THE SURPRISING IMPACTS OF UNIONIZATION: EVIDENCE FROM MATCHED EMPLOYER-EMPLOYEE DATA*

Brigham R. Frandsen †
Brigham Young University

December 24, 2014

Abstract

Using administrative data matching individual worker earnings to employers in a regression discontinuity design based on close union representation elections, this study presents new evidence on the impacts of unionization on establishment and worker outcomes. The paper first shows evidence that close union elections are subject to nonrandom selection, with large discontinuities in pre-election characteristics at the majority threshold. Estimates accounting for this selection show, perhaps surprisingly, that unionization significantly and substantially decreases establishment-level payroll, employment, average worker earnings at the establishment, and the probability of establishment survival. Estimates show the decreases in payroll and earnings are driven by union impacts on the composition of workers at unionization establishments, with older and higher-paid workers more likely to leave and younger workers more likely to join or stay. Worker-level effects on the earnings of workers who stay are small. The distinction between the large negative establishment-level effects and small worker-level effects is interpreted in a model of employer and employee selection into union jobs.
JEL Codes: J01, J51

*This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants: SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The author is indebted to JP Ferguson, Jim Davis, Javi Miranda, and Kristin McCue for help with the data. The author also gratefully acknowledges support from NSF Grant SES-0922355. Word count (text and appendix): 7,973.

†181 Faculty Office Building, Provo, UT 84602, Fax: (801) 422-0194, Tel: (801) 422-4049, E-mail: frandsen@byu.edu.

1 Introduction

The impact of unions is a long-standing and important question in economics. The relative role the dramatic decline in U.S. private sector unionization has played among the institutional forces contributing to increasing inequality over the past three decades depends on the effects unions have on establishments and workers (Acemoglu, Aghion, and Violante, 2002). These effects are also important for understanding the impacts of collective bargaining in settings where unions continue to be prevalent, such as government and service sectors.

Existing evidence on the impact of unions is mixed and conflicting, however. Studies using large worker-level datasets have consistently found large union wage premia (Freeman and Medoff, 1984). More recent data suggests the union wage premium may have fallen since the early 1980s, but remains substantial (Blanchflower and Bryson, 2003). Studies using establishment-level data, however, have typically found small-to-nil effects, in seeming contradiction to the worker-level estimates (Freeman and Kleiner, 1990; Lalonde, Marschke, and Troske, 1996; DiNardo and Lee, 2004).

The discrepancies between worker-level and establishment-level estimates have presented a puzzle to researchers. Offered explanations point to differences in timing (establishment-level studies focus on new unionization, while worker-level studies are driven by long-established unions) or differences in the reliability of research designs (establishment-level studies have used plausibly more credible quasi-experimental designs).

Even aside from differences of data and design, worker-level estimates of the union wage premium and estimates of the effect on establishment-level average wages will differ because they seek to estimate different economic quantities. The worker-level union premium is a price effect; it answers the question: what impact do unions have on the price employers pay for labor of a given skill level? The effect on establishment-level average worker wages conflates this price effect with a composition effect. If the union premium differs across skill levels and unions affect the skill composition of workers at an establishment, then the effect on establishment-level average wages can differ from the average union premium. Disentangling the price and composition effects requires matched data on workers and establishments, which has long been an obstacle, as large, economy-wide datasets matching employers and employees have only recently become available.

This study is the first to exploit economy-wide data matching employers and employees to identify the effects of unionization on establishments and workers. In so doing, it presents evidence for surprising union impacts relative to the prior literature while at the same time unifying worker- and establishment-level estimates by separately identifying effects on establishment survival, payroll, employment, average earnings, and worker composition. Regression discontinuity estimates based on close union elections show union representation sharply reduces establishment payroll, employment, and average earnings of workers at the establishment. These effects are driven not by wage cuts for given workers, but by the effect of unionization on what kinds of workers end up employed at unionizing establishments: older and higher-paid workers are more likely to leave following union certification, while younger workers are more likely to join or stay. Earnings of workers who stay employed at the establishment before and after union certification are little affected on average.

This study also presents new evidence for nonrandom selection in close union certification elections. Close elections swing significantly more frequently in the employer's favor than would be expected were they quasi-random. This sorting leads to substantial discontinuities in average pre-election characteristics as a function of the union vote share across the 50-percent threshold. Narrowly unionizing establishments are also significantly less likely to survive to five to seven years after union certification, which could be both a consequence and a source of nonrandom selection among observed establishments.

Finally, this study makes a methodological contribution. It proposes, tests, and implements strategies for identifying causal effects in a regression discontinuity (RD) design in the presence of confounding selection at the RD threshold such as that described above. The econometric approach uses panel data to control for discontinuities in pre-election characteristics in two ways: (1) identifying effects on first-differenced outcomes in a hybrid of the RD and difference-in-differences designs, and (2) conditioning nonparametrically on and integrating over pre-election outcomes in the RD estimation. The strategies lend themselves to powerful specification tests, and, in this setting, give mutually consistent results.

The next section describes the institutional details of U.S. private sector union representation elections. Section 3 describes the data used in the empirical work. Section 4 lays out the econometric framework for identifying and estimating effects in the regression discontinuity design in the presence of confounding selection. Section 5 presents evidence for non-random selection in private

sector union representation elections as well as estimates of the effects of union representation which account for the selection, and interprets the estimates in terms of a simple model of union selection on the part of employees and employers. Section 6 summarizes the findings and concludes.

2 Background

Since 1935, most U.S. private sector unionization has been governed by the National Labor Relations Act (NLRA), which specifies the rights of unionizing workers. The traditional process by which workers unionize is through a National Labor Relations Board (NLRB) secret ballot election, although an employer may voluntarily bargain with the workers' chosen representative, or in some cases may be required to do so even without an election.[1] The following steps describe the nominal path a group of workers follows to form a union:[2]

1. Petition drive: Union organizers lobby workers, collect signatures expressing a desire for union representation, and submit a petition to the NLRB to hold an election. If the petition is accepted, the NLRB ascertains the scope of the bargaining unit and sets the election time and place, usually the workplace.

2. Election: Eligible workers vote for or against the union, and the union wins if it receives a strict majority of the votes cast.

3. Certification: If the union wins, the NLRB certifies it as the sole authorized representative of the workers in the bargaining unit, and requires the employer to bargain "in good faith" with the union.

In practice, however, the union certification process can be fraught with disputes and delays, before and after the election (Ferguson, 2008). For example, the employer or union may challenge the scope of the bargaining unit on the initial petition, prior to the election. Any disputes on the bargaining unit scope are resolved by the NLRB prior to the election, after hearings involving both parties, if necessary. During the election itself and subsequent vote count, either side may challenge the validity of individual ballots, and if the challenged ballots could be pivotal in the election, the

[1] Secret ballot election has historically been the dominant form of new unionization, although in recent years voluntary recognition through neutrality agreements and card checks have become more common. (Brudney, 2005)

[2] The simple process laid out here follows the procedures described in NLRB (2010).

NLRB rules on whether to count them or throw them out. The employer or union may submit charges of unfair labor practices at any stage, before or after the election. Unfair labor practices are actions on the part of the employer or the union that are deemed to violate the required "laboratory conditions" of a secret-ballot election, and if the charges are found by the NLRB to be of merit, the results of the election are set aside (Moberly, 2002). Unionizing drive disputes are not rare: charges of unfair labor practices were filed in over 21 percent of representation cases between 1999 and 2004 (Ferguson, 2008).

Disputes like these can lead to interventions in the nominal election process that introduce non-random selection even in close elections involving a large number of voters. Interventions after the election that change the voting outcome—for example, enough ballots are successfully challenged and thrown out that the outcome is reversed—can obviously introduce selection. Charges of unfair labor practices may be strategically leveled after a close election loss, also leading to selection in close elections. However, even pre-election interventions can introduce selection if the employer or union can accurately forecast the potential results of an election. More subtly, an unresolved pre-election dispute can lead to ex-post selection when the supervising authority decides to take up the matter after the election only if it may have been determinative. Disputes are, of course, most likely to be determinative if the election is close.

3 Data

3.1 Union Elections

The analysis uses a dataset on the universe of NLRB union representation election results from 1980 to 2009, combining data obtained from Hank Farber, J.P. Ferguson, and Thomas Holmes.[3] Each record in this dataset represents a union certification election held at an establishment, and includes the number of votes cast for and against union representation, the date of the election, and the employer's name and address. The employer name and address information was used to match to Census establishment and individual earnings data (see below).

Table I reports statistics on election characteristics and outcomes for this sample. The sample includes 45,176 representation elections, involving over four million votes cast. The average number

[3]The union data from Thomas Holmes is available on his website, http://www.econ.umn.edu/~holmes/data/geo_spill/

of voters participating in these elections was 93, and the union won 45 percent of them. The average union vote share was 49 percent.

3.2 Payroll and Employment

Establishment-level data on payroll and employment come from the Census Bureau's Longitudinal Business Database (LBD). The LBD includes yearly data on nearly all non-farm private sector employers from 1976 to 2009.[4] An advantage of this database over the Longitudinal Research Database (LRD) used by DiNardo and Lee (2004) is that the LRD is restricted to establishments in the manufacturing sector. Establishments in the LBD are defined as a single physical location where business is conducted. Thus, two places of business owned by the same enterprise are distinct establishments in the LBD. Over 23 million establishments are included in the LBD, although this study focuses only on those where a union certification election was held from 1980 to 2009.

Union election establishments were identified by in the LBD by matching on employer name and address information. The matching was performed by first standardizing the name and address fields in both datasets using an automated procedure and merging on the standardized values. See the appendix for more details on the matching algorithm. This procedure succeeding in identifying 82 percent of the elections in the NLRB dataset.[5]

The LBD includes yearly payroll and employment data for each establishment, derived from administrative payroll tax data from the Internal Revenue Service (IRS). Table II reports statistics on establishment characteristics from the LBD sample. Establishments where the union won the election had larger employment and payroll than where the union lost, both prior to and after the election. Post-election average worker earnings were lower at plants where the union won, however.

3.3 Individual Earnings

Individual-level earnings were obtained from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) database. The LEHD integrates the universe of unemployment insurance-covered (UI) earnings records held by participating state agencies into a cohesive data

[4]The description of the LBD in this section is based on Jarmin and Miranda (2002).
[5]For comparison, the match rate in DiNardo and Lee (2004) was 26 percent.

structure using person and employer identifiers, allowing linkages to other sources of data.[6]

The Employment History Files (EHF) within the LEHD contain quarterly records of individuals' UI-covered earnings. The EHF for each of the 30 covered states contains a record for each employee-employer combination—a job—that produced at least one dollar of wages in that state in each year. The data cover a period as wide as 1985 to 2008, although for most states the data only go back to the early 1990s. The EHF contains more than 2.8 billion records, although I focus on workers employed at establishments where a union representation election was held.

Workers employed at election establishments were identified using the NLRB-LBD match as a starting point and merging on establishment identifiers in the Census Bureau's Business Register Bridge (BRB), which links establishments in the LBD with the LEHD. The overall procedure matched 77 percent of the NLRB elections held in states and years where the LEHD is available, identifying over 1.7 million individuals employed at election establishments.

Table III reports statistics on worker characteristics and earnings from the matched LEHD sample. Employees at establishments where the union won were slightly older, and slightly more likely to be female and nonwhite. The percentage of workers at establishments before a union election who remained at the same establishment following the election was slightly more than 40 percent.[7]

4 Econometric Framework

4.1 Identification

A fundamental obstacle to measuring the effect of unionization is selection bias: outcomes within unionized plants may differ for reasons other than union representation. One promising approach to overcoming these selection issues, first used in this context by DiNardo and Lee (2004), is a regression discontinuity design based on close union representation elections. If plants and workers where the union barely won and barely lost are comparable, then close union elections approximate a randomized experiment, and the resulting difference in the distribution of outcomes provides a

[6]For more details on the construction and uses of the LEHD database, see McKinney and Vilhuber (2008), Lane (2008), Abowd, Haltiwanger, and Lane (2004), and Abowd, Stephens, Vilhuber, Andersson, McKinney, Roemer, and Woodcock (2009).

[7]This would seem low given typical worker turnover rates; however, note that the denominator includes matched workers at any time prior to the election, including those who may have left before the election.

reliable estimate of the causal effect of unionization.

To formalize this idea, let $D = 1(R > r_0)$ be an indicator for a union victory, where R is the union vote share and $r_0 = 0.5$. Let Y_1 be the outcome that would be realized if the union were to win, and let Y_0 be the outcome otherwise, so that the observed outcome is $Y = Y_0 + (Y_1 - Y_0) D$. The impacts of unionization are captured by comparisons of the conditional distributions of Y_0 and Y_1 at the threshold of the running variable, $R = r_0$. Naturally, if the impacts of union representation are heterogeneous, this is a local effect specific to units at the threshold.

The key assumption for identifying the impact of union representation in this setting is that the conditional distribution of potential outcomes as a function of the union vote share is continuous near the threshold of union victory, and thus any jumps in the observed outcome distribution at the threshold are due to the treatment. Formally:

Assumption 1: Local Continuity $F_{Y_d|R}(y|r)$ is continuous in r at $r = r_0$, for $d \in \{0, 1\}$.

This assumption is likely to be satisfied if, for example, unions, workers, and firms are *a priori* uncertain about the outcome of the election when it is close (see Lee, 2008 for a formal proof). The requisite uncertainty may in turn be more plausible for elections involving a relatively large number of voters, which is why studies using this design often restrict to elections involving 20 or more votes (DiNardo and Lee, 2004; Lee and Mas, 2012; Sojourner, Town, Grabowski, Chen, and Frandsen, 2013).

This assumption will be violated, however, if employers or unions can (with positive probability) accurately forecast the voting outcome and take action to influence it, or alter the election results ex post. The institutional mechanisms of ballot challenges and unfair labor practice accusations are plausible channels through which such ex post contamination could occur.

It is far from certain a priori, therefore, that the RD continuity assumption holds in the union election setting. Fortunately, the RD design lends itself to powerful specification tests. Since Y_0 and Y_1 are only observed on one side of the threshold or the other, the continuity assumption cannot be directly tested, but plausible rationales for Assumption 1 will also imply that the density of R and the conditional distributions of pre-treatment characteristics, including lagged outcomes, will be continuous at $R = r_0$. These implications can be tested directly and, depending on their power, can provide compelling evidence for or against Assumption 1 (McCrary, 2008; Imbens and

Lemieux, 2008; Lee and Lemieux, 2010).

If tests show evidence of discontinuities in the conditional distribution of pre-election characteristics at the threshold, then Assumption 1 is questionable. If pre-election outcomes, Y_{-1}, are observed, then alternative identifying assumptions that relax Assumption 1 can be made. The first combines regression discontinuity with difference-in-differences. Define $\Delta_d \equiv Y_d - Y_{-1}$ to be the potential change in the outcome following a union election, with its conditional cdf given by $F_{\Delta_d|R}(y|r)$. One alternative identifying assumption imposes continuity on the conditional distribution of Δ_d:

Assumption A1': Local Continuity in Differences $F_{\Delta_d|R}(y|r)$ is continuous in r at $r = r_0$, for $d \in \{0, 1\}$.

If the conditional distribution of Y_{-1} is continuous at the threshold, then assumption A1 implies A1'. However, A1' may hold even if A1 is violated, if the factors underlying the selection have a time-invariant impact on outcomes.

A similar, but distinct alternative approach to accounting for discontinuities in pre-election characteristics is to control for pre-election outcomes directly in the estimation. Let $F_{Y_d|R,Y_{-1}}(y|r, y_{-1})$ be the conditional cdf of potential outcomes, conditional on $R = r$ and $Y_{-1} = y_{-1}$. This approach relies on the following identifying assumption:

Assumption A1": Local Conditional Continuity $F_{Y_d|R,Y_{-1}}(y|r, y_{-1})$ is continuous in r at $r = r_0$, for $d \in \{0, 1\}$ and all y_{-1} in the support of Y_{-1}.

