

Economic Impacts of Unionization on Private Sector Employers: 1984-2001*

John DiNardo
University of Michigan
and NBER

David S. Lee
UC Berkeley, NBER, and
Center for Advanced Study in the Behavioral Sciences

November 2003

Abstract

Estimating the economic impact of unionization on employers is difficult in the absence of large, representative data on establishments with union status information. It is also confounded by selection bias, because unions may organize at highly profitable enterprises that are more likely to grow and pay higher wages. Using multiple establishment-level data sets that represent establishments that faced organizing drives in the U.S. during 1984-1999, this paper utilizes a regression discontinuity design to estimate the impact of unionization on business survival, employment, output, productivity, and wages. Outcomes for employers where unions *barely* won the election (e.g. by one vote) are compared to those where the unions *barely* lost. Overall, the analysis yields small impacts on all outcomes that we examine, including wages – suggesting union weakness in recent decades.

An earlier version of the paper “The Impact of Unionization on Establishment Closure: A Regression Discontinuity Analysis of Representation Elections” is on-line as *NBER Working Paper #8993*, June 2002. Matthew Butler and Francisco Martorell provided outstanding research assistance. We would like to thank David Card, Robert J. LaLonde, Larry Katz, Enrico Moretti, Morris Kleiner, for helpful discussions and Hank Farber for providing election data. We would also like to thank seminar participants at the Bureau of Labor Statistics, the Federal Reserve Bank of Chicago, and the NBER Labor Economics Summer Institute, and the University of Michigan Labor Workshop for comments on an earlier paper, Ritch Milby at the California Census Data Center, and Thomas Kochan. The views expressed are solely those of the authors and do not represent the views of the NLRB, FMCS, or the Bureau of the Census.

1 Introduction

It is widely understood that unions can raise the cost of labor by raising members' wages above market rates.¹ They can also impose other costs on employers – by limiting discretion in hiring and firing, for example, and altering the structure of pay differentials across skill groups. These constraints can lead to reduced employment, output, and/or productivity, or most dramatically, an accelerated pace of business failures.² Indeed, these effects are often directly acknowledged by employers and employees alike. During union organizing drives, for example, firms routinely threaten to close a plant if the union drive is successful (Bronfenbrenner 1994), and employees seem to take these threats seriously: the risk of plant closure is cited as the leading cause of union withdrawal from organizing attempts.³

A key question for measuring the social costs of unionization is whether the constraints imposed on employers create large or small distortions in the allocation of labor. Today, in the United States, arguments can be made for either case. On the one hand, union membership continues to be a significant determinant of earnings; to the extent that employers are sensitive to the price of labor, this would lead to large employment distortions.⁴ On the other hand, there is a broad consensus that in the past three decades, union power in the U.S. has been on the decline. There has been a decrease in union membership, and new organizing activity,⁵ as well as increased managerial opposition to unionization, and use of permanent replacement workers.⁶ During the 1980s, prominent unions were accepting wage cuts, facing the pressures of the opening of international competition. In light of these developments, the magnitude of the economic impact of the

¹ Of particular note are H. Gregg Lewis's seminal surveys on "union wage gaps" see Lewis (1963), Lewis (1986a), and Lewis (1986b). For a more recent examination, see Blanchflower & Bryson (2003).

² See for example, Abowd (1989), Ruback & Zimmerman (1984), Freeman & Medoff (1984), and Hirsch & Schumacher (1998). For a recent survey and critique, see Hirsch (2004).

³ See Commission for Labor Cooperation (1997). The Commission for Labor Cooperation, a tri-national organization created under the North American Agreement on Labor Cooperation ("NAALC") in response to labor issues related to the North American Free Trade Agreement, called for a study on the impact of plant closings on union organizing drives in the three countries.

⁴ See Mankiw (2004) for example. Even Freeman & Medoff (1984) – who argue that emphasis on the "textbook model" is misplaced – stipulate the existence of *some* employment loss. More recently, in cross-country analysis Nickell & Layard (1999) report that a change from 25 to over 70 percent of the workers covered by collective bargaining is associated with a doubling of the unemployment rate. Lalonde, Marschke & Troske (1996), using a "difference-in-difference" approach with LRD data, find successful organization is associated with significant declines in subsequent employment and output.

⁵ See LaLonde & Meltzer (1991) and Farber (2001) for example.

⁶ According to analysis of data from the Gallup Organization: "Between 1936 and 1965, the period of union-density growth and stabilization, the percentage of the public who approved of labor unions fluctuated between 61% and 75%. After 1965, with the decline in union density, the percentage approving of labor unions ranged between 55% and 66%." The percentage of respondents listing "big labor" as the "biggest threat to the country in the future" fell from 22% to 9% between 1981 and 1985. See Cornfield (1999).

“new unionism” on employers is less certain. Naturally, if union weakness results in modest wage (and non-wage) demands, it would also result in insignificant effects on employment, output, or productivity.

Given the current industrial relations climate in the U.S., when individuals unionize a workplace today, what are the economic impacts on the employer? At least two important challenges hinder the credible measurement of the causal impacts of the unionization. One limiting factor is the absence of large, representative data sets that track establishments over time that provide information on union status.⁷ A second important concern is the fact that unionization is non-random. Depending on the correlation between factors associated with employment, output, and productivity, and higher or lower likelihood of unionization, the observed correlation between union status and employer outcomes may overstate or understate the true effects of unions. Two competing phenomena may induce opposite selectivity biases. On the one hand, unions may tend to organize at highly successful enterprises that are more likely to survive and grow. On the other, a union organizing drive may be more likely to succeed when a firm is poorly managed, or has faced recent difficulties.

In this paper, we present quasi-experimental evidence on the causal effect of unionization on employer business failures/dislocations, employment, output, productivity, and wages, using two large databases representative of U.S. establishments at risk of being unionized. Our analysis is based on the fact that most new unionization occurs as a result of a secret ballot election. By law, a majority vote in favor of the union requires management to recognize the union and bargain “in good faith”. This process creates a natural set of comparisons between establishments that faced elections where the union *barely* won (say, by one vote) and those that faced elections where the union *barely* lost (by one vote). As in other regression-discontinuity designs, the comparison between near winners and near losers eliminates any confounding selection and omitted variable biases, and allows us to devise credible and transparent estimates of the effect

⁷ This has led researchers to use creative data collection methods to examine these questions. For example, Freeman & Kleiner (1990a) conducted on-site interviews of 364 establishments that experienced representation elections in the Boston and Kansas City NLRB districts. Bronars & Deere (1993) construct a dataset of NLRB elections to COMPUSTAT data to construct a panel of 85 firms over a 20-year period. Freeman & Kleiner (1999) also use COMPUSTAT to construct a sample of 319 firms. Lalonde et al. (1996) match NLRB representation elections to a subset of manufacturing establishments that are continuously operating in the LRD to create samples with 500 to about 1100 observations.

of unions on employer outcomes.⁸

We report several findings from analyzing data that spans the 1984-2001 period, and combines information on elections from the National Labor Relations Board (NLRB), on contract expirations from the Federal Mediation and Conciliation Service (FMCS), on subsequent business survival, employment, and output from a commercial database based on telephone listings (InfoUSA), as well as on employment, wages, output, and productivity in the manufacturing sector from the U.S. Census Bureau's Longitudinal Research Database (LRD).

We first document that the outcome of an NLRB election has a binding impact on the collective bargaining process, even among close elections. Where they barely win the election, unions are able to maintain their legal recognition over long time horizons; where they barely lose, there is little evidence of subsequent attempts to organize the workplace. Furthermore, unions who barely win have as good a chance of securing a collective bargaining agreement with the employer as those who win the elections by wide margins. And, as expected, unions who barely lose an election have little chance of ever signing such an agreement. These facts show that – statistically speaking – employers face a minimal risk of ever entering collective bargaining negotiations after a union loses a closely-contested election.

This legally-mandated shift in the bargaining position of the workers, however, does not lead to significant impacts on a number of important indicators of the eventual success of unionized businesses. First, union effects on business survival are small – on the order of -.01 to -.02 on a mean survival rate of .40 over an average of 8 years. Second, in the manufacturing sector, point estimates of the union impacts on employment, output, and productivity are also modest, ranging between -3 and 3 percent for hours, -4 and 4 percent for output, and -2 and 0 percent for output per worker, over 1- to 15-year horizons. The effect sizes are tiny compared to the cross-sectional variability in these outcomes, but modest positive and negative impacts cannot be ruled out due to sampling error. Third, we consider the small estimated impacts on employment and output measures to be plausible in light of our findings on wages: point estimates for union

⁸ Regression discontinuity designs are described in Thistlethwaite & Campbell (1960) and Campbell (1969), and formally examined as an identification strategy recently in Hahn, van der Klaauw & Todd (2001).

impacts on wages range between -2 to 0 percent, with enough precision to easily rule out a 5 percent wage gain after 4-7 years following the election. Finally, we supplement our regression-discontinuity analysis with an “event-study” investigation of union “threat effects” – whereby employers raise wages to avoid the threat of future unionization – by assessing whether wages rise in response to an election, even when the union eventually loses. Point estimates are small and statistically insignificant, ruling out, for example, a 3 percent “union threat” effect, 3 years after the election.

It would seem that these small wage effects are at odds with an enormous literature that has documented substantial union wage premiums. The differing results, however, may be explained by some important differences – other than in research design – in the nature of the data used. First, the modern union wage premium literature typically examines individual-level household survey data, rather than establishment-level data. Freeman & Kleiner (1990a) argue that the latter is more appropriate for directly addressing the direct impacts of a workplace becoming unionized.⁹ Indeed, other establishment-level analyses find small or unmeasurable wage effects (Freeman & Kleiner (1990a); Lalonde et al. (1996)). Second, the data contains information on *recent* unionization (within the past 20 years), while most worker-level data sets have little information on when the union was formed. As noted in Freeman & Kleiner (1990a), existing wage differences between union and non-union workers today average the effects of unions of previous periods and the effects of unionization that occurs today.

It is also tempting to view our results as being at odds with the standard “textbook” treatment of the neoclassical theory of union impacts, which emphasizes the notion of a union as an effective “monopoly” on labor services. There is, however, an older tradition in economics that argues – on a purely theoretical level – that most trade unions are unsuccessful monopolies. Indeed, in his essay, “The Impact of the Union,” Milton Friedman (1950) argued that the ability of unions to raise wage rates at that time was somewhat exaggerated, because most unions could not overcome market forces that would tend to keep

⁹ “While it is common to think of selectivity bias in estimating the union wage effect of the difference between the union (and nonunion) sample and the differential that would result from random organization of a set of workers or establishments, we do not believe that is the most useful way to express the problem. What is relevant is not what unionization would do to a randomly chosen establishment but rather what it would do to establishments with a reasonable chance of being unionized – to firms close the margin of being unionized than to the average nonunion establishment.”

wages aligned with competitive rates. In a published exchange with Paul Samuelson, Friedman explains his reasoning: “I think if [UAW leader Walter] Reuther were to disregard [pressures to moderate wage demands] and if he were to seek – and for the moment let us suppose he is temporarily successful – very radically raised wages, and if that had the effect of grossly reducing employment within the automobile industry you would find opposition building up that would break the union down. Knowing that in advance and being as smart as you and I, he would avoid such action.”

The paper is organized as follows. Section 2 provides some background on the the union recognition process and the industrial relations climate in the U.S. in recent decades. Section 3 describes different definitions of the “impact” of unionization, the regression-discontinuity design for estimating direct impacts of unionization, as well as the identification strategy for assessing indirect, “union threat” effects. We describe the various data sets in Section 4, present the results in Section 5, and discuss the findings in relation to the existing literature in Section 6. Section 7 concludes.

2 The Union Recognition Process and the Industrial Relations Climate

In this section, we provide some background on 1) how workers are “typically” unionized, 2) how examining representation elections is helpful for analyzing union impacts on employers, and 3) why the effects of unions today may differ from those in earlier decades.

In the U.S., the effects of unions cannot be considered without also accounting for the the rights and protections that the law – as specified in the National Labor Relations Act (NLRA) – provides to unionized workers. For example, any group of workers can ignore the provisions of the NLRA, simply announce their membership in a union, and attempt to bargain “collectively” with an employer. But this rarely occurs in practice, since any employer faced with such an attempt would have every right – given the “employment-at-will” doctrine in the U.S. – to simply fire any individual associated with the union. By contrast, if a group of workers gains *legal* recognition as provided for by the NLRA, they are legally protected from being fired for association with a union; furthermore, the law dictates that the employer bargain with the union “in good faith”.

How is *legal* recognition of a union gained in the U.S.? There is no single path to an NLRB election and eventual recognition of the union by the employer, but here is a prototypical scenario:

1. A group of workers decide to try to form a union. These workers contact a labor union and ask for assistance in beginning an organizing drive.
2. In collaboration with the union, the employees begin a “card drive.” The purpose of the card drive is to be able to petition the NLRB to hold an election. Unions generally seek to get cards from at least 50 percent of the workers in the 6 month period of time usually allowed (although only 30% is legally required to be granted an election by the NLRB.)
3. After the cards have been submitted, the NLRB makes a ruling on whether the people the union seeks to represent have a “community of interest” – basically form a coherent group for the purposes of bargaining. The NLRB makes a determination of which categories of employees fall within the union’s “bargaining unit.” Often the parties will differ on the appropriate bargaining unit – employers generally prefer larger and more heterogeneous groupings than do unions.
4. Next, the NLRB holds an election at the work site (with exceptions to account for such things as the vagaries of employment seasonality). A simple majority (50 percent plus 1 vote) for one union is all that is required to win.¹⁰
5. Within 7 days after the final tally of the ballots, parties can file objections to how the election was conducted. In principle, with sufficient evidence that the election was not carried out properly, the NLRB can rule to invalidate the outcome of an election, and conduct another one thereafter. Specific ballots cannot be challenged after the voting is completed.
6. If after this, a union still has a simple majority, then the employer is, in principle, obligated to negotiate “in good faith.”

In practice, employers are generally known to resist organizing drives. With data on firms who faced NLRB elections in the early 1990s, Brofenbrenner (1994) documents that most employers used multiple tactics to delay or deny a collective bargaining agreement. Among the most common are

1. “Captive meetings”. While employers are prohibited from directly firing workers because of lawful union activity, at captive meetings employers are allowed to inform workers of the possible (dire) consequences of unionization, including making the business more susceptible to closure.
2. Firing union activists. While “prohibited,” the penalty imposed on employers, if found guilty, is generally quite minor – reinstatement with back pay. Indeed, the costs have been perceived as so minor that Freeman (1985) observes that the notices that firms are required to post when they engage in illegal firing are referred to as “hunting licenses.”
3. Hire a “management consultant” who advises employers on a variety of tactics to discourage unionization.¹¹
4. Alleging unfair labor practices, disputing the choice of bargaining unit, etc.¹²

There are a number of aspects of this process that are helpful for investigating the causal effects

¹⁰ If two unions split the vote 50-50, they both lose, and neither become certified.

¹¹ For a colorful, albeit idiosyncratic discussion see Levitt (1993).

¹² One example is employers arguing that its employees are not workers but “independent contractors” who are not covered by the provisions of NLRA, see for example, the Dunlop Commission report (1994) and Human Rights Watch (2000).

of unions on employers in the U.S. today. First, the fact that employers are thought to generally oppose organization drives [Kleiner 2001] suggests that both parties perceive to have “something at stake” in the outcome of the election. Higher wages, more benefits, security of a contract, seniority pay scales, or a grievance procedure may be sought by the union, but impose constraints on the employer. By contrast, there is arguably more agreement between the union and management on these issues when employers voluntarily recognize a union without an NLRB election – which does occur, but much less frequently. Thus, our sample – which focuses on recognition gained through elections – is likely to be biased in favor of finding differences in union and non-union workplaces, since the management always has the option of voluntarily recognizing the union without a (costly) NLRB election.

Second, combined with a contentious atmosphere, the secret-ballot nature of the vote undoubtedly generates a certain amount of uncertainty in the outcome of the election, particularly when the vote is expected to be close. Indeed, the NLRB’s 1948 ruling in *General Shoe Corporation* case enshrined the so-called “laboratory condition” doctrine requiring as a matter of law a “nearly ideal” election.¹³ One case where the assumption of some randomness to the vote would likely not be justified would obtain if union certification could be secured through a public (i.e. “non-secret”) petition. For example, if all that was required were 50 percent or more signatures, one could imagine that the sample of establishments/unions where the unions submitted a petition with 51 percent of the signatures would be very different from a (peculiar) group of establishments/unions where the workers submitted signatures that totaled 49 percent. In particular, the incentives to obtain the marginal vote to exceed the 50 percent threshold would induce a discontinuity in the density of the vote shares. No discontinuous jump up in the distribution is apparent in the data; see Appendix Figure 1. We thus find it plausible that the exact vote share is, before the election, 1) somewhat uncertain, and 2) for each individual election, has an *ex ante* probability distribution that is smooth at the 50 percent threshold. These two conditions would imply that the variation in union recog-

¹³ “In election proceedings, it is the Board’s function to provide a laboratory in which an experiment may be conducted, under conditions as nearly ideal as possible, to determine the uninhibited desires of the employees. It is our duty to establish those conditions; it is also our duty to determine whether they have been fulfilled. When, in the rare extreme case, the standard drops too low, because of our fault or that of others, the requisite laboratory conditions are not present and the experiment must be conducted over again.” *General Shoe Corp.*, 77 NLRB 124 (1948). See National Labor Relations Board (2002)

nition status that is isolated by the regression-discontinuity design is as good as that from a randomized experiment (see Lee [2003] for a formal proof).

Finally, our data are drawn from universe of elections that occurred within the past two decades, suggesting that our estimates of union impacts on employers during the period we study (1984–2001) might well be much lower than estimates from establishments who were unionized in the more distant past.

Indeed, there are several reasons to believe that the state of U.S. industrial relations in the last twenty to thirty years has not been favorable for the exercise of union bargaining power. For example, it is widely believed that an important element of a union’s power to achieve improvements in wages and working conditions is the threat of a costly strike. But in recent decades, there has been an increased threat and use of striker replacement workers (LeRoy 1995*a*, Olson 1998).¹⁴ Union leaders believe that President Reagan’s large scale replacement of striking PATCO air traffic controllers had a “chilling effect” on the trade unions in the private sector: the industrial relations climate changed so that employers were much less fearful of employing striker replacements (Donohue 1990).

Some researchers have also argued that there has been an intensification of managerial opposition to unionization (Bronfenbrenner 1994, Dickens & Leonard 1985, Kleiner 2001, LeRoy 1995*b*), with the increased incidence of unfair labor practices (LaLonde & Meltzer 1991, Weiler 1983) and use of management consultants to thwart organizing drives and rid employers of existing unions (Lawler 1990, Lawler 1984). Others have argued that other recent developments in the labor market – innovation in labor-saving technologies, and increased openness to international trade – have contributed to union decline (Farber 2001, Katz & Autor 1999, Acemoglu, Aghion & Violante 2001).

Given this environment, it is no surprise that survey data suggest that there has been a reduction in the “demand” for unionism, in large part, due to the perception that unions are no longer able to provide a significant value to the worker. Farber (1989), for example, reports a substantial decline between 1977 and 1984 in the fraction of non-union workers that believe that unions improve wages and working conditions.

¹⁴ Olson (1998), for example, finds that in all industries excluding construction, the use of striker replacements (as a fraction of strikes) were as high or much higher during the period 1985–1988 than they were during the periods (pre–Wagner Act) periods of 1901–11 and 1921–6.

The perception of union weakness is also evident in polling data from Gallup, which indicates that the fraction of people believing that “big labor” is the “biggest threat” to the future of the country fell from 22 percent in 1981 to 9 percent in 1999.

3 Conceptual Framework and Identification Strategy

This section 1) precisely defines several distinct “union impacts” of interest, and shows which effects can and cannot be identified by our research design, and 2) describes our identification strategy. There are several conceptually distinct notions of a “union impact” on an employer, stemming from the possibility that employers can respond to unionism – even when its employees fail to gain recognition.

Consider, as an example, how unions can affect wages. If labor law prohibited collective bargaining, the employer would offer W_0 , the market wage rate; here, unions – by definition – have no bargaining power. In reality, the NLRA does allow collective bargaining as long as it has support among a majority of the workers; given the ever-present possibility of an organizing drive, the employer might offer a slightly higher wage W_M if doing so placated workers and reduced the chances of an organizing drive. Here, although the workers may never be represented by a union, they nonetheless can benefit from the “threat” of unionizing. If an organizing drive does occur, in order to discourage workers from voting for the union and hence paying the wage W_U , the employer may be willing to offer the wage W_N ($W_N > W_M$), to make voting against the union more attractive to workers. This difference $W_N - W_M$ represents an additional “threat” to the employer; once an election is ordered by the NLRB, there is an increased risk of union recognition. These four different circumstances are summarized in the table below:

| Employer’s offer | Law Allows Unions? | Held Election? | Union Recognized? |
|------------------|--------------------|----------------|-------------------|
| W_0 | No | No | No |
| W_M | Yes | No | No |
| W_N | Yes | Yes | No |
| W_U | Yes | Yes | Yes |

There are thus three different union wage “effects” that can be of interest. $W_U - W_0$ represents how much unionized workers are earning relative to what they would earn if unions were prohibited by U.S. law.

$W_U - W_M$ measures how much they earn relative to what they would if they exogenously lost their rights of collective bargaining, but were nevertheless permitted to re-initiate an organizing drive. Finally, $W_U - W_N$ measures the direct consequence of union recognition, conditional on workers expressing sufficient interest in holding an election.

Ideally, a complete picture of the overall impact of unionism in the U.S. would de-compose $W_U - W_0$ into its constituent parts. Focusing on $W_U - W_M$, however, is sufficient for our goal of understanding the effects of unionization in recent years and within the context of current labor law. We analyze $W_U - W_M$ in two parts: first, the direct impact of union recognition on employers $W_U - W_N$, and then second, the indirect impact of union threat $W_N - W_M$. Apparent from the above table, this two-step approach allows us to consider one exogenous factor at a time. That is, we consider separately the questions: What is the impact of union recognition keeping all other things – including having held an election – equal? and What is the impact of having an election keeping all other things – including failing to gain union recognition – equal?

We consider $W_U - W_N$ to be the first-order effect because if the direct impact of legal recognition is zero, so must be the effect of the *threat* of gaining recognition. Intuitively, $W_N = W_U$ can never be an optimal choice for an employer as long as there is a positive probability that the employer will successfully resist the organizing drive with the lower wage W_M .¹⁵ The converse is not true: $W_N = W_M < W_U$ (no threat effects; positive direct effects) can easily be a profit maximizing solution – even if a marginal increase in W_N would reduce the chance of a union victory.¹⁶

¹⁵ To see this formally, consider that the employer is maximizing expected profits

$$(1 - P(W_N)) \pi(W_N) + P(W_N) \pi(W_U)$$

where $P(\cdot)$ is the probability of a union victory, which is assumed to be a negative function of the employer's offered wage W_N , and $\pi(\cdot)$ is profits as a function of the wage. As long as $P(W_M) < 1$, $W_N = W_M$ will strictly dominate $W_N = W_U$.

¹⁶ The first derivative of the expected profit function in the preceding footnote is

$$\underbrace{(1 - P(W_N))}_{(+)} \underbrace{\frac{\partial \pi}{\partial W_N}}_{(-)} + \underbrace{\frac{\partial P}{\partial W_N}}_{(-)} \underbrace{(\pi(W_U) - \pi(W_N))}_{(-)}$$

Thus, even at $W_N = W_M$, the derivative can be negative. Whether or not the objective function is convex or concave throughout the interval $[W_M, W_U]$. Given that W_M dominates W_U , this would imply that $W_N = W_M$ yields the optimum even while $W_U > W_M$.

3.1 Regression-Discontinuity Design

Our main identification strategy is to exploit an experiment that is embedded in NLRB representation elections via a regression discontinuity design. That is, whether or not a union is legally recognized is a deterministic function of the votes in support of the union, where union status is “switched on” when the vote share crosses the 50 percent threshold. As in other regression-discontinuity designs, we attribute evidence of a discontinuous relation between the vote share and an employer outcome to the causal impact of union recognition.

Formally, using the reduced-form dummy endogenous variable framework of (Heckman 1978) we have the system of equations

$$\begin{aligned} y &= X\gamma + D\beta + \varepsilon \\ D &= 1 \quad V > \frac{1}{2} \\ V &= X\delta + u \end{aligned} \tag{1}$$

where y is the employer outcome (employment, wages, output), D is the indicator of union recognition status, V is the vote share for the union in the representation election, X contains observable variables that determine the vote share or the outcomes, and ε and u are corresponding unobservable determinants. β is the parameter of interest and corresponds to $W_U - W_N$.¹⁷

It is widely understood that the OLS, which essentially computes the difference $E[y|X = x, D = 1] - E[y|X = x, D = 0]$ will be biased for β , since generally, $E[\varepsilon|V > \frac{1}{2}] - E[\varepsilon|V \leq \frac{1}{2}] \neq 0$. On the other hand, if we assume 1) there is some ex ante uncertainty in the vote share, and furthermore that 2) the density of u conditional on X and ε is continuous, then it can be shown that the discontinuity in $E[y|V = v]$ at $v = \frac{1}{2}$ identifies the union effect (Lee 2003). That is,

$$\lim_{\Delta \rightarrow 0^+} E[y|V = \frac{1}{2} + \Delta] - \lim_{\Delta \rightarrow 0^+} E[y|V = \frac{1}{2} - \Delta] = \beta \tag{2}$$

Furthermore, Lee (2003) shows that under these two mild continuity assumptions, the variation in

¹⁷ For expositional purposes, we consider a constant treatment effects model, but the assumption is not important. This can easily be extended to a heterogeneous treatment effect framework.

treatment status has the same statistical properties as a randomized experiment; in particular, the distribution of all elements of X and ε will be approximately the same between the treated and control groups within a small neighborhood of $V = \frac{1}{2}$. This implies that we can test the internal validity of the regression-discontinuity design by assessing whether there are discontinuities in the relation between *any* pre-determined characteristic in X and the vote share. That is, a sharp discontinuity in $E[X|V = v]$ at $v = \frac{1}{2}$ would provide evidence against the “randomization” and hence the research design.

