuckShock$) coefficients are significant at 10% and only one in 5% (less than what is expected in a random setting). A possible question about this test statistical power can arise- the *LuckShock* by construction is very close to zero and has a very small variance (much less than the independent variable). Arguably, more observations are needed to reject the correlation of the instrument with pre-determined covariates. Appendix table C.9 deals with this concern; it contains an equivalent balance table for the much more granular county geographic units (the number of observations is four times larger). Counties are smaller; thus, fewer close elections would occur in each county, and the shock variance would increase. A

Table 3: Regressing the Instrument on Pre-Determined Covariates

	NUW	*Luck Shock*			
	(1)	(2)	(3)	(4)	(5)
δ		2.5%	5%	10%	15%
Running Variable				✓	✓
#Covariates	24	24	24	24	24
#Significant 10%	3	0	1	2	0
#Significant 5%	3	0	1	0	0
#Significant 1%	3	0	0	0	0
R^2	0.16026	0.00235	0.00258	0.00316	0.00269
Adjusted R^2	0.15795	-0.00039	-0.00016	0.00041	-6.01×10^{-5}
F-test	69.320	0.85741	0.94045	1.1512	0.97810
F-test, p-value	5.62×10^{-295}	0.65863	0.54271	0.27926	0.49060
Observations	8,378	8,378	8,378	8,378	8,378

Notes: The unit of observation in this table is CZ in a 4-year presidential election cycle. The table provides an overview of the balance analysis of regressing the instrument *LuckShock* and an independent variable *NUW* against a set of pre-determined observables. For larger bandwidths (10%, 15%) where running variable controls are incorporated into the identification model, the instrument is used after residualizing them. The following covariates are included. Demographics Controls: CZ's population across nine age groups, three racial groups, and the female population share. Industry Composition: percentage of a CZ's workforce in manufacturing and those in medium and high unionization industries defined based on Fortin et al. (2023). Lagged Unionization Elections Controls: Lagged share of newly unionized workers, the 'For Union' vote share, and the share of union wins in elections with a margin of 1 (2, 3) or less. Political Controls: one and two periods lagged Democratic candidate vote share and lagged voter turnout.

similar pattern appears in this table, with only one coefficient different from zero at 10%. In the county-level tests, the coefficients for the instrument are much smaller in magnitude than the coefficients for the independent variable. An indication that the instrument's non-significance is not only due to limited statistical power. Additional tests indicate that the *LuckShock* isn't correlated spatially and serially for different values of δ.

6 Results

6.1 Preliminary Results

I begin my analysis by presenting preliminary results on the relationship between unionization and voting for the Democratic party. Table 4 shows the outcomes from an OLS regression, where the first difference in the vote share for the Democratic candidate is regressed on the flow of newly unionized workers (NUW). This regression aligns with **??**, substituting the shock variable with NUW. All models include Period-fixed effects. In Column 1, which doesn't include any other controls, the results suggest that each newly unionized worker is associated with 0.6 new votes for the Democratic candidate. After adding demographic controls detailed in table 3, the coefficient of column 2 notably decreases. This indicates a potential confounding effect where specific demographics may predict both an increase in unionization and Democratic voting. Column 3 includes all other controls mentioned in the notes of table 3, which does not significantly alter the results. Columns 4 through 6 introduce control for the share of workers involved in unionization attempts (NUA). Intriguingly, all coefficients in these models are substantially larger (ranging from 0.96 to 1.61), suggesting that unionization attempts often occur in regions with diminishing Democratic support. [18]

Table 4: OLS Estimations

	(1)	(2)	(3)	(4)
Panel A- $\Delta DemShare$				
$NewUnions$	0.474** (0.222)	0.332** (0.149)	1.63*** (0.358)	0.985*** (0.266)
$NewAttempts$			-1.88*** (0.370)	-1.08*** (0.362)
Panel B- $\Delta DemSharet - 1$				
$NewUnions$	0.455** (0.202)	0.072 (0.180)	0.508 (0.584)	-0.254 (0.493)
$NewAttempts$			-0.085 (0.923)	0.535 (0.869)
Time Range	1980-2020	1980-2020	1980-2020	1980-2020
Additional Controls		✓		✓
year fixed effects	✓	✓	✓	✓
S.E.: Clustered	by: cznew	by: cznew	by: cznew	by: cznew
Observations	34,015	33,971	34,015	33,971

Notes: Standard errors are clustered at the commuting zone level
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

[18]This trend might be explained by areas that experienced unusually strong support for Democratic politicians in the pre-period, leading to legal easements for unionization (Ellwood and Fine, 1987)

6.2 Main Results

The main results are based on ?? and ?? illustrating the impact of unions on the change in vote share for Democratic presidential candidates. All models are weighted by the county adult population. To account for potential serial correlation, standard errors are clustered at the commuting zone level. The table is segmented based on various bandwidths: Columns 1-2 for a δ of 2.5%, columns 3-4 for 5%, columns 5-6 for 10%, and columns 7-8 for 15%. Additionally, columns 5-8 incorporate controls for running variables. The even-numbered columns also include county-period controls, detailed in table 3 notes.

Table 5: Main Results

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.379*	1.172	1.691***	1.418***	1.885**	1.702**	1.777***	1.521***
	(0.8371)	(0.7469)	(0.5172)	(0.4804)	(0.8128)	(0.7484)	(0.5755)	(0.5278)
δ	2.5%	2.5%	5%	5%	10%	10%	15%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls					✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓
Unzero Luck Observations	2,787	2,787	4,869	4,869	7,459	7,459	9,143	9,143
Close NLRB Elections	4,679	4,679	10,598	10,598	22,967	22,967	33,845	33,845
Observations	34,034	33,971	34,034	33,971	34,034	33,971	34,034	33,971
R^2	0.48459	0.55261	0.48502	0.55291	0.48564	0.55315	0.48553	0.55318
Within R^2	0.00038	0.14089	0.00122	0.14147	0.00242	0.14192	0.00219	0.14198
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports estimates of the effects of the luck shock on the share of the Presidential Election Democratic vote share. Estimations are based on ?? and ??. Columns 5-8 also include running variables controls. Columns in even columns add county-period controls presented in table 3. Standard errors are clustered at the commuting zone level

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

The coefficient of interest- τ is stable between 1.2 and 1.9 in all specifications and is significant for most of them. This coefficient interpretation is the number of new votes the Democratic Nominee is expected to get from one more unionized worker. As one would expect if the instrument is exogenous to the outcome, adding covariates barely affects the coefficient of interest. Note that the treatment effect estimated here is larger than OLS coefficients from table 4 (though the difference is not statistically significant). Possible explanations can be that the share of votes for union formation in areas with declining support for Democrats is smaller or that the effects of unionization are larger following close elections than in cases of landslide elections.

It is noteworthy that the treatment effect estimated here is larger than the OLS coefficients reported in table 4, although this difference does not reach statistical significance. Several factors could account for this discrepancy. One possibility is that the propensity for voting in favor of union formation tends to be lower in areas where declining support for the Democratic party. Alternatively, the impact of unionization might be more significant in scenarios following narrowly won union elections, as opposed to those with landslide union victories

An alternative identification strategy will be to estimate an IV model. Ideally, this model should be estimated using union coverage as the endogenous variable. However, for the analysis period in this paper, no dataset providing union coverage with detailed sub-state geographic information is accessible. In (Borusyak and Kolerman, 2024), we are taking a different approach of using state-industry cells rather than counties and leveraging data from the Current Population Survey (CPS). Notably, the CPS questionnaire has included union membership questions since 1977, enabling us to estimate the first stage equation. This results in a first-stage estimator of approximately 0.8-1 (about 0.3 standard deviations). Such a result supports the interpretation of the main model's coefficient as reflecting the number of new votes per unionized worker.

Using the same dataset used in the analysis, it is possible to estimate the first stage on the flow of newly unionized. Assuming there are no significant motivating impacts of successful unionization on subsequent attempts or on the success rate of unionization in the same CZ, we anticipate the coefficient of this estimate to be 1. This expectation is based on the fact that the *LuckShock* instrument is a component of the newly unionized worker flow, as shown in **??**. Table A.1 details the results of this model estimation, with all coefficients indeed closely approximating 1. Unsurprisingly, the second stage estimates, depicted in table A.2, align closely with the primary findings outlined in table 5.

6.3 Threats to Identification

The primary threat to the identification strategy in this paper arises from potential manipulation in very close unionization elections, as identified in (Frandsen, 2017, 2021) and discussed earlier. If such manipulations correlate with the outcome variable, this could violate the IV independence assumption, thereby undermining the identification strategy.[19] Table 6 presents the main exercise to address this threat, showing coefficients estimated from **??** with various "Donut Holes" (Cattaneo and Titiunik, 2021), which exclude the closest elections based on different criteria. All models include the covariates from the main model. Columns 1 and 5 of table 6 correspond to columns 6 and 8 of table 5, representing 10% and 15% thresholds, respectively. The other columns illustrate different Donut Hole definitions of 1-3 vote margins. Table A.3 presents a similar analysis for 2.5% and 5% bandwidths. The estimations in columns including Donut Holes are marginally larger than those without, providing strong evidence against the hypothesis that manipulated very close elections significantly influence the paper's findings.

Additional indirect approaches to deal with this threat are to directly control measures of unions' strength or to allow ϕ_u, the chance of unions to win close elections to vary based on observables. Table D.13 includes estimations based on both approaches. Regression coefficients are robust to those estimations, providing evidence against a claim that regression coefficients are the main drivers for this paper's results.

Table 6: Robustness to Different Donut Holes Sizes

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.702**	1.976**	2.080***	2.089**	1.521***	1.674***	1.717***	1.719***
	(0.7484)	(0.7770)	(0.8038)	(0.8341)	(0.5278)	(0.5342)	(0.5301)	(0.5485)
Donut Hole	No Tie	1 Vote	2 Vote	3 Vote	No Tie	1 Vote	2 Vote	3 Vote
δ	10%	10%	10%	10%	15%	15%	15%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls	✓	✓	✓	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	19,133	18,650	17,963	17,311	21,253	20,936	20,495	20,120
Close NLRB Elections	22,967	20,784	18,238	15,896	33,845	31,662	29,116	26,774
Observations	33,971	33,971	33,971	33,971	33,971	33,971	33,971	33,971
R^2	0.55315	0.55323	0.55325	0.55323	0.55318	0.55324	0.55325	0.55323
Within R^2	0.14192	0.14209	0.14213	0.14208	0.14198	0.14211	0.14212	0.14208
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.
Source: Source Source Source Source Source Source Source Source Source

[19]It is important to note that I am using a first difference outcome variable and incorporating period fixed effects (FE) in the regression equation, which makes the potential threat a correlation with within-period *trends* in democratic candidate vote shares.

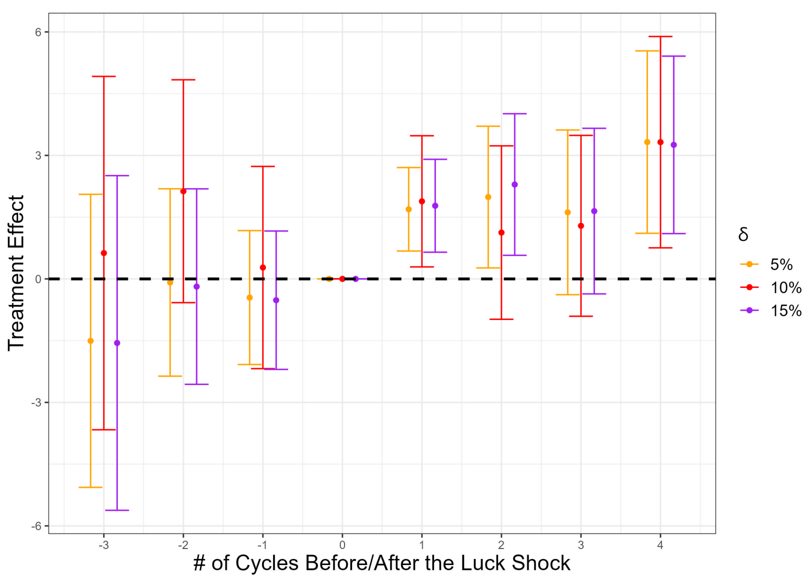
6.4 Local Projection Analysis

The randomness of the $LuckShock$ and the lack of serial correlation allow to use of a local projection estimator to estimate the persistence of the effect over time. The following Reduced-form regression equation will be estimated:

$$Y_{ci,t+\bar{T}} - Y_{cit} = \tau LuckShock_{it}(\delta) + X_{cit}\beta' + \epsilon_{cit} \tag{3}$$

The equation is estimated for different values of $\bar{T}$. For $\bar{T} = 1$, this equation is the same as (**??**). $\bar{T} \neq 1$ represents the effects of New Unionized Workers in period t on outcomes of period $t + \bar{T}$. Coefficients of this regression model are presented in fig. 3. In the X-axis are different values of $\bar{T}$, and in the Y-axis are the values of τ in (3). The color represents the bandwidth used to define the $LuckShock$. All regressions represented in the graph don't include controls other than cycle FE (they correspond to the odd columns of table 5). 95% confidence intervals are presented for each coefficient.

Figure 3: Local Projection-Unions Effect Over Time



Notes: This graph reports a Local Projection analysis to estimate the persistence of the paper's main results. The Y-axis shows the values of τ in (3). The X-axis contains different values of $\bar{T} = 1$ that represent per All regressions represented in the graph don't include controls other than cycle FE. 95% confidence intervals are presented for each coefficient.