The alternative assumptions A1' and A1" are related, but non-nested: one does not imply the other. The empirical analysis compares specifications based on A1' and A1" and reassuringly finds the results are very similar.

If the traditional RD assumption A1 holds, then the average treatment effect at the threshold is identified by comparing the conditional means of post-election outcomes just above and below the threshold: $\delta = \lim_{r \to r_0+} E[Y|R = r] - \lim_{r \to r_0-} E[Y|R = r]$. If alternative assumption A1' holds, then the effect is identified by comparing the conditional means of post-election first differences: $\delta = \lim_{r \to r_0+} E[Y - Y_{-1}|R = r] - \lim_{r \to r_0-} E[Y - Y_{-1}|R = r]$. Finally, if the second alternative assumption

A1" holds, then the effect is identified by integrating over the comparison of means across the threshold conditional on pre-election outcomes: $\delta = E\left[\lim_{r \to r_0+} E\left[Y|R=r, Y_{-1}\right] - \lim_{r \to r_0-} E\left[Y|R=r, Y_{-1}\right]\right]$.

4.2 Estimation and Testing

Following established RD methodology, estimation and testing will consist of graphical analysis accompanied by statistical inference based on local polynomial regressions (Lee and Lemieux, 2010; Porter, 2003). The prototypical RD graphs plot averages of the dependent variable by non-overlapping bins of the running variable that do not straddle the threshold. Local polynomial regressions complement the graphical analysis by allowing formal hypothesis tests and confidence intervals. A typical local linear specification uses weighted least squares to estimate

$$Y_i = \beta_0 + \delta D_i + \beta_1 R_i + \beta_2 R_i D_i + \varepsilon_i$$

with weights $w_i = K\left(\frac{R_i - r_0}{h}\right)$, where K is a kernel function and h is a bandwidth chosen by, say, the procedure in Imbens and Kalyanaraman (2012). The estimate of δ in this specification is the local linear estimator for the average treatment effect on Y at the threshold. If the dependent variable is a pre-treatment characteristic such as a lagged outcome, then a test of $H_0 : \delta = 0$ is a specification check for assumption A1. Testing the continuity of the density of R at the threshold can be carried out as described in McCrary (2008).

If tests of A1 fail, a specification which relies on the weaker assumption A1' and accounts for time-invariant selection at the threshold is

$$Y_i - Y_{i,-1} = \beta_0 + \delta D_i + \beta_1 R_i + \beta_2 R_i D_i + \varepsilon_i, \tag{1}$$

where $Y_{i,-1}$ is a pre-election observation and local linear weighting is also used. This specification is a local linear difference-in-differences specification where the treatment effect δ is identified as the difference between the before-after comparisons across the threshold.

An alternative specification that accounts for failures of A1 and relies instead on A1" also uses weighted least squares to estimate

$$Y_i = \beta_0 + \delta D_i + \beta_1 R_i + \beta_2 R_i D_i + \varepsilon_i,$$

but with weights

$$w_i = K\left(\frac{R_i - r_0}{h}\right) / \hat{P}\left(R_i | Y_{-1i}\right),$$

where $\hat{P}(r|s)$ is a nonparametric estimate of the conditional probability $\Pr(R = r|Y_{-1} = s)$. $\hat{P}(r|s)$ was constructed first discretizing Y_{-1} into decile cells, and then computing the empirical conditional probabilities using the discretized running variables described below.

The running variable, R_i, will be specified in two ways. The primary specification adjusts the union vote share as in DiNardo and Lee (2004) to avoid mechanical discontinuities stemming from differences in the support of the union vote share for elections of different sizes. First, only elections with k or more voters will be included. Second, the raw vote shares will be binned in $100/k$-percent intervals. The support of this adjusted vote share will thus be identical for all included elections. To be precise, if U_i is the number of votes for the union and T_i is the total number of votes cast, the running variable will be constructed as:

$$R_i = \left(\lceil kU_i/T_i \rceil - 0.5\right)/k + .5,$$

where $\lceil \cdot \rceil$ is the ceiling function. Following Dinardo and Lee k will initially be set equal to 20, but in specification tests will vary over a range.

The second running variable specification defines the union margin of victory in terms of the vote tally, not the share. It defines the running variable as the number of union votes minus the number the union needed for victory. Mathematically,

$$R_i^{\text{alt}} = U_i - \left(\lfloor T_i/2 \rfloor + 1\right),$$

where $\lfloor \cdot \rfloor$ is the floor function. In practice these specifications amount to slightly different relative weighting of small and large establishments and different coarseness of the running variable support. The share-based running variable R_i gives large establishments relatively more weight since they are more likely to have close elections in terms of the share, and is quite coarse, with only 20 support points. The tally-based running variable R_i^{alt} gives small establishments relatively more weight since they are more likely to have close elections in terms of the number of votes, and in the dataset has hundreds of support points, allowing for a much finer grid.

For comparison to the previous literature, the main results will be in terms of the share-based running variable, R_i. The tables and Appendix figures report results using the tally-based running variable R_i^{alt} where possible. In some cases the fineness of the tally-based running variable R_i^{alt} resulted in so few observations per cell that Census Bureau confidentiality requirements prevented disclosure of the results, but in all cases (reported and unreported) the results were similar across the two choices for running variable.

Finally, all analyses were performed after collapsing the data on values of the running variable, and weighted by the number of establishments or individuals. The collapsed data satisfy Census Bureau confidential disclosure requirements and, of course, yield identical point estimates to regressions on the original micro data. Heteroskedasticity-consistent inference using the collapsed data is asymptotically equivalent to clustering on the values of the adjusted running variable, as suggested by Lee and Card (2008) to account for specification error arising from a discrete running variable.

5 Results

5.1 Non-random selection in close elections

5.1.1 Vote share density discontinuities

The density of the union vote share shows strong evidence of sorting near the threshold. Figure I shows the distribution of the union vote share (the number of votes for the union divided by the total number of votes) in the sample in 5-percentage point bins. In this figure and elsewhere the analysis is restricted to elections where at least 20 votes were cast, unless otherwise indicated. The mode is around 40 percent, with a significant number of elections in which the union received all votes. The frequency of the bin corresponding to the closest union victories appears to be abnormally low relative to surrounding bins, and a McCrary (2008) test strongly rejects continuity at the threshold (t-stat ≈ 10). The plot suggests the presence of some manipulation or selection which resulted in "too few" establishments experiencing a close union victory. This anomaly is not a function of the way the running variable is defined: Figure A1 in the Appendix shows the distribution of the union margin of victory, defined as the number of union votes minus the number of votes the union needed for victory. The figure shows an anomalous drop in the density at a zero

margin of victory, that is, where the union ended up with exactly the number of votes needed to win.

Sorting in close elections was substantial even in elections with a large number of voters. Previous studies using union elections restricted to elections with at least 20 voters because elections with a larger number of voters were presumably less susceptible to manipulation. However, the union vote share density discontinuity remains large even when restricting to elections with a large number of voters. Figure II plots the density of the union vote share restricting to elections with at least $k \in \{20, 30, 40, 60, 80, 100\}$ voters. The upper-left panel reproduces Figure I with a cutoff of 20 voters. The "hole" in the density corresponding to close union victories persists as larger and larger cutoffs are chosen. The discontinuity is clearly visible even in the lower-right panel which restricts to elections with at least 100 voters. The McCrary density test strongly rejects continuity for all size cutoffs except for 100 where the test becomes less powerful as fewer and fewer elections remain above the cutoff. This evidence suggests that restricting analysis to large elections will not eliminate the selection threat.

5.1.2 Discontinuities in pre-election characteristics

The vote share anomalies documented in the previous subsection suggest that on net close elections have a higher-than-expected probability of favoring the employer. If establishments where close elections favor the employer differ from those where they do not, then the sorting documented in the previous section will lead to discontinuities in pre-election characteristics at the threshold. This section shows evidence for just such discontinuities in pre-election employment, payroll, and worker earnings.

Establishments where the union barely won employed significantly more workers prior to the election than establishments where the union barely lost. Figure III plots average log pre-election employment from the LBD by non-overlapping 5-percent bins in the union vote share. There is a substantial discontinuity at the threshold of union victory.[8] Appendix Figure A2 shows a similarly striking discontinuity in terms of the vote tally, rather than vote share. Local linear estimates in the top row of Table IV show that establishments where the union barely won employed about .13

[8] In Figure III and others there are jumps away from the threshold as well. Much of this is noise: as Figures I and II show, the number of elections drops off quickly at vote shares away from the threshold.

(s.e. = .032) log points more workers prior to the election than establishments where the union barely lost.

Likewise, establishments where the union barely won had larger payrolls prior to the election than establishments where the union barely lost. Figure IV plots average log pre-election payroll from the LBD by non-overlapping 5-percent bins in the union vote share. Establishments where the union barely won have an unexpectedly high payroll. Appendix Figure A3 shows the same discontinuity in terms of the vote tally. Local linear estimates in the second row Table IV show a statistically significant difference of .06 (s.e.=.032).

Finally, establishments where the union barely won also had a more compressed distribution of workers earnings prior to the election than establishments where the union barely lost. Figure V plots the 10th, 50th, and 90th percentiles of log pre-election earnings from the LEHD by non-overlapping 5-percent bins in the union vote share. For the lower-tail percentiles there is a substantial positive discontinuity at the threshold, while for upper-tail percentiles there is substantial negative discontinuity. Local linear estimates in Table IV show that the 10th percentile of pre-election earnings in establishments where the union barely won was .196 (s.e. = .024) log points higher than where the union barely lost, while the 90th percentile was .345 (s.e.=.054) log points *lower*. Establishments where the union barely won therefore had at baseline a much more compressed distribution of workers' earnings than where the union would eventually lose. This compression can be seen more succinctly in the difference between the 90th and 10th percentile of pre-election log earnings. Figure VI plots the 90-10 difference by union vote share, and shows a distinct and large drop at the threshold of union victory. The bottom row of results in Table IV shows the difference is about -.54 log points (s.e.=.034).

Interpreting these differences in light of the vote share discontinuities in the previous section, the evidence suggests that employers were more likely to be able to avert a close union victory at smaller establishments and where workers earnings were more dispersed. Conversely, unions were more likely to be able to avert a close loss at larger establishments and where workers' earnings were more homogeneous.

5.1.3 Discontinuities in post-election establishment survival

Discontinuities in pre-election characteristics are not the only potential source of selection bias: differences in establishment survival probabilities across the union victory threshold could also confound the effects of union representation by inducing compositional changes among observed establishments following a union election. Post-election survival is also an important outcome in its own right.

Narrowly-unionized establishments are significantly less likely to survive following an election than establishments where the union narrowly lost. Figure VII plots the probability of an establishment existing in the LBD—the measure of survival—by union vote share from 4 years prior to the union election to 7 years following the election. Prior to the election there is no discernible discontinuity at the union victory threshold. The probabilities during the year immediately preceding the election are all 100 percent, since the data conditions on establishments that are observed just before a union election. Following the election, however, significant differences begin to emerge by about 3 years after the election, and these differences grow through year 7. Figure A4 in the Appendix plots survival probabilities by vote tally and shows the same pattern. Regression discontinuity estimates in Table V quantify the magnitude of the survival discontinuities. The first row shows that in the year immediately following the election the estimated survival discontinuity is a precisely estimated zero. The second row shows that in year 3, however, the estimated discontinuity in survival probability is a significant 3 to 5 percentage points. At 5 years the estimates are in the 7.5 to 9.5 percentage point range, and the fourth row shows that by 7 years they range from 9 to 11 percentage points, all significant at the 1 percent level. The table's columns show the estimates are robust to the choice of bandwidth, local polynomial specification, and running variable definition.

What do these substantial differences in post-election survival among plants where the union barely won and barely lost mean? First, they mean that estimates of union effects on long-run outcomes that condition on establishments surviving more than a few years out from the election are likely to suffer from an additional source of selection bias. For this reason, the post-election analysis in the remainder of the paper focuses on outcomes observed within the first eighteen months following an election. Second, the differences in survival probability suggest that unionization may increase the probability that an establishment shuts down. This interpretation is consistent with

the negative impacts of unionization on firms' market value found in Lee and Mas (2012). The evidence for sorting and the discontinuities in pre-election characteristics in the previous sections, however, imply that a causal interpretation of the differences in survival probability should be made with caution.

5.2 Effects of union representation on post-election outcomes

The evidence for selection in close elections and the resulting discontinuities in pre-election characteristics of election establishments suggest that level comparisons of outcomes across the threshold of union victory that do not account for these differences would not consistently estimate the causal effect of union representation. Valid inference may still be possible, however, by exploiting the panel nature of the data to account for pre-election differences between establishments on either side of the threshold of union victory. This section presents results that account for pre-election differences in outcomes, suggesting that unionization significantly reduced establishment payroll and employment and shifted the composition of workers remaining at unionizing establishments.

5.2.1 Payroll, worker earnings, and employment

Evidence that takes into account pre-election selection suggests that union representation significantly reduced payroll, employment, and average worker earnings at unionizing establishments. Figure VIII plots the average year-to-year change in log payroll from the LBD by non-overlapping 5-percent bins in the union vote share. The left-hand panel corresponds to the year-to-year change in log payroll just prior to a union representation election, and the right-hand panel corresponds to the year-to-year change straddling the election. The left-hand panel shows no discernible jump in differenced pre-election log payroll at the 50 percent threshold. The continuity in the pre-election differences is consistent with minimal selection on log payroll differences, despite the presence of substantial selection on log payroll levels (cf. Figure IV), and lends credence to Assumption A1' as an alternative to the traditional RD identifying assumption A1. By contrast, the right-hand panel of Figure VIII shows a substantial drop at the 50 percent threshold, implying that unionizing establishments experienced a substantially more negative change in payroll. Appendix Figure A5 shows the same pattern in terms of the vote tally rather than vote share. Table VI reports local linear regression discontinuity estimates corresponding to the figure for log payroll and other out-

comes. The estimates account for pre-election differences in establishments in two ways. The first four columns do so by specifying the dependent variable as a difference, and correspond directly to the figure. The last four columns do so by conditioning nonparametrically on the pre-election levels of the outcome variable as described in Section 4. The first row shows the estimated reductions in log payroll range from minus 8 to 12 log points, and are significant at the 1 percent level. The columns show the estimates are robust to the choice of bandwidth, local polynomial specification, running variable definition, and method of controlling for pre-election outcomes.

RD estimates imply that the negative impact on payroll was driven by a decrease in employment, and to a smaller extent a decrease in average worker earnings at unionizing plants. Figure IX plots differenced log employment by election vote share, with pre-election differences in the left panel and pre- versus post-election differences in the right panel. The left panel reassuringly shows no discontinuity in employment change at the 50 percent threshold. The right-hand panel by contrast shows a clear drop at the threshold. Appendix figure A6 shows the same pattern in terms of the vote tally. The second row of Table VI quantifies the graphical evidence, suggesting that union representation led to a 5 to 9 log point reduction in employment, significant at the one percent level, and robust to alternative specifications.

Average earnings of workers at the plant also appear to be reduced following a union victory. Figure X plots differenced log average earnings by election vote share, with pre-election differences in the left panel and pre- versus post-election differences in the right panel. The right panel shows a clear drop in post-election differenced earnings at the 50 percent threshold, and the third row of Table VI shows the estimated effects generally range from a highly significant 2.5 to a 4 log point reduction, except in the last two columns where the estimates are smaller. The left panel shows a small dip in differenced average earnings prior to the election, which may warrant caution in interpreting the effects on average worker earnings. The evidence below on worker composition is consistent with a reduction in average worker earnings, and lends confidence to these estimates.