In empirically assessing whether there were important discontinuities at the 50 percent threshold, we report our RD results in two ways: 1) graphical plots of $E[y|V = v]$ and $E[X|V = v]$ by 20 vote share categories¹⁸, and 2) approximating the functions $E[y|V = v]$ and $E[X|V = v]$ by fourth order polynomials with an intercept shift at the 50 percent threshold. The first method gives a visual impression of 1) the size of any possible discontinuity relative to the underlying “bumpiness”/curvature in the function and 2) possible approximation errors that could occur from using the polynomial specification. The second method estimates the size of the discontinuity and sampling variability.

Some caution is warranted in making statistical inferences from the polynomial regressions. On the one hand, if the fourth order polynomial functions are “correct”, the estimator is efficiently using data that are both close to and far from the discontinuity threshold. On the other hand, if the true functions do not belong to the class of fourth order polynomials, the discontinuity estimates will in general be biased, and may lead to erroneous inferences of statistical significance. As shown below, based on our graphical analysis, we determined that the fourth-order polynomial was the most parsimonious specification that would not grossly mis-represent the shape implied by the underlying data. For completeness, we also report the results for lower order polynomials.

Finally, some care should be taken in generalizing our RD estimates to a larger population of interest. For example, insofar as there is important unobserved heterogeneity in treatment effects, the effects could be significantly larger among elections that are expected to have 90 percent union support,

¹⁸ 20 vote share bins were chosen because it was the largest number of bins that would accommodate the smallest elections in our sample (20 votes cast). See the Appendix for a detailed explanation of the “integer” problem.

and the RD estimates would “downweight” those larger effects.¹⁹ This would prevent a valid extrapolation of the RD estimates to the larger population of businesses at risk of unionization – particularly if it were the case that most elections resulted in around 90 percent union support. In this particular context, however, this does not seem to be a serious problem. Appendix Figure 1 shows that the bulk of the elections lie in the 25 to 75 percent range of vote shares; a small fraction of the cases are in the 90 percent vote share range. In our manufacturing sample, there are many more elections in the 45 to 55 percent vote share range than the 85 to 95 range – by a factor of 6. So while it is possible that treatment effects may be much larger in the cases where the union is expected to have 90 percent support, such cases do not represent the “typical” election scenario; the median vote share is in the 40-45 vote share range.

3.2 Event-Study Analysis of Threat Effects

Assessing the impact of increased “threat” of unionization while keeping recognition status constant – $W_N - W_M$ – requires a different identification strategy. Our approach is based on the notion that when an employer faces an NLRB representation election, the risk of unionization presumably rises, particularly in the short run, and possibly in the long run. As part of an effort to convince some workers to support management, rather than the union, the employer could offer a higher wage.

To isolate the “threat” effect independently of the direct impact, we employ an “event-study” design to estimate the impact of the occurrence of an NLRB election on the employer – for the sample of elections where we know – ex post – that the union loses and thus fails to gain legal recognition. Thus we keep the actual recognition status constant, but attempt to estimate the impact of the increased “risk” of unionization, as indicated by the event of an NLRB representation election.

We use longitudinal data on manufacturing establishments and estimate the specification

$$w_{it} = \alpha_i + \gamma_t + \sum_{k=-6}^1 D_{it}^k \delta_k \quad (3)$$

where w_{it} is log(average wages) for establishment i in time period t , α_i is a time-invariant fixed effect, γ_t is a year-effect, and D_{it}^k is a dummy variable that takes the value 1 if the election takes place in period

¹⁹ The RD estimand identifies a weighted average treatment effect where the weights are the ex ante probabilities of having a “close” election (Lee 2003).

$t - k$, and 0 otherwise.²⁰ Elections occur every year throughout the period 1984-1999; the specification above simply “re-normalizes” time for each establishment to be relative to the year of the election, in order to provide a picture of the typical before- and after-election experience of an establishment. Assuming that $\delta_{-7} = 0$, the δ_k measure the impact of the event of the election on wages both before and after the election.²¹ If one considers it plausible that the election has no impact on wages more than 2 years before the election event, then the coefficients δ_{-3} , δ_{-4} , etc. should be zero; those coefficients would provide a way to test the over-identifying restrictions of the model.

4 Data Description

Our analysis combines several different data sets: 1) the universe of NLRB representation elections held between 1984 and 1999, 2) the universe of contract expiration notices from the Federal Mediation and Conciliation Service (FMCS) from 1984-2001, 3) business survivorship, employment, and estimated sales volume from a commercial database (InfoUSA) with informatino on population of businesses with a telephone number, as of the year 2001, and 4) detailed employment, output, investment, and wage information from the Census Bureau’s Longitudinal Research Database (LRD) on manufacturing establishments in the U.S. and wage information from 1974 to 1999.

We merge these databases to produce two main estimation samples. The first links the NLRB, FMCS, and InfoUSA data, which are used to examine the impact of unionization on business survivorship, employment, and sales volume for a broadly representative sample of establishments “at risk” of unionization, across all industries. The second links the NLRB, FCMS, and LRD data, which are used to examine the impacts on employment, output, productivity, investment, and wages for a representative sample of manufacturing establishments “at risk” of unionization. Appendix Table 1 and 2 provide a summary of the data sets and provides sample means from the two main datasets. Deferring the details to the Appendix, we summarize here the most important features of the two datasets used.

²⁰ 7 years and earlier are grouped into one category; 11 years and after are grouped together.

²¹ In practice, we omit the dummy D_{it}^0 so the δ_k are all relative to δ_0 . The reported coefficients can easily be re-normalized so that $\delta_{-7} = 0$.

4.1 The NLRB/FMCS/InfoUSA data

We first obtained electronic records on all representation election cases handled by the NLRB in the years from 1984 to 1999. These records have information such as the dates of the filing of the petition, the election, and the closing of the case, as well as the eventual vote tallies, as well as other characteristics such as the size of the voting unit, and the primary industry of the establishment in question. Finally, the records contain the establishment name and exact address.²²

The names and addresses alone were submitted to a commercial marketing database company, InfoUSA, Inc. InfoUSA maintains an annually updated list of all active business establishments (with a telephone listing) in the United States. The basis for their database is the consolidation of virtually all telephone directories in the country. InfoUSA makes a brief call to each establishment at least once a year, to verify their existence, and to update their information on various items such as 1) the total number of employees at the establishment, 2) the estimated sales volume of the establishment, 3) the primary product of the business, and various other characteristics. InfoUSA appended this information to the record for all of the names and addresses that matched to their database (as of May, 2001). InfoUSA was not given any information beyond the name of the business and the street address.

This merged data was then additionally linked to a database of all contract expiration notices between 1984 and February, 2001 – more than 500,000 case records – obtained from the Federal Mediation and Conciliation Service (FMCS) through a Freedom of Information Act (FOIA) request. According to the U.S. Code of Federal Regulations (29 CFR 1425.2)

In order that the Service may provide assistance to the parties, the party initiating negotiations shall file a notice with the FMCS Notice Processing Unit ... at least 30 days prior to the expiration or modification date of an existing agreement, or 30 days prior to the reopener date of an existing agreement...

Thus, parties to collective bargaining agreements are required to file so-called “30-day notices” with the FMCS. Using these data, we added to the NLRB/InfoUSA data information on whether a contract expiration notice was filed from that establishment. This indicator provided our measure of collective

²² Names and addresses are not available for data before 1984.

bargaining “activity” both before and after the election.

Note that these data do not provide outcome measures for more than one year, and in that sense is not a true panel dataset. We only observe “survival”, employment, or sales as of one point in time – in the year 2001. We observe a few “baseline” characteristics from the NLRB election file, but we do not observe employment or sales during the time between the election and the year 2001 since InfoUSA does not retain historical records.

Also note that, since we are measuring employer “survival” as a match (by name and address) in the InfoUSA database, there will undoubtedly be some measurement error. We will inevitably treat some firms as having “died”, when instead InfoUSA was simply unable to match them to their database. On the one hand, this means that that estimates of the *level* of survival rates may be downward biased. On the other hand, the rate of under-matching by InfoUSA is unlikely to be systematically different between close winners and losers, implying that there will not be a *difference* in match rates between the two groups, except if there is a true impact of union certification on survival probabilities.

Similarly, our measure of collective bargaining “activity” will also be downward biased in *levels*. For example, we understate the prevalence of collective bargaining agreements to the extent that our matching algorithm fails to locate a true match in the FMCS data, or that noncompliance with the law (regarding notifying the FMCS when a contract expires) is widespread.²³ Although the levels of this indicator of bargaining “activity” may be biased, it is plausible to assume that these sources of measurement error are not systematically different between close winners and close losers.

4.2 The NLRB/FMCS/LRD Data

The second estimation sample is used to investigate the impact of unionization on other variables that are not available in the InfoUSA database: hours worked, investment, and wages. After combining the NLRB and FMCS databases as described above, the resulting database is linked to the U.S. Census Bureau’s Longitudinal Research Database (LRD). The LRD is a combination of two different data collection efforts:

²³ Some of this downward bias in levels will also be offset by “false positive” matches.

1. Census of Manufactures (CM), which is a census of all manufacturing establishments in the U.S., collected every five years (years ending with “2” and “7”)
2. Annual Survey of Manufactures (ASM), a set of five year panels, in which large firms are surveyed with certainty, and smaller establishments are drawn from the CM, with new samples being generated every 5 years.

The unit of analysis is a “manufacturing establishment” which is generally defined as a single physical location engaged in one of the categories of industrial activity in SIC Division D, Manufacturing. LRD information is confidential and access is limited although available to qualified researchers.²⁴ The survey design is somewhat complicated, but in essence the panel data can be thought of as containing complete annual time-series of information on large manufacturing establishments (greater than 250 employers) and a shorter annual series for a sample of smaller establishments.

5 Results

This section reports 1) RD estimates of the impact on recognition status and collective bargaining activity in the short- and long-run, 2) RD estimates for business survival, employment, and output from the NLRB/InfoUSA data, 3) results for survival, employment, output, productivity, investment, and wages from the NLRB/LRD data, 4) estimates for different time horizons and sub-populations and variable definitions, and 5) event-study estimates of “union threat” effects on wages, and effects of “de-unionization” from an analysis of de-certification elections.

5.1 Impact on Collective Bargaining

We first report evidence that barely winning an election has a lasting impact on legal recognition of the union, and has a measurable impact on the collective bargaining process. If winning a close election had little lasting impact on legal recognition status, there would be no need to assess whether there are any discontinuities in economic outcomes, for there would effectively be no “treatment”. For example, it is possible that unions who barely win are certain to face a de-certification attempt by the employer; or, it is possible that unions who barely lose an initial election are certain to be victorious in a future election. If

²⁴ For more detail on the data, see McGuckin & Pascoe, Jr. (1988) and United States Bureau of the Census (2003).

either of these possibilities are important in practice, there would be little variation in “treatment” among close elections, making it virtually impossible to assess the impact of legal recognition on employer outcomes.

Figure 1A provides evidence on the immediate and lasting impact of winning an initial representation election. The solid squares plot the fraction of elections that result in certification of the union, by the union vote share, in 5 percent vote share groups. There are several dozen cases where a union is not certified even after obtaining more than half the votes, or where a union is certified even without obtaining a majority of the vote; these cases are presumably due to NLRB rulings that over-turn elections due to violations of election procedure, invalid ballots, and so forth. These cases comprise a negligible fraction of all elections. The fraction immediately recognized is essentially zero before the 50 percent threshold, and essentially 100 percent thereafter.

Figure 1A also assesses the hypothesis that unions who barely lose an initial election will inevitably gain recognition in the future. The solid circles represent the probability that a union will eventually gain legal recognition via an election, subsequent to the initial election. Specifically, we focus on elections that take place between 1984 and 1995 and determine whether a union wins an election that is held at those employers, using data from 1984 to 1999.²⁵ The probability rises with the initial election’s vote share, and then drops sharply at the 50 percent threshold, presumably because those who are initially successful do not need to hold another election. Note that some part of the level of the “(2nd) Recognition Later” line is due to a different union gaining recognition at the same employer, as apparent from the right side of the graph. Overall, close losers do seem more likely to eventually gain recognition than unions who lose by a large margin. But close losers in an initial election also have an overwhelmingly lower chance of eventually being recognized than the initial close winners.

Representing the probability that a union will later be de-certified at the employer, the solid triangles in Figure 1A show that while some fraction of unions who barely win an initial election do become de-certified. It is, however, a relatively rare occurrence. The graph implies that about 90 percent of the close

²⁵ So for the 1995 elections, we are allowing 3 years for the union to make another attempt and succeed.

winners maintain their union recognition after three years.²⁶

There is another scenario that would obviate examining employer outcomes at the 50 percent threshold: if employers were just as likely to engage in collective bargaining with both losers and winners of close elections – despite the difference in legal recognition status. That is, the employer may take 49 percent union support to be a signal that it must inevitably negotiate with the union, and hence choose to voluntarily recognize the union and begin collective bargaining negotiations. If this were true, there would be no discontinuous relation between the union vote share and the probability that a collective bargaining agreement is reached subsequent to the election, for example. It would indicate the lack of variability in the treatment of union recognition at the 50 percent vote share cutoff.

Figure 1B rejects this hypothesis. The solid squares plot the probability that an FMCS contract expiration notice is filed at the employer, subsequent to the election.²⁷ The proportion sharply jumps from about 15 to 35 percent at the 50 percent union vote share threshold. Some of the 15 percent threshold will be due to “false positive” matches, as well as the presence of a different union at the employer. This is evident from the open squares, which represents the probability that a contract is observed *before* the election.²⁸ Consistent with the hypothesis that the close winners and losers are otherwise similar, this line shows no visible discontinuity in the proportion at the 50 percent threshold. Representing the effective *change* in collective bargaining activity the *difference* in the two lines is very small leading up to the 50 percent vote share cutoff, jumps sharply, and actually *declines* in the range of 80 to 100 percent. This indicates that there is a greater change in NLRA-induced collective bargaining “activity” for close winners than for unions with overwhelming support.

Figure 1B also illustrates the value in graphing the entire function $E[y|V = v]$, and the danger in relying too heavily on a particular functional form. For the open squares, the jump between the 45-50 and 50-55 vote share categories is not unusual given the jumps between adjacent bins elsewhere in the graph. A

²⁶ Again, we use data from the 1984-1995 period for the initial election, and examine whether a successful de-certification attempt occurs thereafter in the 1984-1999 period.

²⁷ Again, elections from the 1984-1995 period were used, and the dependent variable is whether a contract expiration notices was observed subsequent to the election within the 1984-2001 period.

²⁸ Election data from 1987-1999 are used and it is determined whether a contract expiration notice was filed at the establishment prior to the election using the data from 1984-1999.

fourth-order polynomial fit of the function, however, leads to a statistically significant jump of 0.042 with a standard error 0.01. This is an unbiased estimate – if the fourth order polynomial is the “correct” functional form – but will be biased if the fourth order specification is “incorrect.” For example, a 7th-order polynomial specification gives a statistically insignificant estimate of 0.026 with a standard error of 0.014. Given that – as in all regression discontinuity designs – “statistically significant” results can be sensitive to the choice of functional form, rather than focus singularly on t-statistics, we report point estimates and standard errors from several specifications, and also present the underlying means to provide a visual impression of the expected “bumpiness” one might expect in the function, even in the absence of any true discontinuity.

It should be noted that the observation of a contract expiration notice is not equivalent to whether a union contract is present at the business establishment. In particular, many unions and employers may not comply with the law that requires notifying the FMCS of an impending contract expiration. Evidence on non-compliance or other sources of “under-counting” is provided by an examination of de-certification elections – whereby the employer petitions for an election to determine if a pre-existing union should *lose* its recognition status. In Figure 1B, the open and solid circles present analogous lines for whether a contract expiration notice is observed before and after the de-certification election. As expected, the shape of the two lines mirror that of the certification election, with the probability of observing a contract after the election sharply rising at the 50 percent threshold. The probability of observing a contract before the election ranges from 0.40 to 0.50 (the open circles), but should be 1 – if the contract expiration notice variable were a perfect measure of the presence of a union contract.

It is also important to note that – even apart from the under-counting issue – our contract expiration notice variable also does not perfectly measure the the degree of “bargaining power” of the workers. Just as workers who lose an election can potentially have implicit bargaining power due to the “threat” of unionization, a legally recognized union could pressure an employer to raise wages, even without reaching an actual collective bargaining agreement.

5.2 Analysis of NLRB/InfoUSA: Impact on Survival, Employment, Output

Derived from the NLRB/InfoUSA data, Figures 2 and 3 present our RD estimates of the impact of unionization on survival, employment, sales, and sales per worker. In Figure 2A, the solid squares plot the probability that a business establishment is still in existence as of May, 2001. As mentioned earlier, the employers under consideration are those that held elections between 1984 and 1999. As a result, the business survival effects are averaged over time periods ranging from 2- to 17-year horizons, with more weight given to the longer time-period, since there were more NLRB elections in the mid- to late-1980s.

The solid squares show no visible discontinuity in the survival rates at the 50 percent threshold. Correspondingly, the fourth-order polynomial estimate of the gap yields an effect of -0.012 in probability with a standard error of 0.014. The mean survival rate is about 0.40. The precision of the RD estimates are on the same order of magnitude as the theoretical maximum, since a randomized experiment with a 27560 observations would yield a standard error of the difference of about 0.0041.

The small and potentially null effect on survival is important because it suggests that sample selection bias in an analysis that conditions on survival may be a second order issue. In particular, if the sampling process follows the familiar form of incidental censoring as in:

$$\begin{aligned} y^* &= X\gamma + D\beta + \varepsilon \\ y &= y^* \cdot 1[X\delta + D\phi + v \geq 0] \end{aligned} \tag{4}$$

where the outcome y^* is only observed if the employer remains in business. If (ε, v, X) is independent of D – as in a randomized experiment – and if there is no impact of unionization on survival ($\phi = 0$) then there will be no sample selection bias.²⁹ As argued above, unionization could be thought of as being randomly assigned (among close elections), and Figure 1A is consistent with a zero impact on survival.

Figure 2B presents the RD estimate for $\ln(\text{Employment})$ for the firms that survive as of the year 2001. Again, no visible discontinuity, and the corresponding point estimate is a statistically insignificant positive effect of 0.029 in logs. Given the standard errors, negative impacts of -0.11 can be ruled out at

²⁹ See Lee [2003], for example.

conventional levels of significance. With these data, then, we cannot rule out a moderate negative impact on employment. On the other hand, the impact is small compared to the standard deviation in $\ln(\text{Employment})$ of about 1.64 in logs (see Appendix Table 2).

A similar conclusion is reached from Figure 3A, which presents the result for $\ln(\text{Total Sales Volume})$. Again, the jump from the 45-50 to 50-55 percent vote category is not unusually large. The point estimate is not trivial, but it is statistically insignificant $-.072$. Here, moderately large scale effects cannot be ruled out given the sampling variability of the estimate. The estimates are somewhat more precise when sales is normalized by the size of the employer, as in Figure 3B. The estimated impact on $\log(\text{Sales/Worker})$ is -0.053 with a standard error of 0.049.

A useful feature of our research design is that it generates a number of testable predictions regarding the similarity of employers on either side of the 50 percent union vote threshold. In particular, if in a nearby neighborhood of 50 percent, winning the election is randomly assigned, then *any* pre-determined characteristic should have the same distribution on either side of the cutoff. The NLRB election contain information on industry as well as the number of votes cast, a proxy of the size of the employer at the time of the election. Figure 4A illustrates that these variables are correlated with the vote share, so that employers where unions win are systematically different from those where the union loses. Where the union wins, the employers are less likely to be manufacturing, and more likely to be service sector establishments. Places where the union either wins or loses by a large margin tend to be smaller. But these systematic differences go away as one examines closer and closer elections. Figure 4A illustrates that there are no striking discontinuities at the 50 percent threshold.

Figure 4B is the analogous graph for the sample of surviving establishments as of the year 2001. Another implication of the near randomized variation and sample selection structure assumed above is that if there is no impact of the treatment on sample selection, the distribution of the pre-determined characteristics should be identical – even conditional on survival. Discontinuity gaps at the 50 percent threshold are not apparent in Figure 4B.

5.3 Analysis of NLRB/LRD: Impact on Survival, Employment, Output, Capital,

and Wages

The LRD data produces results similar to that yielded by the InfoUSA data: insignificant impacts of unionization on survival rates, as Figure 5A illustrates. It plots the probability that a manufacturing plant is operating in 1997.³⁰ The point estimate is a statistically insignificant -0.026 in probability, compared to a mean survival rate of about 0.70. As with the InfoUSA data, the effect is small, and the jump at the 50 percent threshold is actually smaller than most differences from adjacent vote share categories away from 50 percent.

Table 1 quantifies these results, showing the results from a number of specifications. The mean difference in survival rates between employers where unions won or lost is about 0.07 to 0.08 in probability, and is statistically significant, as shown in Columns (1) - (4). The importance of functional form is demonstrated in Columns (5) - (8), as the difference disappears as polynomial terms in the vote share are added. The estimate in Column (8) corresponds to the fitted regression line in Figure 5A. As should be the case if unionization is as good as randomly assigned, the estimate does not change significantly as more baseline covariates are added in Columns (9) - (11).

As in the InfoUSA analysis, survival rates are averaged over a range of time horizons, with the elections taking place within the time period 1984-1996.³¹ The solid circles in Figure 5B represent the average number of years that we observe a plant in operation following the election. Again, no apparent discontinuity at the 50 percent threshold, with a point estimate of less than 20 days. The figure shows that we observe, in the LRD data, a plant in operation for about 5 years, on average, after the NLRB election.

Of course, the open circles are means of censored survival times, since the most recent LRD data for this study is 1999. To provide an estimate of the plant's survival time following the election, we estimate an interval regression (Tobit) – for each vote share category, and then with the 4th-order polynomial regression

³⁰ In our analysis of the LRD data, we consider a plant to have “survived” by 1997 if it has a valid, non-zero value for total employment, number of production workers, number of production hours, total payroll, and payroll for production workers in the 1997 Census of Manufactures. A Census year must be used because failure to appear in the 1999 ASM, for example, does not necessarily imply a business failure; the plant may not have been chosen for the ASM sample.

³¹ Since the dependent variable is survival by 1997, the graph does not use elections that take place in 1997 or later.

specification.³² The results are presented as the solid triangles in Figure 5B. There is little evidence of a distinctive change in survival times between the 45-50 and 50-55 vote share categories, and the point estimate of the gap is a statistically insignificant 0.38 (in years). Combined with our results from the InfoUSA data, we conclude that sample selection bias is unlikely to be a first order issue, and we therefore proceed to analyze the sample, conditional on survival.

RD analysis of the LRD data also yields results on bargaining activity, employment, output, and output/worker that are similar to those found with the InfoUSA data, as shown in Figures 6 and 7. In each panel in each figure, three lines are plotted. The solid circles represent the means of the dependent variable – by union vote share category – for establishment-year observations in the years that follow the election. A discontinuity at 50 percent represents our estimate of the causal impact of unionization. The open circles represent the means for observations strictly before the year of the election. For this plot, a significant discontinuity at the 50 percent cutoff would indicate that close winners and losers are systematically different before the election, which would imply a problem with the research design.

Finally, the solid triangles plot averages of the dependent variable after it has been deviated from its pre-election mean. That is, in order to reduce the sampling variability in the discontinuity estimate, each post-election observation is deviated from the overall mean that uses all observations before the election, for each plant. In a randomized experiment, this transformation should not affect the impact estimates, since presumably the pre-election mean is independent of treatment status. Similarly, the RD estimates should not be significantly impacted by this deviated-from-means transformation. But the transformation is likely to alter the shape of the function $E[y|V = v]$, so it provides yet another specification that can be used to probe the sensitivity of the estimates to choices of functional form.

Figure 6A shows a striking discontinuity in the presence of a contract following the election.³³

³² Because the LRD is a combination of a survey and a “census” we can not always identify the precise date that the establishment died, although we can generally pin it down to lying between two different dates where these two dates can be viewed as censoring points. For example, If an establishment is last observed in 1986, then we know it “died” between 1986 and 1987. On the other hand, if last year of the establishment is observed is 1987, then we know it dies between 1987 and 1992; if it is last observed is 1988, then we know it “died” between 1988 and 1992, etc. The dependent variable is then the last year the observation was observed and use the information on the next *Census* year. For 1998 and 1999, it is only right-censored. Our estimation method is the standard extension to the Tobit.

³³ Given the longitudinal structure of the data, an indicator of contract presence was attached to each year. Since contract expi-

This yields the same result as our analysis in Figure 1B, and the magnitude of the gap will be attenuated for the same reasons. This provides evidence that gaining legal recognition is meaningful in the sense that it induces employers to bargain with the union. Figures 6B, 7A, and 7B shows little evidence of a discontinuity in employment ($\ln(\text{total production man-hours})$), output ($\ln(\text{total value of shipments})$), or output/hour ($\ln(\text{total value of shipments/production man-hours})$) at the 50 percent threshold. As expected, the shape of the deviated-from-means plots exhibit comparatively less curvature, and less variability across vote share categories, reflecting a reduction in the variability of the dependent variable.

Figure 8A examines capital assets per worker ($\log(\text{total assets/production workers})$). Again, no visible discontinuity at the 50 percent threshold. On the other hand, the quality of asset data in the LRD is somewhat lower than that for employment and output and highly skewed, which could explain the higher variability in the means by vote share.

By contrast, our estimates for wages ($\log(\text{production payroll/production man-hours})$) appear to be reasonably precise, as shown in Figure 8B. Apparent from the deviated-from-means variable, changes from one vote share category to the next do not exceed 5 percent. Against this background variability, it is difficult to discern any discontinuity at the 50 percent vote share cutoff.

Table 2 presents point estimates of the discontinuity gaps in Figures 6-8, as well as the results from various specifications. Column (1) reports the simple difference in means, comparing the 50-55 percent to the 45-50 percent vote share categories. Column (2) only includes an indicator variable for whether the union won the election, and Columns (3) - (6) add linear, quadratic, cubic, and quartic terms in the union vote share variable. In Column (7), the dependent variable is the post-election outcome minus the pre-election mean of the dependent variable. Column (8) adds the pre-election mean as a regressor, Column (9) adds year dummies, and Column (10) adds 2-digit industry dummies.