Coefficients for $\hat{T} < 0$ represent the effect of a period t's variable on outcomes in earlier periods (each period is a 4-year election cycle). Thus, they can be seen as the results of a placebo test. The model passes this test, as all nine coefficients for $\hat{T} <$ are not significantly different from 0 and very close to it. The coefficients for $\hat{T} > 1$ indicate that the effect of Newly Unionized Workers on voting in the Presidential Elections is persistent over time. The standard errors are increasing in $\hat{T}$, but coefficients are relatively stable between 1.3-3.1 during the whole period. The 1.6 effect that was found for the first period is inside the confidence intervals of all estimations for $\hat{T} > 0$.

6.5 Robustness & Heterogeninty Analsys

Various robustness tests were carried out and are shown in appendix D. A summary of them will appear below. The robustness appendix begins with two un-obvious choices in the variables definition and sample selection. The first is the choice of the denominator of the instrument, the dependent, and the independent variables. In the main specification, they were all calculated as shares of total votes in the presidential elections. Three other denominators are offered, and estimations are robust to each of them. The second choice is the inclusion of observations with zero close unionization elections. Although this inclusion is valid econometrically, it is not obligatory and can seem unintuitive. The appendix shows that excluding those observations has a negligible effect on the estimations.

Robustness to an alternative regression model is also tested in the appendix section: using an IV model and replacing the using NUW^c that represents only workers unionized through close NLRB elections as the independent variable. The estimation of such a model is very similar to this paper's main results and to the IV model that includes NUW as the endogenous variable. Lastly, the appendix shows that results are robust to the exclusion of each state and each period.

Appendix E contains several heterogeneity tests. The first test estimates unions' effect separately for four time ranges (1980-1984, 1988-1996, 2000-2008, 2012-2020). The estimators create a U-shape. In the first two periods, the effect is strong and significant. In the third, it becomes very close to zero and insignificant. In the last period, the effect increased but is still insignificant. This pattern aligns with various historical accounts of the prevalence of the moderate center wing of the Democratic Party in the 1990s and early 2000s. Nevertheless, one should be careful from drawing strong conclusions from those estimations due to big standard errors and small samples.

Most variation in the *LuckShock* comes from CZs with a small population. In large CZs, there will be closer NLRB elections. Due to the law of large numbers, close wins and losses will tend to balance each other, and the *LuckShock* will lean to 0. Table E.18 show that the effect size in more populated counties is larger than the average effect. Thus, the main effect found in the paper can be seen as a lower bound of the average effect. Last heterogeneity analyses show that the effect is higher for unionization elections that took place 2-4 years before the Presidential Elections than elections conducted 0-2 years before. A possible explanation for this disparity is that unions' dues are beginning to be charged only after signing a collective agreement. The negotiation could take several months or years, and only after signing can unions charge dues and transfer them to political goals.

6.6 Effect Size

Since 1976, union density in the US has experienced a substantial decline, dropping from 22% to 9%. This decrease is primarily due to the dissolution of existing unions and a marked reduction in new unionization efforts, as depicted in fig. A.2. To illustrate the magnitude of this effect, fig. 4 presents an estimate of voting outcomes under hypothetical scenarios where successful unionization attempts in each state remained at their 1976 levels. The horizontal axis of fig. 4 represents Presidential election cycles, while the vertical axis denotes the electoral college votes in these elections. The actual Electoral College outcomes are shown by the orange line, whereas the red and purple lines are based on estimations from this study. In the purple line scenario, it is assumed that the union's impact on voting, as estimated in close elections, is applicable to all elections.[20] Conversely, the red line uses the estimated effect to assess the impact of only close elections.[21] An analogous graph depicting the Democratic candidate's popular vote shares is presented in Appendix fig. A.4.

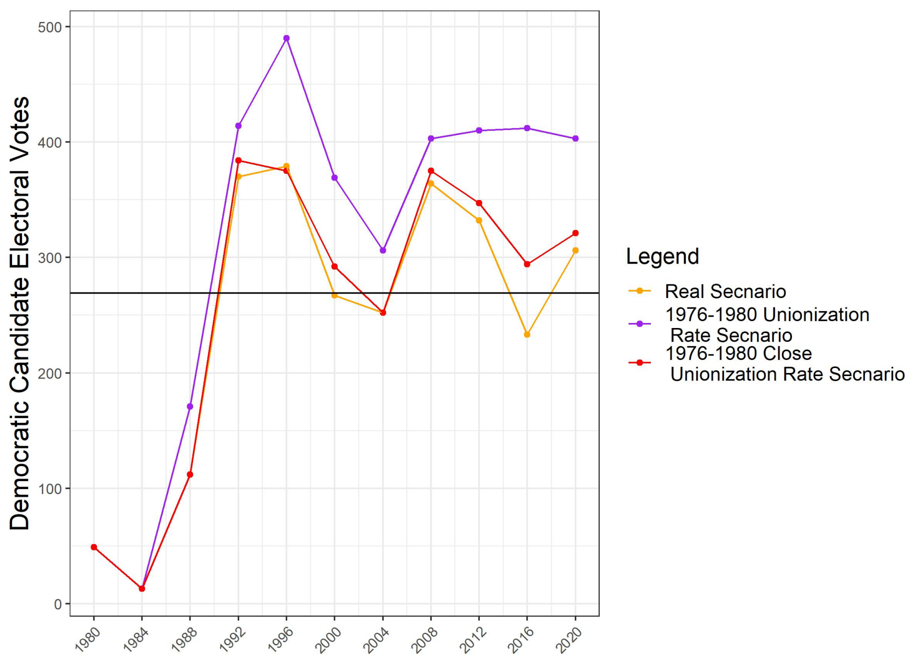
There are significant disparities between the extrapolated scenario (purple line) and the actual outcomes. In every election, the Democratic candidate secures an increasing number of electoral votes, leading to landslide victories since 1988. The differences between the red line scenario, which does not involve extrapolation, and actual outcomes are less pronounced but still noteworthy, suggesting that Trump would not have been elected in 2016 under this scenario.

[20] To estimate voting in this context, the main effect identified in this study is multiplied by the difference between the states' expected unionization rates if unionization had stayed at its 1976 rates and actual unionization rates. This product is added to the real vote shares, and electors are allocated accordingly.

[21] Since the effects for small bandwidths are estimated based on the RDD randomization approach that assumes that the chance of winning close elections is constant, this exercise doesn't require extrapolation

However, it is important to acknowledge that the scenarios depicted in these graphs are not realistic counterfactuals. In a world where states maintained their 1976 unionization rates, the ideological stances of major parties and certain cultural factors would likely have been different, influencing voting behavior through various mechanisms. Nonetheless, this simplified analysis underscores the significant influence of declining union density on the US political landscape over the past decades based on the effects estimated previously. The potential electoral shifts may underscore the necessity for the Democratic Party to engage with new voter bases since the late 1980s, potentially explaining the rise of the New Democrats movement and the third-way agenda in this period Kuziemko et al. (2023).

Figure 4: Estimated Electoral College for Scenario in which States' Union Density Remained at the 1976 Levels



Notes: This graph shows the electoral votes of the Democratic candidate in real and counterfactual scenarios. The counterfactual scenarios are estimated based on fixing each state's close/total unionization on its 1976 levels and estimating counterfactual voting based on the union's effect found in this paper (fourth column of table 5).

7 Congress Results

Table 7 presents the results for the impact on Congress elections. The table columns align with those in table 5. These findings replicate the main results and demonstrate that each newly unionized worker contributes to an increase of between 1.3 to 3.3 votes for the Democratic candidate in Congress elections. Although coefficients exhibit a somewhat larger magnitude, the corresponding standard errors also increase, preventing a meaningful inference about the difference in effect sizes. These outcomes serve as robust evidence for the reliability of the main results, indicating that unions influence voting behavior in diverse settings. It is noteworthy that the sample deviates from the main sample, encompassing elections in non-presidential years[22] and excluding districts that changed between election cycles due to redistricting.

Table 7: Congress Results

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.728**	1.777**	1.477**	1.303	2.355**	3.329**	1.716*	2.110**
	(0.6812)	(0.7927)	(0.7506)	(0.9756)	(1.102)	(1.384)	(0.9073)	(0.8607)
δ	2.5%	2.5%	5%	5%	10%	10%	15%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls					✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓
Unzero Luck Observations	13,800	13,800	18,203	18,203	22,293	22,293	24,410	24,410
Close NLRB Elections	5,216	5,216	11,884	11,884	25,902	25,902	38,027	38,027
Observations	37,531	37,074	37,531	37,074	37,531	37,074	37,531	37,074
R^2	0.16651	0.17904	0.16656	0.17884	0.16809	0.18017	0.16794	0.17969
Within R^2	0.00083	0.01270	0.00089	0.01247	0.00272	0.01406	0.00254	0.01348
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

A heterogeneity analysis of the effect is presented in fig. 5. For this analysis, I will divide the sample into 'Solid Democrat' congressional districts and the rest. 'Solid Democrat' districts are defined as those where the Democratic candidate received at least 2/3 of the combined votes of the two main parties in the pre-period.[23] A primary

[22]The inclusion of non-presidential congressional election with a shock that is aggregated over a 4-year period creates mechanical correlation between two consecutive elections in the same area. The standard errors are clustered at the commuting zone level and account for this.

[23]Due to the exclusion of districts affected by redistricting, it's important to note that for many

reason for this categorization is the alignment of US unions with the Democratic Party, as established in this paper and various others (Dark, 2001; Feigenbaum et al., 2019; Matzat and Schmeißer, 2022). Solid Democrat districts are those where strong connections between congressmen and unions are possible, thus making it intriguing to estimate their influence.

Figure 5 displays estimations from this heterogeneity analysis, with each bar representing the effect on congressional elections for different sub-samples based on varying bandwidths. Similar to other analyses in this paper, the broader bandwidths of 10% and 15% encompass additional controls for the running variables. Standard control variables, featured in the even columns of table 5, are incorporated into all estimations[24] Coefficients in the solid Democrat districts are significant and notably large (ranging from 2.5 to 5.3), whereas coefficients in other districts are considerably smaller (-0.3 to 1.6) and not statistically differ from zero. Tests for the difference between coefficients indicate that the gap between them is different from zero with p values ranging from 3% to 25%.[25] The more pronounced effect in solid Democrat districts may suggest a strategic allocation of union resources in these areas, possibly due to the anticipation of gaining from this support.

The Congress sample enables us to assess the impact on Congressmen's ideology, as expressed in the AFL-CIO yearly scoreboard of their roll calls. This measure further justifies dividing the sample into solid-Democrat districts and others, aligning with the methodology of this paper. The outcome variables used in our analysis reflect differences over a four-year span. The primary variation in these variables emerges from districts that have experienced party shifts in Congressional representation. [26], such variation with less interest as it basically captures the main effect on voting that is already documented in this paper. Focusing on solid-Democrat districts leads to significantly smaller standard errors as party switches to Republican Congressmen in these districts are relatively rare, occurring in only about 6% of cases.

Figure 5 displays coefficients obtained from estimations of the unions' effects on the AFL-CIO score in the subsamples of solid Democrat districts and all others. Panel A illustrates the impact of unionization on the first difference in the AFL-CIO

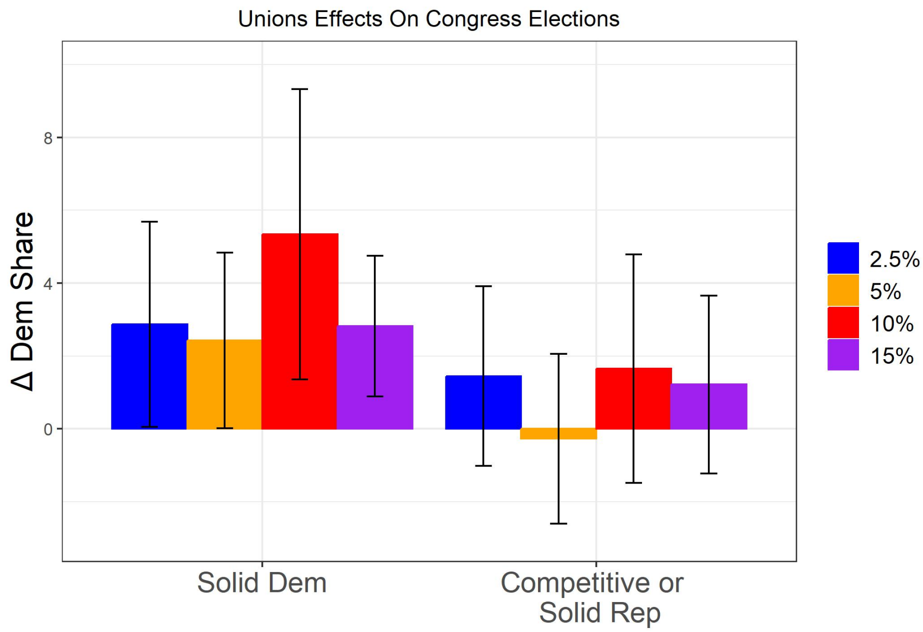
districts, data for several pre-periods is not available. Therefore, defining solid Democrat districts based on multiple pre-periods, as in Autor et al. (2020), is not feasible in this context.

[24]I exclude 2- and 3-period lagged democratic vote shares due to their absence in districts undergoing redistricting processes

[25]Significant at 25% for the 2.5% bandwidth, at 2.9% for the 5% bandwidth, at 6.6% for the 10% bandwidth, and at 14.8% for the 15% bandwidth.