5.2.2 Worker composition

The effects on payroll, worker earnings, and employment are starkly counter to conventional wisdom on the effects of union representation. Theories of unionism suggest that unions may trade off pay increases with employment, but these results suggest otherwise. Does union representation

really decrease worker pay while also decreasing employment? The answer is no. The reason is that these effects are driven by the union impact on the composition of workers. A union victory shifts workforce composition towards younger and lower-paid workers, with little impact on average earnings for workers who stayed.

Evidence from the LEHD suggests a union victory shifted the employee composition at an establishment toward younger workers. Figure XI plots the average age of workers by union vote share both pre-election (left panel) and post-election (right panel). The left panel shows little evidence of selection on age, and the right panel shows a noticeable drop in average worker age at the 50 percent threshold. The top row of Table VII quantifies the drop at about -1.8 years (s.e.=.21).

The shift toward younger workers was driven by a relative increase in the hiring rate of younger workers in response to a union victory. Figure XII plots the average age of employees who came to the establishment following a representation election by union vote share, and shows a substantial drop at the 50 percent threshold in the average age of workers leaving the establishment. Table VII shows the estimated effect is about -1.3 years (s.e.=.1).

The workers who were hired following a union victory were also on average lower paid. This can be seen by looking at the effect of union victory on the earnings of workers who came to the establishment post-election. Figure XIII plots median log annual earnings of workers who came following a union election by the union vote share. The plot shows a substantial drop at the 50 percent threshold, implying that a union victory increased the relative hiring rate of lower-paid workers. The third row of results in Table VII estimates a highly significant difference of -.211 log points (s.e. = .052).

The shift toward lower-paid workers was also driven by increased exit of higher-paid workers. This can be seen by looking at the effect of union victory on the pre-election earnings of workers who leave the establishment after a union election. Figure XIV plots average log annual earnings of workers who leave following a union election by the union vote share. The plot shows a substantial jump at the 50 percent threshold, implying that a union victory increased the relative exit rate of higher-paid workers. The fourth row of results in Table VII estimates a large and highly significant difference of .585 (s.e. = .088).

Much of the negative effect on payroll and average worker earnings, then, is largely driven by

the effect of unionization on who leaves and who comes in response to a union victory. Younger, lower-paid workers are more likely to be hired following a union victory, while higher-paid workers are more likely to leave. The result is that average worker earnings at establishments where the union won an election are lower.

But were the earnings of workers who stayed affected by union representation? Figure XV plots the average pre- versus post-election change in log earnings for workers who remained employed at election establishments by the union vote share. The figure shows little evidence of a substantial effect on the earnings of those who stayed. Table VII confirms the graphical evidence, with the corresponding estimate of .003 (s.e.=.016). On average, there is little evidence that union representation had a large impact on the earnings of workers who remained at election establishments.

5.2.3 Interpretation

The single largest establishment-level impact of union representation appears to be the effect on what kinds of workers end up employed at unionized plants with little impact on the average earnings of workers who stay. Together with findings elsewhere that unions compress the distribution of individual-level earnings (Frandsen, 2012), the evidence is consistent with models of unionization that imply that the set of workers at unionized plants is determined by two-sided selection on the part of both the employer and potential workers (Abowd and Farber, 1982; Card, 1996).

A simplified example of this kind of model, adapted from Card (1996), endows a worker i with general productivity g_i, which is also her wage in the nonunion sector:

$$w_i^n = g_i.$$

The return to skill in the union sector, however, is lower, so the worker's wage in a union job would be:

$$w_i^u = \theta_0 + \theta_1 g_i,$$

where $\theta_0 > 0$ and $0 < \theta_1 < 1$. After a successful unionizing drive, workers decide to stay at the union plant (if they were there already) or queue for a union job (if they weren't) if the difference between the union and nonunion wage exceeds ρ_i, the person-specific preference cost of working in

a union job. The worker's decision to stay or join is therefore determined by:

$$g_i < \frac{\theta_0}{1-\theta_1} - \frac{\rho_i}{1-\theta_1}. \tag{2}$$

The newly-unionized employer decides to keep a worker (or hire her from the queue) if the sum of the worker's general productivity and match-specific productivity ω_i exceeds the union wage. The employer's criterion to keep or hire a worker is therefore:

$$g_i > \frac{\theta_0}{1-\theta_1} - \frac{\omega_i}{1-\theta_1}. \tag{3}$$

The worker's and the employer's decision rules have opposing influences on the composition of workers who end up employed at the establishment: given the depressed return to skill, while the employer would prefer to hire and retain higher-skilled workers, they prefer to leave in response to a union victory. The model predicts that when the shift in the union wage scale (captured by θ_0) is small, the worker's decision rule (2) is more likely to bind, shifting the composition of workers who end up employed at the unionized plant toward the lower end of the productivity distribution.

The evidence in this paper supports this prediction. The lack of an overall effect on the earnings of workers who remained at the establishment post-election suggests θ_0 is quite small. Consistent with the model's prediction, both the finding that workers who join following a union victory are younger and lower paid and the finding that workers who leave establishments following a union victory are higher-paid suggest that at the hiring and exit decisions the employee selection rule is more binding. The model also suggests an interpretation of the employment effects reported above. In terms of the model, when θ_0 is small, the fraction of workers who prefer to stay following a union victory will also be small. The modest negative effect on employment may therefore reflect primarily voluntary separations rather than firings in response to the union victory.

6 Summary and Conclusions

This paper presented new evidence on the impacts of unionization on establishments and workers. First, close union representation elections exhibit substantial nonrandom selection. The distribution of election outcomes showed unexpectedly few close union victories. The potential manipulation of

close elections led to significant discontinuities in the underlying characteristics of establishments at the majority-rule threshold of union victory, casting doubt on causal estimates of union impacts based on level comparisons of establishments where the union barely won or barely lost.

Estimates exploiting the panel nature of the data to account for pre-election selection are quite different from previous results. They suggest that union representation led to a decrease in establishment payroll, employment, and average worker earnings. The decrease in payroll and earnings was primarily driven by changes in workforce composition in response to a union victory. Higher-paid workers were more likely to leave after a union victory, and younger, lower-paid workers were relatively more likely to come. Thus the remaining pool of workers was younger and lower paid. There is no evidence of any impact on the earnings of workers who remained at the plant. Thus, the single largest effect of union representation in these results is the impact on the mix of workers remaining employed at the establishment, consistent with models of unionization in which the set of workers at union plants is subject to selection on the part of both the employer and employees.

As is generally true of discontinuity-based research designs, the causal impacts reported here apply to establishments that were near the margin of union victory or defeat. The modal representation election is close, so in this sense the findings here apply to typical unionizing drives, but certainly not to all. Apart from differences in sample and research design, this could partially explain the contrast between the sizeable union premium found in much of the previous literature and the lack of evidence of an individual earnings effect in this study.

These findings also differ from previous election-based RD studies of union effects, which found little impact on a range of establishment-level outcomes. Part of the difference is explained by pre-election selection in close union elections that the evidence here accounts for. The evidence here also includes a broader set of industries, a wider time frame, and better name-and-address match quality, which may also explain the contrasting results.

The findings reported here sound a cautionary note, but perhaps not a dealth knell for RD designs based on close elections, even when there is evidence for selection. First, the findings here underscore the need for pre-treatment "balance" tests and specification checks. But the very panel data used to detect potentially confounding selection can also be used to construct estimators that account for pre-existing differences in a hybrid of regression discontinuity and difference-in-

difference research designs.

Author affiliation

Brigham R. Frandsen

Brigham Young University, Department of Economics

Appendix

Construction of the dataset

As described in the text, the dataset used in this paper consists of NLRB certification election results matched to confidential establishment-level datasets maintained by the Census Bureau.

The union certification election records were collected by the NLRB, and in large part maintained by the AFL-CIO. John-Paul Ferguson, Thomas Holmes, and Hank Farber obtained the election records from the NLRB, and made them available for this research. The complete data set covers the period 1963-2009, although the main sample used in the analysis covers the years 1980-2009. The raw data contains results from elections stemming from several different type of petitions, including cases where a union seeks to be certified (RC), an employer seeks an election to remove an existing union (RM), or employees seek to remove a union (RD). I restrict to RC-cases, where a union seeks certification. The dataset contains many duplicate records. In some cases they are true duplicates: one election generated multiple records in the database. In these cases I simply delete the redundant entries. In other cases, multiple entries arise from more than one union being on the ballot. In these cases the relevant union vote share is the largest one; I therefore retain the entry with the largest vote share, and delete the others. Finally, in some cases multiple elections were held at the same establishment because, for example, different groups of workers constituted different bargaining units. Since I can't distinguish between workers in different bargaining units, the relevant vote share is the largest, so again I keep only the entry corresponding to the election where the union received the highest vote share.

The second data component consists of the Census Bureau's Longitudinal Business Database (LBD). As described in the text, the LBD contains longitudinally-linked establishment-level panel

data on payroll and employment for virtually the universe of U.S. private-sector employers. The names and addresses for these establishments are available from the Standard Statistical Establishment List (SSEL), also known as the Business Register (BR), which provides the sampling frame for various economic censuses and surveys.

The matching process to combine these two data sources is as follows. First, employer name and address information from both the NLRB dataset and the Census Bureau's Business Register (BR) were cleaned and standardized using the SAS Data Quality Server standardization functions. NLRB election records were then matched to BR records by several combinations of state, county, city, employer name, street address, and industry code. The match was performed iteratively in descending order of strictness. The cutoff level of strictness was determined by hand checking matches from each iteration, and stopping once match quality dipped below 95 percent. This procedure successfully matched 82 percent of the NLRB election records, much higher than the 26 percent match rate in DiNardo and Lee (2004).

Finally, some of the analysis required individual-level data on employees at election establishments. Individual-level earnings data came from the Employment History Files (EHF) within the LEHD database. The EHF contains employee, employer, and earnings data for each employment relationship that generated at least one dollar of wages. The EHF includes a state employer identification number (SEIN) with each record, and in some cases an identifer for the establishment within the employer, which is important for multi-unit employers. For the cases where there is no establishment identifier, the LEHD provides a Unit-to-Worker (U2W) imputation to assign workers to establishments. The Census Bureau maintains a crosswalk of firm- and establishment-level identifiers (the Federal Employer Identification Number, or EIN, county, and industry) which was used to merge the election establishments identified in the LBD with the LEHD. The overall match rate between the NLRB records that correspond to states and years where the LEHD is available was only slightly lower (77 percent).

References

Abowd, John M., and Henry S. Farber (1982): "Job Queues and the Union Status of Workers," *Industrial and Labor Relations Review*, 35, pp. 354–367.

Abowd, John M., John Haltiwanger, and Julia Lane (2004): "Integrated Longitudinal Employee-Employer Data for the United States," *American Economic Review Papers and Proceedings*, 94, 224–229.

Abowd, John M., Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock (2009): "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in *Producer Dynamics: New Evidence from Micro Data*, ed. by T. Dunne, J.B. Jensen, and M.J. Roberts, pp. 149–230. University of Chicago Press for the National Bureau of Economic Research, Chicago.

Acemoglu, Daron, Phillipe Aghion, and Gianluca Violante (2002): "Technical Change, Deunionization, and Inequality," in *Carnegie-Rochester Conference Series on Public Policy*.

Blanchflower, David, and Alex Bryson (2003): "What effect do unions have on wages now and would 'What do unions do?' be surprised?," NBER Working Paper No. 9973.

Brudney, James J. (2005): "Contractual Approaches to Labor Organizing: Supplanting the Election Paradigm?," in *Labor and Employment Relations Association Series, Proceeding of the 57th Annual Meeting*, pp. 106–113.

Card, David (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–979.

DiNardo, John, and David S. Lee (2004): "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001," *Quarterly Journal of Economics*, 119, 1383–1441.

Ferguson, John-Paul (2008): "The Eyes of the Needles: A Sequential Model of Union Organizing Drives, 1999-2004," *Industrial & Labor Relations Review*, 62, 3–21.

Frandsen, Brigham R. (2012): "Why Unions Still Matter: The Effects of Unionization on the Distribution of Employee Earnings," unpublished working paper.

Freeman, Richard B., and Morris M. Kleiner (1990): "The Impact of New Unionization on Wages and Working Conditions," *Journal of Labor Economics*, 8, pp. S8–S25.

Freeman, Richard B., and James L. Medoff (1984): *What Do Unions Do?* Basic Books, New York.

Holmes, Thomas J. (2006): "Geographic Spillover of Unionism," Federal Reserve Bank of Minneapolis Research Department Staff Report 368.

Imbens, Guido, and Karthik Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 79, 933–959.

Imbens, Guido W., and Thomas Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635.

Jarmin, Ron S., and Javier Miranda (2002): "The Longitudinal Business Database," Center for Economic Studies Discussion Paper CES-WP-02-17.

Lalonde, Robert J., Grard Marschke, and Kenneth Troske (1996): "Using Longitudinal Data on Establishments to Analyze the Effects of Union Organizing Campaigns in the United States," *Annals of Economics and Statistics / Annales d'conomie et de Statistique*, pp. pp. 155–185.

Lane, Julia (2008): "Linking Administrative and Survey Data," Society of Labor Economists 2008 Annual Meeting Presentation.

Lee, David S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142, 675 – 697, The regression discontinuity design: Theory and applications.

Lee, David S., and David Card (2008): "Regression discontinuity inference with specification error," *Journal of Econometrics*, 142, 655 – 674.

Lee, David S., and Thomas Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Lee, David S., and Alexandre Mas (2012): "Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961-1999," *The Quarterly Journal of Economics*, 127, 333–378.

McCrary, Justin (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714.

McKinney, Kevin L., and Lars Vilhuber (2008): "LEHD Infrastructure files in the Census RDC - Overview Revision 219," U.S. Census Bureau Technical Paper.

Moberly, Michael D. (2002): "Corrections Before Representation Elections: Restoring "Laboratory Conditions" By Repudiating Unfair Labor Practices," *University of Pennsylvania Journal of Labor and Employment Law*, 4, 375–416.

National Labor Relations Board (2010): "Procedures Guide," Available on the world wide web at http://www.nlrb.gov/publications/Procedures_Guide.htm.

Porter, Jack (2003): "Estimation in the Regression Discontinuity Model," unpublished paper, Department of Economics, Harvard University.

Sojourner, Aaron, Robert J. Town, David C. Grabowski, Michelle M. Chen, and Brigham R. Frandsen (2013): "Impacts of Unionization on Quality and Producitivity: Regression Discontinuity Evidence from Nursing Homes," Unpublished manuscript.

Table I: NLRB Certification Elections Summary Statistics

	All	Union loss	Union victory
Pr(Union victory)	45%	0%	100%
Union vote share, average	49%	33%	70%
Number of voters, average	93	100	84
Number of voters, total	4,191,075	2,494,768	1,696,307
Number of elections	45,176	24,974	20,202

Data source: NLRB certification election records.

Table II: Average Establishment Characteristics from the Longitudinal Business Database

	All	Union Loss	Union Victory
Number of employees			
pre-election	254	236	277
post-election	259	242	281
Payroll (millions)			
pre-election	$8.87	$8.27	$9.59
post-election	$9.13	$8.62	$9.75
Average worker earnings			
pre-election	$36,100	$35,511	$36,803
post-election	$35,219	$36,319	$33,903

Notes: Payroll and worker earnings in year 2000 dollars.

Table III: Average Employee Characteristics from the LEHD

	All	Union Loss	Union Victory
Age at time of election	40	39	41
female	44%	42%	45%
Nonwhite	38%	35%	40%
Quarterly earnings	$8,097	$8,204	$7,975
Stay at establishment following election	42%	43%	41%
N	1,774,975	946,607	828,368

Notes: Pre-election means. Quarterly earnings are in year 2000 dollars.