The first row shows that the estimated impact of a union victory on the filing of a contract expiration notice is consistently large and statistically significant across specifications. The simple difference in

ration notices will not be filed every year, the dependent variable for the LRD data is whether or not a contract expiration notices was filed in the present year or within the last 3 years. Note also, to keep the estimation sample constant, 0 was assigned to all observations in the LRD before 1984, even though we have no NLRB or LRD data before 1984.

Column (1) is about 0.22 and is about 0.18 in the full specification in Column (10). By contrast, the second and third rows (production hours, output) reports estimates that are much more sensitive to the choice of functional form. They illustrate the danger in relying on one particular specification, and the benefit from showing a larger number of specifications and reporting graphically the means for each vote share category, as in Figures 6-8. For example, the coefficient on the union victory in the hours equation is -0.20 and is statistically significant in the quadratic specification. On the other hand, adding a cubic term changes the coefficient to 0.085 with a standard error of 0.08. That estimates of the discontinuity gap could be so sensitive to functional form may not be surprising, given the significant curvature of the function, as illustrated in Figure 6B.

We favor the specifications in Columns (7) - (10) – particularly for the second and third rows – because there is much less curvature in the functions for the “de-meanned” variables, as apparent in Figures 6B and 7A. The point estimates for production hours range from -0.024 to 0.028; the most precise estimate (Column 10) implies that a 8 percent decline in hours can be statistically ruled out. For output, the estimates range from -0.043 to 0.011 with the most precise estimate ruling out a 10 percent negative impact.

For output/hour – our measure of “productivity” – the estimates are relatively more stable across specifications, and the estimates in Columns (7) to (10) are also more precise. Point estimates range from a impact of -0.015 to -0.019 with a standard error of 0.035. The results are somewhat less precise for assets/worker, with estimates ranging from -0.136 to -0.029, all statistically insignificant.

The last row of Table 2 shows that the estimated wage effects are small, with relatively small standard errors. In fact, all but one of the estimates are negative. Among Columns (7) to (10) the estimates range from -0.026 to -0.016. Consistent with the overall picture presented in Figure 8B, the 15 to 20 percent union wage effect that is typically estimated from household survey data is easily statistically rejected in the LRD data. Among these estimates, the largest positive wage effect that is within two standard errors of the point estimate is about 0.014.

5.4 Estimates for Different Time Horizons, Sub-populations, and Outcomes

Using the full specification in Column (10), Table 3 reports the estimates for different time horizons following the election. Columns (1), (2), and (3) include observations that are 0-3 years, 4-7 years, and 8 or more years after the election, respectively.³⁴ For the observation of a contract expiration notice, the effect is 0.077 in the 0-3 year period – which is to be expected since a contract less likely to expire within the first three years following the election. The estimate rises sharply to 0.307 for year 4 to 7, as shown in Column (2).

The second through fifth rows show results similar to those in Table 2, for each time period: small, positive, and statistically insignificant impacts on hours and output, and small, negative, and statistically insignificant impacts on output/hour or assets/worker. The point estimates are relatively stable across different time horizons.

Finally, the wage effects are also relatively stable. 4 to 7 years after the election, the unionization impact is estimated to be -0.025 with a standard error of 0.021. After 8 years, the estimated impact is -0.005 with a standard error of 0.028. Again, effects of 15-20 percent can be rejected, and the largest positive wage effects that are within two standard errors of the point estimates are 0.017 in years 0 to 3, 0.017 in years 4-7 and 0.051 in year 8 or later. Freeman & Kleiner (1990*b*) note that as far back as Douglas (1930) it has been conjectured that union wage effects were most likely to be found in first contracts than during the “later and more mature years of union development”.

Table 4 reports an analysis of the impacts by different sub-populations. We do this to investigate whether or not our results in the aggregate are masking important effects that vary by important observable dimensions. For reference, Column (1) repeats the estimates from Column (10) in Table 2. In Columns (2) and (3) we split the sample by the number of votes cast. The effects for plants with greater than 75 voters are more positive – for all outcomes except assets/worker – but they are not statistically different from those for the smaller plants.

³⁴ Pre-election observations are kept for constructing the “pre-election mean”.

In Columns (4) and (5) the sample is stratified by whether the NLRB and LRD street address strings exactly matched. To the extent that our matching algorithm – which attempts to match names and addresses even when the strings from the databases are not exactly the same – generates false positive matches, true negative impacts on hours and output, or true positive impacts on wages, could be attenuated towards zero. To the contrary, however, estimates are more positive for hours and output, and more negative for wages in Column (4), compared to Column (5). More direct evidence on the quality of the match between the NLRB and LRD is provided in Appendix Table 3. Regressions of $\ln(\text{production workers})$ from the LRD on $\ln(\text{eligible voters})$ from the NLRB yield slope coefficients of about 0.95 to 0.98 with standard errors of about 0.017, with no important difference between the plants that did or did not exactly match to an NLRB record on the basis of the street address.

Finally, in Columns (6) and (7) the plants are divided into “high-wage” and “low-wage” industries.³⁵ There is a large difference in estimates for assets/worker, but it is not statistically different from zero at conventional levels, given the sampling variability. There do not appear to be important differences for the other outcomes.

As a final robustness check, we estimated effects for different measures of labor “inputs” (total employment, non-production employment), “outputs” (accounting for changes in inventories), “productivity” (different ratios of outputs to inputs), “capital/labor” (different employment measures as denominator), and “wages” (annual earnings for all workers, non-production workers, including supplementary labor costs). Deriving from using the specification in Column (10) from Table 2, these estimates are reported in Appendix Table 4. The table shows that for each broad category, the estimates are stable across different definitions of the outcome.

5.5 Threat Effects and De-Unionization Effects

Figure 9A reports estimates of the “threat effect” – denoted $W_N - W_M$ in Section 3. It plots the estimated

³⁵ More specifically, every establishment was assigned the mean $\ln(\text{production wage, 2000 dollars})$ of its industry. Then the industries were categorized into “high” or “low” groups based on whether the mean $\ln(\text{wage})$ was greater or less than the median “assigned” wage in the sample.

coefficients δ_k from Equation 3, which represents the effect of an election taking place at an establishment on wages, while keeping the non-union status constant. That is, Figure 9A uses only establishments where the union ultimately lost the certification election, and plots the level of wages in years preceding and following the election – including plant fixed effects, and year effects. Each “year-relative-to-election” coefficient is reported relative to year 0, the year of the election, and the upper and lower 95% confidence bounds are plotted as well.

The figure shows that wages are relatively stable in the 7 years leading up to the election. They fall by less than 3 percent in that 7-year pre-election period. The relative stability of the wages before the election provide support for this “event-study” design. Wages are also relatively stable for up to 11 years after the election, with point estimates of wage growth in the post-election period ranging from 0 to 2 percent over the entire period. A 3 percent wage increase by year 3 can be statistically rejected.

For completeness, we include Figure 9B, the analogous picture for the plants where the union won the election. It shows, by contrast, declining wages both after and *before* the election; thus, where unions win, the specification does not pass the over-identifying restriction implied by the “event-study” design.

The analysis thus far has primarily focused on certification elections, because they represent unionization activity that has occurred in recent decades. A similar analysis, however, can be conducted for decertification elections, which represent cases where the employer is attempting to eliminate union recognition. The two main drawbacks with such an analysis is that 1) decertifications are comparatively less frequent, and 2) the data here do not contain detailed information on the history of the union at the plant. Thus, it is difficult to assess the extent to which effects estimated from decertifications today actually represent the effect of unionized workplaces that have been organized many years ago. Nevertheless, for completeness, we report the RD estimates in Figures 1 to 3 for the InfoUSA data and Appendix Tables 5, 6, and 7 for the LRD data. Overall, they show similar results on employment, capital/worker, and wages with point estimates being less than 7 percent in absolute value. The main difference is that for the LRD data, the point estimates are much larger in magnitude for output, and consequently output/hour: they range from negative 15 to 25 percent. The standard errors for all of these estimated effects are substantially larger,

and thus moderate positive and moderate negative impacts cannot be statistically rejected for many of the outcomes.

6 Interpretation and Relation to Previous Literature

In this section, we discuss the interpretation of our five principal findings. First, we conclude from Figures 1A and 1B, that the impact of winning a “close” election on union recognition is immediate and lasting, and that unionization causes an employer to respond in at least one way – by raising the probability of reaching a collective bargaining agreement with the union. That the treatment of unionization is lasting, even at the 50 percent vote share threshold is perhaps not surprising. First of all, by law, if a union loses a representation election, an “election bar” prohibits an election for at least a year under NLRB rules if the union loses.³⁶ Second, during this intermediate period, the workers *do not* enjoy any of the rights of unionized workers provided to them under the NLRA; the employer has considerable leeway to recruit workers who can be expected to have less sympathy with the unionization attempt. Third, union activists report that certification drives are costly and a even a close failure to unionize often leads to significantly diminished support for a second attempt to win collective bargaining rights.

Second, two independent data sets (InfoUSA and LRD) suggest that the impact of unionization on employer survival – the “extensive margin” of employment – is small. Point estimates are close to zero and statistically insignificant. It is possible that the effects are “moderate” (e.g. negative 0.04 on a mean survival rate of 0.40) but that we are unable to statistically detect an effect of such magnitude. This interpretation would be particularly plausible if we found substantial union effects on the “intensive margin” of employment. Indeed, it would be somewhat puzzling to find large impacts on employment with no impacts on employer survival. Our estimates from both the InfoUSA and LRD data, however, can rule out moderate employment effects of 11 or 12 percent or larger.

Third, the two data sets yield small point estimates of the impacts on employment (workers or hours), output, and measures of productivity. From the InfoUSA data, estimates of the impact on employ-

³⁶ If the union wins, a “certification bar” prohibits a decertification election for at least one year.

ment, output, and output/worker range from positive 3 percent to negative 8 percent. For the LRD data, point estimates range from positive 3 percent to negative 2 percent. There are two plausible interpretations of these results. On the one hand, unionization may induce a moderate employment effect, but sampling variability in the estimates prevents us from detecting such magnitudes. For example – although our point estimate is 0.029 – we cannot reject a 10 percent negative employment response to unionization. This is easily consistent with a 15-20 percent union wage effect and a 0.5 to 0.66 labor demand elasticity.³⁷ On the other hand, the impact on employment may actually be somewhat smaller than 10 percent, potentially because unionization may not lead to a 15-20 percent wage increase in the first place. For example, a 5 percent wage increase, with a 0.5 or 0.66 labor demand elasticity would lead to a reduction in employment of about 2.5 to 3.5 percent, which is also consistent with the point estimates presented in this section.

Our fourth finding provides evidence in favor of the latter interpretation. In particular, we find somewhat small wage effects, and can rule out wage effects as small as positive 4 percent. Even with an labor demand elasticity as large as 1, if employers are “on the demand curve”, then this would imply that the estimates are able to statistically reject an employment effect as small as 4 percent at conventional levels.

Finally, we find little evidence that workers benefit from increased wages when an employer faces an increased risk of unionization. Focusing on cases where the union is known to eventually lose the election, after accounting for permanent heterogeneity in wages across employers, year effects, there appears to be little change in wages in the years leading up to and following a certification election. While we deem these “threat effect” estimates to be less reliable than our estimates of the direct impact of unionization, the “event-study” specification does pass the specification test of finding no impact of the election on wages that are determined several years prior to the election event.

The finding that appears most at odds with the existing literature is our estimates of the wage effect. One reason for the discrepancy is that the the data we examine are for employers that switched from non-union to union status during the 1984-1999 period. By contrast, estimates of union wage premium typically utilize data on the current “stock” of union and non-union employers. Even if the union wage premium were

³⁷ This would be assuming that the only impact of unionization on employment occurs through raising the wage.

zero for recently unionized firms, it may be averaged with union wage premiums for older, more-established employers that were organized many decades ago.

Another reason for the difference is that we examine establishment-level data, while most union wage studies examine individual worker-level data. Thus, some and potentially a large part of the observed cross-sectional wage gap in household survey data may be attributable to plant-level heterogeneity in the level of wages. Indeed, using establishment-level data, Freeman & Kleiner (1990*b*) and Lalonde et al. (1996) either find wage effects much smaller than the typical 15-20 percent premiums found in individual-level data, or no effect at all.

A final, related possibility is that the existing literature that is based on household survey data may be identifying a conceptually distinct union “treatment effect”. That is, as articulated in Heckman (1990), the typical target of estimation is How much would a randomly-chosen individual gain is moved from the non-union sector to the union sector? Implicit in this question, is the presumption that the “randomly-chosen” individual is moved from a “randomly-chosen” non-union employer to a “randomly-chosen” union employer. In the present analysis, we are asking, instead, How much more does a randomly-chosen employer have to pay its workers when it becomes unionized? To the extent that unions form at high-wage employers, it is quite possible to estimate a large union wage gap, without estimating a large causal impact of unionization on wages.

As far as theoretical models of the impact of unions are concerned, there is a long history upon which to draw.³⁸ It is interesting to observe that the mere existence of a “union wage premium” was the subject of considerable debate until the 1960s. Indeed, the view that unions did not raise wages was once routinely attributed to “economists.”³⁹ Writing in 1950, Milton Friedman argued that “Unions . . . have not had an extremely important effect, to date, on the structure of wage rates” arguing *inter alia* that in a competitive economy, the labor demand elasticity facing labor unions would vitiate any attempt on their

³⁸ Arguably, the causal effect of unionization on employers one of the oldest concerns in modern economics. Humphrey (1992) observes that this question was the impetus for one of the earliest published uses of the “supply and demand” figure in Jenkin (1870)

³⁹ Consider the following from Jenkin (1868): “Let us now hear what men in unions claim to have accomplished . . . As to wages, the men say: – ‘We *have* raised wages; if political economy says this is impossible, so much the worse for political economy; we know that unions do raise wages, and our employers know it, and this is one reason why they are hostile to unions.’”

part to raise wages. Paul Samuelson, on the other hand, argued strongly against such a conclusion citing evidence that unions such as the UAW did in fact significantly raise real wages and for their members and arguing that Friedman had failed to identify a proper “quasi-experiment”⁴⁰ Moreover, most models of the effect of collective bargaining on wages depend on the elasticity of non-unionized employment to the establishment including models based on “imperfect competition” in the labor market (see Manning (2003) for example.)

7 Conclusion

This study meets two important challenges of credibly estimating the magnitude of the causal effects of unions. First, we have constructed a large data set that represents a virtual universe of establishments facing potential unionization, linked to a comprehensive database on survivorship on businesses. Using over 27,000 observations, we have also exploited a feature of the NLRB election process to produce quasi-experimental estimates that are likely to be free of selection and omitted variable biases. Our results suggest that the causal effect of the “new unionism” on establishment survival is quite small. In manufacturing, we generally find very, small wage effects – we can not rule out small (5 to 6 percent) positive (or negative) effects. Likewise, our estimates of the effect of unionization on manufacturing employer outcomes suggest small effects.

It would be also interesting to examine other outcomes which affect workers. For instance, although they find small effects on most employers outcomes Freeman & Kleiner (1990*b*) find enormous impacts of unionization on “industrial jurisprudence” – such as the existence of grievance procedures, seniority provisions, the written posting of promotion opportunities and profit sharing – and as both Freeman & Medoff (1984) and Slichter, Healy & Livernash (1960) suggest, these “voice” effects could be ultimately more important for workers than possible “monopoly union” effects.

⁴⁰ See Milton Friedman (and discussants) (1950). Friedman was quite explicit about the need for an appropriate “counterfactual” and attempted to construct one using crude aggregate figures (the only ones available to him at the time.) Samuelson’s response: “How do you solve these terrible problems of empirical identification, concerning which analytic effect you are observing, by these simple historical comparisons?”

- Abowd, J. M. (1989), 'The effect of wage bargains on the stock market value of the firm', *American Economic Review* **79**(4).
- Acemoglu, D., Aghion, P. & Violante, G. L. (2001), 'Deunionization, technical change, and inequality', *Carnegie-Rochester Conference Series on Public Policy* **55**, 229–264.
- Blanchflower, D. G. & Bryson, A. (2003), 'What effect do unions have on wages now and would 'what do unions do' be surprised?', Working Paper 9973, National Bureau of Economic Research, Cambridge, MA.
- Bronars, S. G. & Deere, D. R. (1993), 'Union organizing activity, firm growth, and the business cycle', *American Economic Review* **83**, 203–220.
- Bronfenbrenner, K. (1994), 'Employer behavior in certification elections and first contracts: Implications for labor law reform', in S. Friedman, R. Hurd, R. Oswald & R. Seeber, eds, 'Restoring the Promise of American Labor Law', ILR Press, Ithaca, New York, pp. 75–89.
- Campbell, D. T. (1969), 'Reforms as experiments', *American Psychologist* **24**(4), 409–429.
- Commission for Labor Cooperation (1997), 'Plant closings and labor rights: A report to the council of ministers on the effects of sudden plant closings on freedom of association and the right to organize in Canada, Mexico, and the United States'.
- Commission on the Future of Worker–Management Relations (1994), 'Fact finding report', U.S. Department of Labor, U.S. Department of Commerce, Washington D.C.
- Cornfield, D. B. (1999), 'Shifts in public approval of labor unions in the United States, 1936–1999', Guest Scholar Poll Review. www.gallup.com/poll/guest_scholar/gs990902.asp.
- Dickens, W. T. & Leonard, J. S. (1985), 'Accounting for the decline in union membership, 1950–1980', *Industrial and Labor Relations Review* **38**(3), 323–334.
- Donohue, T. (1990), 'Testimony of Thomas Donohue, Secretary Treasurer of the AFL-CIO', in 'Preventing Replacement of Economic Strikers 1990: Hearing on Senate 112 Before the Subcommittee on Labor of the Senate Committee on Labor and Human Resources, 101st Congress, Second Session'.
- Douglas, P. H. (1930), *Real Wages in the United States, 1890–1926*, Houghton Mifflin, Cambridge, Mass.
- Farber, H. S. (1989), 'Trends in worker demand for union representation', *American Economic Review* **79**(2). Papers and Proceedings of the Hundred and First Annual Meeting of the American Economic Association.
- Farber, H. S. (2001), 'Accounting for the decline of unions in the private sector, 1973–1998', *Journal of Labor Research* **22**(3).
- Freeman, R. B. & Kleiner, M. (1990a), 'The impact of new unionization on wages and working conditions', *Journal of Labor Economics* **8**(1), S8–S25.
- Freeman, R. B. & Kleiner, M. M. (1990b), 'The impact of new unionization on wages and working conditions', *Journal of Labor Economics* **8**(1, Part 2), S8–S25.
- Freeman, R. B. & Kleiner, M. M. (1999), 'Do unions make enterprises insolvent', *Industrial and Labor Relations Review* **52**(4), 510–527.
- Freeman, R. B. & Medoff, J. L. (1984), *What Do Unions Do?*, Basic Books, New York.
- Friedman, M. (1950), 'Some comments on the significance of labor unions for economic policy', in D. M. Wright, ed., 'The Impact of the Union: Eight Economic Theorists Evaluate the Labor Union Movement', Harcourt, Brace and Company, New York. Institute on the Structure of the Labor Market, American University, Washington D.C.
- Hahn, J., van der Klaauw, W. & Todd, P. (2001), 'Identification and estimation of treatment effects with a regression–discontinuity design', *Econometrica* **69**(1), 201–209.
- Heckman, J. (1978), 'Dummy endogenous variables in a simultaneous equation system', *Econometrica* **46**, 931–960.

- Heckman, J. (1990), 'Varieties of selection bias', *American Economic Review* **80**, 313–318.
- Hirsch, B. T. (2004), What do unions do for economic performance?, in J. T. Bennett & B. E. Kaufman, eds, 'What Do Unions Do? The Evidence Twenty Years Later', chapter 6.
- Hirsch, B. T. & Schumacher, E. J. (1998), 'Unions, wages, and skills', *Journal of Human Resources* **33**(1), 201–219.
- Human Rights Watch (2000), 'Unfair advantage: The freedom of association in the united states under international human rights standards'.
- Humphrey, T. (1992), 'Marshallian cross diagrams and their uses before alfred marshall: The origins of supply and demand geometry', *Economic Review* pp. 3–23.
- Jenkin, F. (1868), Trade–unions: how far legitimate?, in S. C. Colvin & J. A. Ewing, eds, 'Papers, Literary, Scientific, &c by the late Fleeming Jenkin', Vol. 2, Longmans, Green & Co., London. Originally published in the *North British Review* March, 1868.
- Jenkin, F. (1870), The graphic representation of the laws of supply and demand, and their application to labour, in S. C. Colvin & J. A. Ewing, eds, 'Papers, Literary, Scientific, &c by the late Fleeming Jenkin', Vol. 2, Longmans, Green & Co., London.
- Katz, L. F. & Autor, D. H. (1999), Changes in the wage structure and earnings inequality, in O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3, North Holland, Amsterdam.
- Kleiner, M. (2001), 'Intensity of management resistance: Understanding the decline of unionization in the private sector', *Journal of Labor Research* **22**(3).
- Lalonde, R. J., Marschke, G. & Troske, K. (1996), 'Using longitudinal data on establishments to analyze the effects of union organizing campaigns in the united states', *Annales D'Economie et de Statistique* **41/42**, 155–185.
- LaLonde, R. J. & Meltzer, B. D. (1991), 'Hard times for unions: Another look at the significance of employer illegalities', *University of Chicago Law Review* **58**, 953–1014.
- Lawler, J. J. (1984), 'The influence of management consultants on the outcome of union certification elections', *Industrial and Labor Relations Review* **38**(1), 487–501.
- Lawler, J. J. (1990), *Unionization and Deunionization: Strategy, Tactics, and Outcomes*, University of South Carolina Press, Columbia, SC.
- Lee, D. S. (2003), Randomized experiments from non-random selection in u.s. house elections, Unpublished manuscript, University of California, Berkeley.
- LeRoy, M. H. (1995a), 'The changing character of strikes involving permanent striker replacements, 1935-1990', *Journal of Labor Research* **16**, 423–437.
- LeRoy, M. H. (1995b), 'Regulating employer use of permanent striker replacements: Empirical analysis of nlr and rla strikes 1935–1991', *Berkeley Journal of Employment and Labor Law* **16**(1), 169–208.
- Levitt, M. J. (1993), *Confessions of a Union Buster*, Crown, New York, NY. With Terry Conrow.
- Lewis, H. G. (1963), *Unionism and relative wages in the United States*, University of Chicago Press, Chicago.
- Lewis, H. G. (1986a), Union relative wage effects, in O. Ashenfelter & R. Layard, eds, 'Handbook of Labor Economics', Vol. 2, North Holland, Amsterdam, chapter 20, pp. 1139–1182.
- Lewis, H. G. (1986b), *Unionism relative wage effects: A Survey*, University of Chicago Press, Chicago.
- Mankiw, G. N. (2004), *Principles of Economics*, third edn, Thomson Southwestern, Mason, OH. Chapter 26.
- Manning, A. (2003), *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press, Princeton, NJ.
- McGuckin, R. H. & Pascoe, Jr., G. A. (1988), The longitudinal research database (Ird): Status and research possibilities, CES Working Paper 88-2, Center for Economic Studies, U.S. Bureau of the Census,

Upper Marlboro, MD.

National Labor Relations Board (2002), 'An outline of law and procedure in representation cases', Published by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. Prepared by the Office of the General Council.

Nickell, S. & Layard, R. (1999), Labor market institutions and economic performance, in O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3C of *Handbook in Economics*, Elsevier Science, New York, chapter 9, pp. 473–521.

Olson, C. A. (1998), The use of strike replacements in labor disputes: Evidence from the 1880s to the 1980s. Paper presented at the 1999 Winter meetings of the IRRA, New York City.

Ruback, R. & Zimmerman, M. (1984), 'Unionization and profitability: Evidence from the capital market', *Journal of Political Economy* **92**, 1134–1157.

Slichter, S. H., Healy, J. J. & Livernash, E. R. (1960), *The Impact of Collective Bargaining on Management*, Brookings Institution, Washington, D.C.

Thistlethwaite, D. & Campbell, D. (1960), 'Regression–discontinuity analysis: An alternative to the ex post facto experiment', *Journal of Educational Psychology* **51**, 309–317.

United States Bureau of the Census (2003), 'Longitudinal research database', Accessed October 10, . www.census.gov/e

Weiler, P. (1983), 'Promises to keep: Securing workers' rights to self-organization under the nlra', *Harvard Law Review* **96**, 1769, 1772–1773.

8 Appendices

8.1 Data Appendix

8.1.1 The NLRB to InfoUSA match

First, electronic records on all representation election cases handled by the NLRB in the fiscal years 1984 to 1999 were obtained. These records contain information such as the dates of the filing of the petition, the election, and the closing of the case, as well as the eventual vote tallies, as well as other characteristics such as the size of the voting unit, and the primary industry of the establishment in question. Most importantly the file contains information on the name and street address at which the representation election was held.

These 139,881 records were then matched by name and address to a commercial marketing database company called InfoUSA, Inc. Before being sent to InfoUSA, however, the address fields were first “standardized” using a program called “Mailers +4 Postal Automation Software.” For example, “1 Broad Street” was changed to “1 BROAD ST”. This was done to facilitate matching the NLRB data to the data from InfoUSA.

As discussed in the text, InfoUSA maintains an annually updated list of all business establishments (with a telephone listing) in the United States. The basis for their database is the consolidation of virtually all telephone books in the country. InfoUSA makes a brief call to each establishment at least once a year, to verify their existence, and to update their information on various items such as 1) the total number of employees at the establishment, 2) the estimated sales volume of the establishment, 3) the primary product of the business, and various other characteristics.

We submitted the name and address information from our “address standardized” NLRB data to InfoUSA who matched as many of the submitted records to their current database (as of May, 2001) and then appended their information to the record. Apart from the name and address information, no other information was given to InfoUSA.