[26]the standard deviation of the AFL-CIO score among same party districts is 12.7 in comparison to 64.6 in party switcher districts

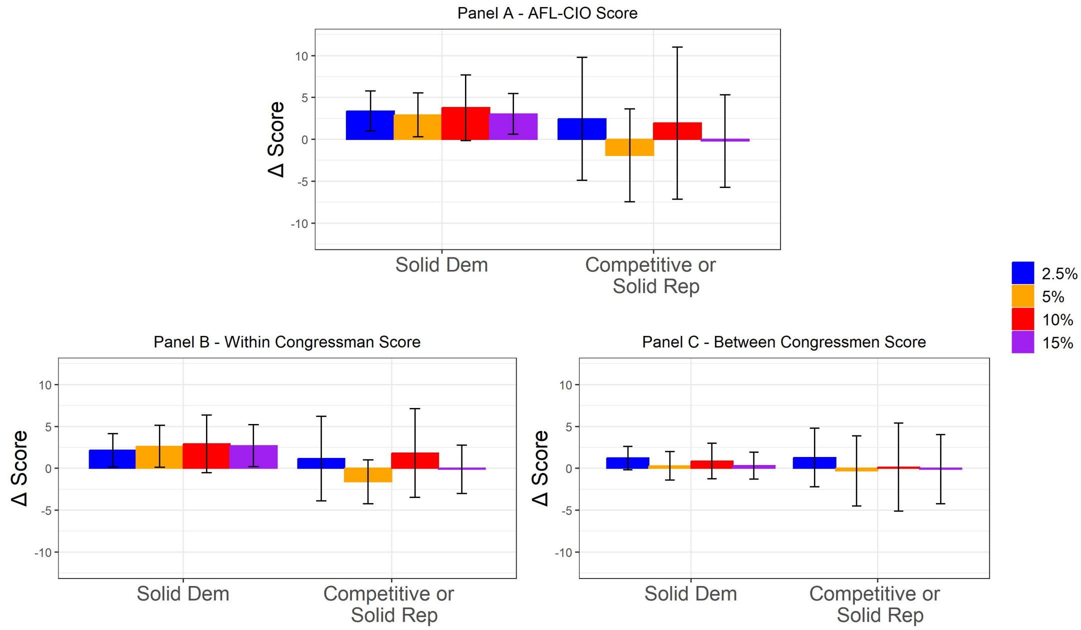
Figure 5: Heterogeneity of The Unions' Effect on Congress



Notes:

score. Coefficients are significantly positive in the Solid Democrat districts, indicating that in those districts, if 1% of the voter population will join unions through close elections, the score is expected to increase by 2.9-3.8 points. Very large standard errors don't allow for determining unambiguous conclusions in the non-solid democrat sample. Panels B and C decompose the ideological impact into effects 'within Congressmen' (measured by the district-period score minus the congressman's career average) and effects resulting from congressman replacement. In solid Democrat districts, the predominant influence (66%-90%) is attributed to 'within congressman' effects, which are statistically significant in most specifications (coefficients ranging from 2.1 to 2.9, significant at 10% or lower). These findings suggest that unions in solid Democrat districts actively collaborate with Democratic Congress members, possibly explaining their motivation to pay back on election day.

Figure 6: Unions effect on AFL-CIO Score



Notes:

8 Mechanisms

The main effect on presidential and congressional elections previously estimated is substantial, exceeding a magnitude of 1. Such a large effect cannot be attributed solely to the direct effect of unionization, where individuals alter their voting preferences post-unionization. Considering many unionized workers might already favor Democrats pre-unionization, the direct effects account for an even smaller part of the overall impact. In appendix F, I examine unionization's direct effect using the "Cooperative Election Study", which has been conducted annually since 2005 and includes data on 372,242 individuals with union membership and national voting questions. Employing OLS with extensive demographic controls and matching methods, I find the maximum potential household effect of unionization peaks at 0.15, just 10% of the main effect. These results align with similar findings in other datasets (Freeman, 2003; Silver, 2011).

In the rest of this section, I will try to indicate possible indirect mechanisms that can drive the large effect identified in this paper.

8.1 Effect on Political Contributions

In this subsection, I will exploit the Database on Ideology, Money in Politics, and Elections (DIME) to estimate the effect of unions on the share of political contributors - individuals who contribute to political goals. Federal and various state campaign contribution laws prohibit mandatory union dues dollars from being used for political campaign contributions. Many unions thus offer their members to voluntarily donate to political action committee (PAC) funds or specific candidates. The proportion of individual contributors can help explain the rise in support for the Democratic Party in two ways. Firstly, contributors are likely to be more politically engaged citizens who volunteer for campaign activities and persuade people in their social circle. Therefore, the proportion of contributors can be regarded as a measure of voters' political activity. Secondly, many of the contributions go to local branches of the party or local candidates. Thus, a correlation between contributions and campaign spending is expected, and more contributors imply an increase in spending that could explain the rise in voting.

Such analysis is in line with Matzat and Schmeißer (2022) that used the DIME database and matched it to specific workplaces where union elections were conducted using information about employer names and worker addresses. Using the DID method, they estimate the effect of unions on the total sum of contributions at the workplace level. They found that unions significantly increased contributions to democratic party candidates at the expense of contributions to the GOP among workers as well as among managers.

This paper's analyses will use the same source of variation– unionization elections and similar outcome variables. The main difference is that I will aggregate the results of the unionization elections into CZ-level shock and use (aggregate) county-period measures of contributions rather than workplace-level measures. An advantage of the aggregation approach taken here is that it doesn't require matching workplaces and individual contributors. Thus, it is more transparent and, by design, can cover all relevant workplaces and workers (as long as the worker resides and works in the same CZ). The downside of this paper's method is that in the aggregation process, close wins cancel out close losses due to the law of large numbers, thus reducing the treatment variation massively. Another disadvantage is that aggregate contributions-based outcome measures tend to be very noisy and have heavy-tailed distributions. Due to the former reason, the main outcome variables will be shares of unique contributors. In the early stages of the analysis, the total and the number of contributions per capita were tested as outcome variables. Very big donors significantly influenced both yielding variables with many outlier observations that could not be used effectively in a regression model with a limited sample.

For consistency reasons, the analysis will focus on contributors to federal political candidates and entities (Presidential, Senate, and House candidates, Federal committees, and 527 organizations) that contributed from 1988-2016, in which a uniform inclusion criterion of donations in the database was in place. [27]. A unique contributor ID is assigned to each individual in the database based on his biographical information, allowing the creation of the following measure of political contributors in each county in each 4-years election cycle:

$$Y_{cit} = \frac{\#Contributors_{cit}}{Voters_{cit}} \tag{4}$$

Where $\#Contributors_{cit}$ is the number of unique contributors in county cit. As additional outcome variables, I will use the share of contributors to Democratic candidates and the share of left and far-left contributors, which I will define using the ideology score estimated in Bonica (2014) based on the donations history of each individual.

The estimation results are presented in table 8, where each column corresponds to a different bandwidth and includes the covariates from even columns of table 5.

In Panel A, we observe the estimated effect on the share of contributors. The results indicate a notable impact: each newly unionized worker joining through close elections is worth an increase of 0.12-0.21 in new contributors. This effect is substantial. Referring to the summary statistics in the appendix B, a 15% increase in unionization via close elections could elevate a commuting zone's contributors' share from the 25% percentile to the 75% percentile.

Panel B shifts focus to the share of Democrat contributors. Although the effects are still significant, they are 20%-40% smaller than the effects on total contributions, which is somewhat smaller. This discrepancy is intriguing. One might expect, based on the estimates of this paper, a more concentrated effect on contributions to the Democratic party. Several factors could account for this difference, such as statistical noise from including Republicans and other contributors or a "general equilibrium" effect where increased Democratic party contributions lead to heightened fundraising efforts by the Republican Party. Unfortunately, the available statistical power is insufficient to pinpoint these mechanisms.

table A.5 delves into the effects on left and far-left contributors, with ideological classifications based on Bonica (2014)'s ideology measure. Left donors are defined as those with ideology scores below the median in each cycle, and far-left donors as those below the 25% percentile. While the coefficients for left contributors remain

[27] Between 1975-1988, a contribution will be included if the individual's election-cycle amount is \$500 or more. In 1989-2016, A contribution will be included if the amount is \$200 or more.

significant, they are slightly lower, suggesting larger effects on non-left contributors or, alternatively, that unions prompt right moderates to donate to the Democratic party. However, caution is advised in interpreting these results. The methodology used by Bonica (2014) for determining donor ideologies estimates simultaneously all donors' ideologies based on information from the full sample of donations. Since unions increase donations in general, they may also affect the identification process itself and lead to bias in the ideology identification.

The analysis also reveals substantial effects on far-left donors, potentially reflecting a strong alignment with union causes. Yet, as with the previous findings, these results should be approached with caution.

Lastly, Table A.6 includes a placebo test for all the estimations I mentioned above. This test includes applying the same models with outcomes replaced by period $t-1$ outcomes. None of the 16 coefficients showed significant differences from zero. This strongly suggests that the original model effectively captures the causal effect of unions on the share of unique contributors.

Table 8: Unions Effect on Contributions

	(1)	(2)	(3)	(4)
Panel A- $\Delta Contributors(\%)$				
Luck Shock	0.217*** (0.065)	0.118** (0.047)	0.170** (0.073)	0.118** (0.057)
Panel B- $\Delta Dem\ Contributors(\%)$				
Luck Shock	0.127** (0.044)	0.094** (0.032)	0.124* (0.053)	0.080* (0.041)
δ	2.5%	5%	10%	15%
Time Range	1992-2016	1992-2016	1992-2016	1992-2016
Running Variable Controls			✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	6,299	8,909	11,605	12,921
Close NLRB Elections	2,506	5,640	12,165	17,752
year fixed effects	✓	✓	✓	✓
Observations	21,278	21,278	21,278	21,278

Notes: This table reports estimations of the union status effect on election-cycle unique contributors. Contribution data is from the Database on Ideology, Money in Politics and Elections (DIME). All estimates contain the covariates that are included in the even columns of table 5. Robust standard errors are in parenthesis.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

8.2 Other Potential Mechanisms

Spending by unions themselves can be another mechanism behind the large effect of unions on voting. Although they can't directly use members' dues as campaign contributions, they can direct other resources to campaign or to invest dues indirectly in politics by lobbying, campaigning, or advertising by themselves. A simple analysis of unions' financial reports can provide evidence for the potential of this spending to impact voting. Every US trade union that includes private-sector workers must file an annual financial report. All reports since 2000 are available online. Starting in 2005, unions with a yearly expenditure of at least 250,000$ [28] are obliged to include their "Political Activities and Lobbying" disbursements. [29]

A simple examination of the available reports reveals that since 2005, unions have, on average, dedicated approximately 6.8% of their resources to political objectives. This translates to an average expenditure of 53.1$ (in real 2020 dollars) per unionized worker annually or 212.4$ over a four-year election cycle. However, conducting a more detailed analysis of these data to establish a causal relationship between winning close elections and a rise in political spending presents significant challenges. One major difficulty lies in linking political expenditures, primarily made by union central headquarters, with elections in specific workplaces. Yet, the large spending on each unionized worker indicates that this is a possible channel, and further studies can try to identify it.

A different potential mechanism by which unions might influence politics is through their impact on local Democratic politicians. Specifically, unions may encourage these politicians to adopt more pro-labor stances, which in turn could attract more voters in general elections. The influence of unions on the ideological leanings of congressmen is explored in section 7. Furthermore, additional evidence supports this part of the mechanism. For instance, Sojourner (2013) reveals that areas with higher unionization rates in certain low-status occupations tend to have a higher representation of legislators from these occupations. In a similar vein, Becher and Stegmueller (2021) shows that congressmen representing districts with strong union

[28] 96.9% of total unions expenditure are from unions above the 250,000$ threshold

[29] Disbursements associated with, but not limited to, the following: (1)Political disbursements or contributions. (2) Dealing with the executive and legislative branches of the federal, state, and local governments. (3) Advance the passage or defeat of existing or potential laws or the promulgation or any other action with respect to rules or regulations (including litigation expenses). (4) Influence the selection, nomination, election, or appointment of anyone to public office or office in a political organization. (5) Support for or opposition to ballot referenda. (6) Communications with members (or agency fee-paying nonmembers) and their families for registration, get-out-the-vote, and voter education campaigns. (7) Establishing, administering, and soliciting contributions to union-segregated political funds (or PACs)

presence are more inclined to represent the views of their economically disadvantaged constituents.

However, the question of whether pro-labor positions inherently attract more voters remains open for further investigation. If this is indeed true, then the influence of unions on the political leanings of local politicians could represent another significant way in which unions impact voting patterns.

8.3 Effects on Turnout

This subsection evaluates unions' impact on voter turnout. I aim to discern if unions influence voting by attracting new voters or shifting existing voter preferences. The analysis utilizes voter registration data from Leip (2022) and employs the main models of this paper, **??** and **??**. The focus is on elections from 1992 onwards, including non-presidential years (2006, 2010, 2014, and 2018) to broaden the dataset. As in other cases, in these years, we define the outcome variable as the first difference from the result four years earlier. To adjust the independent variable and the instrument to the outcome, we modify their denominator to the number of registered voters instead of the number of voters used for other estimations in this paper.

Table 9 presents the findings. The columns mirror those in table 5, showing all positive coefficients ranging from 0.8 to 2.4, a similar range to the range of the main effect on Democratic candidate vote share, hinting at unions' role in boosting Democrat-leaning voter turnout. However, the large standard errors and lack of significant coefficients means definitive conclusions cannot be drawn.

9 Conclusion

This paper introduces the novel Regression Discontinuity Aggregation (RDA) Method, applying it to assess the impact of unionization on Democratic Party vote shares in presidential elections at the Commuting Zone level. The findings indicate that each newly unionized worker contributes an additional 1.2-1.9 votes to Democratic Party candidates. These results are robust, consistent over time, and validated by multiple placebo tests, with similar effects observed in congressional elections.