Table IV: Discontinuities in pre-election establishment and worker characteristics

Dependent Variable:	(1)	(2)	(3)
log employment	0.134 ***	0.129 ***	0.228 ***
	(0.032)	(0.037)	(0.062)
log payroll	0.060 *	0.059	0.133 ***
	(0.032)	(0.043)	(0.045)
log worker quarterly earnings percentile:			
10th	0.196 ***	0.240 ***	0.350 ***
	(0.024)	(0.034)	(0.111)
20th	0.156 ***	0.190 ***	0.282 ***
	(0.023)	(0.030)	(0.091)
50th	0.047	0.102 **	0.133
	(0.031)	(0.040)	(0.085)
80th	-0.081 **	-0.019	-0.019
	(0.037)	(0.039)	(0.074)
90th	-0.345 ***	-0.308 ***	-0.368 ***
	(0.054)	(0.032)	(0.057)
90-10 difference	-0.541 ***	-0.548 ***	-0.718 ***
	(0.034)	(0.052)	(0.136)
local polynomial type	linear	linear	quadratic
bandwidth	0.15	0.2	0.25

Notes: Estimates and robust standard errors for the coefficient on an indicator for union victory in local polynomial regressions of the dependent variable in the left-hand column on an indicator for union victory interacted with a polynomial of the indicated degree in the union vote share. Local regressions restrict to observations within the indicated bandwidth on either side of the 50 percent threshold. The data were collapsed by the union vote share variable.

Table V: Discontinuities in n establishment survival

	(1)	(2)	(3)	(4)
Survival through:				
1 year	-0.009	-0.013	-0.009	-0.009
	(0.006)	(0.009)	(0.009)	(0.008)
3 years	-0.040 ***	-0.049 **	-0.033 ***	-0.034 ***
	(0.010)	(0.021)	(0.007)	(0.009)
5 years	-0.076 ***	-0.095 ***	-0.078 ***	-0.088 ***
	(0.013)	(0.023)	(0.006)	(0.010)
7 years	-0.088 ***	-0.111 ***	-0.099 ***	-0.104 ***
	(0.010)	(0.020)	(0.008)	(0.015)
running variable	vote share	vote share	vote count	vote count
local polynomial type	linear	quadratic	linear	quadratic
bandwidth	0.15	0.2	5	10

Notes: Estimates and robust standard errors for the coefficient on an indicator for union victory in local polynomial regressions of an indicator for establishment presence in the LBD as of the indicated year following a union representation election on an indicator for union victory interacted with a polynomial of the indicated degree in the normalized union vote tally or share. Local regressions restrict to observations within the indicated bandwidth on either side of the threshold. The data are means by the union vote variable.

Table VI: Effects of union representation on establishment outcomes

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log payroll	-0.077 ***	-0.083 ***	-0.120 ***	-0.124 ***	-0.075 ***	-0.093 ***	-0.098 ***	-0.079 ***
	(0.010)	(0.010)	(0.032)	(0.029)	(0.021)	(0.023)	(0.013)	(0.022)
log employment	-0.053 ***	-0.050 ***	-0.084 ***	-0.093 ***	-0.049 ***	-0.051 ***	-0.077 ***	-0.067 ***
	(0.006)	(0.013)	(0.027)	(0.027)	(0.014)	(0.020)	(0.018)	(0.025)
log average worker earnings	-0.025 ***	-0.033 **	-0.036 ***	-0.031 ***	-0.026 ***	-0.042 ***	-0.021	-0.012
	(0.006)	(0.013)	(0.011)	(0.012)	(0.007)	(0.007)	(0.016)	(0.017)
differenced?	Y	Y	Y	Y	N	N	N	N
control for lagged outcome?	N	N	N	N	Y	Y	Y	Y
running variable	vote share	vote share	vote count	vote count	vote share	vote share	vote count	vote count
local polynomial type	linear	quadratic	linear	quadratic	linear	quadratic	linear	quadratic
bandwidth	0.15	0.2	5	10	0.15	0.2	5	10

Notes: Estimates and robust standard errors for the coefficient on an indicator for union victory in local polynomial regressions of the dependent variable in the left-hand column on an indicator for union victory interacted with a polynomial of the indicated degree in the normalized union vote tally or share. In columns (1)-(4) the dependent variable is the observation from the year following the election minus the observation from the year prior to the election. The specifications in columns (5)-(8) control nonparametrically for the pre-election value of the dependent variable. Local regressions restrict to observations within the indicated bandwidth on either side of the threshold. The data are means by the union vote variable.

Table VII: Effects on Worker Composition and Stayer's earnings, LEHD

Dependent Variable:	(1)	(2)	(3)
age (all)	-1.842	-1.507	-1.687
	(0.214)	(0.187)	(0.300)
age (comers)	-1.335	-1.559	-1.651
	(0.104)	(0.140)	(0.280)
log quarterly earnings (comers)	-0.211	-0.193	-0.235
	(0.052)	(0.029)	(0.058)
log quarterly earnings (leavers)	0.585	0.478	0.764
	(0.088)	(0.107)	(0.061)
log quarterly earnings (stayers, differenced)	0.003	0.002	0.004
	(0.016)	(0.019)	(0.019)
bandwidth	0.15	0.2	0.25
local polynomial type	linear	linear	quadratic

Notes: Estimates and robust standard errors for the coefficient on an indicator for union victory in local polynomial regressions of the dependent variable in the left-hand column on an indicator for union victory interacted with a polynomial of the indicated degree in the union vote share. The differenced log quarterly earnings of stayers consists of the individual's average quarterly earnings in the year following the election minus the individual's average quarterly earnings in the year prior to the election. Local regressions restrict to observations within the indicated bandwidth on either side of the 50 percent threshold. The data were collapsed by the union vote share variable.

Table VIII: Match success rates between NLRB and Census datasets

	Number of elections	Percent
NLRB dataset, all	88,356	100%
Matched to LBD	72,424	82%
NLRB dataset, LEHD states and years	26,307	100%
Matched to LEHD	20,160	77%

Notes: Match frequencies and rates from name-and-address matching procedure between the NLRB elections dataset and the Census datasets (LBD and LEHD). The match rate denominator for matching to the LEHD includes only elections in 30 states covered by the LEHD during the years in which data are available.

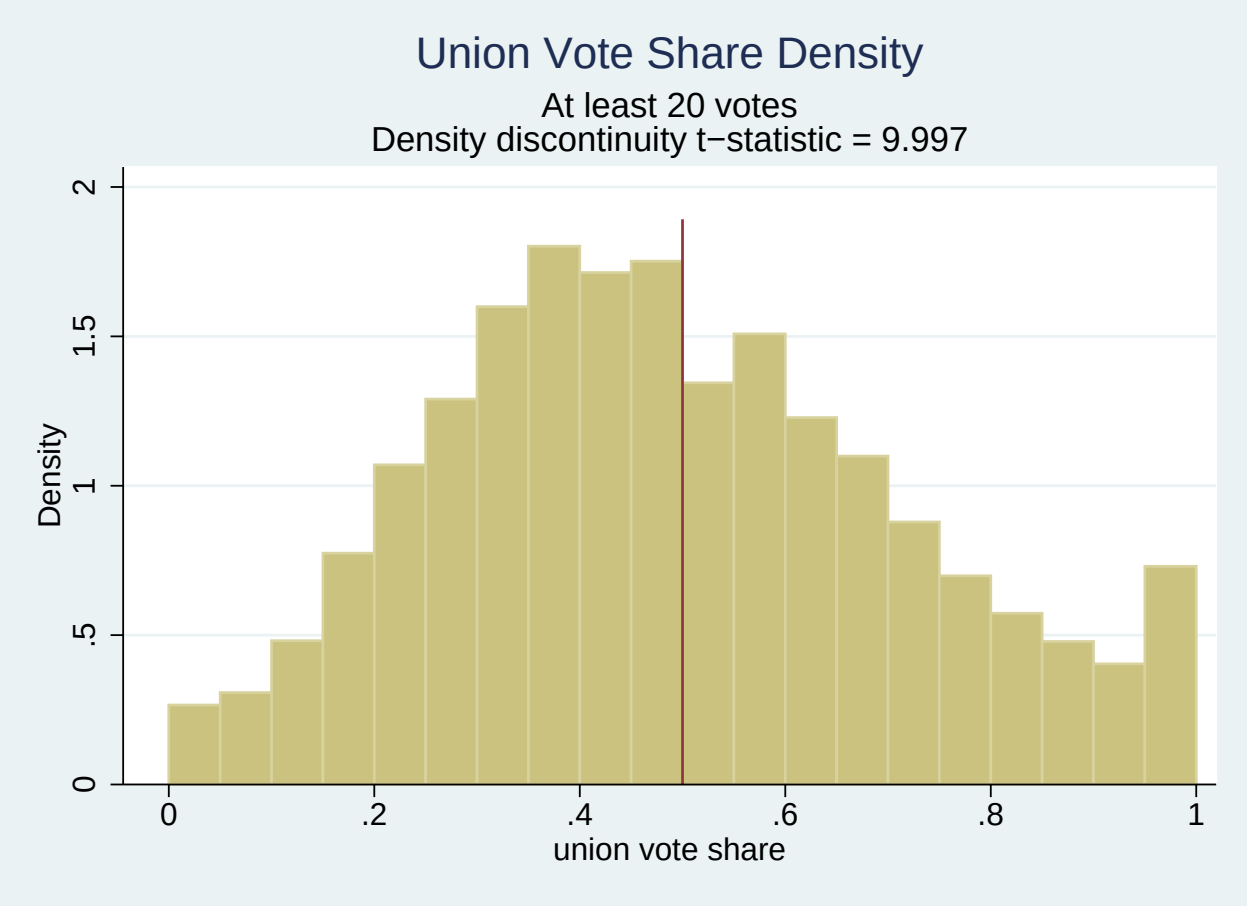

Figure I: Density of binned union vote share for elections with at least 20 voters. The density discontinuity t-statistic is from a McCrary (2008) test using a triangular kernel with a bandwidth of .3. Data are from NLRB election records.

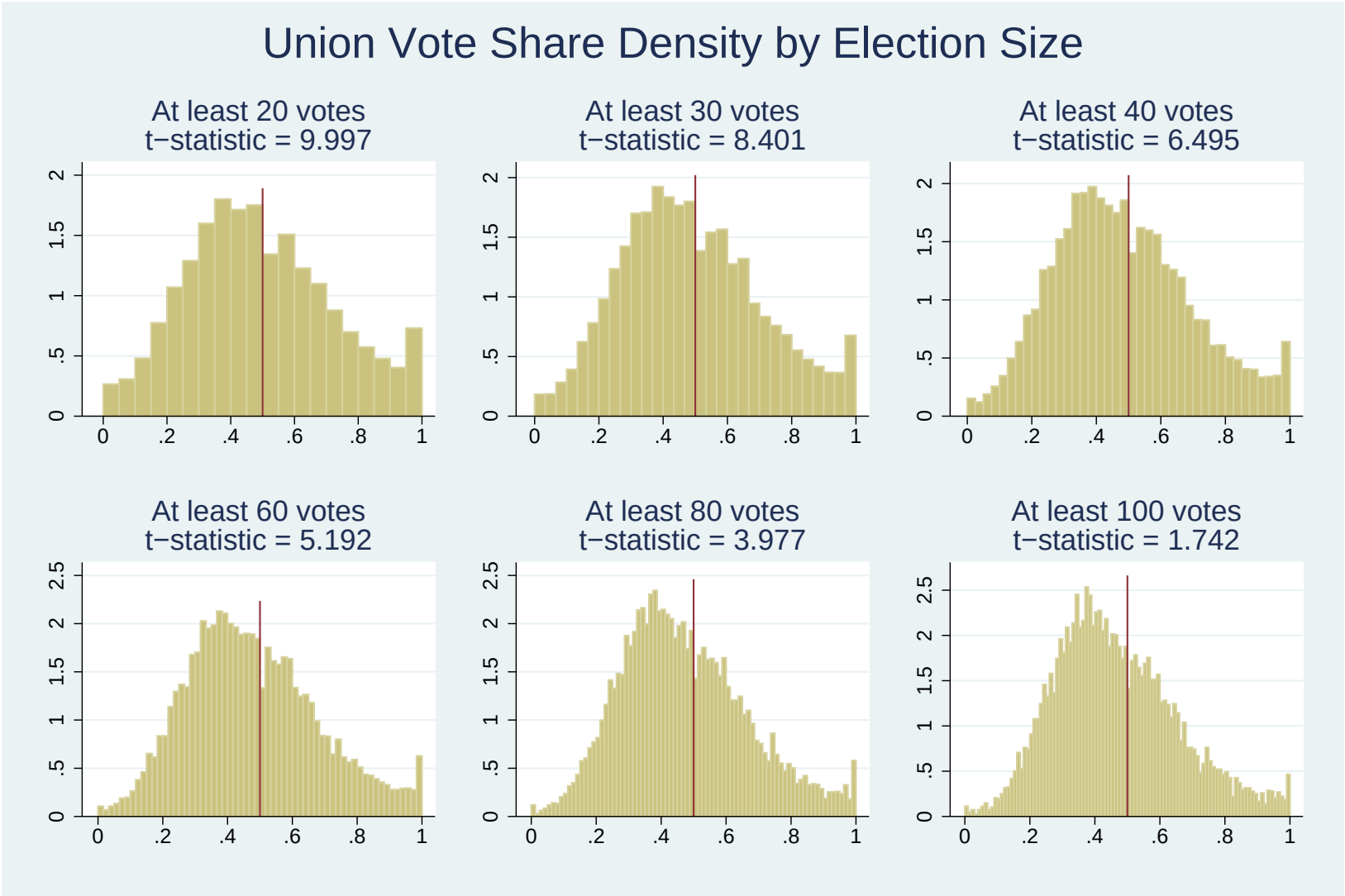

Figure II: Density of binned union vote share for elections with at least the indicated number of voters. The reported t-statistics are from McCrary (2008) density discontinuity tests using a triangular kernel with a bandwidths of .1 to .3. Data are from NLRB election records.