Before merging this data to our data from the Federal Mediation and Conciliation Service (described below) the data were cleaned for duplicates. There were three types of duplicates: 1) genuine

duplicates – more than one NLRB case with a specific employer, 2) duplicates which were an artifact of the fact that our NLRB data came in two files: one contained data from 1977 to 1991 and the other contained the records for 1984-1999. Most of the duplicate pairs therefore occurred for the years 1984–1991, although there were some duplicate pairs in other years because of the fiscal/calendar year distinction; 3) a very small number of duplicates where two records containing exactly the same information.

8.1.2 Matching Algorithm: NLRB to FMCS and NLRB to LRD

Matching records between the NLRB and FMCS and between the NLRB and LRD data involved the following procedure. Matching was done on the basis of the company name, street address, city, and state, which are elements in each of the three datasets. First, all datasets were stripped of special characters such as @, #, %, &, *, etc. Then, “common” words were stripped from both the name and address fields. For example, “COMPANY”, “CORPORATION”, “INC”, “CO”, “STREET”, “AVENUE”, etc. What was considered “common” was based on a complete cataloging of all words in the NLRB and FMCS database. The words were ranked by their frequency, and the most frequent ones were considered “common”. For each case in the NLRB, the algorithm isolated the subset of records in the FMCS (LRD) that matched exactly on city name and USPS state abbreviation. Within this smaller subset, the NLRB record’s name and address were compared to the corresponding entries in the FMCS (LRD). For each comparison, a “spelling distance” (the function SPEDIS in SAS) was created. The spelling distance is a number from 0 to 100 indicating the difference between two strings. 0 means an exact match, and 100 indicates no similarity. An index that combined the name and address spelling distances were constructed, and all comparisons where the index was below a particular threshold was considered a “match”.

In particular, the coefficients for the name and address spelling distances, as well as the threshold was determined in the following way. 100 records were chosen randomly from the NLRB database. We conducted a manual search for matches in the FMCS database using regular expression searching (“grep” in Unix). We called the matches that resulted from this non-automated procedure “true” matches. Then a dataset was constructed that attached to each of the randomly chosen 100 records from the NLRB, all the

records from the FMCS in the same city. Each observation in this “expanded” dataset, then, is assigned a “1” if the pairing is a “true match” (as determined by the manual search), and “0” otherwise. Using this dataset, a probit was run using this indicator variable as the dependent variable and the spelling distances for name and address as the independent variables. The coefficients from this probit (including the constant) are used to construct the index described above; inclusion of the constant in this index implies that 0 is the optimal threshold for deciding what is a match.

Note that in order to make the algorithm invariant to the order in which NLRB records are matched, the matching algorithm is done “with replacement”. That is, multiple NLRB records could potentially match to the same FMCS (LRD) record.

With the LRD data, each establishment can have multiple names (due to spelling differences, orders of names, and ownership changes), each corresponding to one year within the longitudinal database.⁴¹ Thus, NLRB records tended to match to several different names, and it was necessary to pick the single “best” match. This was done by first eliminating all names of establishments for years that were equal to or greater than the year of the election. Then, among the remaining matches - which all had indices greater than 0 - the highest match index was chosen. The single establishment that had that highest index was assigned to the NLRB record. Again, note that it is possible, given matching “with replacement” that more than one NLRB record can match to the same establishment. Therefore, all standard errors are clustered at the level of the establishment.

8.1.3 Sample Selection and Variable Construction

First, in order to minimize measurement error in our survival indicator, we keep only those NLRB records that have a non-missing street address, and those addresses that were successfully standardized using our address standardization software. In addition, we keep all elections in which 20 or more votes were cast. Appendix Table I reports the means and standard deviations of the variables for the entire merged database and the restricted sample used for the estimation. The table shows that the means are reasonably similar

⁴¹ More specifically, the LRD itself does not contain names and addresses. It must be merged on from the Standard Statistical Establishment List (SSEL) - via a unique identifier that links the two data sets.

between the two samples, except in the variables that reflect the scale of the employer (number of eligible voters, votes cast, employment, sales volume).

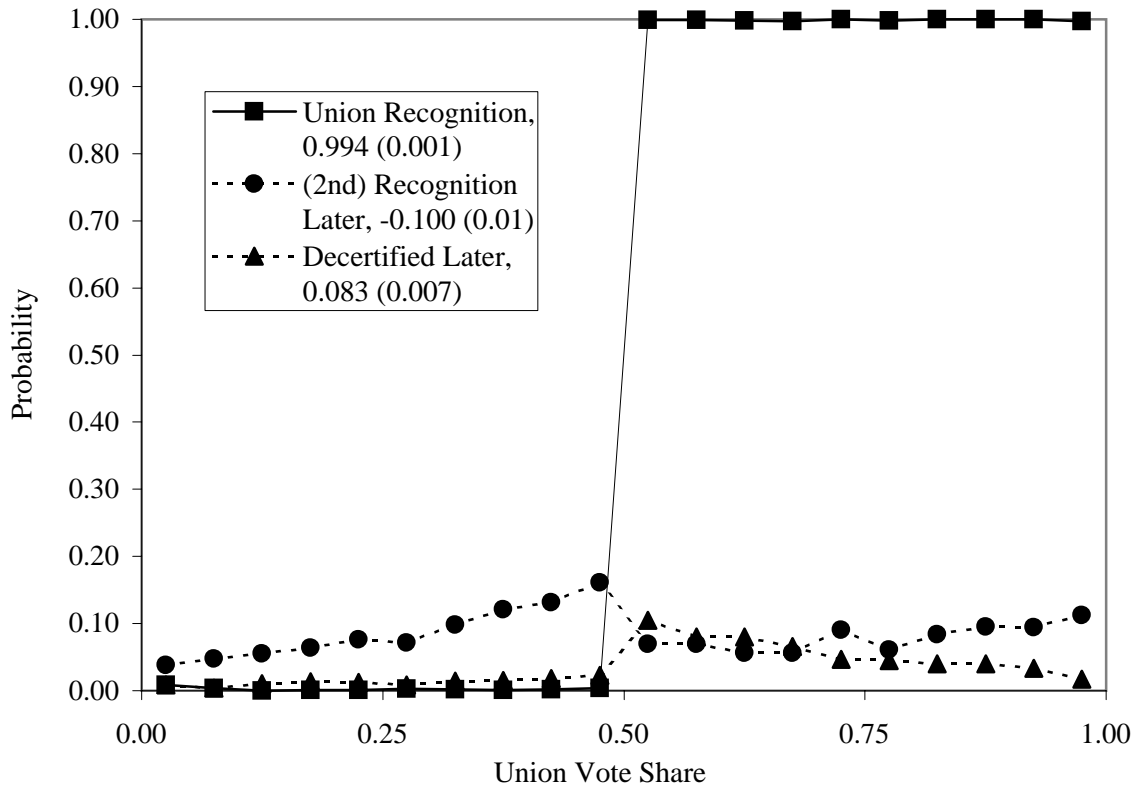
Finally, great care was taken to construct the vote share for the union. There is a problem with simply computing the ratio of the number of votes for the union to the total number of votes. This is because there is substantial variability in the number of votes cast, and a union victory is secured by obtaining strictly more than 50 percent of the vote (plus 1 vote). Consider all elections where an even number of votes are cast. Elections with any number of votes cast could result in exactly 50 percent of the vote. However, it is *impossible* that an election with less than 100 votes cast could have a vote share between 50 and 51 percent of the vote. This *mechanically* induces a discontinuity in the size distribution at the 50 percent threshold of the vote share, and this is entirely an artifact of the fact that the vote share is not *literally* continuous, and instead has finite and discrete support, with the support changing with the number of votes cast.

Thus, we made a minor adjustment to the vote share variable in order eliminate this problem. For every case where there was an *even* number of votes cast, an amount equal to $0.5/(\# \text{ votes cast})$ was subtracted from the vote share. For example, if 25 out of 50 votes were for the union, the vote share became $0.50 - 0.01 = .49$; 26 out of 50 votes meant a vote share of $0.52 - .01 = 0.51$. Cases where an odd number of votes cast were unadjusted. This minor adjustment restores symmetry in the support for vote share, and the new “vote share” variable still possesses the property that strictly more than 0.50 implies a union victory. Finally, the vote share was “binned” so that all vote shares between 0.50 and 0.55 were assigned the vote share of 0.525, shares between 0.45 and 0.50 were assigned the share of 0.475, and so forth. In this way, vote shares were standardized to the support for the elections with the smallest number of votes cast (20).

A completely different approach is to abandon the use of the vote share completely, by focusing on the *absolute* vote count, and comparing elections in which the union either won or lost by literally 1 vote. This eliminates this “integer problem”, but at the same time tends to push larger establishments away from the threshold that determines victory (generating a pronounced U-shape in the average size of the establishment, with respect to the absolute vote margin of victory/loss). This was the approach used in DiNardo and Lee [2001]; it should be noted that the results reported there are qualitatively and

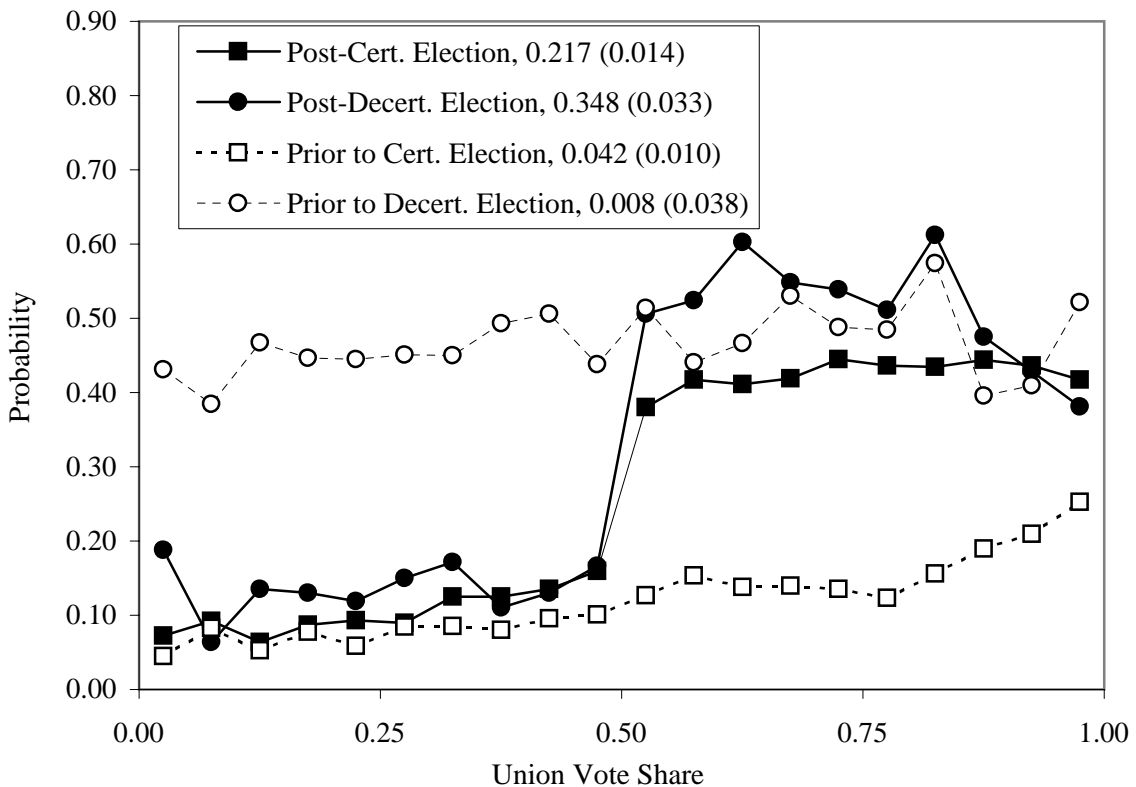
quantitatively similar to the results in this paper, suggesting that our findings are not sensitive to the method used to address the integer problem.

Figure 1A: Recognition, Subsequent Certification or Decertification, by Union Vote Share



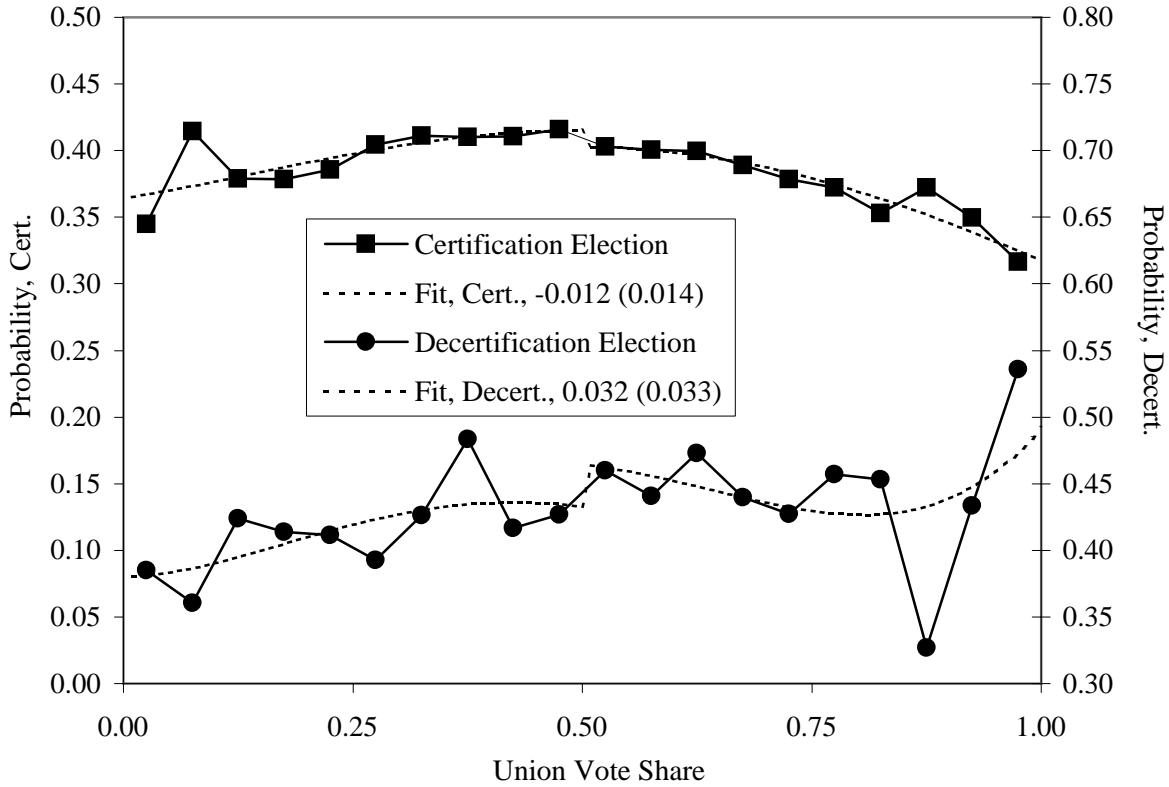
Note: 21405 Observations

Figure 1B: Contract Expiration Notice Filed, Prior to and Post-Certification or Decertification Election, by Union Vote Share



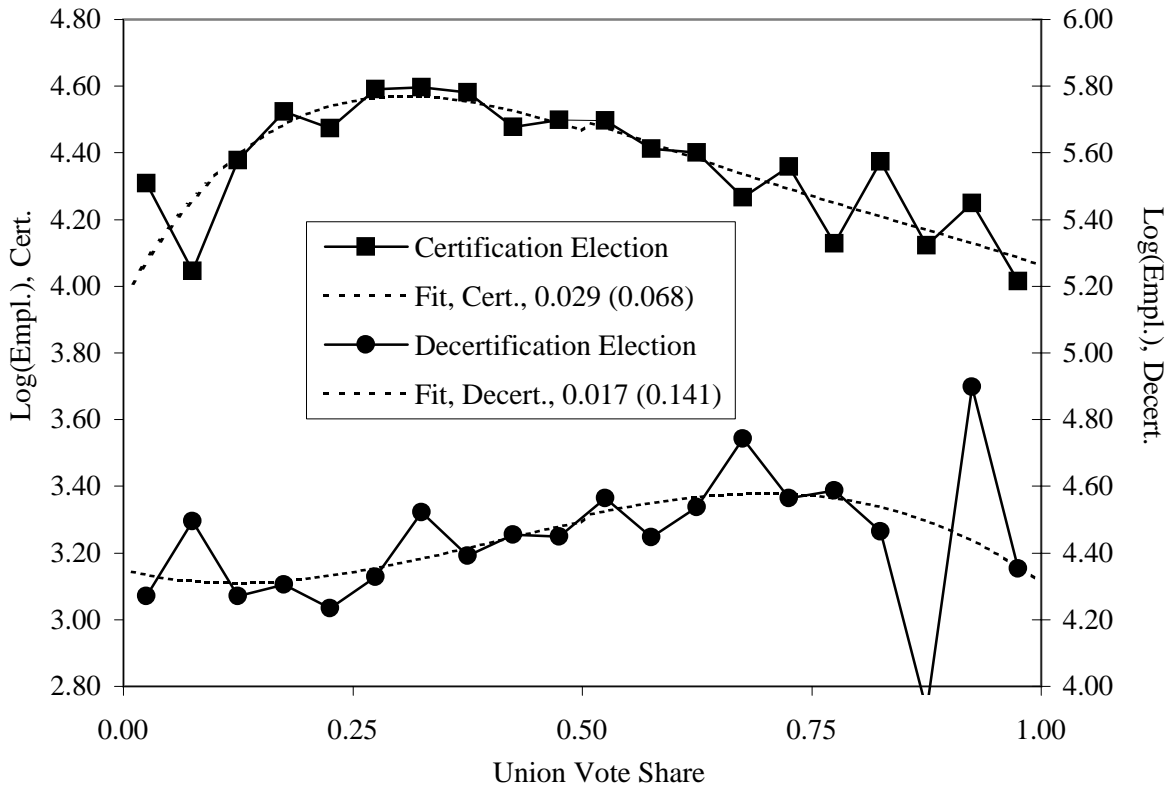
Note: 21405, 3785, 21457, and 3445 Observations, respectively.

Figure 2A: Survival at Year 2001, by Union Vote Share



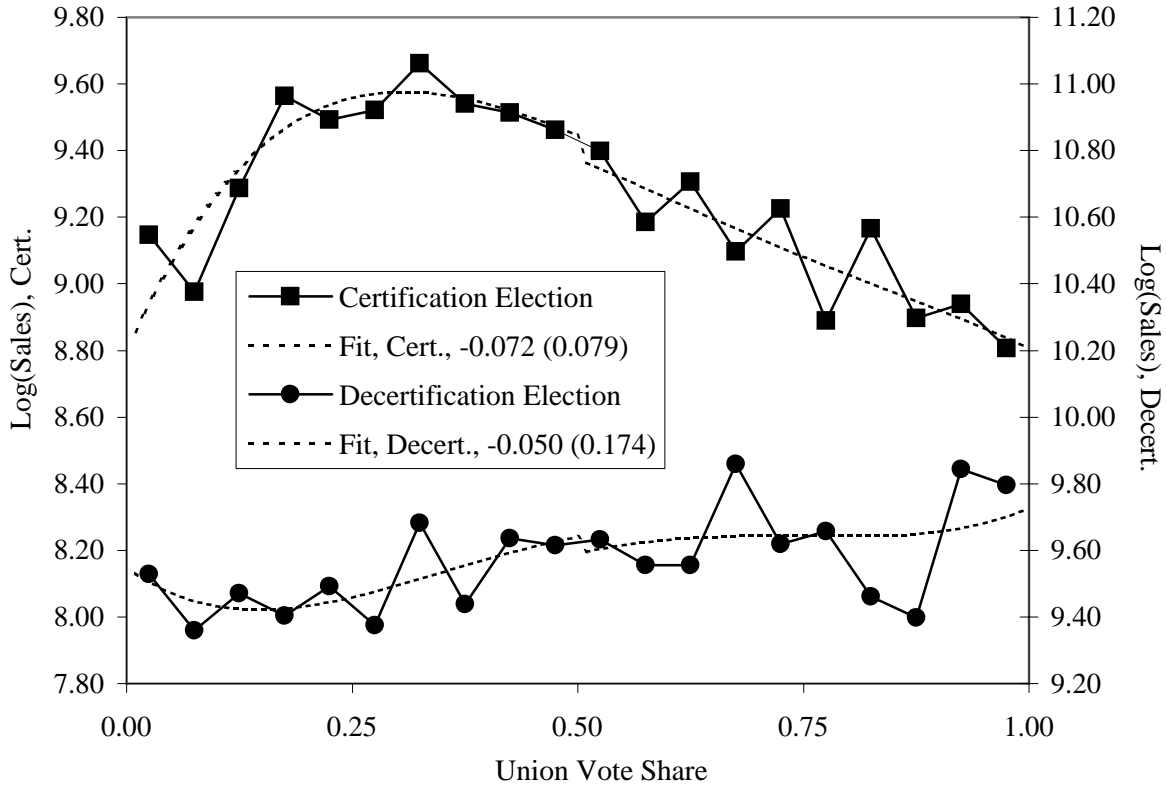
Note: 27560 and 4638 observations, respectively.

Figure 2B: Log(Employment), Year 2001, by Union Vote Share



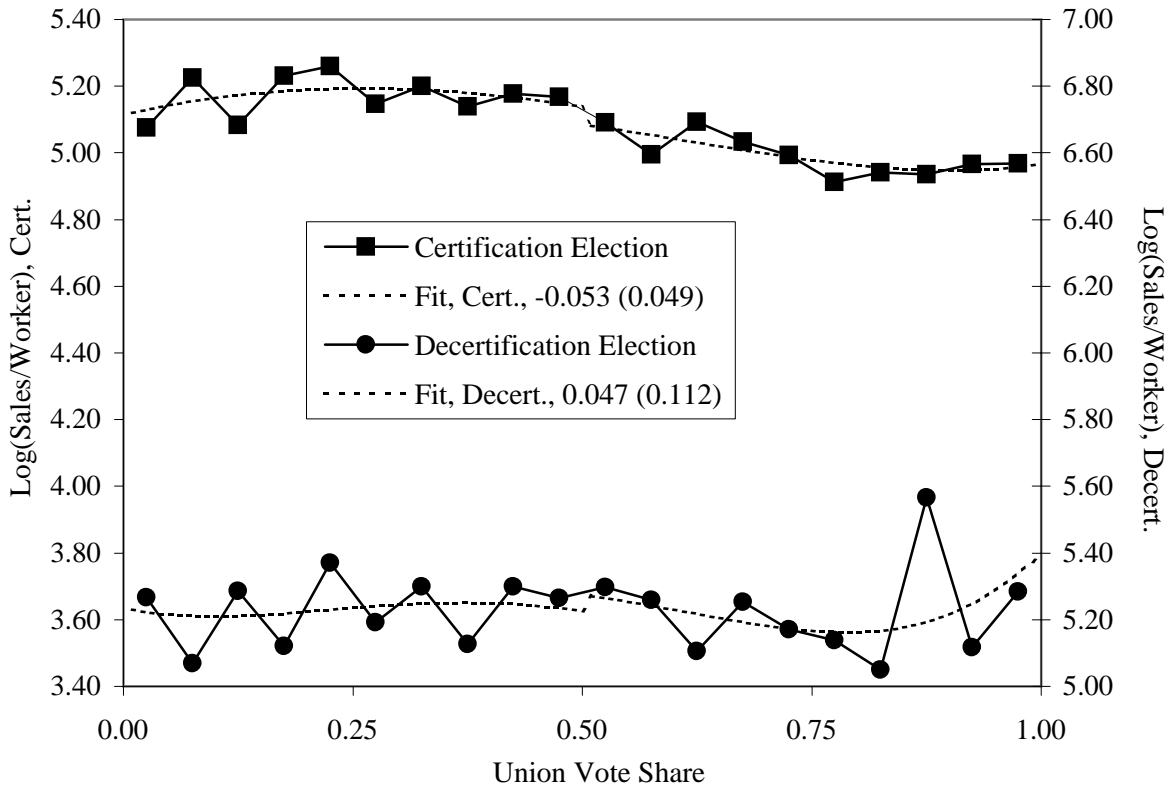
Note: 9792 and 1857 observations, respectively.

Figure 3A: Log(Sales), by Union Vote Share



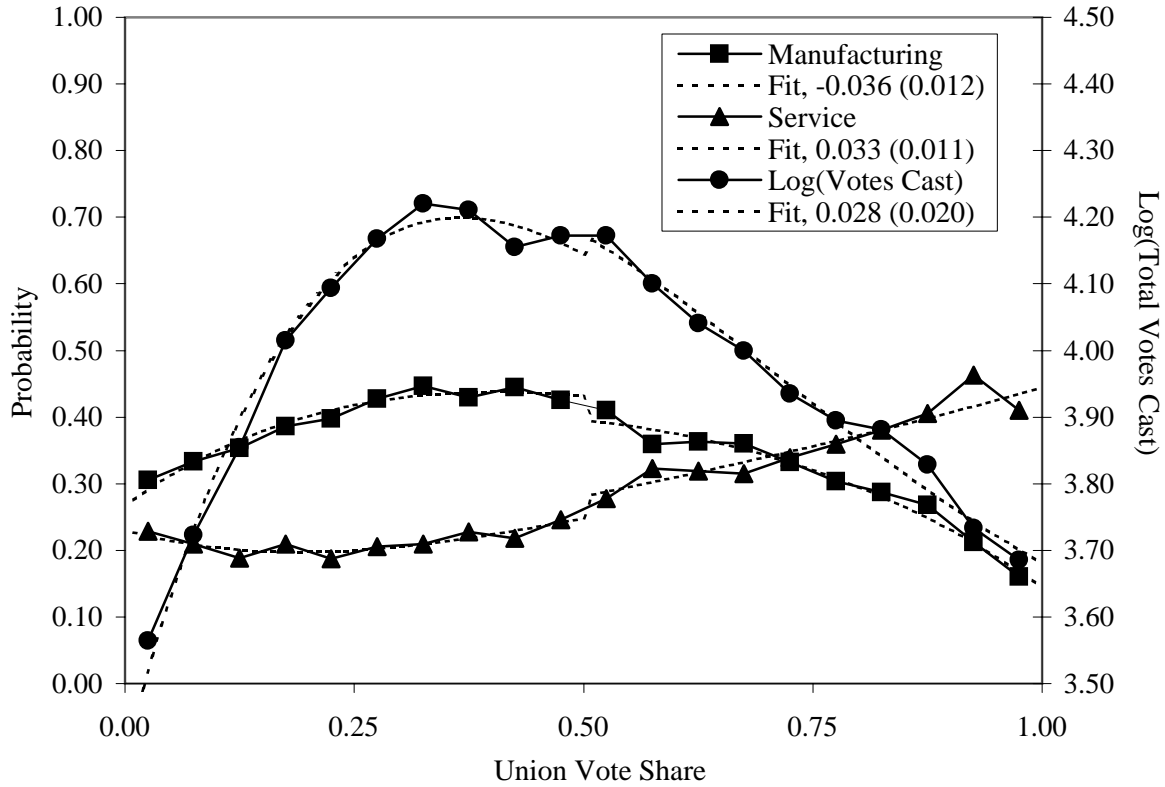
Note: 9125 and 1738 observations, respectively.