I present evidence that unions increase contributions, strategically allocating resources in solid democratic areas with new members and shifting politicians to the left as potential indirect mechanisms that may explain the substantive effect.

From 1976 to 2020, a period characterized by a large drop in unionization rates; there's an implied total decline of 4.4 percentage points in the potential Democratic

Table 9: Unions Effect on Turnout

	$\Delta Turnout$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.666	1.239	1.721	1.374	2.355	2.278	1.054	0.8258
	(1.770)	(1.696)	(1.398)	(1.371)	(2.074)	(2.054)	(1.443)	(1.382)
δ	2.5%	2.5%	5%	5%	10%	10%	15%	15%
Time Range	1992-2020	1992-2020	1992-2020	1992-2020	1992-2020	1992-2020	1992-2020	1992-2020
Running Variable Controls					✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓
Unzero Luck Observations	6,880	6,880	9,727	9,727	12,758	12,758	14,279	14,279
Close NLRB Elections	2,658	2,658	5,978	5,978	12,849	12,849	18,732	18,732
Observations	31,015	30,824	31,015	30,824	31,015	30,824	31,015	30,824
R^2	0.55468	0.56068	0.55470	0.56070	0.55470	0.56072	0.55475	0.56074
Within R^2	7.18×10^{-5}	0.01202	0.00013	0.01205	0.00012	0.01210	0.00025	0.01215
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports estimations of the union status effect on the turnout calculated as the number of votes in national elections out of the total registered voters. Both denominator and nominator are from Dave Leip's election atlas. Tables columns are equivalent to table 5. Unlike the paper's other results, the number of registered voters is the independent variable and the instrument's denominator. This change is required to make them comparable to the outcome variable.

Robust standard errors are in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

Party's vote share, attributable to the dynamics explored in this study. This decline might be linked to the Democratic Party's strategic shifts in agenda and outreach to new demographics, as well as the Republican Party's efforts to attract voters from previously unionized groups. These strategies could illuminate various political and cultural shifts witnessed in recent decades.

References

Allern, E. H. and Bale, T. (2017). *Left-of-centre parties and trade unions in the twenty-first century*. Oxford University Press. 3

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*, 110(10):3139–3183. 14, 16, 27

Baumgartner, F. R., Breunig, C., and Grossman, E. (2019). *Comparative policy agendas: Theory, tools, data*. Oxford University Press. 15

Becher, M. and Stegmueller, D. (2021). Reducing unequal representation: The impact of labor unions on legislative responsiveness in the us congress. *Perspectives on Politics*, 19(1):92–109. 4, 33

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, 58(2):367–386. 31, 32

Bonica, A. (2016). Database on Ideology, Money in Politics, and Elections (DIME). 15

Borusyak, K. and Hull, P. (2020). Non-random exposure to exogenous shocks: Theory and applications. Technical report, National Bureau of Economic Research. 5

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213. 5, 9

Borusyak, K. and Kolerman, M. (2024). Regression discontinuity aggregation, with an application to the union effects on inequality. 5, 8, 9, 10, 20

Cattaneo, M. D., Frandsen, B. R., and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference*, 3(1):1–24. 4, 9

Cattaneo, M. D. and Titiunik, R. (2021). Regression discontinuity designs. *arXiv preprint arXiv:2108.09400*. 4, 9, 10, 21

Cattaneo, M. D. and Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, 14:821–851. 5

Clots-Figueras, I. (2011). Women in politics: Evidence from the indian states. *Journal of public Economics*, 95(7-8):664–690. 9

Collins, W. J. and Niemesh, G. T. (2019). Unions and the great compression of wage inequality in the us at mid-century: evidence from local labour markets. *The Economic History Review*, 72(2):691–715. 5

Dark, T. E. (2001). *The unions and the Democrats: An enduring alliance*. Cornell University Press. 27

David, H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, 103(5):1553–97. 11

David, H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American economic review*, 103(6):2121–68. 11

DiNardo, J., Fortin, N., and Lemieux, T. (1995). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. 2

DiNardo, J. and Lee, D. S. (2004). Economic impacts of new unionization on private sector employers: 1984–2001. *The Quarterly Journal of Economics*, 119(4):1383–1441. 3, 6, 8

Ebbinghaus, B. (1995). The siamese twins: citizenship rights, cleavage formation, and party-union relations in western europe. *International review of social history*, 40(S3):51–89. 3

Eckert, F., Fort, T. C., Schott, P. K., and Yang, N. J. (2021). Imputing missing values in the us census bureau's county business patterns. Technical report, National Bureau of Economic Research. 13

Ellwood, D. T. and Fine, G. (1987). The impact of right-to-work laws on union organizing. *Journal of Political Economy*, 95(2):250–273. 8, 18

Farber, H. S., Herbst, D., Kuziemko, I., and Naidu, S. (2021). Unions and inequality over the twentieth century: New evidence from survey data. *The Quarterly Journal of Economics*, 136(3):1325–1385. 2, 5

Feigenbaum, J., Hertel-Fernandez, A., and Williamson, V. (2019). From the bargaining table to the ballot box: Downstream effects of right-to-work laws. Technical report, Working Paper. 3, 4, 27

Ferguson, J.-P. (2016). Results of nlrb union-representation elections, 1965-1998. https://https://github.com/jpfergongithub/nlrb_old_rcases. 48

Ferguson, J.-P., Dudley, T., and Soule, S. A. (2018). Osmotic mobilization and union support during the long protest wave, 1960–1995. *Administrative Science Quarterly*, 63(2):441–477. 8

Fortin, N., Lemieux, T., and Lloyd, N. (2022). Right-to-work laws, unionization, and wage setting. Technical report, National Bureau of Economic Research. 5, 13

Fortin, N. M., Lemieux, T., and Lloyd, N. (2021). Labor market institutions and the distribution of wages: The role of spillover effects. *Journal of Labor Economics*, 39(S2):S369–S412. 2

Fortin, N. M., Lemieux, T., and Lloyd, N. (2023). Right-to-work laws, unionization, and wage setting. In *50th Celebratory Volume*, pages 285–325. Emerald Publishing Limited. 17

Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression discontinuity designs*. Emerald Publishing Limited. 7, 10, 11, 20

Frandsen, B. R. (2021). The surprising impacts of unionization: Evidence from matched employer-employee data. *Journal of Labor Economics*, 39(4):861–894. 3, 6, 7, 8, 11, 20

Freeman, R. B. (2003). What do unions do... to voting? 4, 29, 65

Gethin, A., Martínez-Toledano, C., and Piketty, T. (2021). *Political cleavages and social inequalities: A study of fifty democracies, 1948–2020*. Harvard University Press. 3

Juravich, T. and Shergold, P. R. (1988). The impact of unions on the voting behavior of their members. *ILR Review*, 41(3):374–385. 4

Kim, S. E. and Margalit, Y. (2017). Informed preferences? the impact of unions on workers' policy views. *American Journal of Political Science*, 61(3):728–743. 4

Knepper, M. (2020). From the fringe to the fore: Labor unions and employee compensation. *Review of Economics and Statistics*, 102(1):98–112. 3, 6, 11, 48

Kuziemko, I., Marx, N. L., and Naidu, S. (2023). "compensate the losers?" economic policy and partisan realignment in the us. Technical report, National Bureau of Economic Research. 3, 25

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697. 4

Leip, D. (2022). Dave leip's atlas of the u.s. presidential elections. 15, 34

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2023). Voteview: Congressional roll-call votes database. `https://voteview.com/`. 15

Madestam, A., Shoag, D., Veuger, S., and Yanagizawa-Drott, D. (2013). Do political protests matter? evidence from the tea party movement. *The Quarterly Journal of Economics*, 128(4):1633–1685. 8

Matzat, J. and Schmeißer, A. (2022). Do unions shape political ideologies at work? *arXiv preprint arXiv:2209.02637*. 3, 5, 27, 30

National Cancer Institute, DCCPS, S. R. P. r. F. (2022). Surveillance, epidemiology, and end results (seer) program populations (1969-2020). 13

Nellis, G. and Siddiqui, N. (2018). Secular party rule and religious violence in pakistan. *American Political Science Review*, 112(1):49–67. 9

Padgett, S. and Paterson, W. E. (1991). *A history of social democracy in postwar Europe*. Longman Publishing Group. 3

Rosenfeld, J. (2014). *What unions no longer do*. Harvard University Press. 7

Silver, N. (2011). The effects of union membership on democratic voting. 4, 29, 65

Sojourner, A. J. (2013). Do unions promote members' electoral office holding? evidence from correlates of state legislatures' occupational shares. *ILR Review*, 66(2):467–486. 33

Sojourner, A. J., Frandsen, B. R., Town, R. J., Grabowski, D. C., and Chen, M. M. (2015). Impacts of unionization on quality and productivity: Regression discontinuity evidence from nursing homes. *ILR Review*, 68(4):771–806. 3, 6

Tolbert, C. M. and Sizer, M. (1996). Us commuting zones and labor market areas: A 1990 update. Technical report. 11