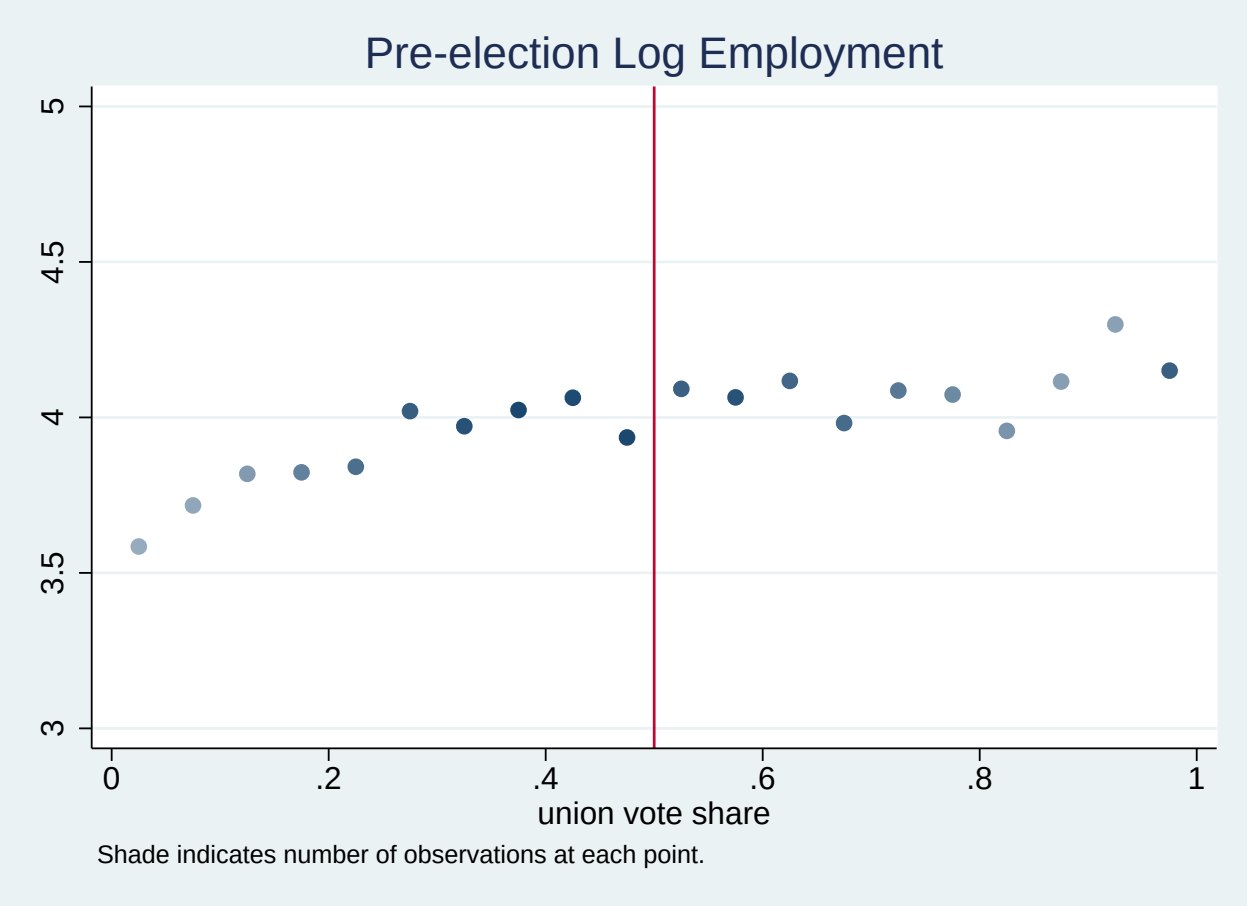

Figure III: Pre-election log employment by union vote share. The figure plots raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

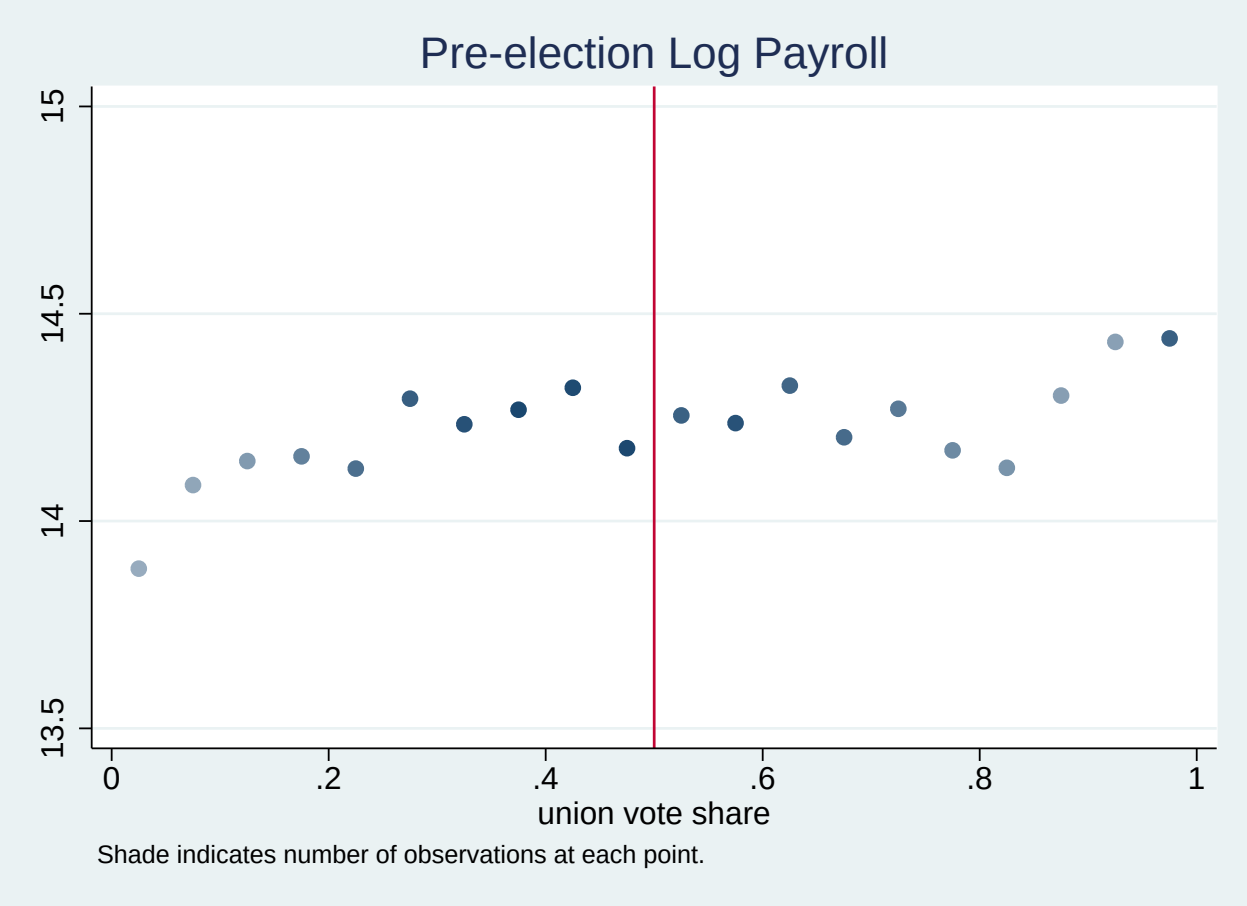

Figure IV: Pre-election log payroll by union vote share. The figure plots raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

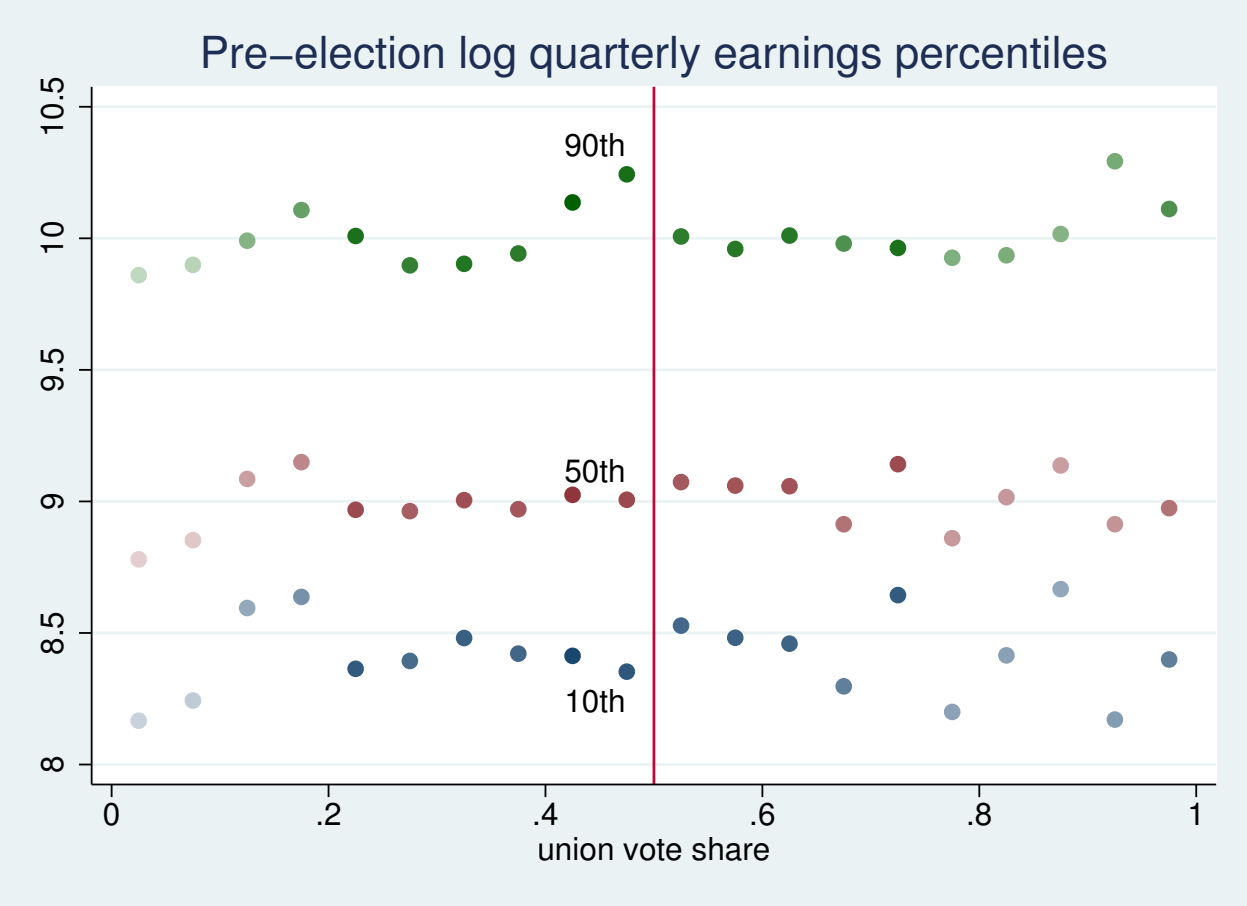

Figure V: 90th, 50th, and 10th percentiles of pre-election log quarterly earnings by union vote share RD graph. The points are raw percentiles by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

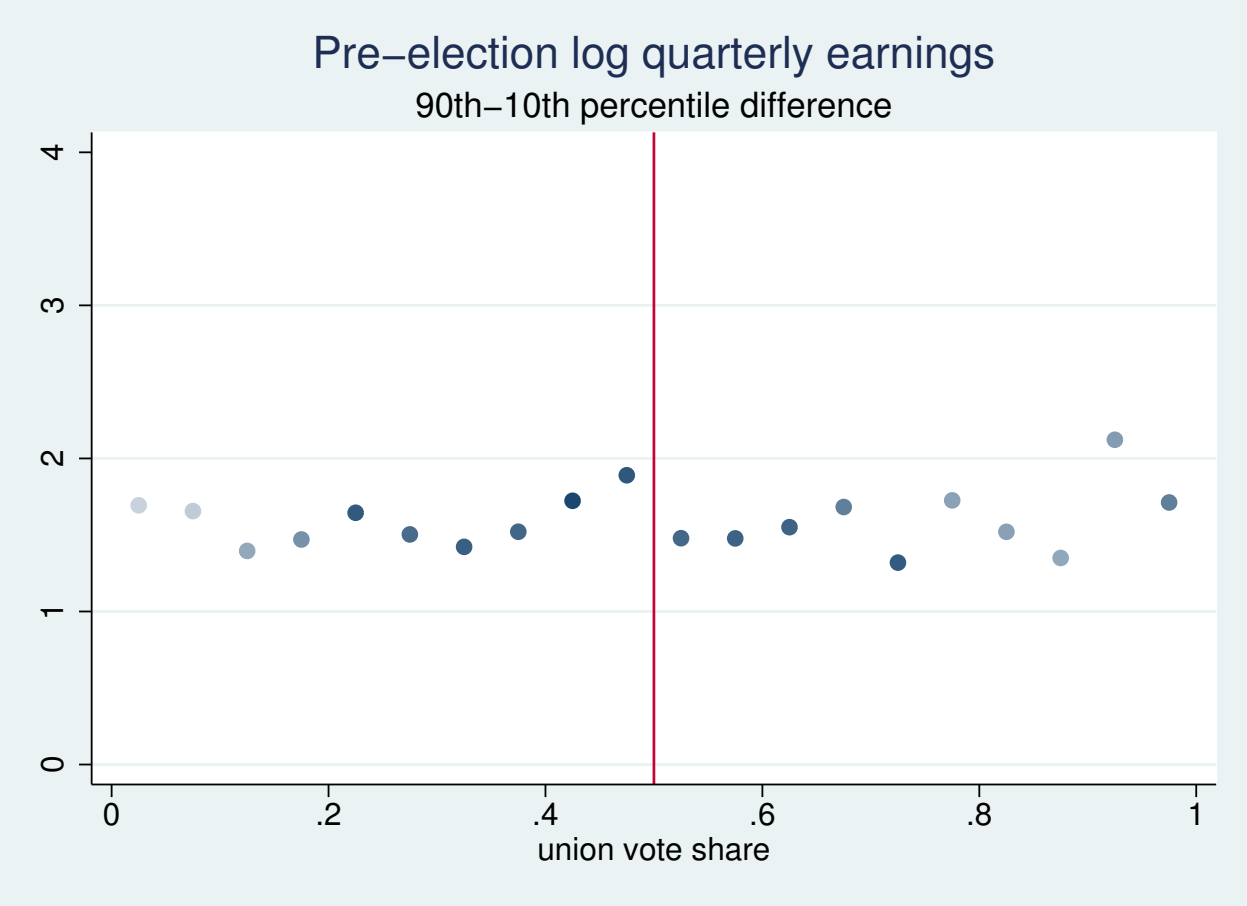

Figure VI: Difference between the 90th and 10th percentile of pre-election log quarterly earnings by union vote share. The points are differences in raw percentiles by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

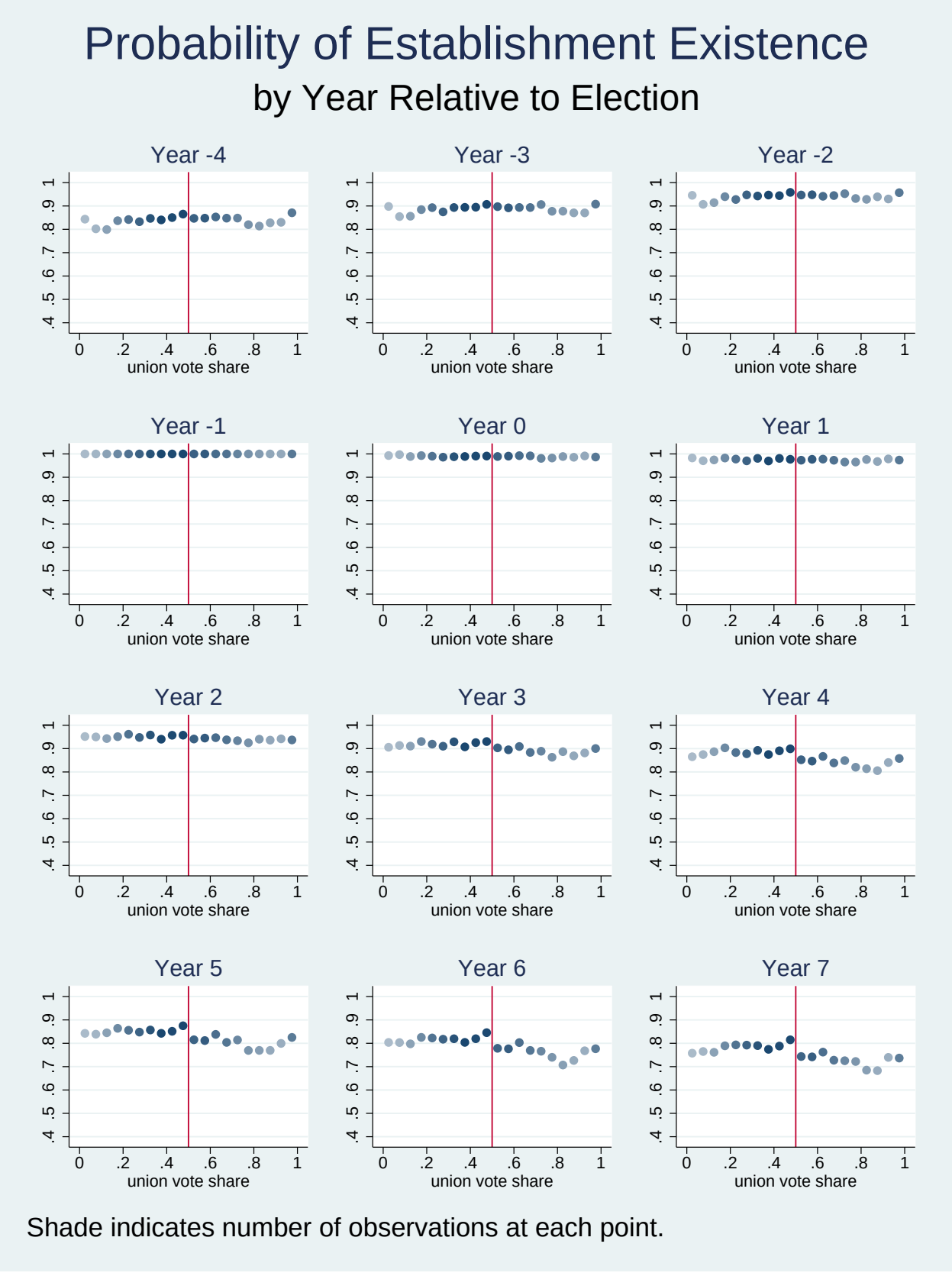

Figure VII: Probability of establishment existing in the Longitudinal Business Database (LBD) for the given year relative to the union election. Data are from the LBD NLRB election records.