Figure 3B: Log(Sales/Worker), by Union Vote Share



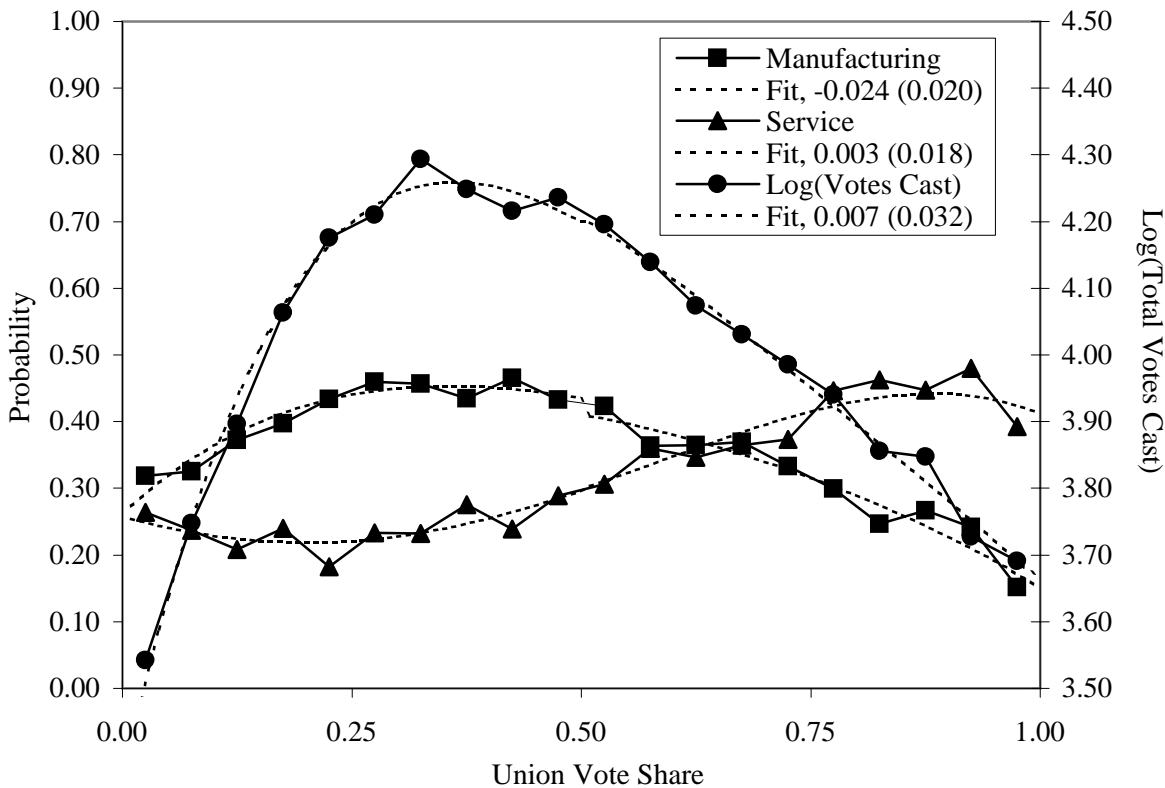
Note: 8634 and 1674 observations, respectively.

Figure 4A: Baseline Characteristics at Time of Election, by Union Vote Share



Note: 32198 observations

Figure 4B: Baseline Characteristics at Time of Election, Conditional on Survival at Year 2001, by Union Vote Share



Note: 13062 observations

Figure 5A: Survival at Year 1997, by Union Vote Share, LRD

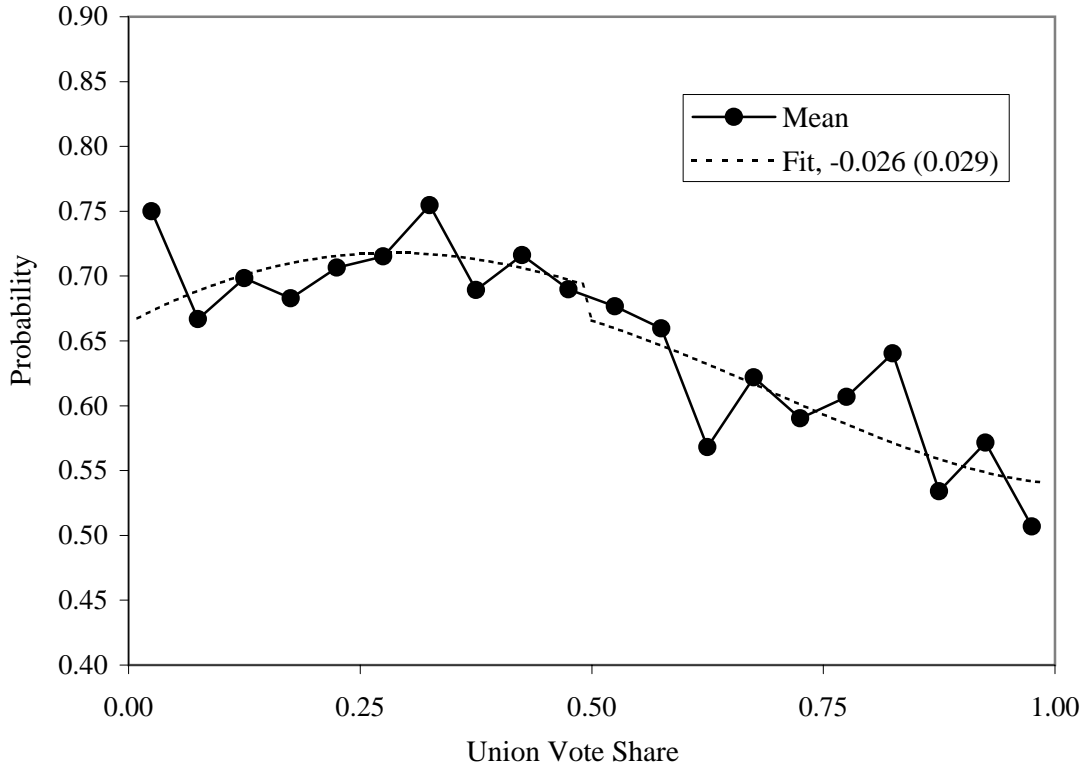
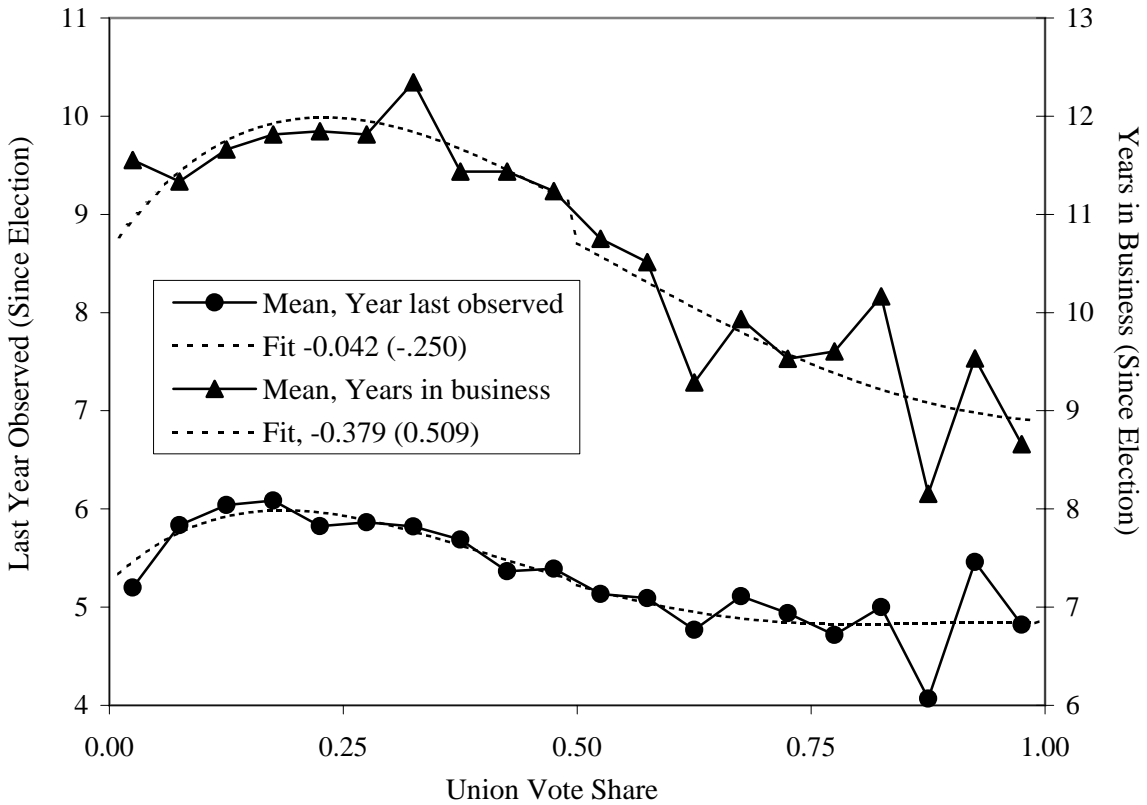


Figure 5B: Last Year Observed and Years in Business Since Election, by Union Vote Share, LRD



Note: 5608 Observations, 4816 Establishments

Figure 6A: Observation of a Contract Expiration Notice, Pre- and Post-Election, by Union Vote Share, LRD

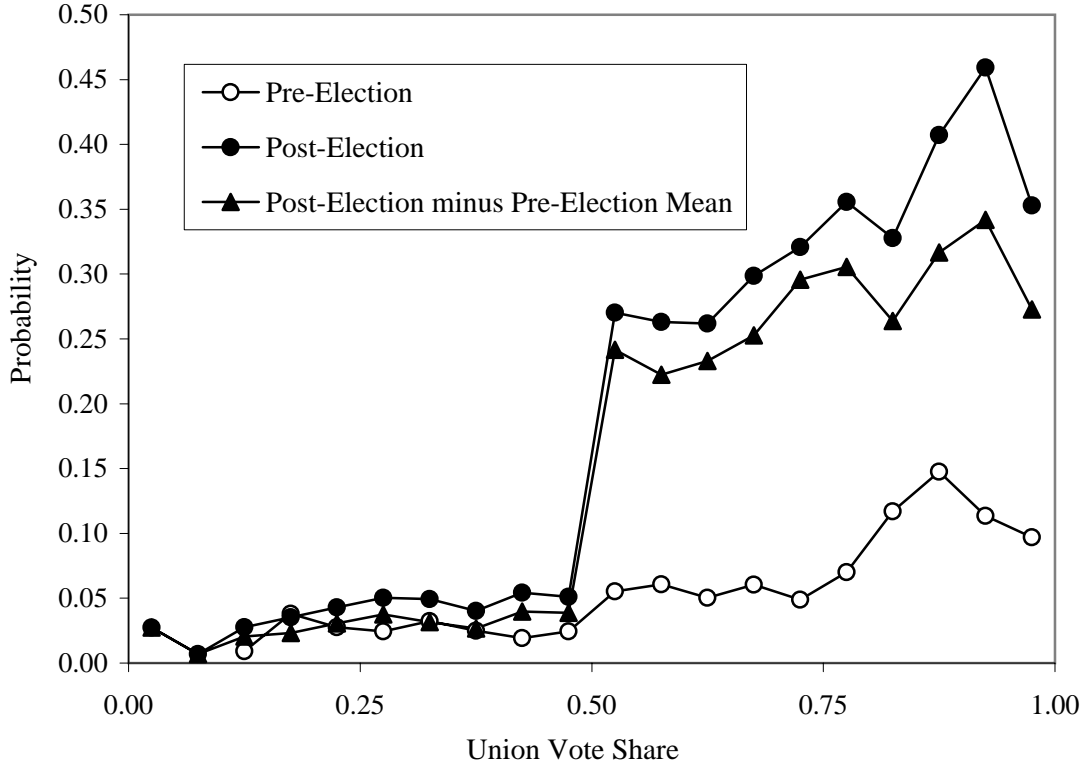
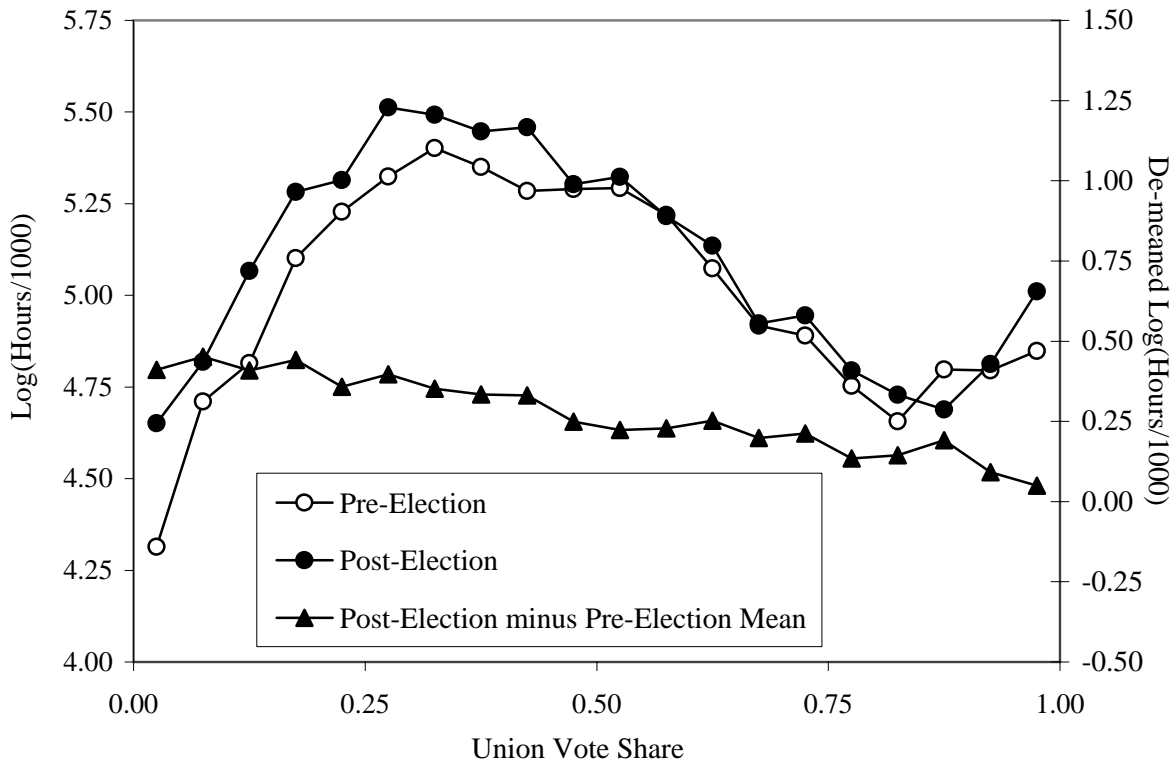


Figure 6B: Log(Production Hours/1000), Pre- and Post-Election, by Union Vote Share, LRD



Note: Observations: Pre-Election 38870, Post-Election 28929, Post-Election minus Pre-Election Mean 28790

Figure 7A: Log(Output/1000), Pre- and Post-Election, by Union Vote Share, LRD

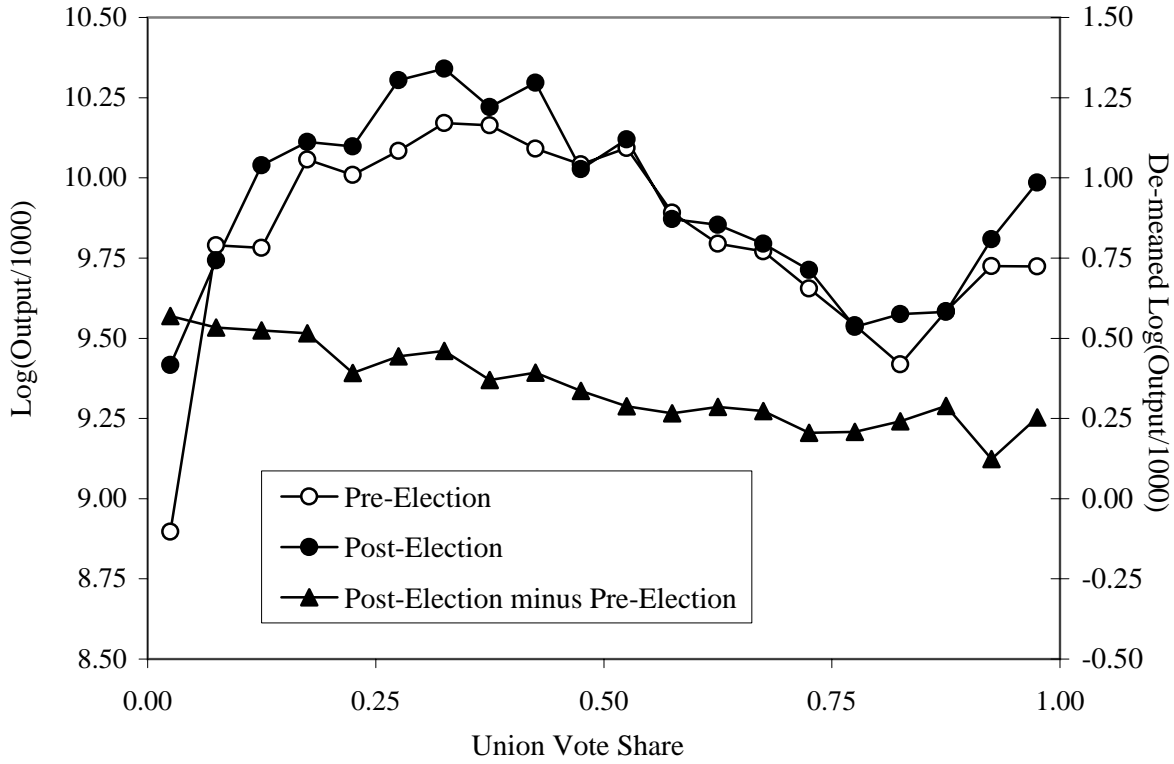
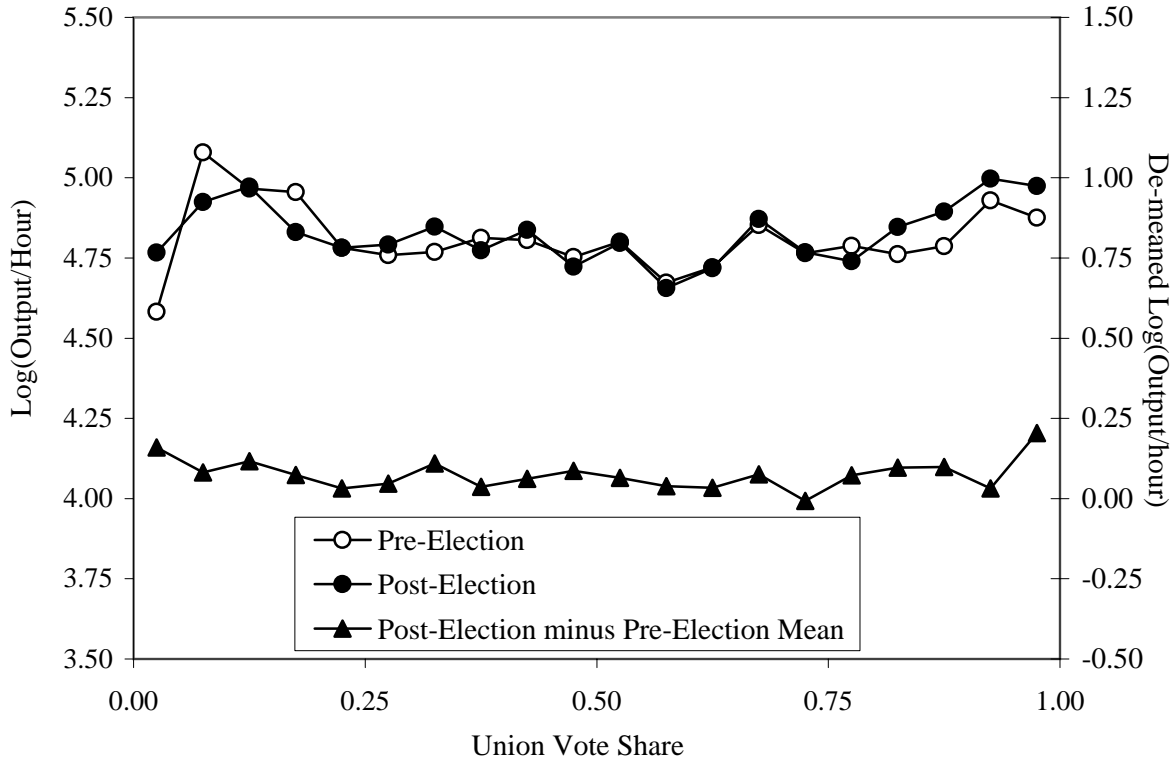


Figure 7B: Log(Output/Hour), Pre- and Post-Election, by Union Vote Share, LRD



Note: Observations: Pre-Election 38854, Post-Election 28918, Post-Election minus Pre-Election Mean 28785

Figure 8A: $\text{Log}((\text{Assets}/\text{worker})/1000)$, Pre- and Post-Election, by Union Vote Share, LRD

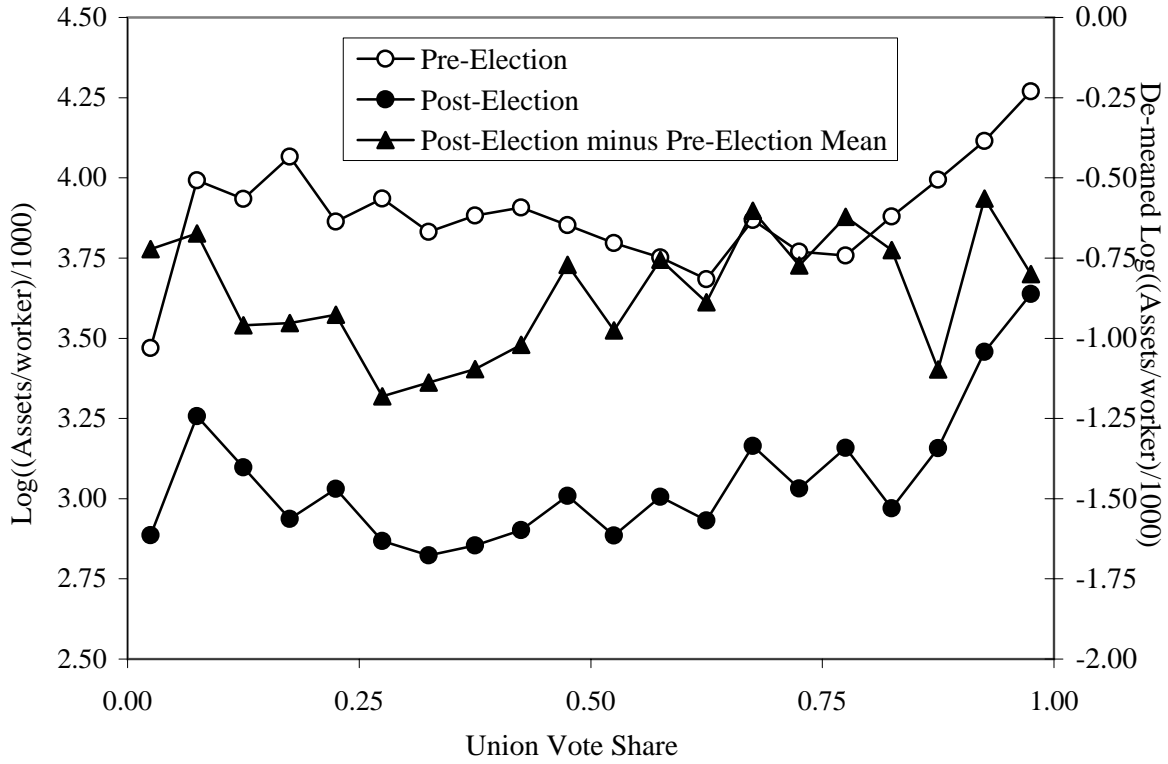
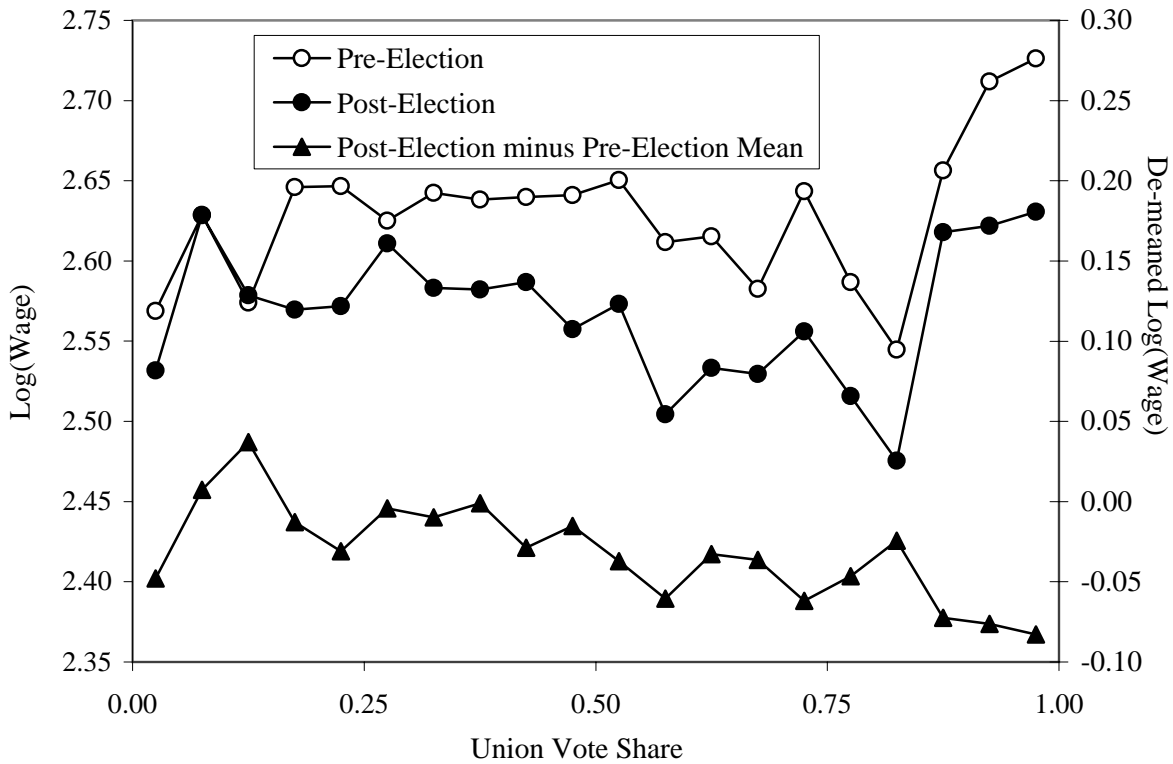