A Supplementary Figures and Tables Noted in The Text

Figure A.1: Union Density and Democratic Nominee vote Share 1976-2020

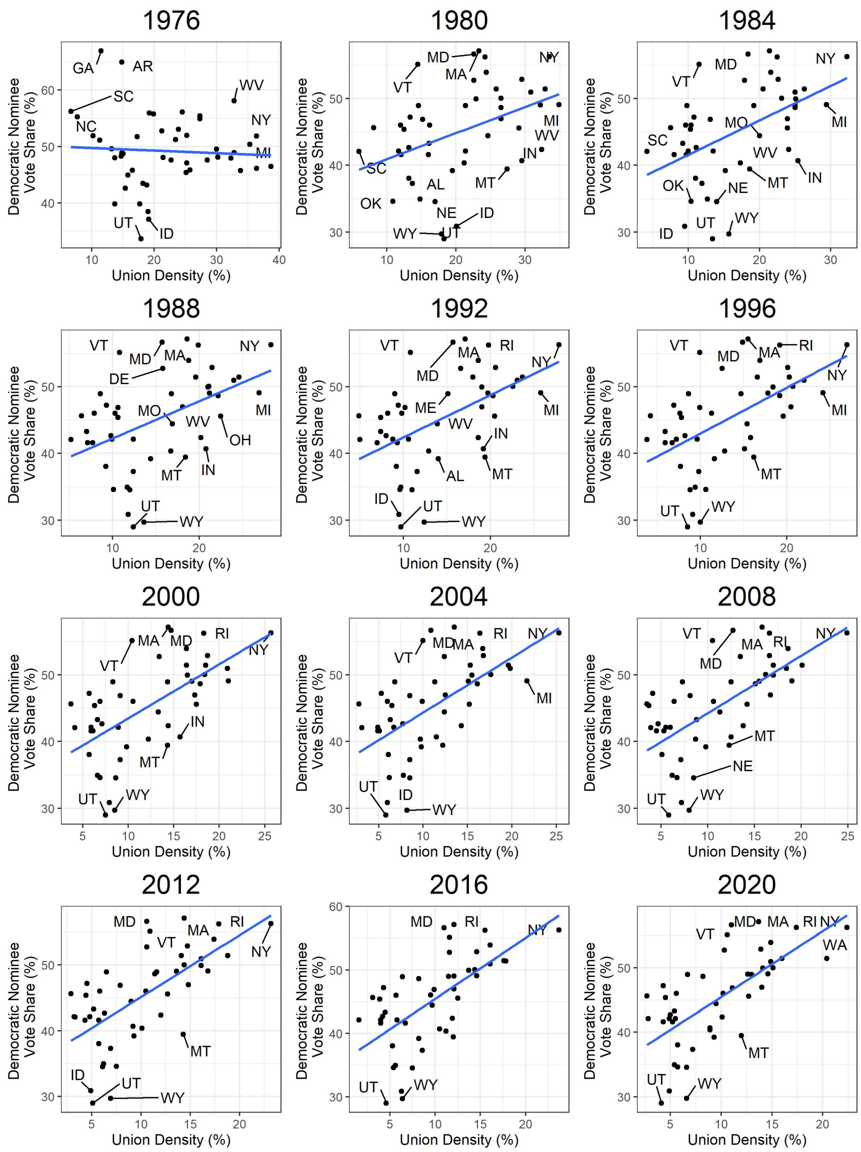


Figure A.2: Yearly Representation Cases

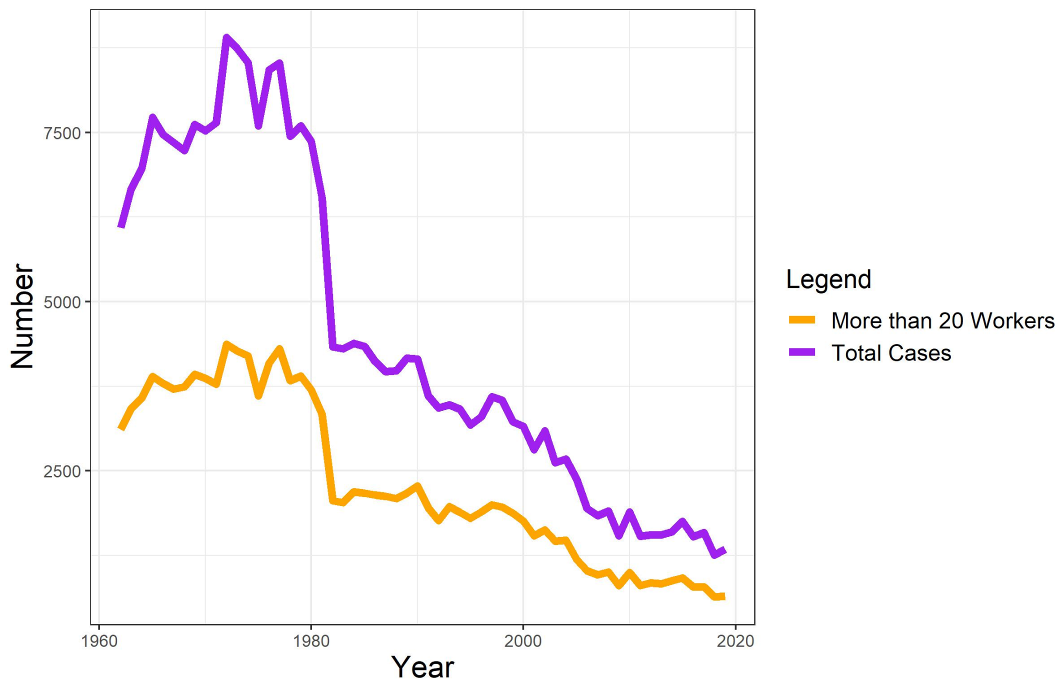


Figure A.3: Unions Expenses per Member

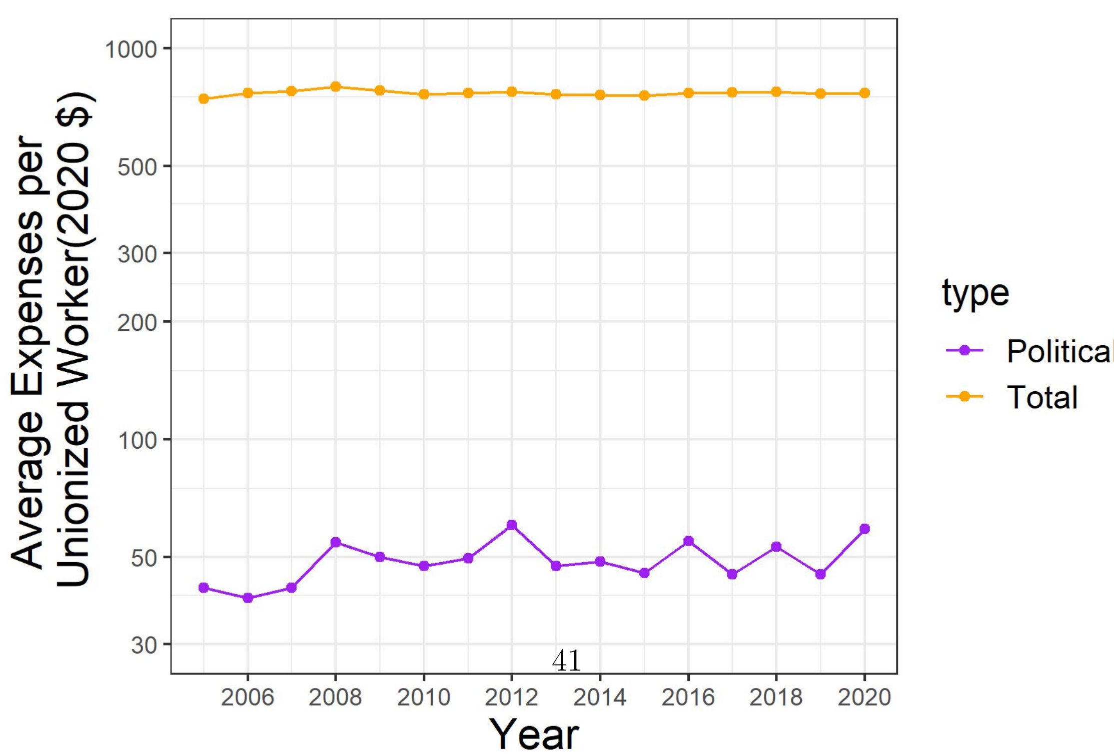
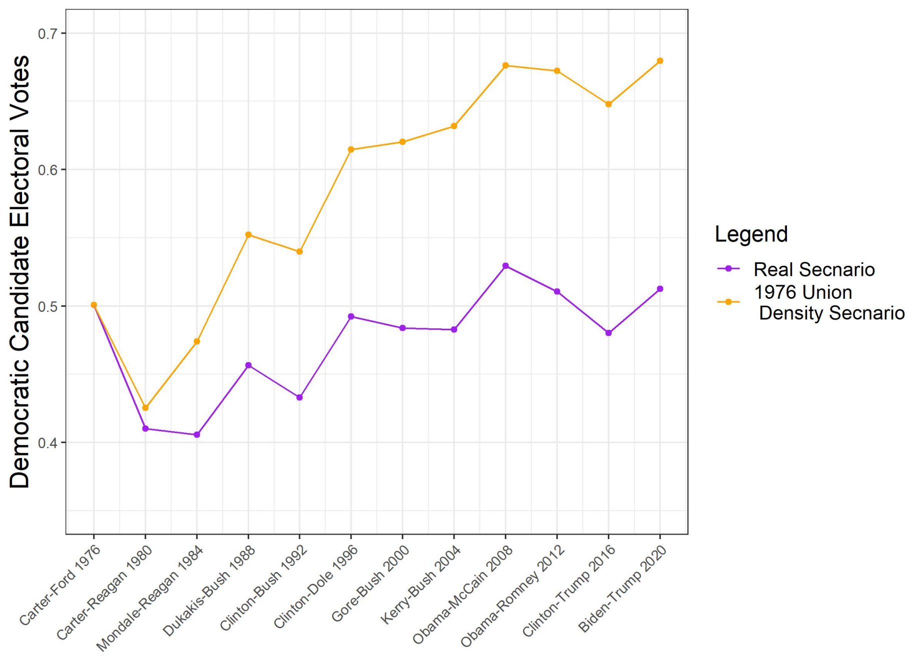


Figure A.4: Estimated Vote Share for Scenario in which States' Union Density Remained at the 1976 Levels



Notes: This graph shows the vote share of the Democratic candidate in a real and a counterfactual scenario. The 1976 union density scenario is estimated based on fixing each state's union density on its 1976 levels and estimating counterfactual voting based on union's effect found in this paper (six column of table 5).

Table A.1: Main Results IV- First Stage

	NUW							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.07***	1.09***	1.12***	1.14***	1.03***	1.06***	1.10***	1.10***
	(0.214)	(0.205)	(0.156)	(0.151)	(0.167)	(0.161)	(0.111)	(0.104)
δ	2.5%	2.5%	5%	5%	10%	10%	15%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls					✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓
Unzero Luck Observations	10,660	10,660	14,960	14,960	19,133	19,133	21,253	21,253
Close NLRB Elections	4,679	4,679	10,598	10,598	22,967	22,967	33,845	33,845
Observations	34,034	33,971	34,034	33,971	34,034	33,971	34,034	33,971
R^2	0.16	0.22	0.18	0.24	0.35	0.39	0.43	0.47
Within R^2	0.02	0.10	0.05	0.13	0.25	0.30	0.34	0.38
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors are in parenthesis.

Source: Source Source Source Source Source Source Source Source Source

Table A.2: Main Results IV- Second Stage

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NUW	1.288	1.078	1.511***	1.248***	1.822**	1.610**	1.618***	1.379***
	(0.8869)	(0.7466)	(0.5514)	(0.4702)	(0.8267)	(0.7366)	(0.5439)	(0.4884)
δ	2.5%	2.5%	5%	5%	10%	10%	15%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls					✓	✓	✓	✓
Additional Controls		✓		✓		✓		✓
Unzero Luck Observations	10,660	10,660	14,960	14,960	19,133	19,133	21,253	21,253
Close NLRB Elections	4,679	4,679	10,598	10,598	22,967	22,967	33,845	33,845
Observations	34,034	33,971	34,034	33,971	34,034	33,971	34,034	33,971
R^2	0.48372	0.55165	0.48174	0.55043	0.48302	0.54918	0.48444	0.55087
Within R^2	-0.00131	0.13905	-0.00515	0.13670	-0.00267	0.13430	9.18×10^{-5}	0.13755
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors are in parenthesis.

Source: Source Source Source Source Source Source Source Source Source

Table A.3: Robustness to Different Donut Holes Sizes (2.5% and 5% bandwidths)

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock	1.172	1.477*	1.556*	1.535*	1.418***	1.553***	1.606***	1.591***
	(0.7469)	(0.7748)	(0.8235)	(0.8558)	(0.4804)	(0.4849)	(0.4779)	(0.4930)
Donut Hole	No Tie	1 Vote	2 Vote	3 Vote	No Tie	1 Vote	2 Vote	3 Vote
δ	2.5%	2.5%	2.5%	2.5%	5%	5%	5%	5%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls								
Additional Controls	✓	✓	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	10,660	7,855	5,892	4,563	14,960	13,758	11,979	10,581
Close NLRB Elections	4,679	2,496	1,523	1,038	10,598	8,415	5,869	4,505
Observations	33,971	33,971	33,971	33,971	33,971	33,971	33,971	33,971
R^2	0.55261	0.55267	0.55268	0.55266	0.55291	0.55298	0.55300	0.55297
Within R^2	0.14089	0.14102	0.14104	0.14100	0.14147	0.14161	0.14165	0.14158
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: * Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

Table A.4: Robustness to Including Tie Elections

	$\Delta DemShare$			
	(1)	(2)	(3)	(4)
Luck Shock	0.8784	1.278***	1.399*	1.344**
	(0.7278)	(0.4783)	(0.7297)	(0.5250)
δ	2.5%	5%	10%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls			✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	11,739	15,485	19,363	21,392
Close NLRB Elections	5,885	11,804	24,173	35,051
Observations	33,971	33,971	33,971	33,971
R^2	0.55255	0.55283	0.55306	0.55310
Within R^2	0.14078	0.14132	0.14175	0.14183
year fixed effects	✓	✓	✓	✓

Notes: * Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

Table A.5: Unions Effect on Left Contributions

	(1)	(2)	(3)	(4)
Panel C- $\Delta Left\ Contributors(\%)$				
Luck Shock	0.108*** (0.040)	0.063** (0.029)	0.090* (0.047)	0.045 (0.036)
Panel D- $\Delta Far\ Left\ Contributors(\%)$				
Luck Shock	0.088*** (0.031)	0.053** (0.025)	0.096*** (0.036)	0.031 (0.032)
δ	2.5%	5%	10%	15%
Time Range	1992-2016	1992-2016	1992-2016	1992-2016
Running Variable Controls			✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	6,299	8,909	11,605	12,921
Close NLRB Elections	2,506	5,640	12,165	17,752
year fixed effects	✓	✓	✓	✓
Observations	21,278	21,278	21,278	21,278

Notes: This table reports placebo estimations of the union status effect on election-cycle unique contributors. Contribution data is from the Database on Ideology, Money in Politics and Elections (DIME). All estimates are of the second stage equation (**??**) and contain the variables that are included in the last column of table 5. The estimations are of newly unionized workers in period t on outcomes of period t-1. Thus this estimation is a placebo mode.

Robust standard errors are in parenthesis.

Source: Source Source Source Source Source Source Source Source Source

Table A.6: Unions Effect on Contributions- Placebo

	(1)	(2)	(3)	(4)
Panel A- $\Delta Contributors(\%)$				
Luck Shock	-0.075 (0.063)	0.029 (0.058)	0.050 (0.077)	-0.082 (0.068)
Panel B- $\Delta Dem\ Contributors(\%)$				
Luck Shock	0.002 (0.056)	0.073 (0.050)	0.104 (0.068)	-0.010 (0.060)
Panel C- $\Delta Left\ Contributors(\%)$				
Luck Shock	-0.022 (0.045)	0.047 (0.042)	0.063 (0.057)	-0.035 (0.047)
Panel D- $\Delta Far\ Left\ Contributors(\%)$				
Luck Shock	-0.018 (0.036)	0.003 (0.035)	0.017 (0.045)	-0.048 (0.038)
δ	2.5%	5%	10%	15%
Time Range	1996-2020	1996-2020	1996-2020	1996-2020
Running Variable Controls			✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	5,712	8,071	10,644	11,974
Close NLRB Elections	2,139	4,795	10,317	15,041
year fixed effects	✓	✓	✓	✓
Observations	21,241	21,241	21,241	21,241

Notes: This table reports placebo estimations of the union status effect on election-cycle unique contributors. Contribution data is from the Database on Ideology, Money in Politics and Elections (DIME). All estimates are of the second stage equation (**??**) and contain the variables that are included in the last column of table 5. The estimations are of newly unionized workers in period t on outcomes of period t-1. Thus this estimation is a placebo mode.

Robust standard errors are in parenthesis.

Source: Source Source Source Source Source Source Source Source Source

Table A.7: Turnout Out of Registered Voters Analyses

	d_per_reg_elig (1)	d_per_dem_elig (2)	d_per_rep_elig (3)	d_per_reg_elig (4)	d_per_dem_elig (5)	d_per_rep_elig (6)
New Unionized Workers	-0.4646*	0.8063	-0.7948***	-0.4426**	0.7107	-0.8747***
	(0.2451)	(0.5819)	(0.2178)	(0.2007)	(0.4860)	(0.2272)
δ	7.5%	7.5%	7.5%	10%	10%	10%
Additional Controls	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	2,283	1,164	1,164	2,609	1,341	1,341
Close NLRB Elections	7,018	4,159	4,159	9,810	5,806	5,806
Observations	7,702	3,994	3,994	7,702	3,994	3,994
R^2	0.21187	0.22493	0.26136	0.21210	0.22509	0.26068
Within R^2	0.02196	0.06876	0.07096	0.02223	0.06896	0.07010
year fixed effects	✓	✓	✓	✓	✓	✓

Notes:
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

B Data Appendix

B.1 Low-Level Data

The low-level dataset is the universe of NLRB elections in the US between 1976 and 2020. It is based on three data sets obtained from open sources:

- NLRB elections that took place between 1976 to 1990 from Ferguson (2016)
- NLRB elections that took place between 1991 to 2008 from Knepper (2020)
- NLRB elections that took place between 2009 to 2020 were obtained directly from the NLRB open database.