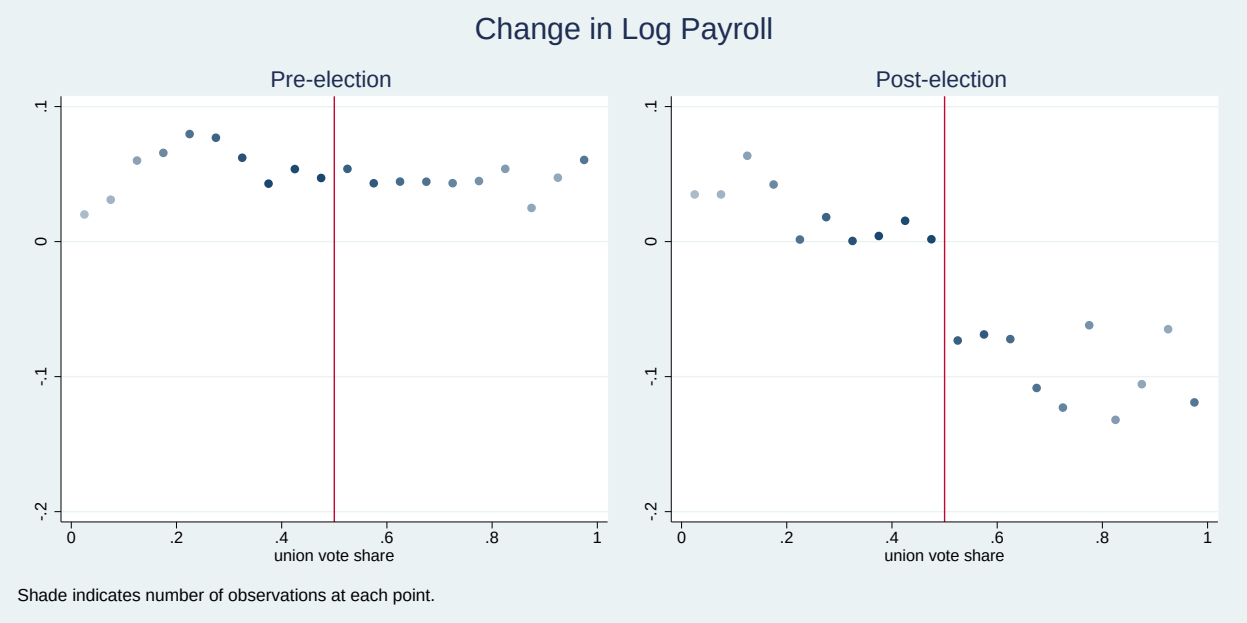

Figure VIII: Change in log payroll by union vote share. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

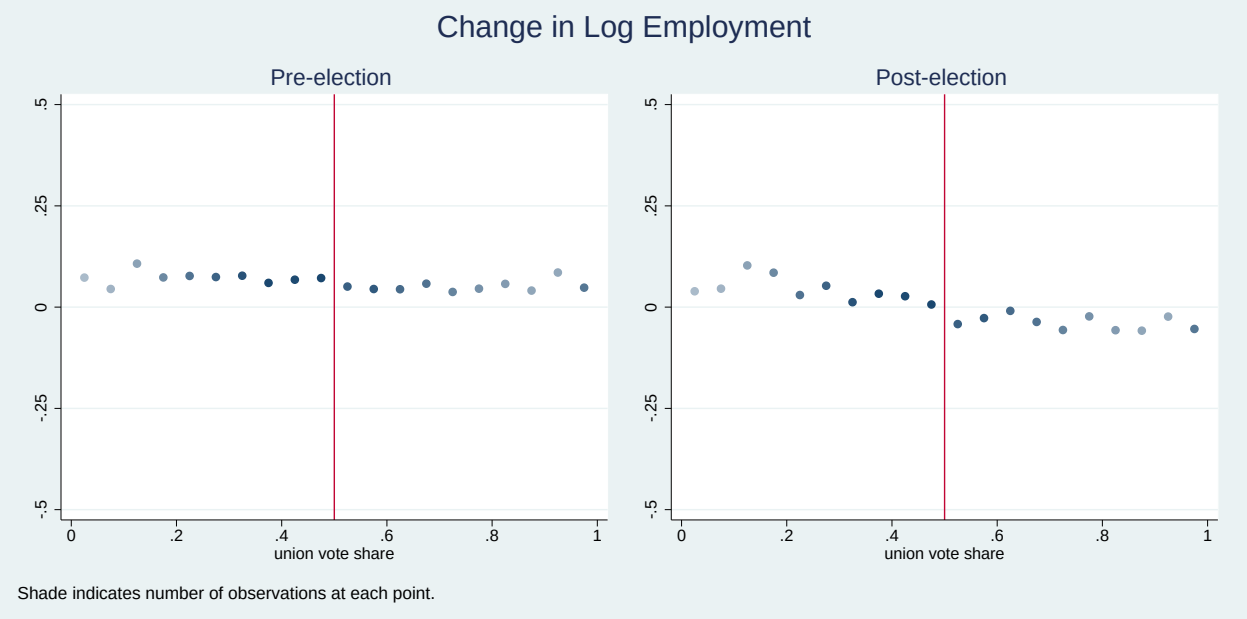

Figure IX: Change in log employment by union vote share. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

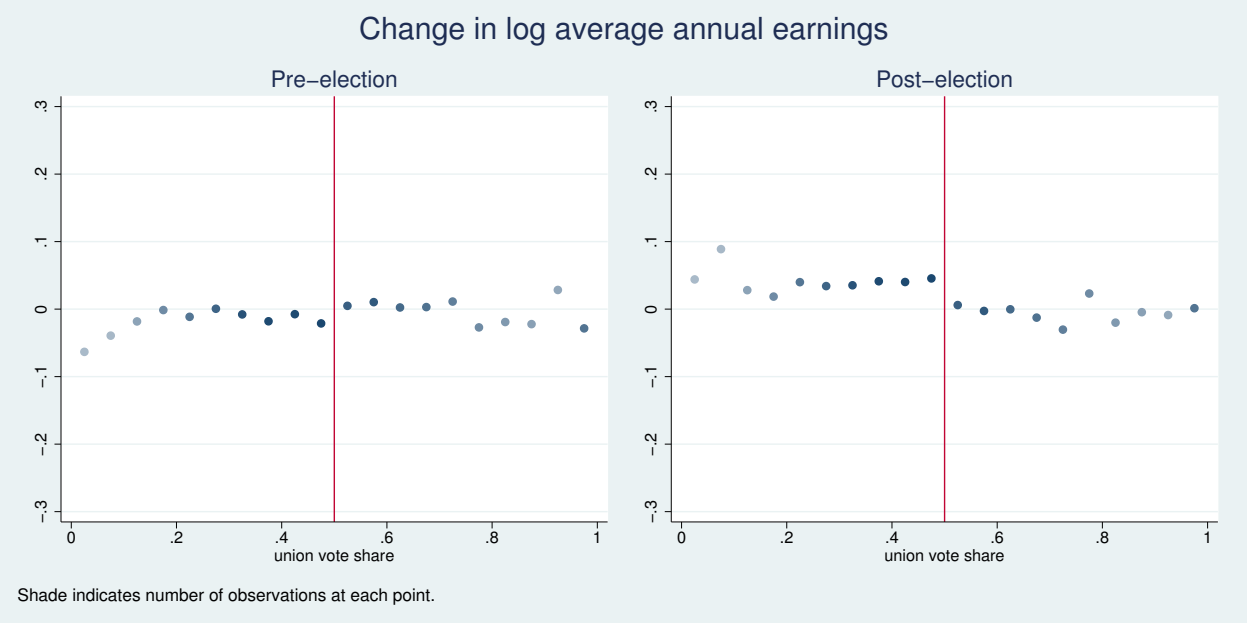

Figure X: Change in log average annual earnings by union vote share. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

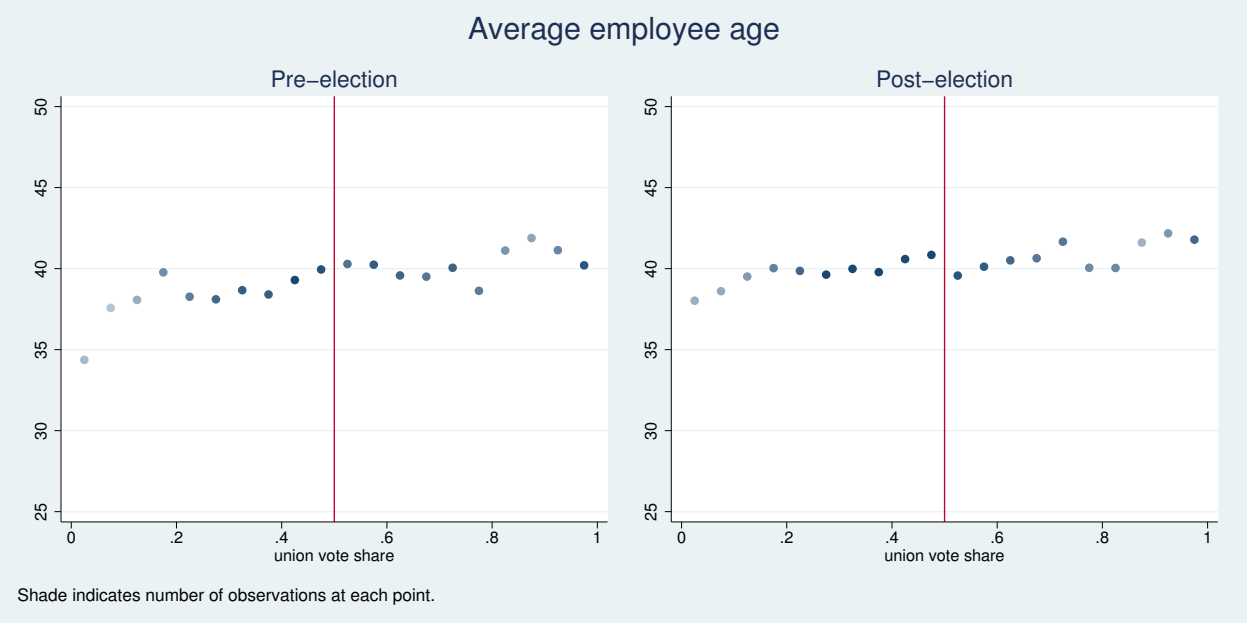

Figure XI: Average employee age by union vote share. The left panel includes employees during the year prior to the election. The right panel includes employees during the year following the election. The points are raw averages by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

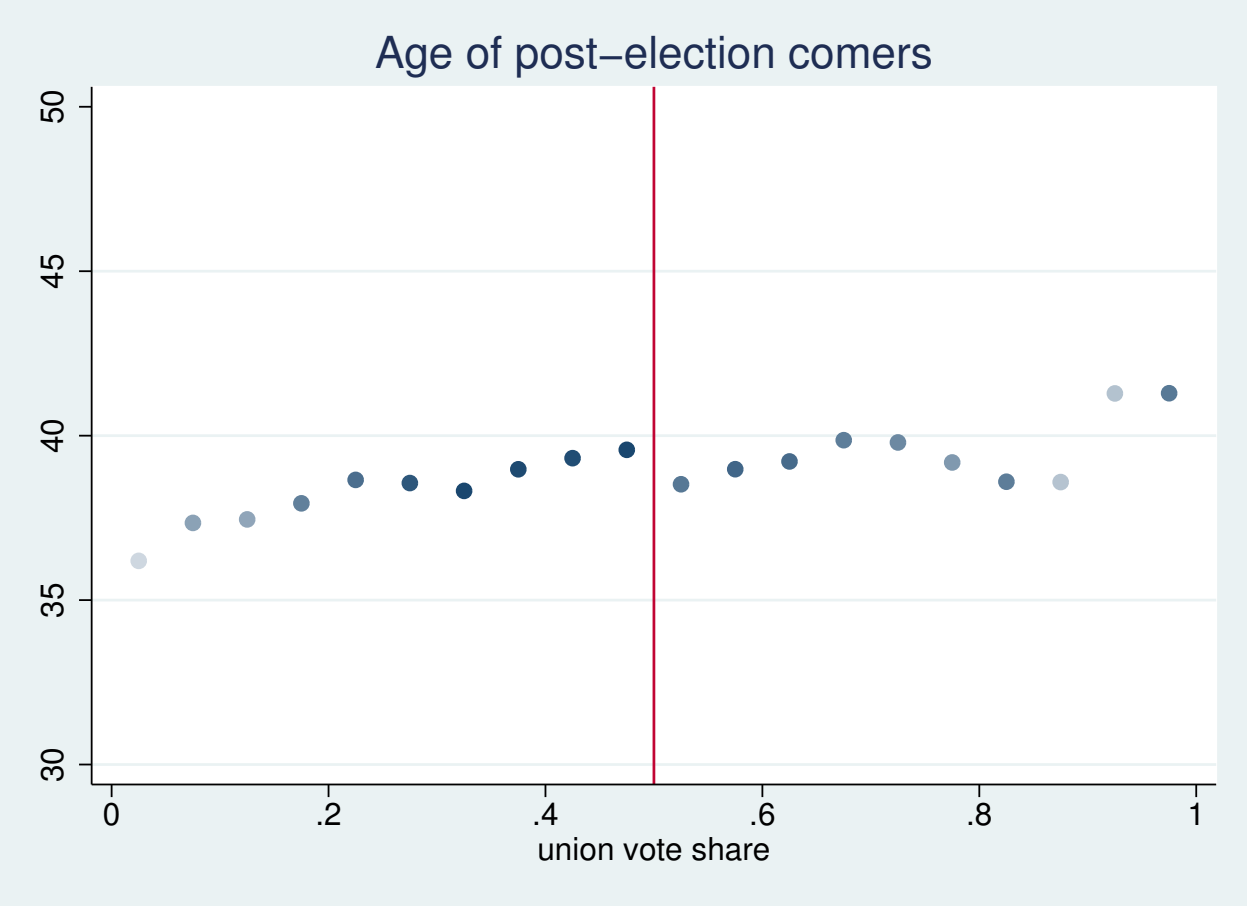

Figure XII: Average age of employees who came to an establishment post-election by union vote share. The points are raw averages by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