Figure 8B: $\text{Log}(\text{Production Hourly Wage})$, Pre- and Post-Election, by Union Vote Share, LRD



Note: Observations: Panel A, Pre-Election 37005, Post-Election 20505, Post-Election minus Pre-Election Mean 20346
 Panel B, Pre-Election 38870, Post-Election 28929, Post-Election minus Pre-Election Mean 28790

Figure 9A: Event Study Estimates of Election Impact on Log(Wage), Union Loses

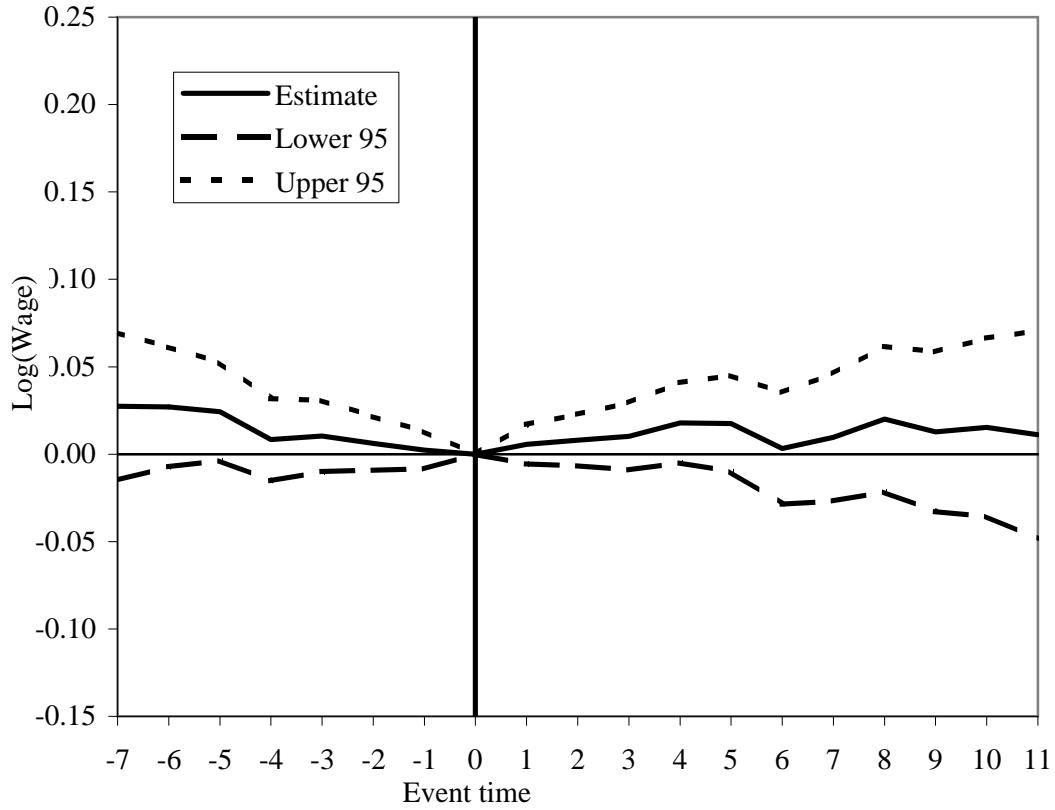
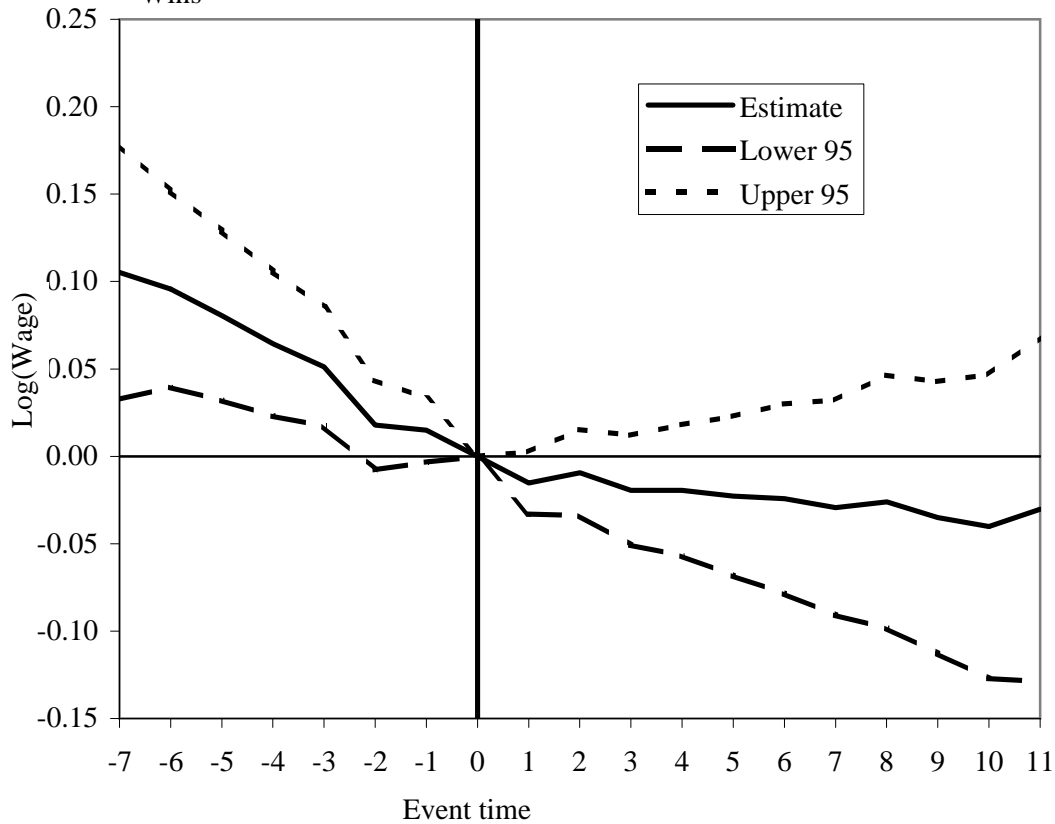


Figure 9B: Event Study Estimates of Election Impact on Log(Wage), Union Wins



Note: Panel A, 32538 Obs, 3584 Establishments; Panel B, 14899 Obs, 1891 Establishments

Table 1: OLS and Regression-Discontinuity Estimates, Impact of Union Recognition on Business Survival

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Union Won | -0.089 (0.013) | -0.092 (0.013) | -0.078 (0.013) | -0.073 (0.013) | -0.046 (0.022) | -0.041 (0.023) | -0.027 (0.029) | -0.026 (0.029) | -0.021 (0.028) | -0.025 (0.028) | -0.021 (0.028) |
| Vote Share | -- | -- | -- | -- | -0.130 (0.055) | 0.182 (0.134) | 0.499 (0.398) | 0.379 (0.732) | 0.552 (0.720) | 0.120 (0.721) | -0.062 (0.730) |
| (Vote Share) ² | -- | -- | -- | -- | -- | -0.322 (0.130) | -1.110 (0.946) | -0.630 (2.670) | -1.405 (2.618) | -0.533 (2.615) | 0.143 (2.638) |
| (Vote Share) ³ | -- | -- | -- | -- | -- | -- | 0.516 (0.617) | -0.210 (3.886) | 0.862 (3.804) | 0.415 (3.797) | -0.551 (3.818) |
| (Vote Share) ⁴ | -- | -- | -- | -- | -- | -- | -- | 0.365 (1.947) | -0.106 (1.905) | -0.133 (1.902) | 0.331 (1.907) |
| Log(Votes Cast) | -- | -- | 0.061 (0.008) | 0.065 (0.008) | -- | -- | -- | -- | -- | 0.059 (0.008) | 0.064 (0.008) |
| Year dummies? | No | Yes | Yes | Yes | No | No | No | No | Yes | Yes | Yes |
| Industry dummies? | No | No | No | Yes | No | No | No | No | No | No | Yes |
| R-Squared | 0.0084 | 0.0614 | 0.0729 | 0.0893 | 0.0094 | 0.0106 | 0.0107 | 0.0107 | 0.0637 | 0.0743 | 0.0907 |

Note: Sample size is 5608. Some elections match to same establishment, so standard errors are clustered on the establishment level (4816 independent establishments). Includes all elections that occurred before 1997 that has the same name and address as a pre-election-year entry in the LRD. Dependent variable is whether establishment appears in the 1997 Census of Manufactures. Least squares point estimates reported.

Table 2: Least-squares Regression-Discontinuity Estimates of Union Effects

| Dependent Variable | Coefficient on Won Election | | | | | | | | | |
|---------------------|-----------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Contract Expiration | 0.220 (0.021) [4733] | 0.252 (0.011) [28796] | 0.198 (0.016) [28796] | 0.191 (0.017) [28796] | 0.202 (0.020) [28796] | 0.198 (0.022) [28796] | 0.181 (0.020) [28796] | 0.182 (0.020) [28796] | 0.181 (0.020) [28796] | 0.179 (0.020) [28796] |
| Log(Hours) | 0.009 (0.087) [4733] | -0.318 (0.036) [28796] | -0.260 (0.063) [28796] | -0.203 (0.063) [28796] | 0.085 (0.080) [28796] | 0.097 (0.080) [28796] | -0.024 (0.056) [28796] | 0.015 (0.051) [28796] | 0.018 (0.050) [28796] | 0.028 (0.049) [28796] |
| Log(Output) | 0.079 (0.094) [4730] | -0.347 (0.042) [28785] | -0.293 (0.072) [28785] | -0.254 (0.073) [28785] | 0.067 (0.090) [28785] | 0.080 (0.091) [28785] | -0.043 (0.055) [28785] | -0.010 (0.050) [28785] | -0.004 (0.050) [28785] | 0.011 (0.049) [28785] |
| Log(Output/worker) | 0.072 (0.063) [4730] | -0.028 (0.029) [28785] | -0.032 (0.048) [28785] | -0.051 (0.048) [28785] | -0.018 (0.060) [28785] | -0.016 (0.061) [28785] | -0.019 (0.035) [28785] | -0.019 (0.034) [28785] | -0.018 (0.034) [28785] | -0.015 (0.034) [28785] |
| Log(Assets/worker) | -0.121 (0.108) [3379] | 0.122 (0.049) [20346] | 0.020 (0.082) [20346] | -0.020 (0.082) [20346] | -0.059 (0.102) [20346] | -0.048 (0.103) [20346] | -0.136 (0.104) [20346] | -0.090 (0.093) [20346] | -0.064 (0.075) [20346] | -0.029 (0.072) [20346] |
| Log(Wage) | 0.015 (0.025) [4733] | -0.039 (0.011) [28796] | -0.041 (0.019) [28796] | -0.044 (0.020) [28796] | -0.005 (0.024) [28796] | -0.002 (0.024) [28796] | -0.026 (0.017) [28796] | -0.018 (0.016) [28796] | -0.018 (0.016) [28796] | -0.016 (0.015) [28796] |
| Sample | +/- 5% | All | All | All | All | All | All | All | All | All |
| Polynomial Terms | 0 | 0 | 1 | 2 | 3 | 4 | 4 | 4 | 4 | 4 |
| Dependent Variable | Level | Level | Level | Level | Level | Level | De-meaned | De-meaned | De-meaned | De-meaned |
| Include Base Mean? | No | No | No | No | No | No | No | Yes | Yes | Yes |
| Year dummies | No | No | No | No | No | No | No | No | Yes | Yes |
| Industry dummies | No | No | No | No | No | No | No | No | No | Yes |

Note: Within-election clustered standard errors in parentheses. Number of observations in brackets. Each entry is the estimated coefficient on the "Union won" indicator in a least squares regression. "Base Mean" is the average of the dependent variable for years strictly before the election year. "De-meaned" denotes that the dependent variable is the outcome minus the "base mean". "+/- 5%" sample are elections where the union vote share is between 45 and 55 percent.

Table 3: Least-squares Regression-Discontinuity Estimates of Union Effects, by Time after Election

| Dependent Variable | (1) | (2) | (3) |
|---------------------|------------------------------|-----------------------------|-----------------------------|
| Contract Expiration | 0.077 (0.018) [13240] | 0.307 (0.032) [8745] | 0.214 (0.041) [6811] |
| Log(Hours) | 0.030 (0.042) [13240] | 0.032 (0.066) [8745] | 0.041 (0.094) [6811] |
| Log(Output) | 0.001 (0.044) [13235] | 0.030 (0.064) [8742] | 0.040 (0.092) [6808] |
| Log(Output/worker) | -0.027 (0.031) [13235] | -0.004 (0.046) [8742] | 0.008 (0.061) [6808] |
| Log(Assets/worker) | -0.043 (0.070) [10332] | -0.035 (0.114) [6167] | 0.029 (0.133) [3847] |
| Log(Wage) | -0.013 (0.015) [13240] | -0.025 (0.021) [8745] | -0.005 (0.028) [6811] |
| Sample | 0-3 Years post-El. | 4-7 Years post-El. | 8+ Years post-El. |

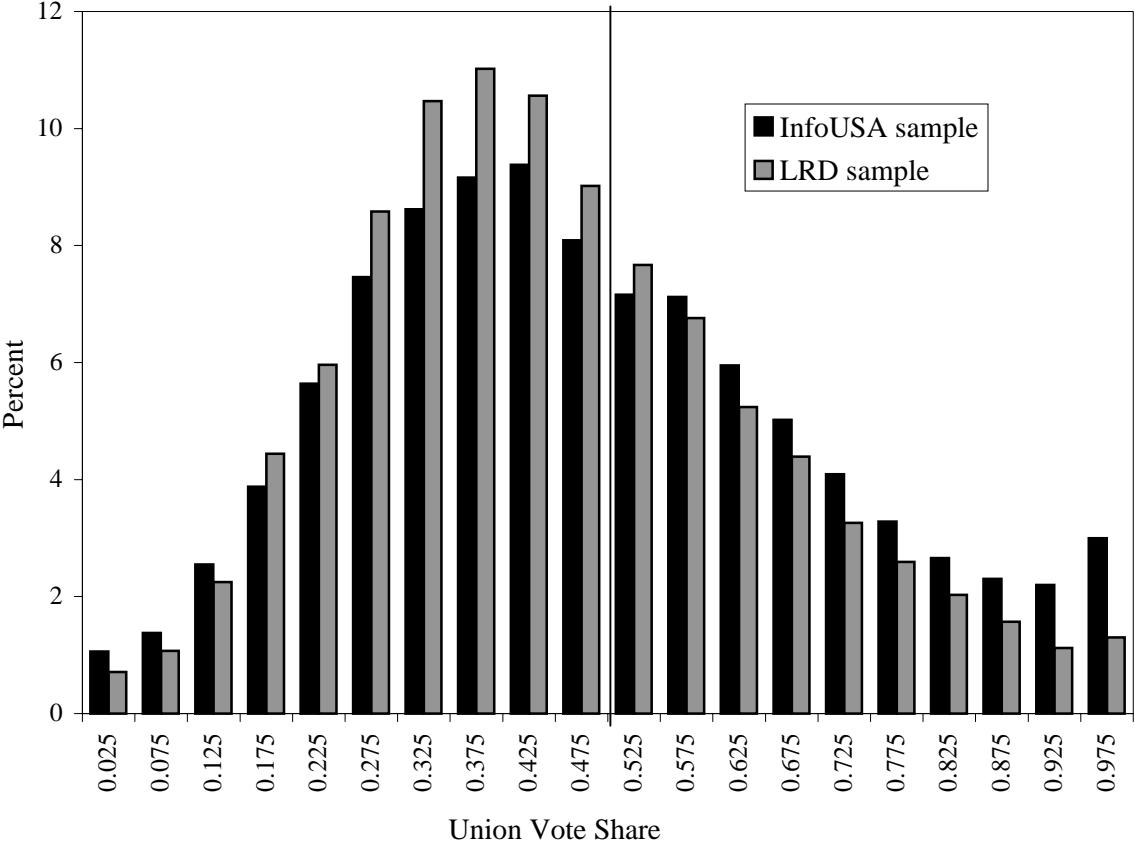
Note: Within-election clustered standard errors in parentheses. Number of observations in brackets. Each entry is the estimated coefficient on the "Union won" indicator in a least squares regression. Specification is the same as Col. (10) in Table 2. Each column restricts the sample to the observations that are within the first three years, between four to seven years, and eight years or later, relative to the election year.

Table 4: Least-squares Regression-Discontinuity Estimates of Union Effects, by Sub-Sample

| Dependent Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---------------------|------------------------------|------------------------------|------------------------------|-----------------------------|------------------------------------|------------------------------|------------------------------|
| Contract Expiration | 0.179 (0.020) [28796] | 0.139 (0.028) [11500] | 0.205 (0.028) [17296] | 0.196 (0.037) [6609] | 0.171 (0.024) [22187] | 0.166 (0.028) [14620] | 0.191 (0.028) [14176] |
| Log(Hours) | 0.028 (0.049) [28796] | 0.027 (0.063) [11500] | 0.056 (0.064) [17296] | 0.129 (0.109) [6609] | 0.000 (0.054) [22187] | 0.024 (0.073) [14620] | 0.028 (0.066) [14176] |
| Log(Output) | 0.011 (0.049) [28785] | -0.050 (0.070) [11495] | 0.077 (0.063) [17290] | 0.040 (0.117) [6608] | 0.000 (0.053) [22177] | 0.036 (0.074) [14614] | -0.018 (0.065) [14171] |
| Log(Output/worker) | -0.015 (0.034) [28785] | -0.068 (0.048) [11495] | 0.027 (0.045) [17290] | -0.079 (0.069) [6608] | 0.001 (0.039) [22177] | -0.003 (0.047) [14614] | -0.033 (0.049) [14171] |
| Log(Assets/worker) | -0.029 (0.072) [20346] | 0.061 (0.093) [8090] | -0.072 (0.100) [12256] | -0.016 (0.141) [4769] | -0.051 (0.085) [15577] | 0.076 (0.107) [10054] | -0.129 (0.096) [10292] |
| Log(Wage) | -0.016 (0.015) [28796] | -0.021 (0.023) [11500] | -0.016 (0.020) [17296] | -0.035 (0.031) [6609] | -0.011 (0.017) [22187] | -0.029 (0.022) [14620] | -0.002 (0.021) [14176] |
| Sample | All | <75 Prod. Workers | 75+ Prod. Workers | Exact match on Str. Add. | Non-Exact match on Str. Add. | "Low-wage" Industry | "High-wage" Industry |

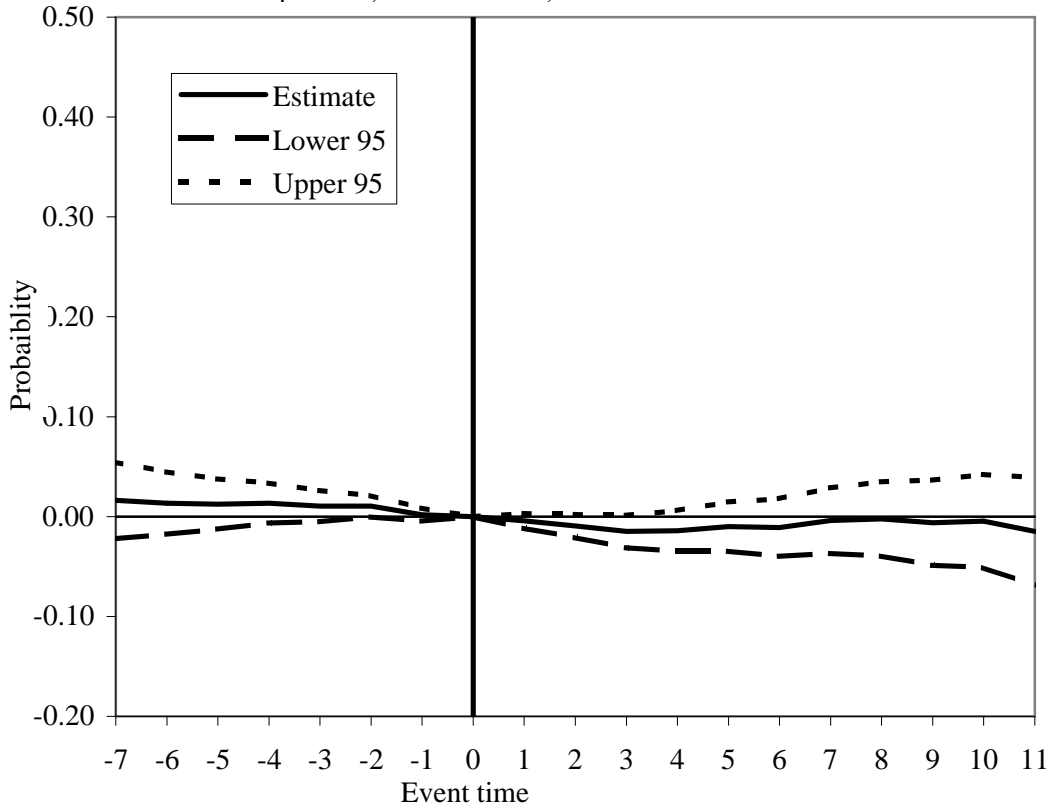
Note: Within-election clustered standard errors in parentheses. Number of observations in brackets. Each entry is the estimated coefficient on the "Union won" indicator in a least squares regression. Specification is the same as Col. (10) in Table 2. Each column uses a different sub sample. (2) and (3) split the sample by size of the voting unit; (4) and (5) by whether there is an exact string match between the street address entry from the NLRB and LRD data sets; and (6) and (7) by "high-wage" and "low-wage" industries.