All observations contain information about the name of the employer and the union, the number of workers eligible to vote, NLRB election results, and key dates (filling election petition date, election date, and case closing date).

The county of the workplace variable was missing for the years 1991-1999 and 2009-2020. For 1991-1999, county name and county code (3 last digits of the FIPS county 5 digits code) were provided. In most cases, both can identify exactly one county. For the minority of counties that both indicate more than one possible county (in the case that there are two counties in different states with the same name and number), the two missing letters were supplied based on the state of the local union that filled the unionization request.

For the years 2009-2020, the county variable was found based on three steps algorithm:

- The municipality and the state of the workplace.
- If the municipality variable was missing or if the municipality split between several counties, the employer's zip code was used to match a county.
- If the previous two steps didn't yield a single county, the workplace address and name were searched in google maps by google maps API to find exact location and match it to county.

The algorithm found a single county for 94.6% of NLRB elections between 2009-2020. Finally, a county variable is available for 99.2% of observations on the full sample.

Unions and employers can challenge votes in NLRB elections; in rare cases, a re-election process could occur. Thus, getting the majority (minority) votes does not guarantee the union's victory (defeat). Figure B.5 shows the relationship between the vote share for the union and the final result of NLRB elections. Each point in the graph represents a 0.5% width bin of NLRB election results. The graph indicates that the difference in chances of final win just below and just above the cutoff is 87%.

Figure B.5: Shares of Union win, 200 bins

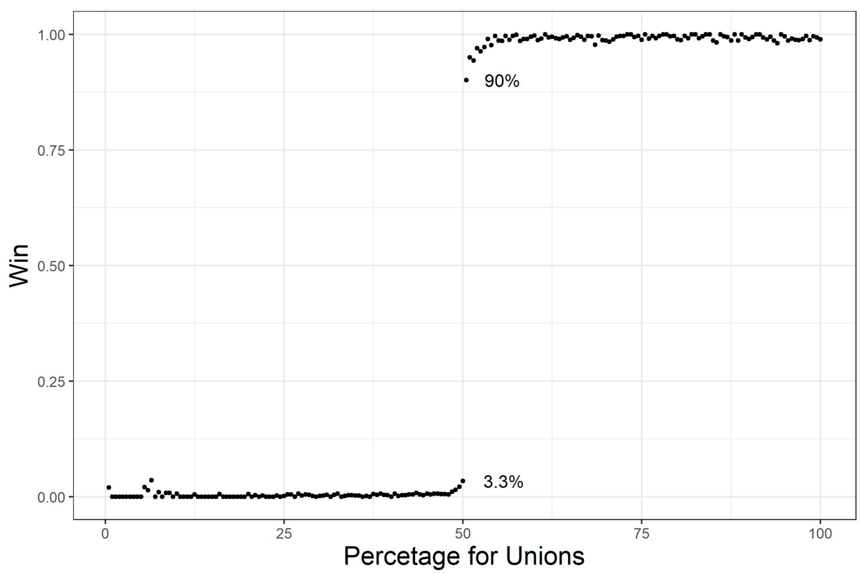


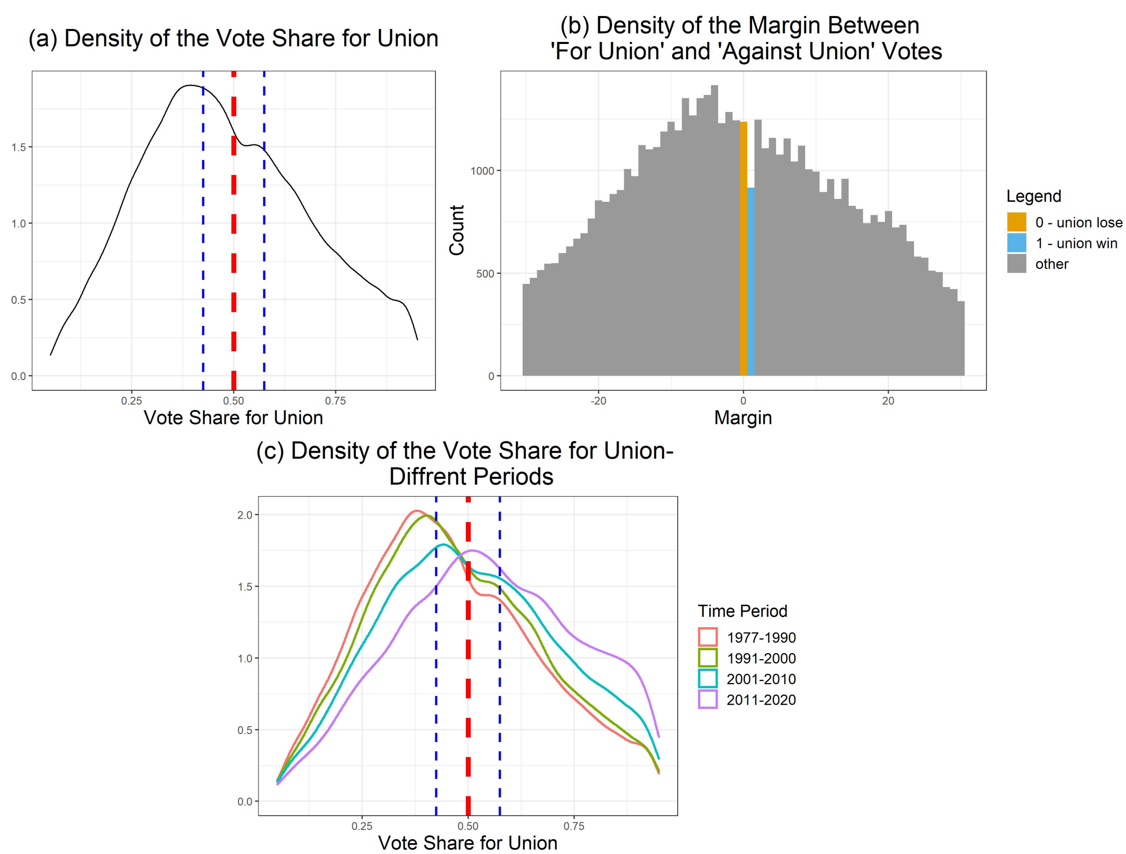
Figure B.6 shows densities graphs of the voting for the union in NLRB elections. In panel (a) the running variable is the continuous vote share for the union. In panel (b), it is the discrete margin between votes for and against union formation. Both indicate that the center of the density is skewed to the left, with more elections ending in close losses than close wins. In panel (a) the blue vertical lines represent a 7.5% bandwidth definition of close elections. The average chance of wining close elections- ϕ is calculated as the weighted (by number of votes) ratio between close wins and close elections, for 7.5% bandwidth it is equal to 45.7%.

Panel (b) presents evidence of manipulations in NLRB election results. The orange bar shows the number of NLRB elections that ended in a tie (in this case, the union loses), and the blue line presents the number of elections that ended in the union winning by exactly one vote. The visible gap between the bars is an intuitive indication of deviation from results expected in a clean setting.

Formally, the McCrary density discontinuity test for manipulations in the running variable reports t-statics of -3 when using the discrete running variable and statics of -10 when using the continuous running variable. Both tests were conducted using triangular kernel and quadratic polonium. A 7.5% bandwidth used in the main specification was chosen for the continuous variable. A 23 votes margin was selected by a standard data-driven process for the discrete variable.

Panel (c) shows changes in the density function over time. Each line represents a 10-year period (the first red line represents 14 years) and shows the density function of the vote share for unions in NLRB elections. The graph indicates that the density is becoming more centered over time; as a result, the chance of winning close elections- ϕ_u is a bit higher for later NLRB elections. Based on a bandwidth of 7.5%, the average chance increases from 44.4% in the first period (1977-1990) to 51.1% in the last period (2011-2020). To absorb bias that stems from this slight increase, all regression specifications in the paper include election-cycle fixed-effects. In addition, in the paper, I show that results are robust to change in the ϕ_u over time and other observable characteristics.

Figure B.6: Density Functions



B.2 High-Level Data

C Additional Balance Tests

C.1 Instrument in the RHS

Table C.8 presents estimations of the following regression models:

$$W_{cit} = \tau NUW_{it} + \sigma_t + \eta_{cit} \tag{C.1}$$

$$W_{cit} = \tau LuckShock_{it} + \sigma_t + \eta_{cit} \tag{C.2}$$

W_{cit} is a pre-determined covariate, σ_t is a period fixed effect. Column 3 presents the τ coefficients of (C.1), while columns 4-6 present the τ coefficients of (C.2) under different close NLRB elections bandwidths. Table C.9 presents estimations of the same model with the shock, and the independent variables are calculated at the county level rather than the CZ level ($LuckShock_{cit}$ and NUW_{cit}).

Table C.8: Presidential Elections Summary Statistics

Variable	Mean	NUW	LuckShock- $\delta = 2.5\%$	LuckShock- $\delta = 5\%$	LuckShock- $\delta = 7.5\%$
(1)	(2)	(3)	(4)	(5)	(6)
Panel A. - Covarites					
Percentage Male	0.489	-0.146**	-0.037	0.093	0.006
		(0.063)	(0.158)	(0.118)	(0.109)
Percentage Black	0.129	0.83*	2.209	2.132	1.797
		(0.45)	(1.927)	(1.365)	(1.162)
Percentage Elders	0.119	-0.391***	-0.531	-0.656**	-0.457*
		(0.113)	(0.443)	(0.307)	(0.248)
Panel B. - Lagged Political Outcomes					
Diff Turnout t-1	0.002	-0.112	-0.206	-0.304	-0.333
		(0.172)	(0.648)	(0.463)	(0.403)
Diff Democratic Votes Share t-1	0.009	0.481**	1.253	-0.163	-0.641
		(0.192)	(1.387)	(1.041)	(0.801)
Diff Democratic Votes Share Congress t-1	-0.006	0.56	-2.351	1.291	0.747
		(0.649)	(2.989)	(2.231)	(1.721)
Diff Contributors Per Capita Left	0.002	0.097***	-0.03	0.002	-0.002
		(0.023)	(0.039)	(0.026)	(0.023)
Diff Contributors Per Capita	0.004	0.092***	-0.079	-0.005	-0.019
		(0.03)	(0.052)	(0.041)	(0.036)
Panel C. - Lagged Unionization Outcomes					
Per New Unions Prev	0.494	4.612***	1.082	0.619	1.754
		(0.93)	(2.078)	(1.434)	(1.266)
Per Yes Prev	0.007	0.784***	-0.049	0.016	-0.034
		(0.097)	(0.217)	(0.169)	(0.141)
Non Zero Variable Observations		4809	1740	2594	3084

Notes: This table reports sample means and coefficients from regressing several pre-determined covariates on independent and instrument variables. Column 3 presents the τ coefficients of (C.1). Columns 4-6 present the τ coefficients of (C.2) under different close NLRB elections bandwidths. Regression models are displayed in (C.1) and (C.2). Robust standard errors appear in parenthesizes.

Source: Source Source Source Source Source Source Source Source Source

Table C.9: County-Level Presidential Elections Summary Statistics

Variable	Mean	NUW	LuckShock-$\delta = 2.5\%$	LuckShock-$\delta = 5\%$	LuckShock-$\delta = 7.5\%$
(1)	(2)	(3)	(4)	(5)	(6)
Panel A. - Covarites					
Percentage Male	0.49	-0.099***	0.024	0.021	0
		(0.026)	(0.06)	(0.041)	(0.034)
Percentage Black	0.132	1.4***	-0.082	0.53	0.612
		(0.292)	(0.783)	(0.523)	(0.402)
Percentage Elders	0.125	-0.092***	-0.13	-0.05	-0.043
		(0.035)	(0.115)	(0.081)	(0.066)
Panel B. - Lagged Political Outcomes					
Diff Turnout t-1	0.001	-0.001	-0.022	-0.055	-0.005
		(0.063)	(0.191)	(0.13)	(0.107)
Diff Democratic Votes Share t-1	0.011	0.216**	0.51	0.087	0.032
		(0.107)	(0.418)	(0.281)	(0.22)
Diff Democratic Votes Share Congress t-1	-0.003	0.371	0.92	0.02	0.127
		(0.227)	(0.826)	(0.55)	(0.445)
Diff Contributors Per Capita Left t-1	0.002	0.046***	0.006	0.016	0.012
		(0.015)	(0.018)	(0.013)	(0.012)
Diff Contributors Per Capita t-1	0.004	0.044**	-0.009	0.016	0.002
		(0.017)	(0.025)	(0.02)	(0.015)
Panel C. - Lagged Unionization Outcomes					
Per New Unions Prev	0.449	2.959***	0.286	-0.124	0.704*
		(0.54)	(0.659)	(0.496)	(0.418)
Per Yes Prev	0.007	0.771***	-0.077	0.136	0.179
		(0.129)	(0.219)	(0.142)	(0.148)
Non Zero Variable Observations		11276	2758	4818	6172

Notes: This table reports sample means and coefficients from regressing several pre-determined covariates on independent and instrument variables. Column 3 presents the τ coefficients of (C.1). Columns 4-6 present the τ coefficients of (C.2) under different close NLRB elections bandwidths. Regression models are displayed in (C.1) and (C.2). Robust standard errors appear in parenthesizes.