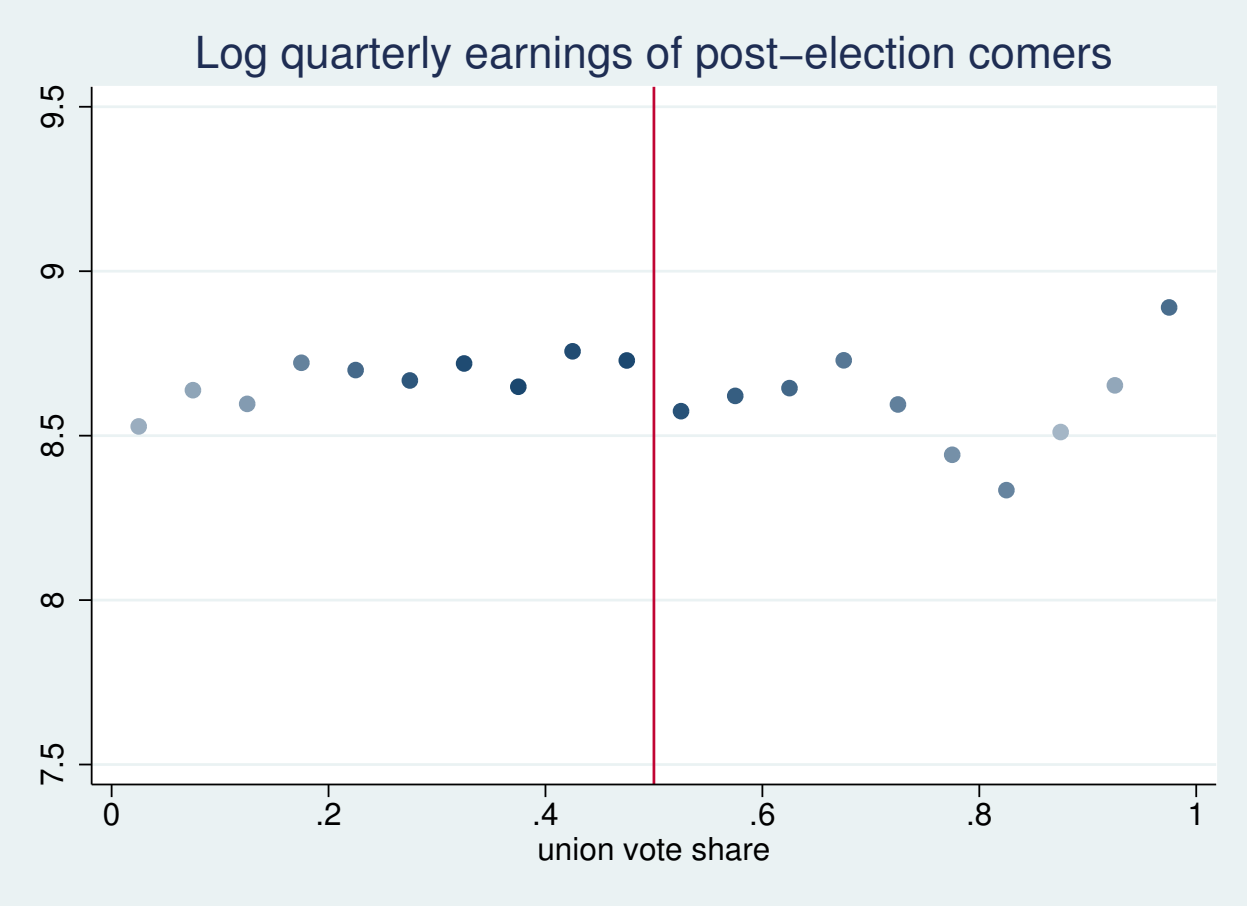

Figure XIII: Log quarterly earnings of workers who came to an establishment post-election. The points are raw medians by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

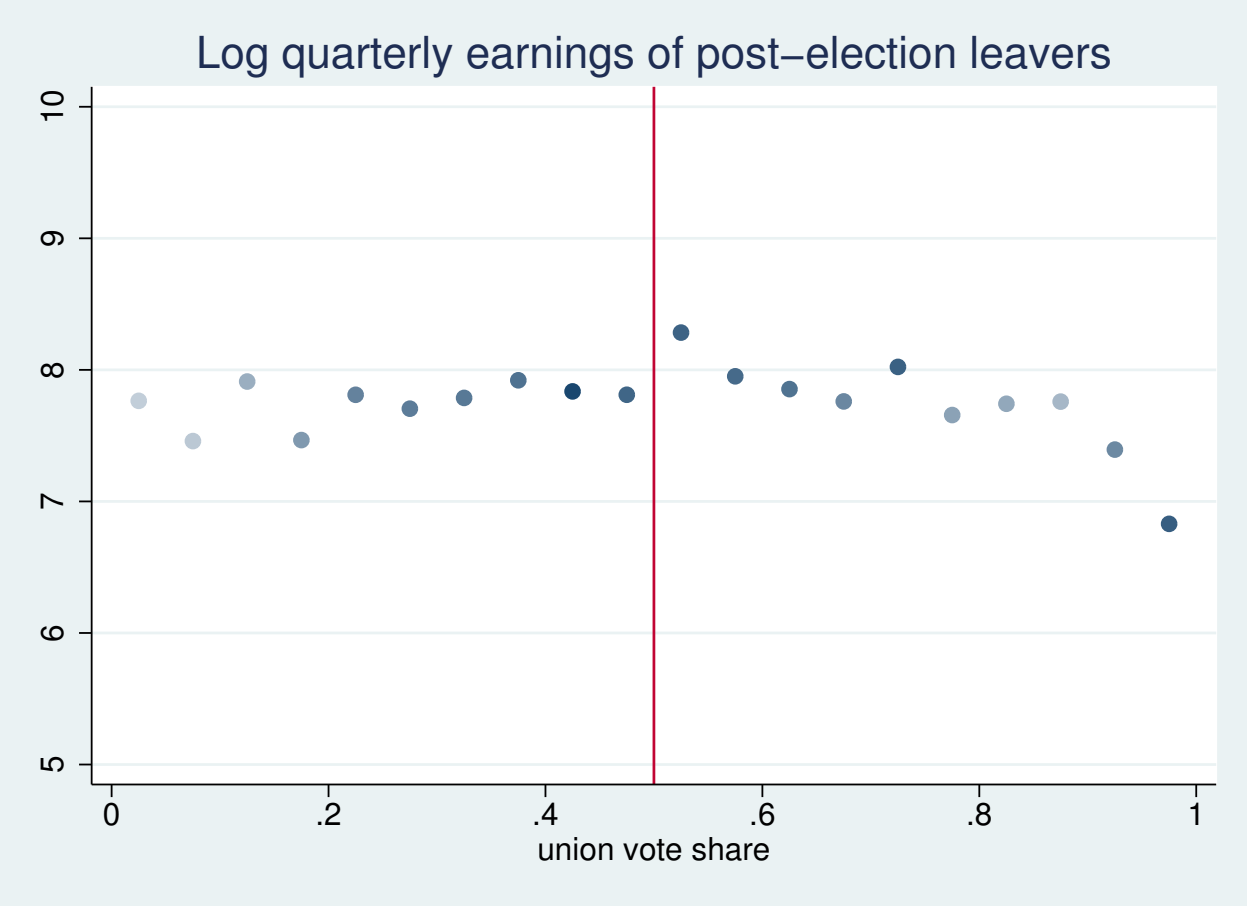

Figure XIV: Average log pre-election quarterly earnings of leaving employees by union vote share. The points are raw averages by non-overlapping vote share bins of width .05. Data are from the LEHD and NLRB election records.

Figure XV: Change in log quarterly earnings of employees who stay by union vote share. The left panel is the change from two years prior to the election to one year prior to the election. The right panel is the change from the year prior to the election to the year after the election. The points are raw averages by non-overlapping vote share bins of width .05. Data are from the LBD and NLRB election records.

APPENDIX

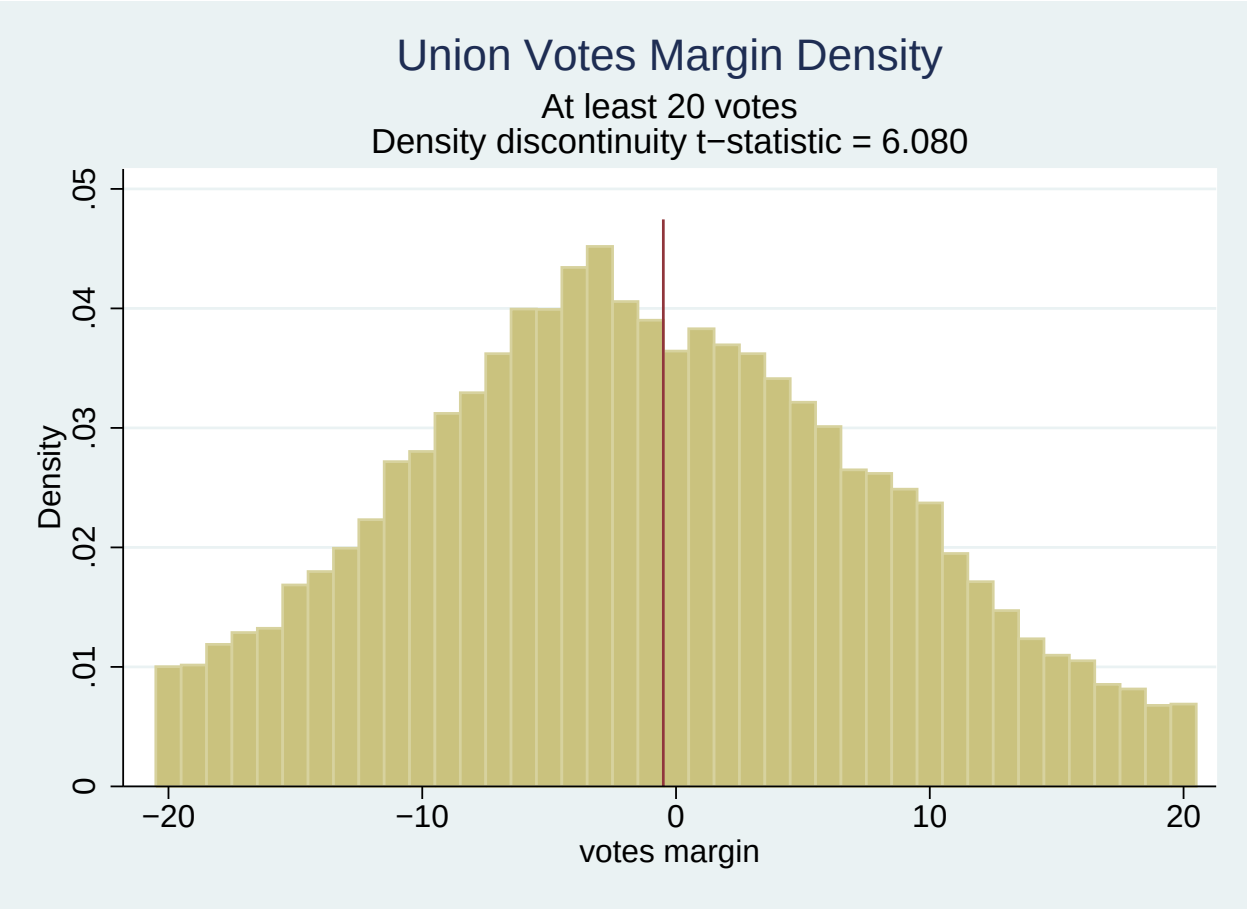

Figure A1: Density of the union votes margin for elections with at least 20 voters. The density discontinuity t-statistic is from a McCrary (2008) test using a triangular kernel with a bandwidth of 15 votes. Data are from NLRB election records.

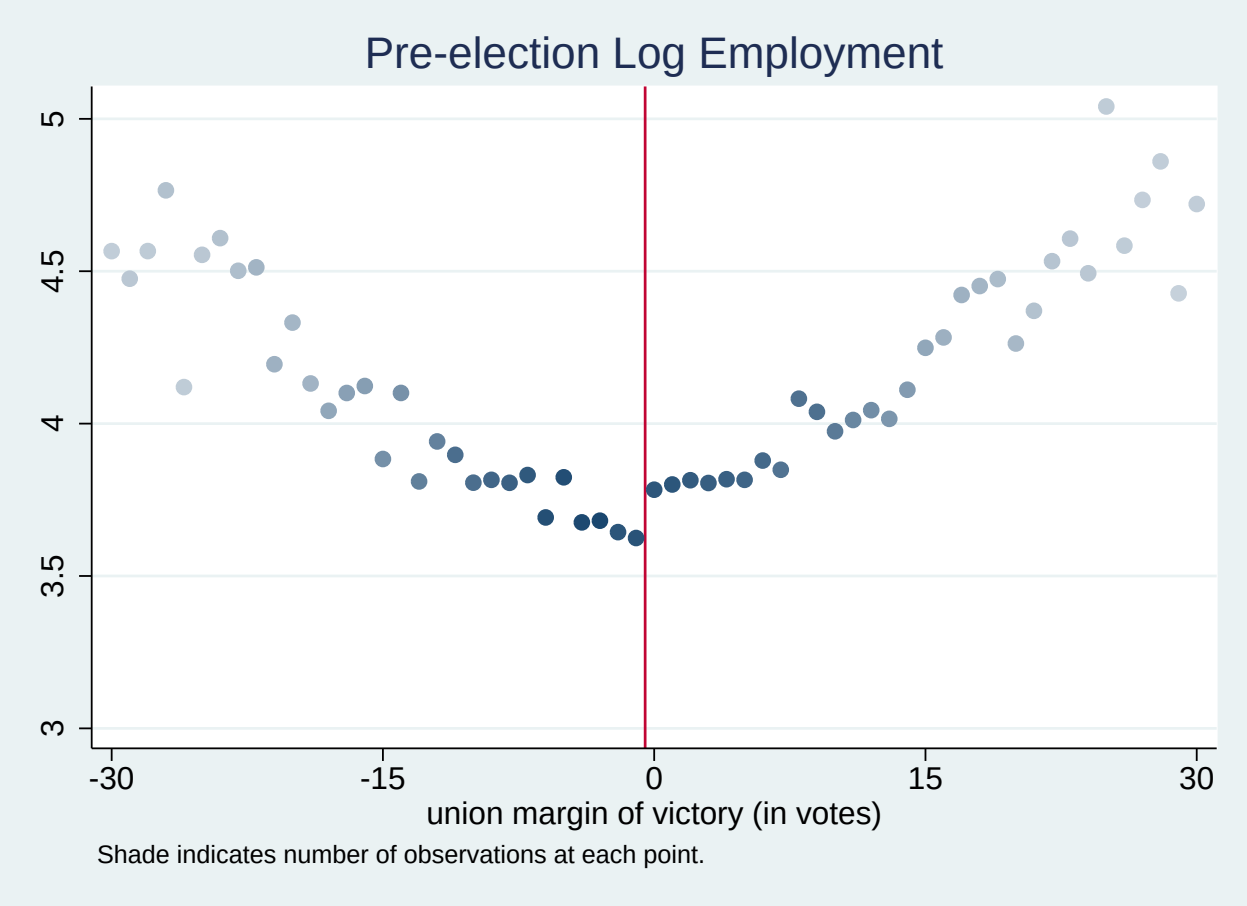

Figure A2: Pre-election log employment by union vote tally margin. The figure plots raw averages by individual tally count. Data are from the LBD and NLRB election records.