Appendix Figure 1: Frequency Distribution of Union Vote Share, All Certification Elections, InfoUSA and LRD samples

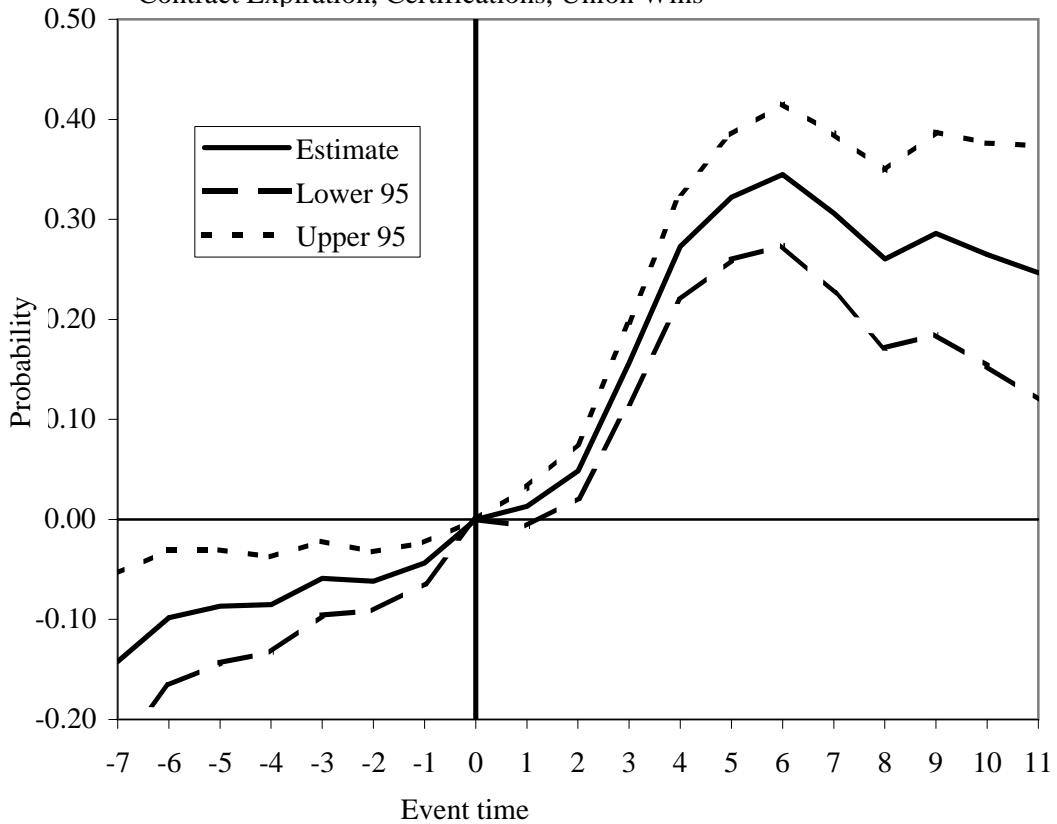


Note: InfoUSA sample: 27560 observations, LRD sample; 5608 observations

Appendix Figure 2A: Event Study Estimates of Election Impact on Contract Expiration, Certifications, Union Loses

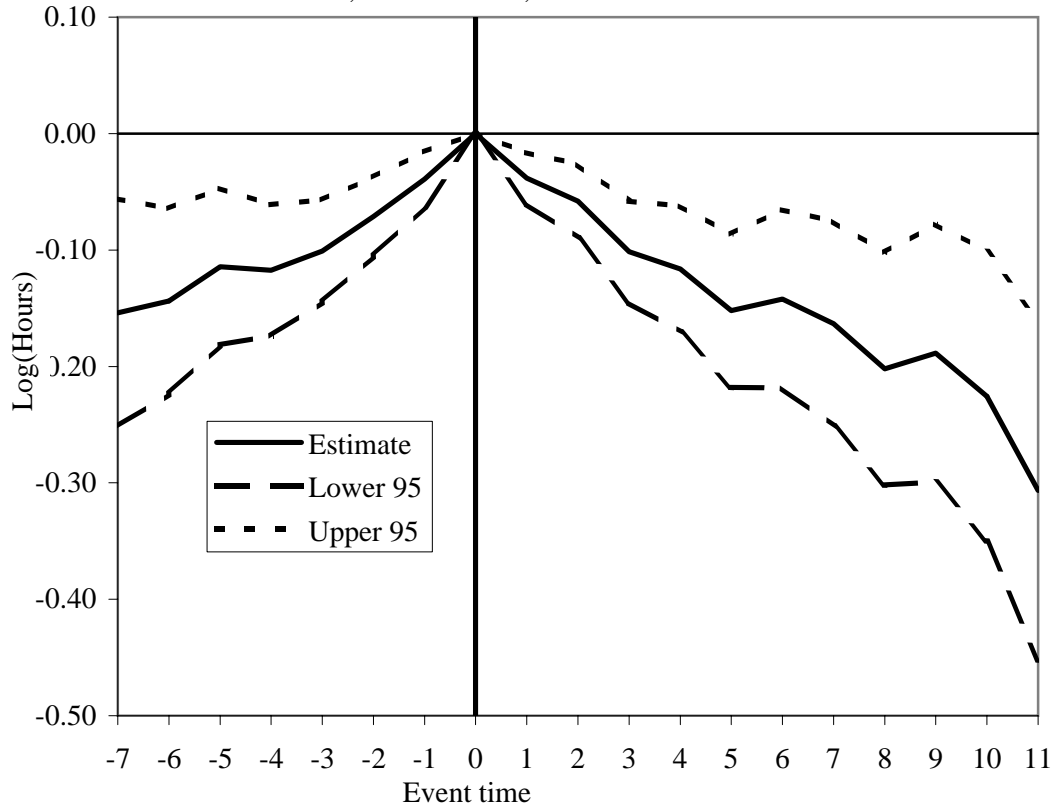


Appendix Figure 2B: Event Study Estimates of Election Impact on Contract Expiration, Certifications, Union Wins

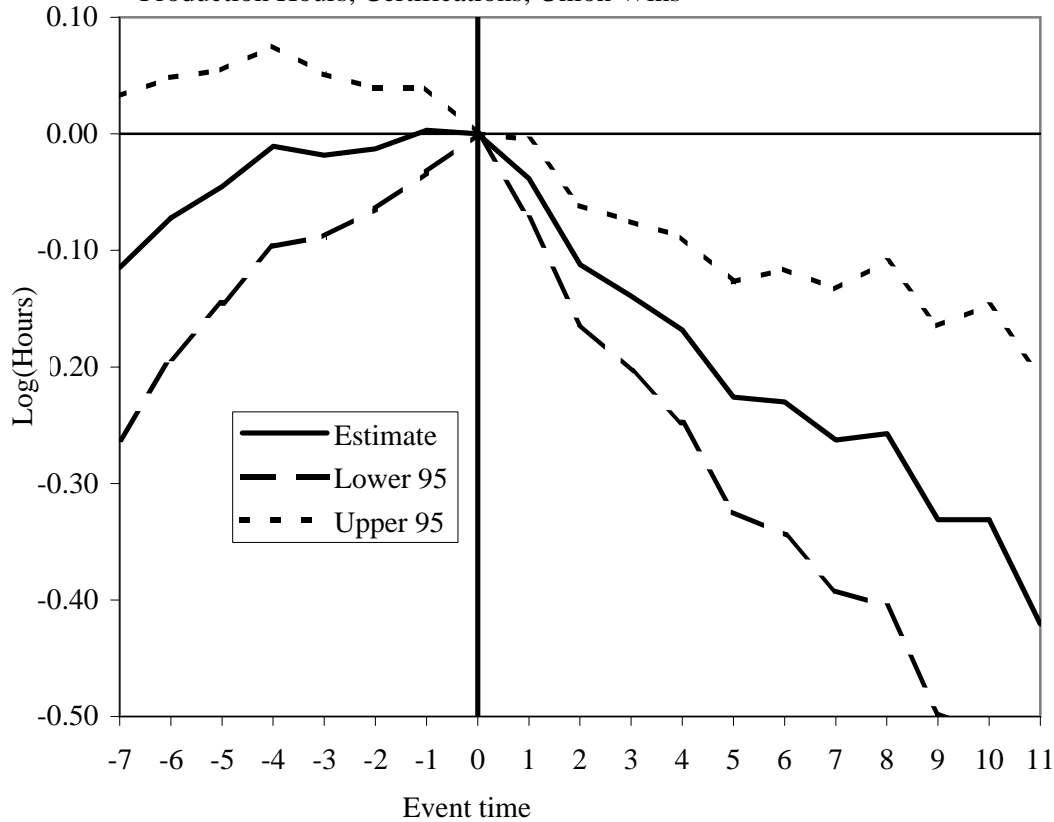


Note: Panel A, 32538 Obs, 3584 Establishments; Panel B, 14899 Obs, 1891 Establishments

Appendix Figure 3A: Event Study Estimates of Election Impact on Production Hours, Certifications, Union Loses

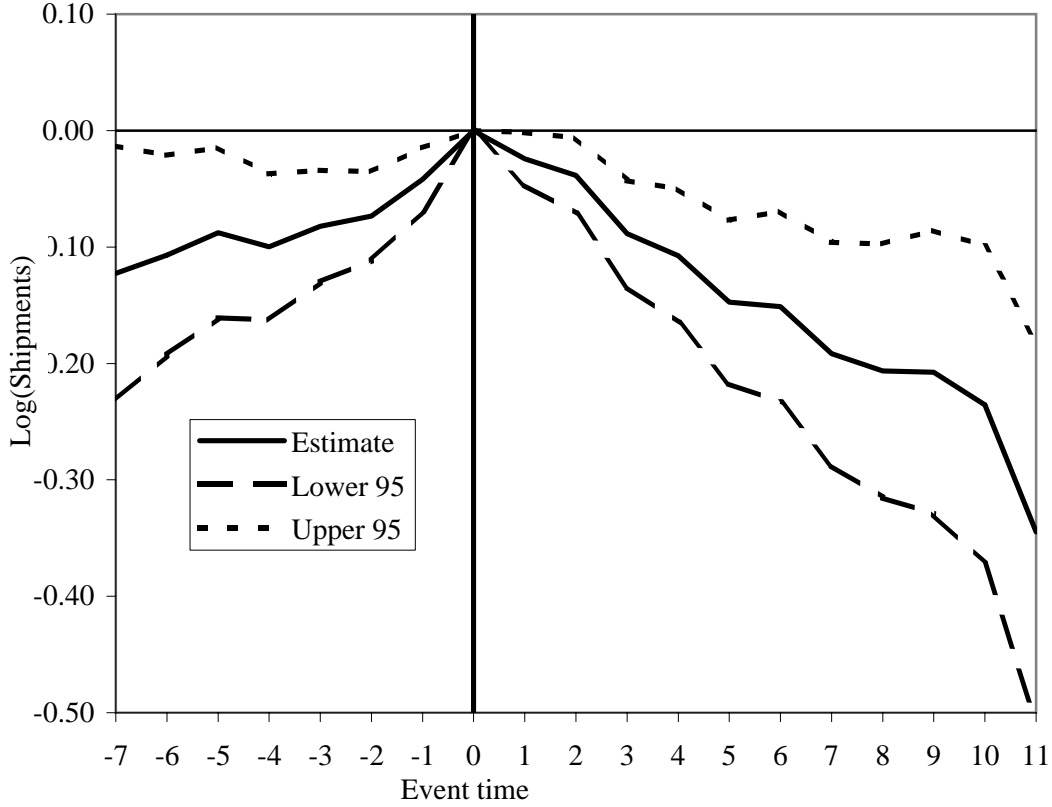


Appendix Figure 3B: Event Study Estimates of Election Impact on Production Hours, Certifications, Union Wins

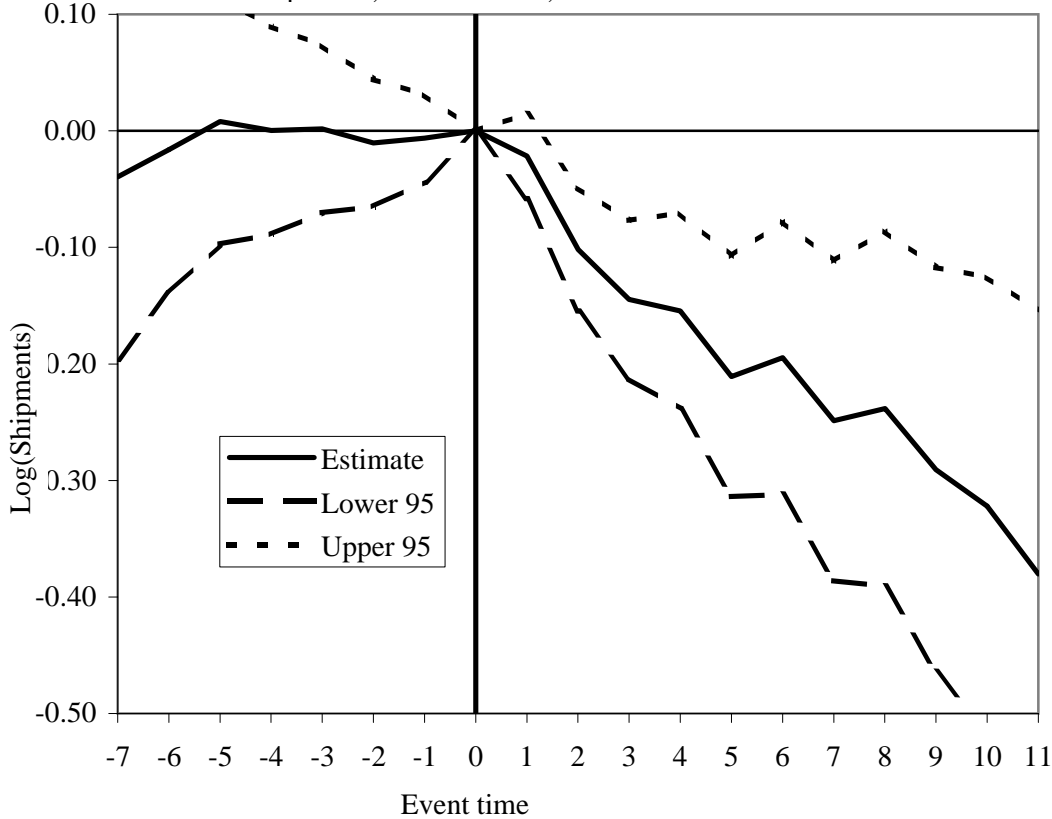


Note: Panel A, 32538 Obs, 3584 Establishments; Panel B, 14899 Obs, 1891 Establishments

Appendix Figure 4A: Event Study Estimates of Election Impact on Total Value of Shipments, Certifications, Union Loses

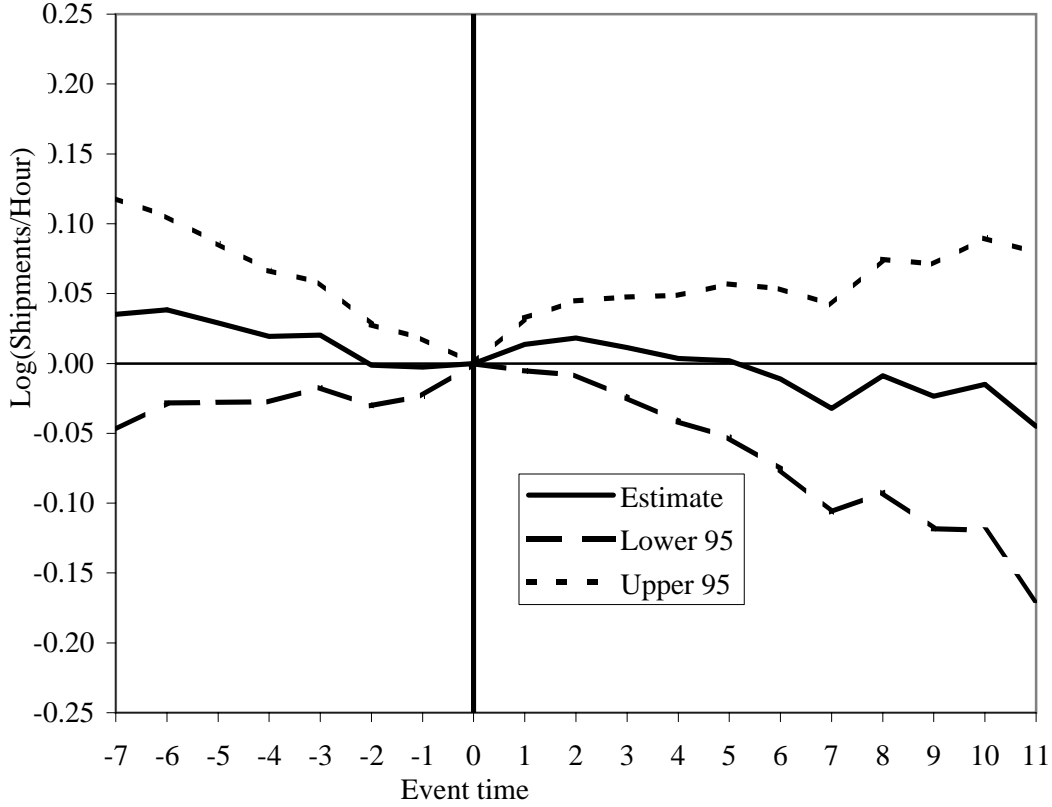


Appendix Figure 4B: Event Study Estimates of Election Impact on Total Value of Shipments, Certifications, Union Wins

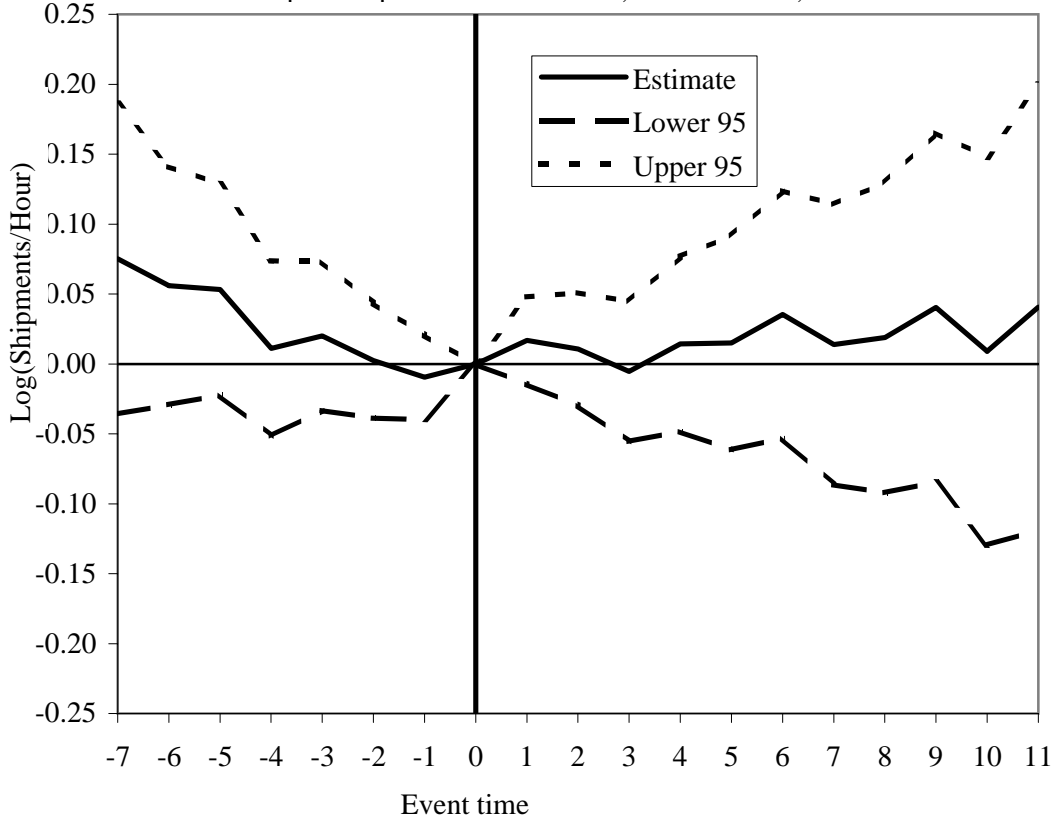


Note: Panel A, 32523 Obs, 3584 Establishments; Panel B, 14897 Obs, 1891 Establishments

Appendix Figure 5A: Event Study Estimates of Election Impact on Total Value of Shipments per Production Hour, Certifications, Union Loses

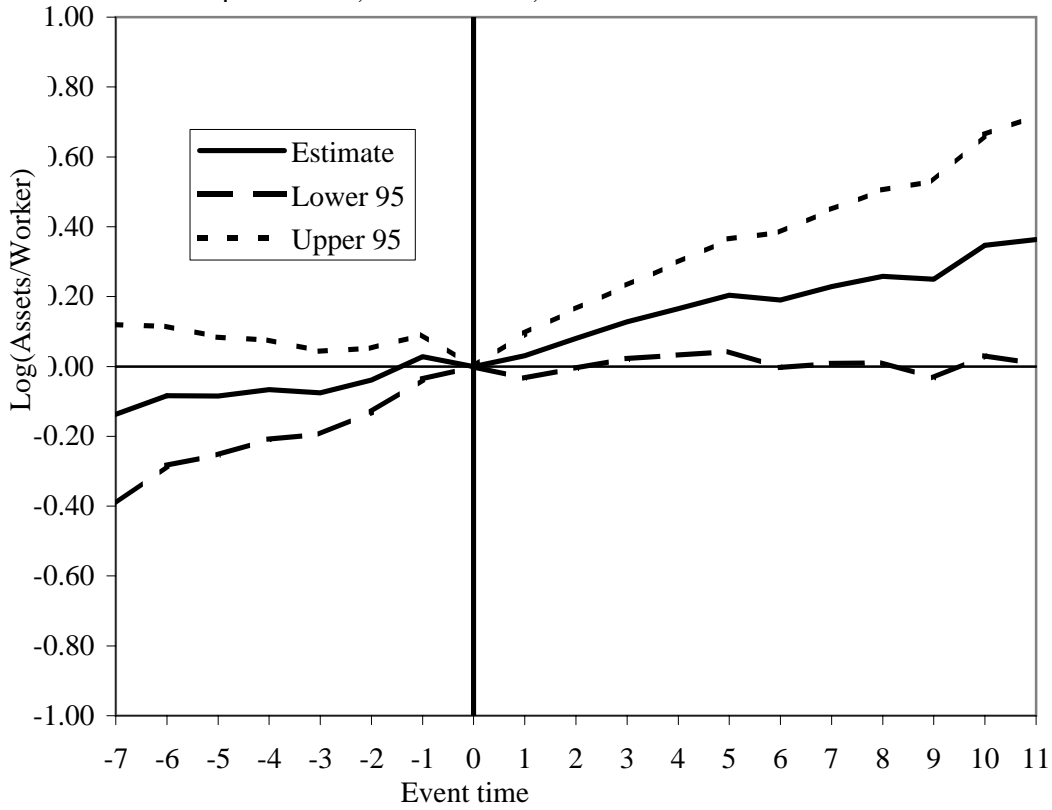


Appendix Figure 5B: Event Study Estimates of Election Impact on Total Value of Shipments per Production Hour, Certifications, Union Wins

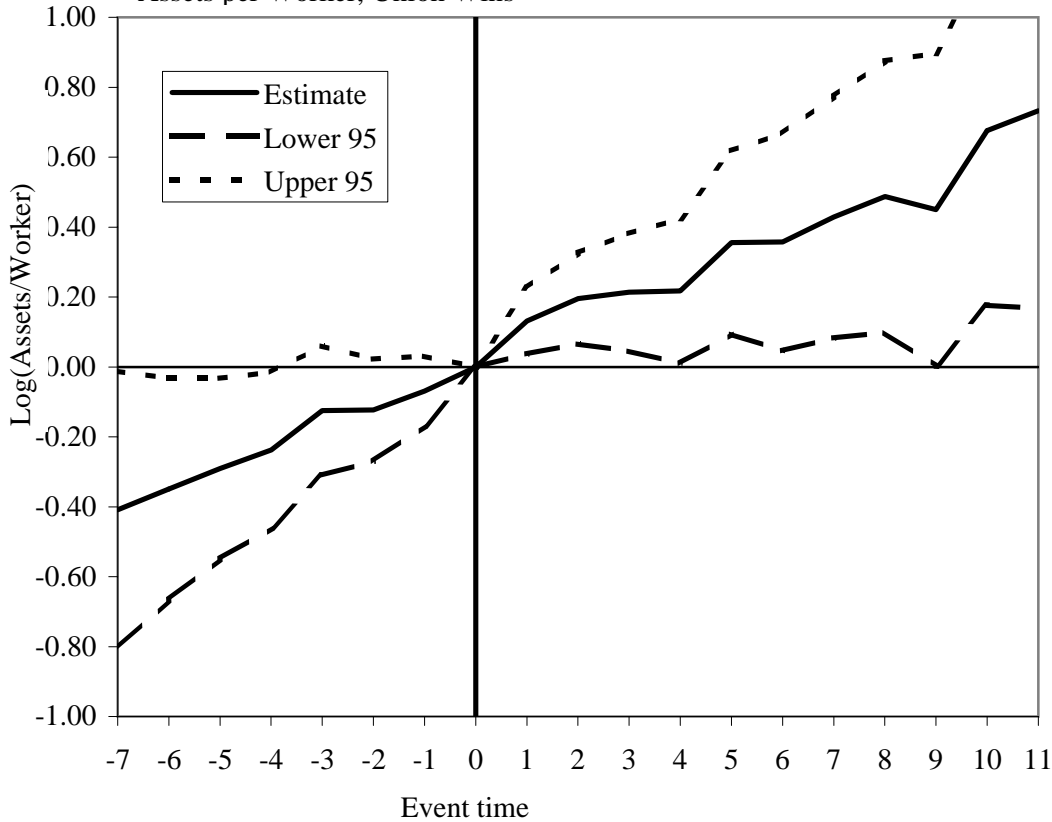


Note: Panel A, 32523 Obs, 3584 Establishments; Panel B, 14897 Obs, 1891 Establishments

Appendix Figure 6A: Event Study Estimates of Election Impact on Total Assets per Worker, Certifications, Union Loses

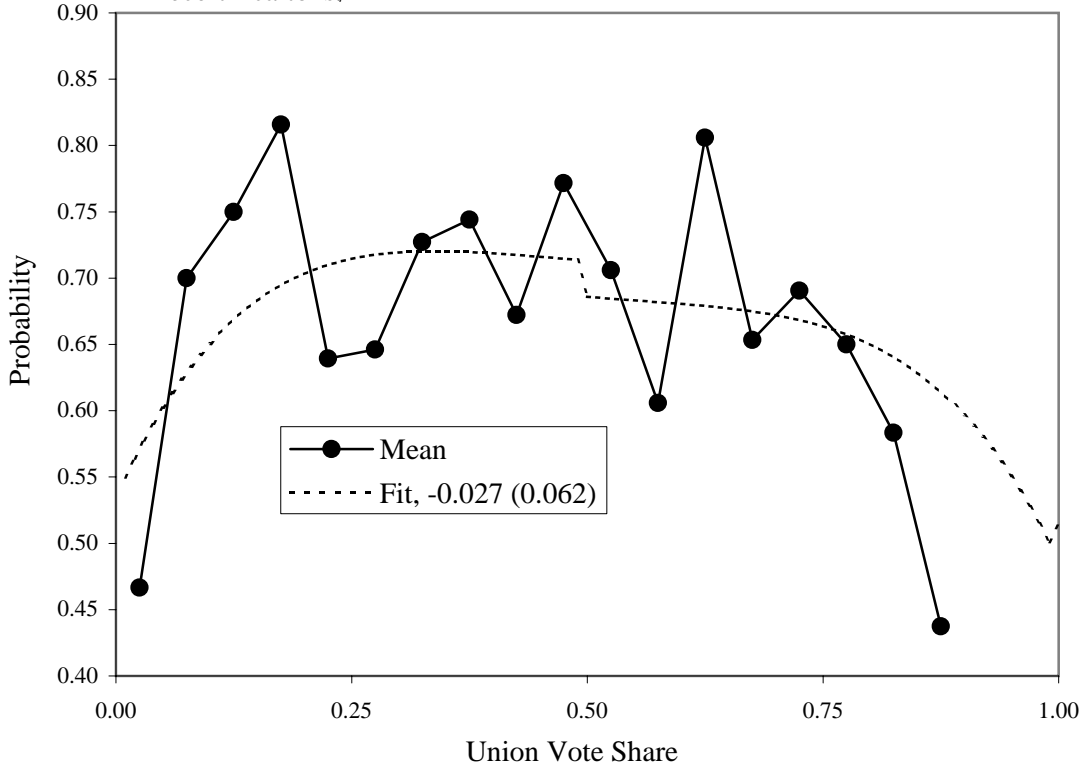


Appendix Figure 6B: Event Study Estimates of Election Impact on Total Assets per Worker, Union Wins