Source: Source Source Source Source Source Source Source Source Source

C.2 Serial Correlation Test

A serial correlation of the *LuckShock* will indicate that part of the shock is due to counties' characteristics; these characteristics could be correlated with the outcome variable leading to a violation of the exclusion restriction. A least-squares regression of a simple one-period lagged AR(p) model is conducted to test for such serial correlation.

$$LuckShock_{it} = LuckShock_{i,t-1} + e_{it} \tag{C.3}$$

Both variables were residualized to period FE and the running variable. The test was conducted at different bandwidths- 2.5%, 5%, 7.5%. An equivalent test was conducted for the dependent variable- NUW. Results of the test are shown in table C.10 and indicate that while the dependent variable is significantly auto-correlated ($\beta = 0.21$, $R^2 = 0.17$ $t = 4.6$), the shock isn't significant for any specification with very small $\beta's$ and negligible R^2.

Table C.10: Serial Correlation Test

	NUW	LuckShock- $\delta = 2.5\%$	LuckShock- $\delta = 5\%$	LuckShock- $\delta = 7.5\%$
	(1)	(2)	(3)	(4)
Per New Unions Prev	0.2313***			
	(0.0517)			
LuckShock Prev		0.0322	0.0436	0.0519*
		(0.0370)	(0.0303)	(0.0306)
Observations	8,382	8,382	8,382	8,382
R^2	0.18136	0.00095	0.00185	0.00268
Adjusted R^2	0.18136	0.00095	0.00185	0.00268

Notes: This table reports .
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

C.3 Spatial Correlation Test

This appendix will test the *LuckShock* spatial correlation. The test is based on a simple regression model:

$$LuckShock_{it} = LuckShock_{it}^{N} + e_{it} \tag{C.4}$$

$LuckShock_{it}^{N}$ denotes the *LuckShock* in period t of CZ neighbor to CZ i. Table C.11 shows results for a sample that includes all CZ pairs for shocks based on different bandwidths; an equivalent test was conducted for the dependent variable- NUW; standard errors are clustered at the CZ level. Similar to the serial correlation test, the spatial correlation test indicates that while the independent variable -NUW is spatially correlated ($\beta = 0.03$), the instrument coefficients are very small, and there is no significant positive spatial correlation for any of the definitions of the $LuckShock$. In the only place with a slightly significant correlation(column 2), the correlation is negative, which indicates that it probably stems from random noise.

Table C.11: Spatial Correlation Test

	NUW	LuckShock- $\delta = 2.5\%$	LuckShock- $\delta = 5\%$	LuckShock- $\delta = 7.5\%$
	(1)	(2)	(3)	(4)
Per New Unions Neighbor	0.0307***			
	(0.0107)			
LuckShock Neighbor		-0.0075**	-0.0013	-0.0060
		(0.0035)	(0.0043)	(0.0039)
Observations	118,052	118,052	118,052	118,052
R^2	0.00185	0.00013	-4.22×10^{-6}	8.99×10^{-5}
Adjusted R^2	0.00185	0.00013	-4.22×10^{-6}	8.99×10^{-5}

Notes: This table reports .
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

D Robustness

This appendix includes several robustness tests conducted on the main model. Commuting zones that contain counties from several states were excluded from the sample in all tests (20% of the CZ, 25% of the total population).

D.1 Other Denominators

In the main specification, the instrument, the independent variable, and the outcome variable are all calculated as shares of the total votes in the presidential elections. Table D.12 reports estimations of the paper's main effects after replacing the denominator in three other measures- the CZ population, the CZ adult population, and the CZ's votes for the two major parties. The first column is identical to column 5 of the main results table 5, and columns 2-4 report the results for the same regression's specification for the three denominators mentioned above, respectively. The different specifications yield very similar results. The minor differences between those estimations can stem from the union's effects on turnover and on voting for minor parties.

Table D.12: Robustness to Different Denominators

	Delta Dem Share			
	(1)	(2)	(3)	(4)
New Unionized Workers	1.678***	1.300***	1.162**	1.216**
	(0.4513)	(0.3892)	(0.5617)	(0.5644)
δ	7.5%	7.5%	7.5%	7.5%
$\overline{R^C_{it}}$	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	3,084	3,084	3,084	3,084
Close NLRB Elections	16,343	16,343	16,343	16,343
Denominator	Votes	Two Main Parties Votes	Eligable to Vote Population	Total Population
Observations	8,382	8,382	8,382	8,382
R^2	0.55722	0.57726	0.59141	0.57629
Within R^2	0.02442	0.03655	0.04839	0.05148
year fixed effects	✓	✓	✓	✓

D.2 Controlling for Union Strength

A potential counterargument to this paper's identification strategy might suggest that in areas where unions are gaining strength, their likelihood of winning close unionization elections increases, coinciding with regions where the Democratic party is also gaining momentum. This overlap could be attributed to manipulations and contested votes in very close elections.

The first five columns of table D.13 deal with such a story by adding to the standard specification controls for two measures of unions' strength: (1) The average share of votes for union formation in all unionization elections that will be denoted by $\overline{R_{it}}$ (2) The share of workers involved in New Unionization Attempts out of the workforce; this variable is denoted by NUA. These variables can be derived easily from the NLRB dataset. Both variables weren't included in the main specification because they can be seen as bad controls. [30] Table D.13 presents the results of this

[30] For example, suppose that immediately after the Presidential election, a close union win in one workplace led to more unionization attempts (not necessarily close) in other workplaces in the

exercise. The first column is the same as column 6 of table 5. Columns 2-5 add different sets of the two variables. The coefficients are almost identical, suggesting that the main results do not stem from the *LuckShock* being an approximation for union strength.

D.3 Allow for Variations in ϕ_u

A more comprehensive way to deal with the general threat that the *LuckShock* is correlated with pre-conditions is to allow the estimated chance of winning close elections ϕ_u to vary based on NLRB election characteristics. By allowing ϕ_u to vary based on observables, the threat to identification is reduced to the option of unobservable pre-conditions that affect the results of close NLRB elections and voting in Presidential elections.

Instead of setting every ϕ_u to be a constant $\hat{\phi}$, it will be estimated (based on actual close elections results) as a function of observable election characteristics:

$$\phi_u = f(Z_u)$$

Two characteristics will be used: the workplace's 2-digit industry and the NLRB regional office that handles the unionization attempt (a total of 26 regions). Those characteristics are available for every NLRB election[31].

Columns 6-10 of table D.13 present the results of this exercise. In column 6 ϕ_u is calculated separately for each region as the share of close elections ending in a union win. In column 7, ϕ_u is allowed to vary linearly over time in each region. Columns 8-9 are equivalent for cells of two-digit industries. In column 10, ϕ_u may vary based on both the NLRB region and the workplace industry.

Results in the table are fairly stable, indicating that the main effect probably doesn't stem from *LuckShock* correlation with observable pre-conditions. The slight reduction in coefficient size can be due to the over-fitting of ϕ_u values.

same CZ. Later, those attempts may increase voting for the Democratic party candidate in the next election.

[31]The industry variable is missing for elections after 2009 and will be completed later (MK)

D.4 Exclusion of Observations with Zero Close Unionization-Elections

The main sample includes observations with zero close elections and, thus, a luck shock of precisely zero. This inclusion is valid econometrically[32] and allows me to use a balanced panel with effect in the regression models. Yet, including such observations aren't obligatory and distance the estimations used in the paper from canonical RD models that include only observations that had a possibility of receiving treatment. Table D.14 shows that the paper's main estimations are robust to such exclusion. Even columns represent regression models with the sample and controls included in column 5 of table 5 for close elections bandwidths of 5%&7.5%, odd columns represent estimations with the same controls for a sample that excludes observations with no close elections. The coefficient is very similar. Estimation for the excluded sample yields a bit higher standard error, probably due to nosier estimates of the control covariates.

D.5 Different Model Specifications

The independent variable throughout the paper is NUW- the share of newly unionized workers. In this sub section, I will show that the results are robust to the following two specifications:

1. Replacing NUW in NUW^C– the share of works unionized through close elections.
2. Using a reduced form equation– estimating the luck shock effect on voting in Presidential election directly.

The first specification will narrow the mechanisms through which the instrument-$LuckShock$ can effect new unionization. In the main specification, the $LuckShock$ can affect unionization directly through two mechanisms: (1) directly through close NLRB elections; (2) indirectly through increasing new unionization attempts or increasing the vote share for unions in landslide elections. The first specification proposed here excludes the later mechanism.

The second specification ignores the fuzzy design of NLRB elections–getting a majority of the votes does not guarantee union certification due to the (slight) chance

[32] The luck Shock definition in **??** is proper for counties with no close election. In addition, note that $E[LuckShock|\#CloseElection > 0] = E[LuckShock|\#CloseElection = 0] = 0$

Table D.13: Robustness of Main Results

	Delta Dem Share									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
New Unionized Workers	1.550***	1.477***	1.635***	1.509***	1.652***	1.538***	1.527***	1.475***	1.458***	1.316***
	(0.4591)	(0.4787)	(0.4456)	(0.4490)	(0.4873)	(0.5191)	(0.5041)	(0.4782)	(0.4726)	(0.4601)
δ	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%
Additional Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	3,084	3,084	3,084	3,084	3,084	3,084	3,084	3,082	3,084	3,084
Close NLRB Elections	16,338	16,338	16,338	16,338	16,338	16,338	16,338	16,338	16,338	16,338
$\overline{R_{it}}$		✓		✓	✓					
NUA			✓	✓	✓					
$NUA : \overline{R_{it}}$					✓					
ϕ_u	Average	Average	Average	Average	Average	Region	Region*Time	Industry	Industry*Time	Region+Industry
Observations	8,382	8,382	8,382	8,382	8,382	8,382	8,382	8,382	8,382	8,382
R^2	0.55352	0.55543	0.56386	0.56632	0.56777	0.55371	0.55386	0.55460	0.55483	0.55662
Within R^2	0.01565	0.01985	0.03843	0.04386	0.04705	0.01606	0.01639	0.01804	0.01853	0.02249
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes:
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

Table D.14: Robustness to the Exclusion of Zero Close-Elections Observations

	Delta Dem Share			
	(1)	(2)	(3)	(4)
New Unionized Workers	1.524***	1.553***	1.567***	1.678***
	(0.5607)	(0.5208)	(0.4736)	(0.4513)
δ	5%	5%	7.5%	7.5%
$\overline{R_{it}^C}$	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓
Unzero Luck Observations	2,594	2,594	3,084	3,084
Close NLRB Elections	10,523	10,523	16,343	16,343
Observations	2,594	8,382	3,084	8,382
R^2	0.58344	0.55831	0.57893	0.55722
Within R^2	0.01241	0.02681	0.01108	0.02442
year fixed effects	✓	✓	✓	✓

of challenging election results. The main specification controls the option of decertification by using the share of eventually unionized workers as the independent variable in the second stage. Thus, a reduced-form version of the main regression equation measures an intent-to-treat version of the main effect estimated in this paper. Table D.15 shows the main specification of this paper (column 6 of table 5) in addition to equivalent specifications that are based on the two proposed specifications. All coefficients are very similar, which doesn't allow to conclude the correct model or the consequences of the differences between the models.

D.6 Excluding each State and each Period

Figure D.7 show that empirical results do not stem from only one state or only one period (a 4-years election cycle). The graph shows the main specification of this paper (column 6 of table 5) for filtered samples. The vertical axis of panel (a) indicated an excluded state, and the horizontal axis indicated the estimated effect based on the filtered sample; the red line is the paper's main effect. Panel (b) displays equivalent estimations for samples with excluded periods. The panels indicate that the results' significance is robust to the exclusion of each state and the exclusion of both. The exclusion of 1984's election cycle lowers the estimated effect by almost a third Such a decline is consistent with Reagan's intense activity against labor unions in his first term, the exclusion of Georgia lowers the effect by approximately 15%.