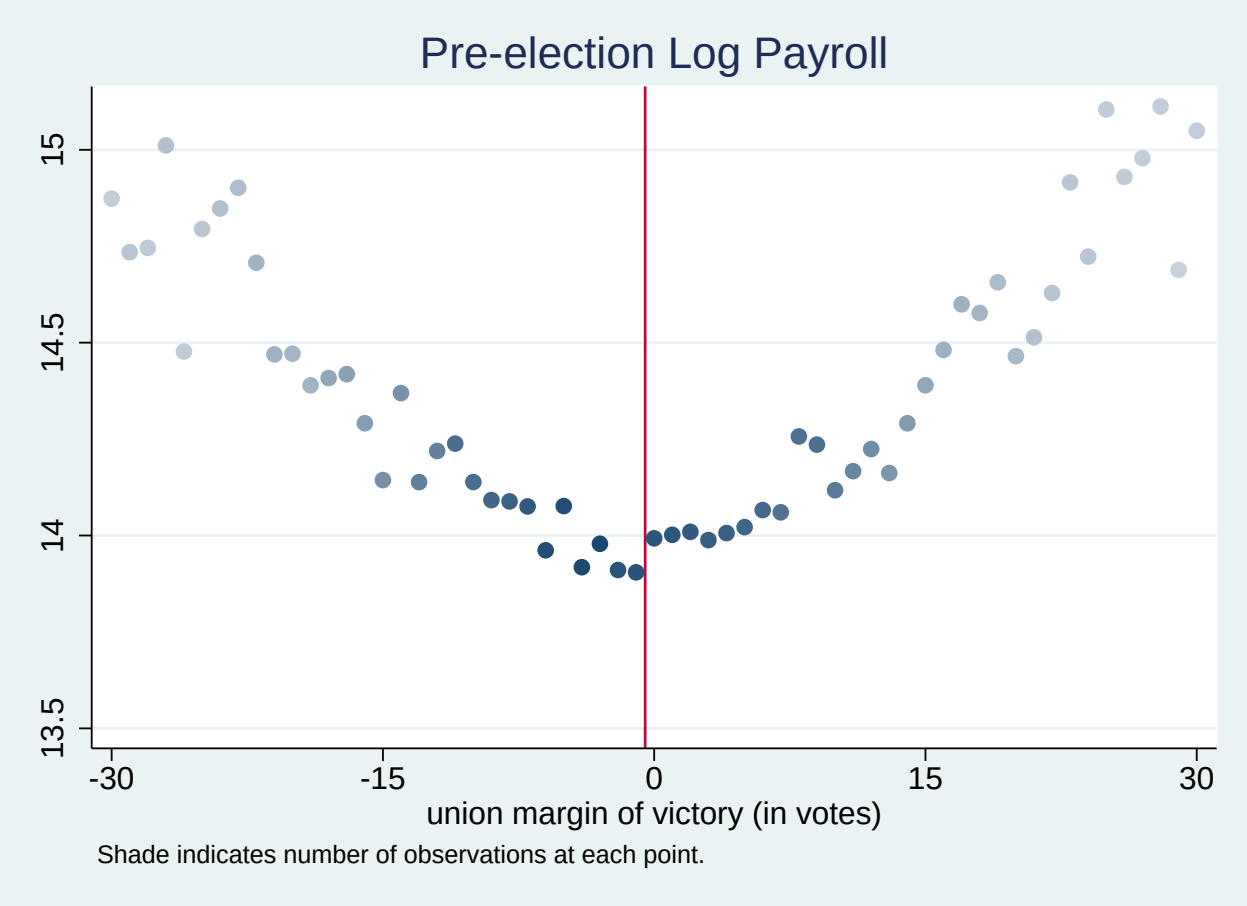

Figure A3: Pre-election log payroll by union vote tally margin.The figure plots raw averages by individual tally count. Data are from the LBD and NLRB election records.

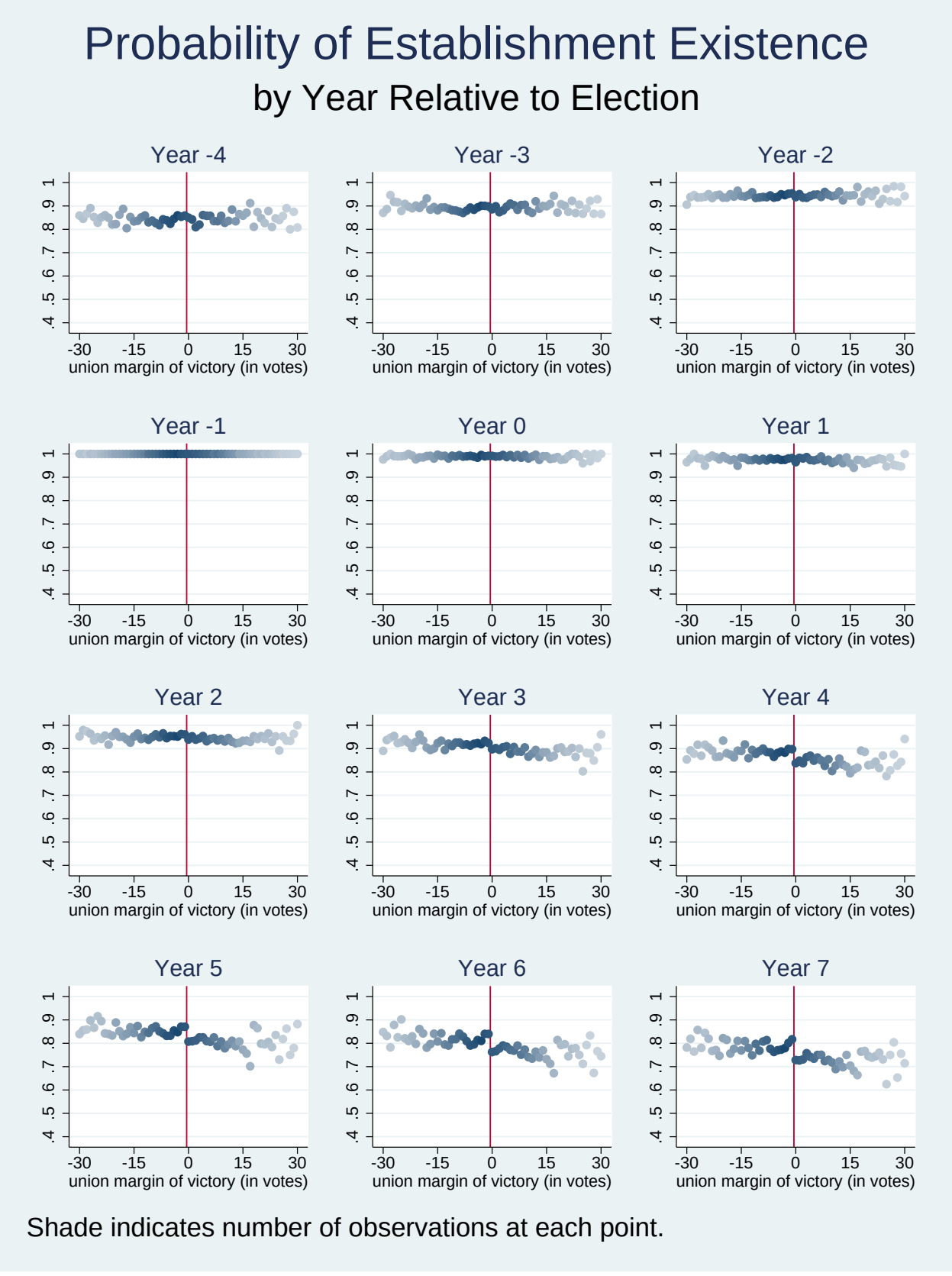

Figure A4: Probability of establishment existing in the Longitudinal Business Database (LBD) for the given year relative to the union election. Data are from the LBD NLRB election records.

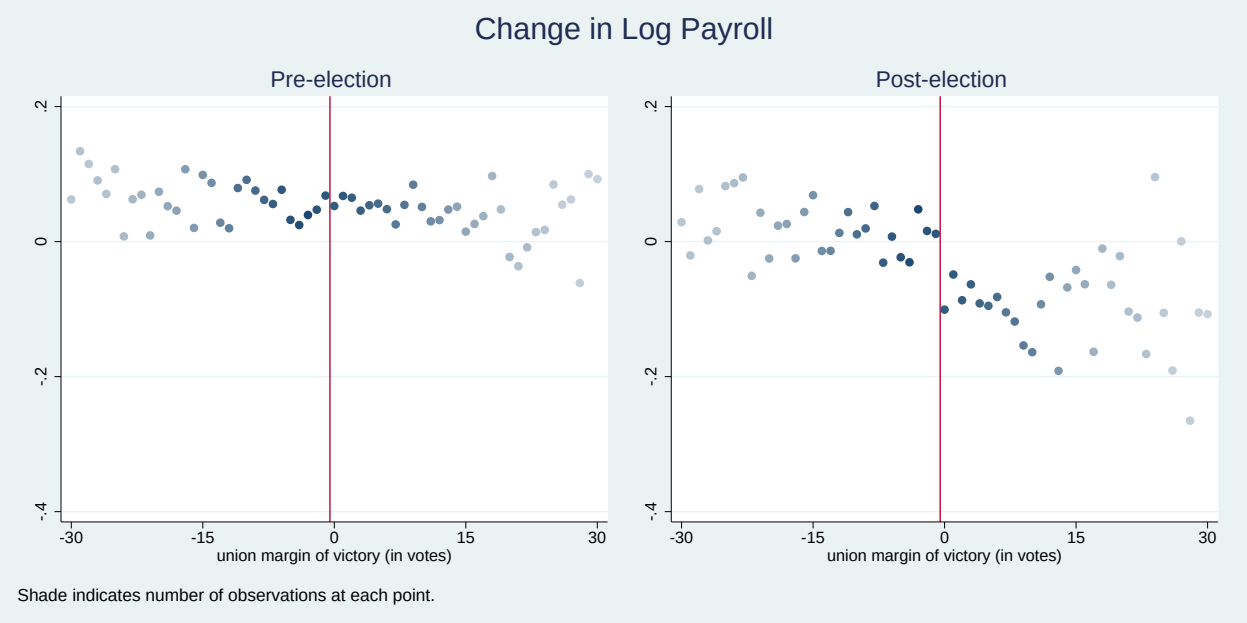

Figure A5: Change in log payroll by union vote tally margin. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by individual tally count. Data are from the LBD and NLRB election records.

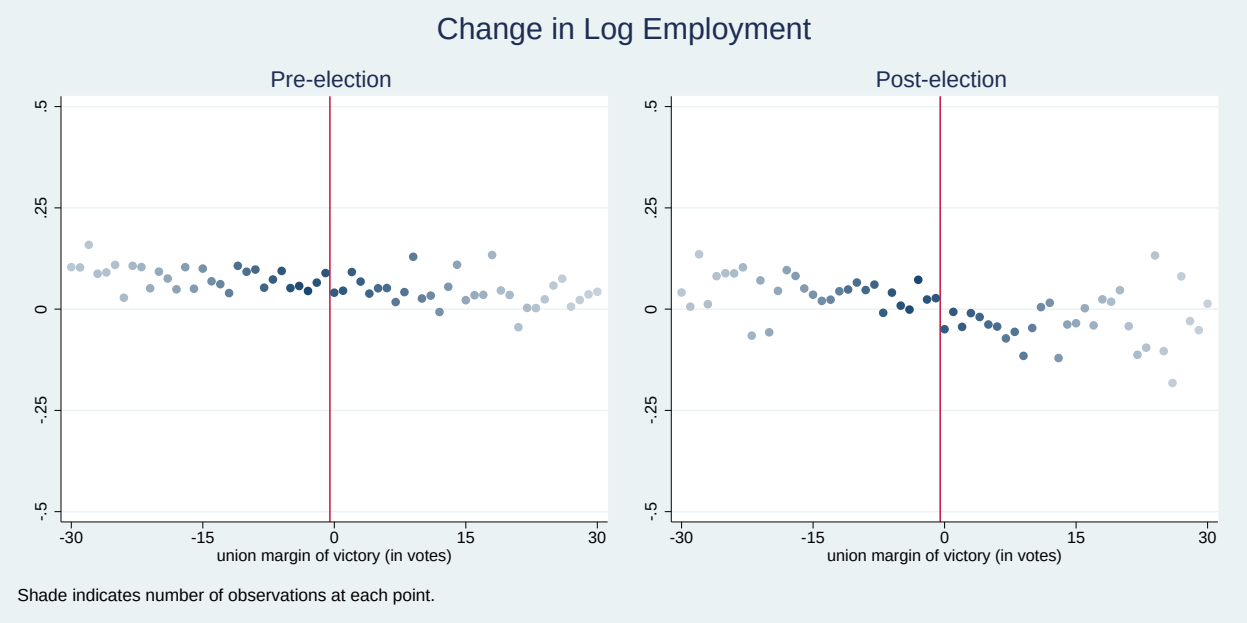

Figure A6: Change in log employment by union vote tally margin. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by individual tally count. Data are from the LBD and NLRB election records.

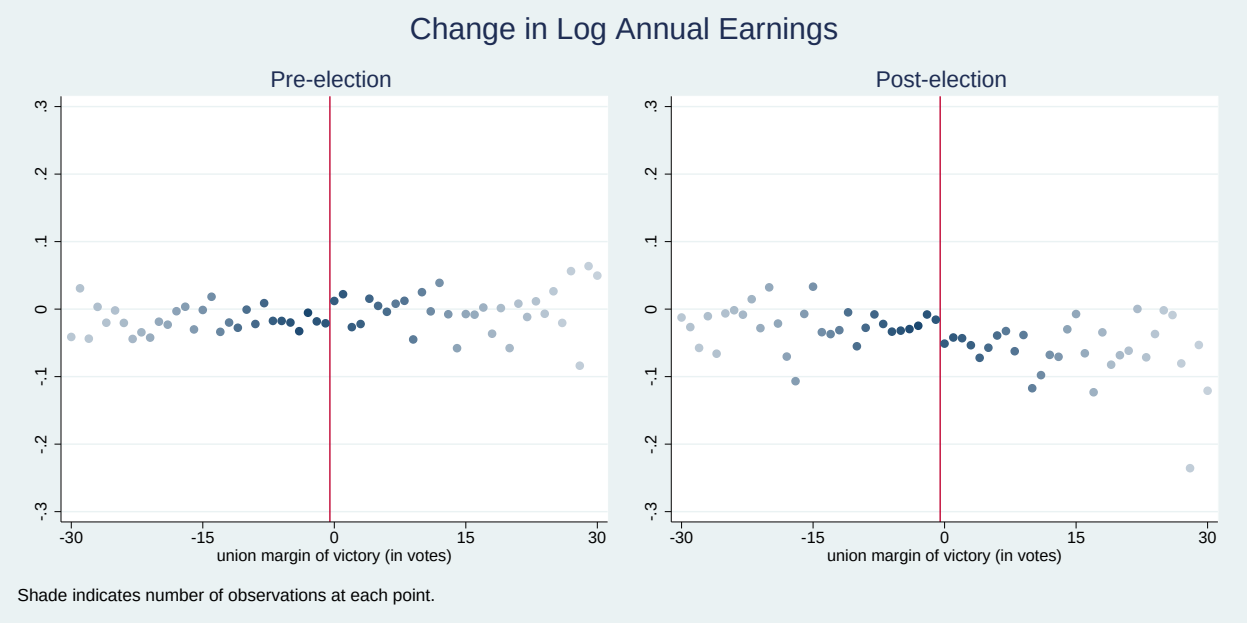

Figure A7: Change in log average annual earnings by union vote tally margin. The left panel plots the change from two years prior to the election to one year prior to the election. The right panel plots the change from the year prior to the election to the year after the election.The figure plots raw averages by individual tally count. Data are from the LBD and NLRB election records.