Table D.15: Robustness to Different Model Specifications

	Delta Dem Share		
	(1)	(2)	(3)
New Unionized Workers	1.678***		
	(0.4513)		
New Close Unionized Workers		1.893***	
		(0.5075)	
Luck Shock			1.770***
			(0.4733)
δ	7.5%	7.5%	7.5%
$\overline{R^C_{it}}$	✓	✓	✓
Additional Controls	✓	✓	✓
Unzero Luck Observations	3,084	3,084	3,084
Close NLRB Elections	16,343	16,343	16,343
Second Stage Indpendent Variable	NUW	NUW^c	-
Regression Model	IV	IV	OLS
Observations	8,382	8,382	8,382
R^2	0.55722	0.56301	0.56322
Within R^2	0.02442	0.03718	0.03762
year fixed effects	✓	✓	✓

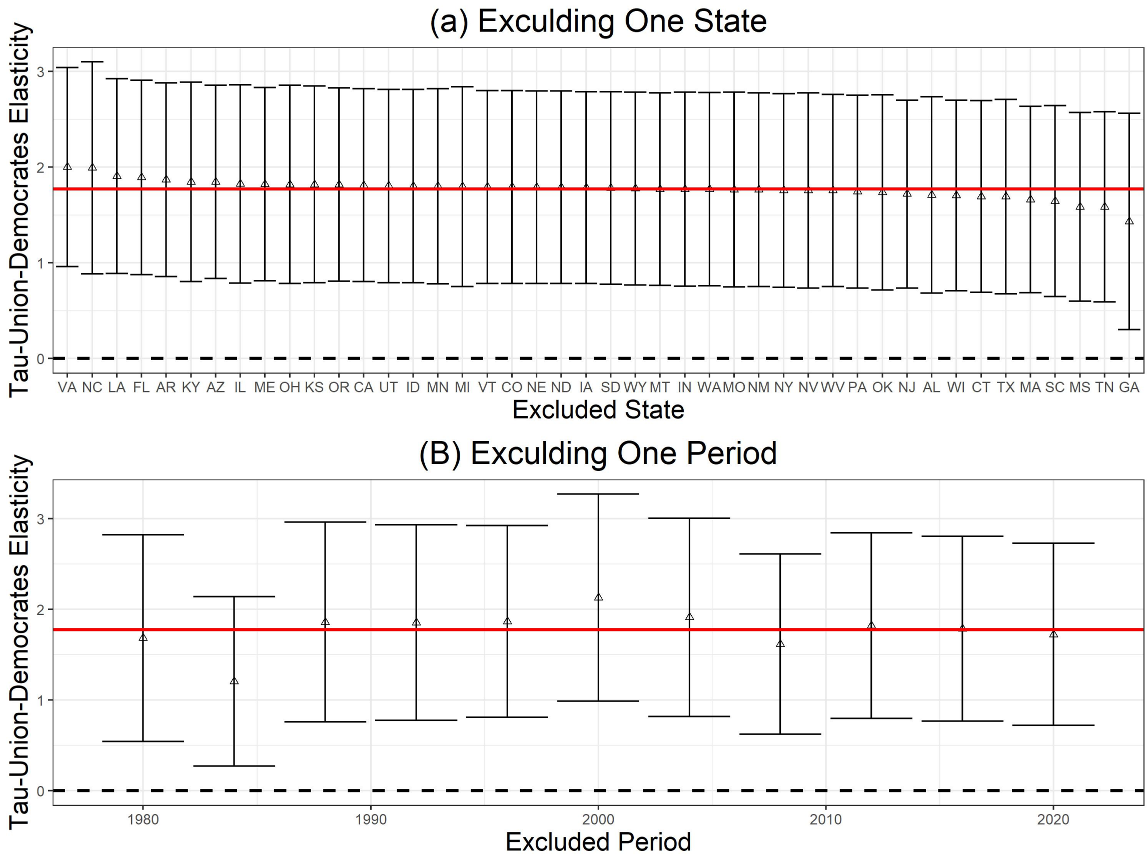
E Heterogeneous Effect

E.1 Heterogeneity by periods

Table E.16 presents estimations equivalent to this paper's main results for different time ranges. Each column represents two or three election cycles (estimation of separate effect for each election cycle is problematic and yields very high standard errors, presumably due to the small variation in the treatment). In each column, the sample is restricted to include only observations from the relevant cycles. Controls variable are the controls in column 6 of table 5.

For the first time range of 1980-1988, point estimates are significant and a bit higher than the paper's estimated effect. The two following periods yield estimators smaller than the main effect and close and undistinguished from zero. A Smaller effect in this period fits well with historical narratives which present the Democratic Party of those years as more centrist. The point estimate is pretty big in the last period (2016-2020). Still, standard errors are also big due to the negligible rate of new unionizations, preventing the ability to conclude about unions' effect.

Figure D.7: Robustness to Exclusion of Each State and Election Cycle



E.2 Heterogeneity by Time since Unionization

In the main analyses, the instrument ($LuckShock$) and the independent variable ($LuckShock$) are based on aggregating counties' NLRB election results in four-year election-cycle intervals. In this sub-section, I estimate this paper's main results with an instrument and independent variables based on one-year intervals. This allows me to estimate heterogeneous effects by time since unionization. The following version of **??** and **??** are used to this estimation:

$$Y_{i,t+1} - Y_{it} = \tau^T NUW_{it}^T + \theta^T \overline{R_u^{CT}} + X_{it}\beta' + \eta_{it} \tag{E.5}$$

$$NUW_{it}^T = \gamma_1^T LuckShock_{it}^T + \gamma_{2t}^T \overline{R_u^{CT}} + X_{it}\gamma_3' + \nu_{it} \tag{E.6}$$

T represents years before Presidential elections (NUW_{it}^4 is the share of newly

Table E.16: Heterogeneity by periods

	ΔDemShare					
	(1)	(2)	(3)	(4)	(5)	(6)
Luck Shock	1.271	1.840*	0.4534	-0.0430	5.928***	5.375**
	(0.7964)	(0.9797)	(0.6526)	(0.6617)	(2.159)	(2.597)
δ	5%	15%	5%	15%	5%	15%
Time Range	1980-1988	1980-1988	1992-2004	1992-2004	2008-2020	2008-2020
Running Variable Controls		✓		✓		✓
Additional Controls	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	5,295	7,063	5,903	8,288	3,762	5,902
Close NLRB Elections	4,620	15,113	4,065	13,110	1,913	5,622
Observations	9,259	9,259	12,374	12,374	12,338	12,338
R^2	0.66582	0.66608	0.50639	0.50850	0.55275	0.55367
Within R^2	0.25151	0.25207	0.17825	0.18176	0.23061	0.23218
year fixed effects	✓	✓	✓	✓	✓	✓

unionized workers through NLRB elections that took place four to three years before the Presidential elections). Each column in table E.17 presents results for one of the four possible values of T. Estimates for T=3,4 are significant and higher than the paper's main effect. Estimates for T=1,2 are insignificant and smaller than the paper's main effect. The relatively high points estimates for T=3,4 indicate a significant influence that passes through the indirect mechanism of union allocating dues to political campaigns. Unions begin to charge dues only after signing a collective agreement; negotiating such an agreement can take several months to a few years. Thus, the union dues channel is expected to be more influential for NLRB elections that took place earlier (relative to the Presidential Elections).

E.3 Heterogeneity by CZ size

Most variation in the *LuckShock* comes from CZs with a small population. In large CZs, there will be more close NLRB elections, and the share of workers out of the entire CZ voters population in each close NLRB election will be smaller. Due to the law of large numbers, close wins and close losses will tend to balance each other in large areas, and the *LuckShock* will lean to 0.

If there is heterogeneity in unions' effect between large and small CZs, the RDA method will give more weight to small CZs. Table E.18 check if overweighting small CZs drives the results identified before. Column 1 is the same as Column 6 of table 5. Columns 2-7 restrict the sample to include only observations with an average adult

Table E.17: Time since Unionization

	$\Delta DemShare$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Luck Shock(t=1)	1.809	2.548*						
	(1.507)	(1.521)						
Luck Shock(t=2)			1.997**	1.247				
			(0.7928)	(1.029)				
Luck Shock(t=3)					1.553*	1.858*		
					(0.9135)	(1.073)		
Luck Shock(t=4)							1.024	0.7328
							(1.298)	(1.540)
δ	5%	15%	5%	15%	5%	15%	5%	15%
Time Range	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020	1980-2020
Running Variable Controls		✓		✓		✓		✓
Additional Controls	✓	✓	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	14,960	21,253	14,960	21,253	14,960	21,253	14,960	21,253
Close NLRB Elections	2,483	7,951	2,217	7,403	2,227	7,028	1,816	5,702
Observations	33,971	33,971	33,971	33,971	33,971	33,971	33,971	33,971
R^2	0.55260	0.55281	0.55270	0.55318	0.55262	0.55282	0.55251	0.55251
Within R^2	0.14088	0.14127	0.14106	0.14199	0.14092	0.14130	0.14069	0.14069
year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

population larger than some threshold. Each column contains a different threshold, beginning at 5,000 and ending at 200,000. Statistics regarding the number of CZs and the rate of the U.S. population living in them are presented in each column. The table clearly shows that coefficients are increasing in the sample restriction criteria.[33] This increase indicates that unions' effects on voting are probably larger in larger CZs. Those CZs get reduced weight in this paper identification method. The general population's average treatment effect is probably greater than the effect found in the paper. Thus, the estimators in the paper should be interpreted as lower bound estimations.

F Direct Effects of Unionization

The aim of this appendix is to estimate the direct impact of unionization, which refers to the union status effect at the individual (and household) level. Such estimation is in line with two previous works (Freeman, 2003; Silver, 2011) and can shed light on one mechanism of the main effect estimate above.

The estimation is based on the "Cooperative Election Study" dataset - a national stratified sample survey administered by YouGov conducted yearly since 2005 and deals mainly with political issues. The dataset contains 372,242 observations for the

[33]A possible explanation for such an increase is that in large CZs, local unions will use dues in political activity within the CZ borders, while in smaller CZ, the money will spill over to other CZs.

Table E.18: Sample Restriction by CZ's Population

	Delta Dem Share						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
New Unionized Workers	1.663***	1.664***	1.700***	1.699***	1.948***	2.434***	2.485**
	(0.4507)	(0.4509)	(0.4554)	(0.4623)	(0.5602)	(0.7513)	(0.9842)
δ	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%
$\overline{R^C_{it}}$	✓	✓	✓	✓	✓	✓	✓
Additional Controls	✓	✓	✓	✓	✓	✓	✓
Unzero Luck Observations	3,084	3,081	3,066	3,036	2,777	2,306	1,638
Close NLRB Elections	16,343	16,339	16,324	16,291	15,955	15,184	13,726
Min Average Pop	0	5,000	10,000	20,000	50,000	100,000	200,000
Number of Counties	762	713	660	596	442	305	185
Share of Total Popualtion	100%	99.9%	99.7%	99.3%	96.6%	91.7%	83%
Observations	8,382	7,843	7,260	6,556	4,862	3,355	2,035
R^2	0.55630	0.55637	0.55596	0.55569	0.55573	0.55828	0.57907
Within R^2	0.02780	0.02784	0.02650	0.02581	0.01837	-0.00441	-0.00933
year fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports this paper's main results for different criteria of minimum average size of the CZ. All estimates are of the second stage equation (**??**) and contain the same covariates as column 6 of table 5. Column 1 identical to column 6 of table 5. Columns 2-7 restrict the sample to include only observations with an average adult population larger than some threshold. Robust standard errors are in parenthesis.
* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source

five Presidential Elections between 2004 and 2020. Besides demographic and voting variables, the dataset contains union status and union family status variables that indicate if another person in the individual household is a union member. More information about the dataset construction is available in the data appendix B.

Two regression models will estimate the union effect on individual voting. The first is the OLS model with demographic controls. The second exploits the large dataset and the rich demographic information using a matching method that matches each union member or member in a union household individual identical in all demographic parameters besides the union status. The following regression equations will be used:

$$Dem_j = \beta_1 Union_j + \beta_2 UnionHH_j + X_j \gamma' + \epsilon_j \tag{F.7}$$

$$Dem_j = \beta_1 Union_j + \beta_2 UnionHH_j + \sum_x d_{jx}\alpha_x + \epsilon_j \tag{F.8}$$

j is an index for individual, Dem_j indicates if j voted for the Democratic President Nominee. $Union_j$ indicates if j is a union member. $UnionHH_j$ indicates that j is not a union member, but someone in his household is. X_j is a vector of demographic characteristics- Gender, Education Group[34], Race, Marriage status, Age, Year of the election, State. d_{jx} is a dummy variable that indicates $X_j = x$. [35] The sample contains all individuals who vote. Survey weights were inflated so that the total weight of each year would be equal.

Table F.19 presents the estimated effects for the pooled dataset, appendix **??** present the same effects for each survey year separately. The Third row in the table presents the estimated effect of one Unionized Worker on Democratic Party votes. It is calculated as the sum of the "Union Member" effect and the "Family Union Member" effect multiplied by the ratio between members in union households and the number of unionized individuals.[36] The individual effect is 0.07-0.08, and the full Unionized Worker effect is 0.1-0.12. Effects are quite stable over the years, while the last elections yielded a bit smaller coefficients.

[34] Highest level of education, six groups- No high school, High school graduate, Some college, 2-year of college, 4-year of college, Post-grad

[35] For matching, 10-year age groups are used instead of exact age

[36] Unionized Worker Effect = Union Effect + Union Family Effect $* \frac{\#FamilyUnionMember}{\#UnionMember}$

Table F.19: Unions Effects- Individual Level

	dem						Delta Dem Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
unionTRUE	0.1078***	0.0803***	0.0799***	0.0825***	0.0578***	0.0720***		
	(0.0146)	(0.0132)	(0.0128)	(0.0047)	(0.0044)	(0.0057)		
union_hhTRUE				0.0664***	0.0358***	0.0409***		
				(0.0044)	(0.0042)	(0.0052)		
New Unionized Workers							1.550***	1.645*
							(0.4591)	(0.8570)
One Unionized Worker Effect	0.1078	0.0803	0.0799	0.1393	0.0884	0.1070	1.550	1.645
Max One Unionized Worker Effect Effect	0.1364	0.1061	0.105	0.1496	0.0982	0.1198		
Years	2004-2020	2004-2020	2004-2020	2008-2020	2008-2020	2008-2020	1980-2020	2004-2020
Method	OLS	OLS	Matching	OLS	OLS	Matching	RDA	RDA
Observations	372,242	372,242	372,242	369,321	369,321	369,321	8,382	3,810
R^2	0.00370	0.16769	0.64882	0.00433	0.17150	0.57797	0.55352	0.51481
Within R^2		0.00235	0.00211		0.00198	0.00284	0.01565	0.10372
gender fixed effects		✓			✓			
educ fixed effects		✓			✓			
race fixed effects		✓			✓			
marriage status fixed effects		✓			✓			
age fixed effects		✓			✓			
year fixed effects		✓			✓		✓	✓
state fixed effects		✓			✓			
demographics_cell fixed effects			✓			✓		

Notes: This table reports estimations of the union status effect on individual and family members. Estimations are based on the "Cooperative Election Study" for presidential elections from 2004 to 2020. Column 1 is based on (F.7) and Column 2 is based on (F.8). Each demographics cell contains unique combination of the folowing variables: gender, educ ,race, marriage status, agen group, year and state. Robust standard errors are in parenthesis

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Source: Source Source Source Source Source Source Source Source Source