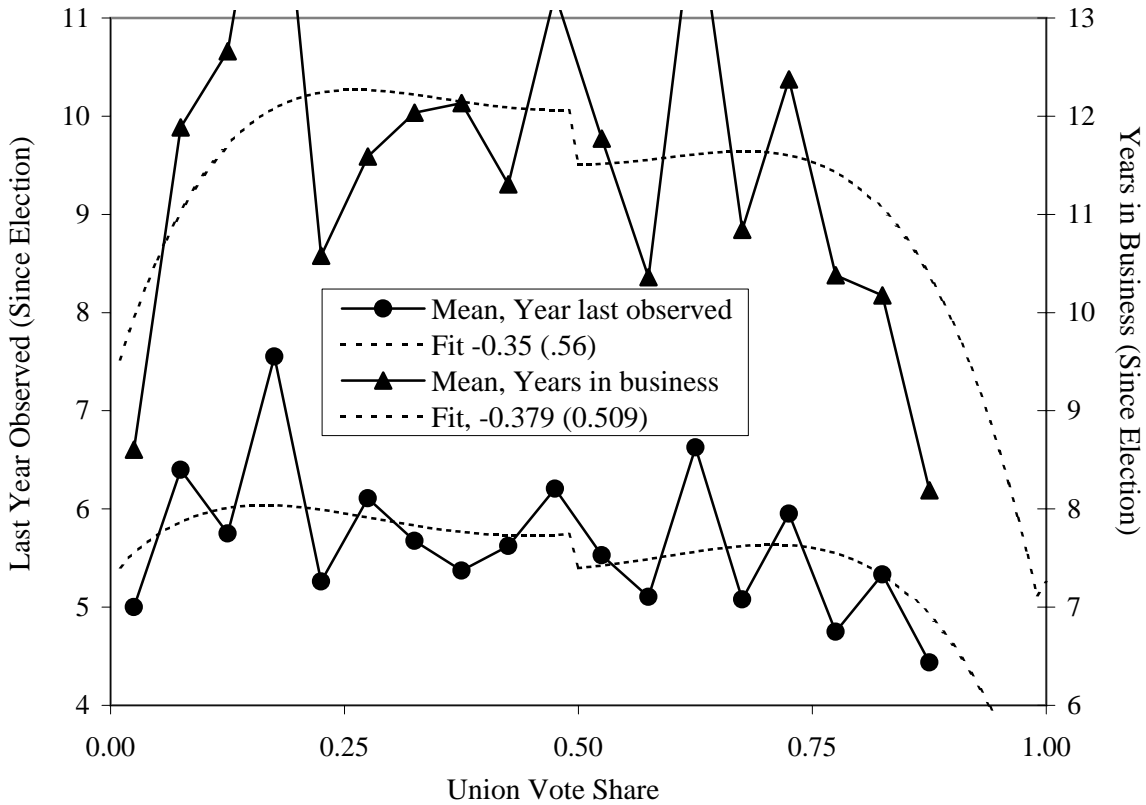


Note: Panel A, 25555 Obs, 3580 Establishments; Panel B, 11668 Obs, 1887 Establishments

Appendix Figure 7A: Survival at Year 1997, by Union Vote Share, Decertifications, LRD

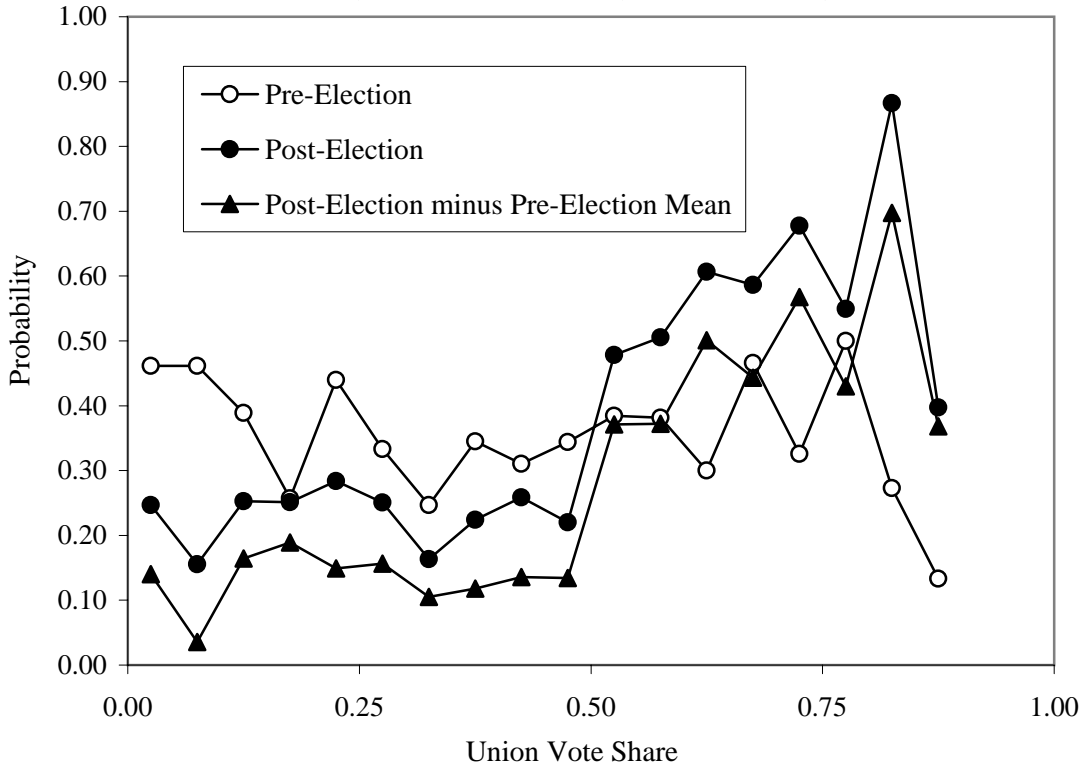


Appendix Figure 7B: Last Year Observed and Years in Business Since Election, by Union Vote Share, De-certifications LRD

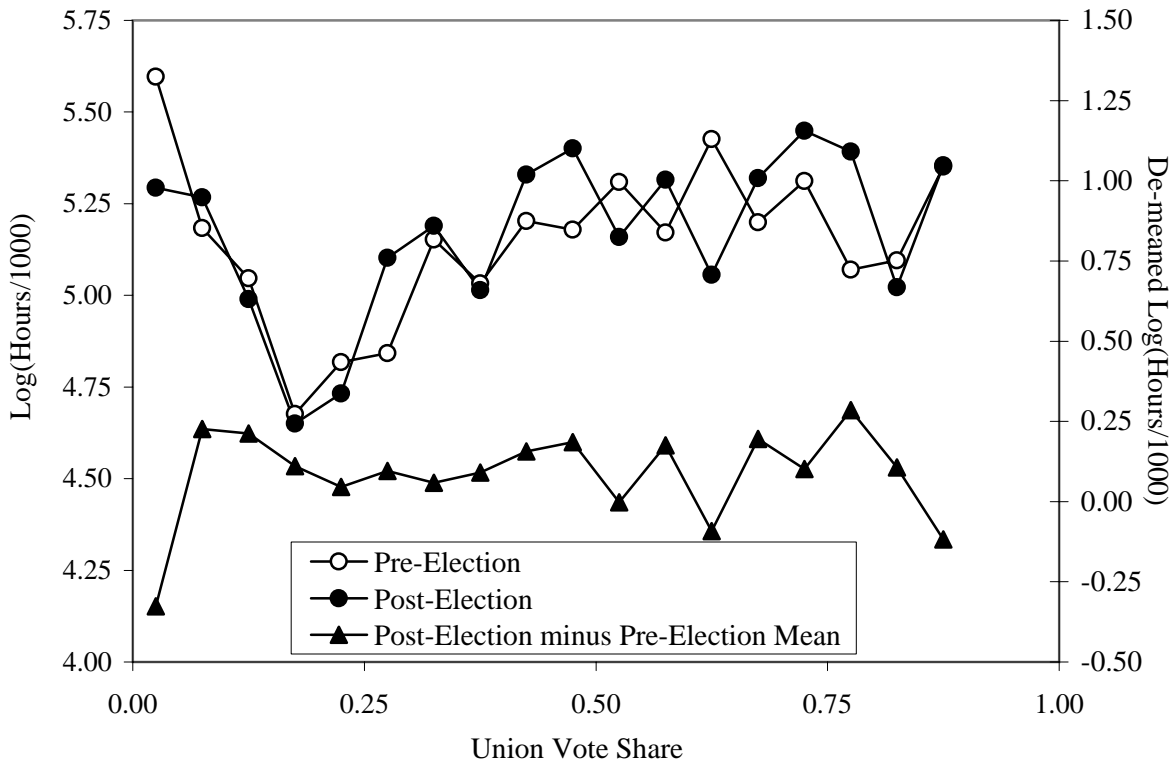


Note: 1085 Observations, 998 Establishments

Appendix Figure 8A: Observation of a Contract Expiration Notice, Pre- and Post-Election, by Union Vote Share, Decertifications, LRD

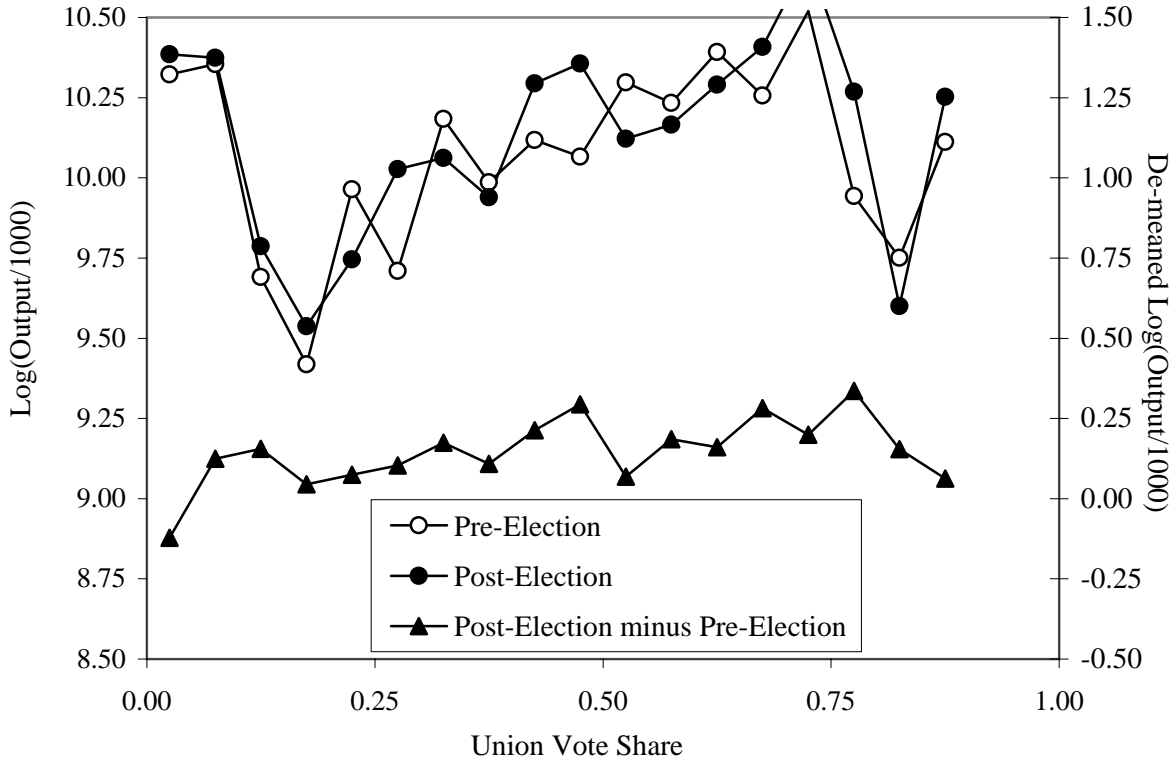


Appendix Figure 8B: Log(Production Hours/1000), Pre- and Post-Election, by Union Vote Share, Decertifications, LRD

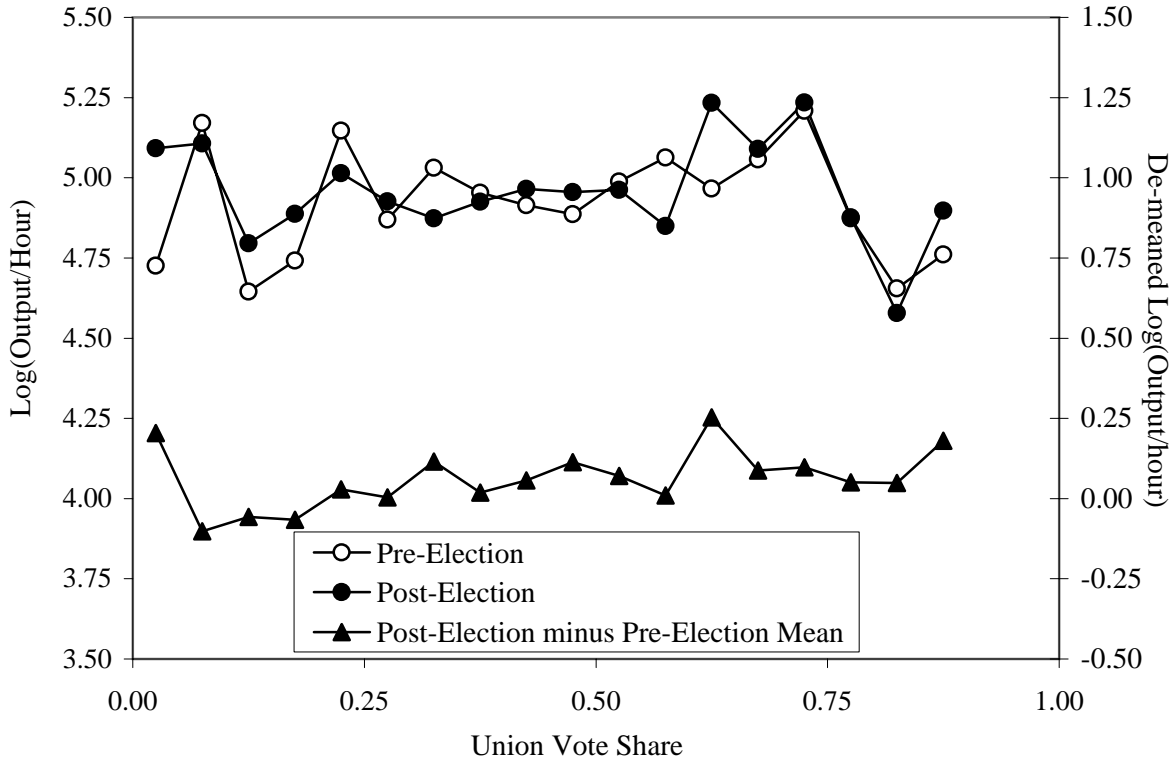


Note: Observations: Panel A, Pre-Election 1043, Post-Election 5957, Post-Election minus Pre-Election Mean 5954
 Panel B, Pre-Election 9188, Post-Election 5957, Post-Election minus Pre-Election Mean 5954

Appendix Figure 9A: Log(Output/1000), Pre- and Post-Election, by Union Vote Share, Decertifications, LRD



Appendix Figure 9B: Log(Output/Hour), Pre- and Post-Election, by Union Vote Share, Decertifications, LRD



Note: Observations: Pre-Election 9185, Post-Election 5954, Post-Election minus Pre-Election Mean 5951

Appendix Table 1: Description of Data Sets

| Data Set | Description | Main Variables |
|-------------------|---|--|
| NLRB | National Labor Relations Board Election Database: Contains universe of representation elections (certification and decertification) that occurred in the U.S. between 1984 and 1999. Total observations: 62415. | Date of election, # eligible voters, # votes cast, # votes for the union |
| FMCS | Federal Mediation and Conciliation Service Database: Contains all contract expiration notices filed with the FMCS between 1984 and 2001. Total Observations: 563565. | Date of Contract Expiration |
| InfoUSA | InfoUSA Database: Contains all business establishments that appear in Phone Directories, as of May, 2001. | Employment, Sales Volume |
| LRD | Longitudinal Research Database: Comprised of two components. 1) Census of Manufactures (1977, 1982, 1987, 1992, 1997): Universe of all manufacturing establishments in the U.S., and 2) Annual Survey of Manufactures (1974-1999): Universe of all large manufacturing establishments plus probability sample of smaller manufacturers. | Wages, Hours, Output, Total Value of Shipments, Assets |
| NLRB/FMCS/InfoUSA | Total number of elections (62415). 20+ votes cast (33503), <1000 votes cast (33382), certification and de-certification elections only (32855), valid election date (32712). Company name and establishment street address used to match to FMCS and InfoUSA data. 13062 of the 32712 were found in the InfoUSA database (as of May, 2001). | Variables from NLRB, FMCS, and InfoUSA |
| NLRB/FMCS/LRD | Company name and establishment street address used to match NLRB/FMCS data (32712 observations) (see above) to LRD. Kept only matches where the year from the LRD is strictly less than the year of the election. With multiple matches, single "best" match was kept. 8533 of the 32712 matched. Kept cases where the ratio of eligible voters (NLRB) to production workers (LRD) is at least half (7743 of 8533). | Variables from NLRB, FMCS, and LRD |

Appendix Table 2: Sample Means, InfoUSA/NLRB/FMCS and LRD/NLRB/FMCS data sets

| Variable | InfoUSA/NLRB/FMCS | | | | LRD/NLRB/FMCS | | | |
|-----------------------------------|-------------------|---------|----------|------------|---------------|---------|----------|------------|
| | Obs | Mean | Std. Dev | Std. Error | Obs | Mean | Std. Dev | Std. Error |
| Union Wins Election | 27560 | 0.429 | 0.495 | (0.003) | 5608 | 0.360 | 0.480 | (0.006) |
| Number of Votes Cast | 27560 | 88.6 | 107.6 | (0.6) | 5608 | 111.9 | 121.9 | (1.6) |
| Log(Votes Cast) | 27560 | 4.091 | 0.806 | (0.005) | 5608 | 4.327 | 0.840 | (0.011) |
| Total Employment | 16987 | 172.8 | 353.0 | (2.7) | 5608 | 149.6 | 192.0 | (2.6) |
| Log(Employment) | 16987 | 4.057 | 1.645 | (0.013) | 5608 | 4.412 | 1.197 | (0.016) |
| Total Value of Shipments (\$1000) | 16200 | 32192.3 | 72308.4 | (568.1) | 5608 | 32217.9 | 71991.9 | (961.3) |
| Log(Total Value of Shipments) | 16200 | 8.984 | 1.832 | (0.014) | 5605 | 9.434 | 1.478 | (0.020) |
| Log(Shipments/Employment) | 8634 | 5.111 | 0.966 | (0.010) | 5605 | 5.021 | 0.772 | (0.010) |
| Number of Production Workers | --- | --- | --- | --- | 5608 | 112.1 | 151.9 | (2.0) |
| Log(Production Workers) | --- | --- | --- | --- | 5608 | 4.077 | 1.239 | (0.017) |
| Production Man-hours (1000s) | --- | --- | --- | --- | 5608 | 228.9 | 308.8 | (4.1) |
| Log(Production Hours) | --- | --- | --- | --- | 5608 | 4.762 | 1.281 | (0.017) |
| Shipments per Production Hour | --- | --- | --- | --- | 5608 | 140.9 | 207.9 | (2.8) |
| Log(Shipments/Hour) | --- | --- | --- | --- | 5605 | 4.671 | 0.833 | (0.011) |
| Employee Annual Earnings (\$1000) | --- | --- | --- | --- | 5608 | 27.4 | 11.1 | (0.1) |
| Log(Annual Earnings) | --- | --- | --- | --- | 5608 | 3.383 | 0.374 | (0.005) |
| Production Wage | --- | --- | --- | --- | 5608 | 11.9 | 6.1 | (0.1) |
| Log(Production Wage) | --- | --- | --- | --- | 5608 | 2.538 | 0.377 | (0.005) |

Notes: Certification election cases only. Total Employment (I/N/F: reported; L/N/F: sum of non-production workers during quarter that contains March 12 and production workers). Total Value of Shipments (I/N/F and L/N/F): dollar value of products and services sold. Production workers: average of production workers in 4 quarters containing the 12th of March, May, August, and November. Production Man-Hours: sum of 4 quarters of plant man-hours of production workers. Employee Annual Earnings: Sum of earnings paid to production and non-production workers, divided by total employment. Production Wage: Earnings paid to production workers, divided by production man-hours. Dollar amounts are in 2001 dollars for InfoUSA/NLRB/FMCS, and are (CPI-adjusted) 2000 dollars for LRD/NLRB/FMCS.

Appendix Table 3: Correlation Between Similar Measures across Datasets

| Data Set | Dependent Variable | Mean | Slope Coefficient on Alternative Measure | | |
|----------|----------------------|---------|--|--------------------|-----------------|
| | | | All | Exact Street Match | Non-Exact Match |
| InfoUSA | Log(Employment) | 4.298 | 0.913 | 0.910 | 0.906 |
| | | (0.024) | (0.011) | (0.017) | (0.020) |
| | | [4181] | [4181] | [2692] | [1489] |
| NLRB | Log(Eligible Voters) | 4.273 | 0.237 | 0.233 | 0.248 |
| | | (0.013) | (0.010) | (0.013) | (0.015) |
| | | [4181] | [4181] | [2692] | [1489] |
| LRD | Log(Prod. Workers) | 4.605 | 0.972 | 0.959 | 0.976 |
| | | (0.016) | (0.010) | (0.017) | (0.012) |
| | | [3583] | [3583] | [841] | [2742] |
| NLRB | Log(Eligible Voters) | 4.678 | 0.743 | 0.824 | 0.722 |
| | | (0.014) | (0.008) | (0.015) | (0.009) |
| | | [3583] | [3583] | [841] | [2742] |

Note: Standard errors in parentheses. Number of observations in brackets. Upper panel reports median regression slope coefficients. Lower panel reports OLS slope coefficients. Each entry is slope coefficient on the alternative measure (e.g. 0.913 is the slope coefficient from a median regression of Log(Employment) from InfoUSA on Log(Eligible Voters) from NLRB. The entry below is the reverse median regression). "Exact Street Match" indicates that the street address matched exactly between the two data sets.

Appendix Table 4: Coefficients on Won Election, Alternative Measures of Employer Outcomes

| "Labor Inputs" | | "Capital" | |
|----------------------------------|------------------------------|--------------------------------------|------------------------------|
| Log(Total Empl.) | 0.039 (0.045) [28796] | Log(Total Assets) | -0.012 (0.077) [20346] |
| Log(Non-prod. Empl.) | 0.027 (0.051) [28096] | Log(Assets/Total Empl.) | -0.028 (0.069) [20346] |
| Log(Production Empl.) | 0.037 (0.048) [28796] | Log(Assets/Production Empl.) | -0.029 (0.072) [20346] |
| Log(Production Hours) | 0.028 (0.049) [28796] | Log(Assets/Production Hours) | -0.029 (0.072) [20346] |
| "Output" | | "Wages" | |
| Log(Total Value of Shipments) | 0.011 (0.049) [28785] | Log(Annual Earnings) | -0.028 (0.015) [28796] |
| Log(Adjusted Value of Shipments) | 0.016 (0.050) [28778] | Log(Production Annual Earnings) | -0.027 (0.016) [28796] |
| | | Log(Production Hourly Wage) | -0.016 (0.015) [28796] |
| "Productivity" | | | |
| Log(TVS/Total Empl.) | -0.027 (0.030) [28785] | Log(Non-production Annual Earnings) | -0.009 (0.024) [27991] |
| Log(Adj. TVS/Total Empl.) | -0.024 (0.030) [28778] | Log(Non-prod/Prod. Annual Earnings) | 0.010 (0.028) [27991] |
| Log(TVS/Production Empl.) | -0.026 (0.033) [28785] | Log(Adjusted Annual Earnings) | -0.025 (0.015) [28796] |
| Log(Adj. TVS/Prod. Empl.) | -0.023 (0.034) [28778] | Log(Adj. Production Annual Earnings) | -0.024 (0.016) [28796] |
| Log(TVS/Production Hours) | -0.015 (0.034) [28785] | Log(Adj. Production Hourly Wage) | -0.013 (0.016) [28796] |
| Log(Adj. TVS/Prod. Hours) | -0.012 (0.034) [28778] | Log(Adj. Non-prod. Annual Earnings) | -0.006 (0.025) [27991] |

Note: Entries are coefficients from specification (10) from Table 2). Clustered (on establishment) standard errors in parentheses, number of observations in brackets. Adjusted TVS: TVS + net change in finished goods inventory + net change in work-in-process inventory. Adjusted wage variables add total supplementary labor costs to the payroll before dividing by employment (or hours). For production and non-production "wage" variables, total supplementary labor costs are first multiplied by the fraction of the payroll that is paid to production or non-production workers, before added to the respective payrolls.

Appendix Table 5: OLS and Regression-Discontinuity Estimates on Business Survival, Decertifications

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Union Won | -0.030 (0.029) | -0.036 (0.029) | -0.047 (0.029) | -0.035 (0.029) | -0.058 (0.048) | -0.035 (0.049) | -0.019 (0.061) | -0.027 (0.062) | -0.031 (0.062) | -0.026 (0.061) | -0.021 (0.062) |
| Vote Share | | | | | 0.086 (0.120) | 0.536 (0.266) | 0.815 (0.738) | 1.590 (1.332) | 1.560 (1.309) | 1.847 (1.307) | 1.755 (1.296) |
| (Vote Share) ² | | | | | | -0.561 (0.300) | -1.338 (1.911) | -4.887 (5.356) | -5.005 (5.234) | -6.423 (5.248) | -6.040 (5.239) |
| (Vote Share) ³ | | | | | | | 0.546 (1.316) | 6.414 (8.356) | 6.866 (8.118) | 8.949 (8.169) | 8.440 (8.196) |
| (Vote Share) ⁴ | | | | | | | | -3.138 (4.414) | -3.495 (4.269) | -4.465 (4.304) | -4.228 (4.338) |
| Log(Votes Cast) | | | 0.053 (0.017) | 0.050 (0.018) | | | | | | 0.052 (0.017) | 0.049 (0.018) |
| Year dummies? | No | Yes | Yes | Yes | No | No | No | No | Yes | Yes | Yes |
| Industry dummies? | No | No | No | Yes | No | No | No | No | No | No | Yes |
| R-Squared | 0.001 | 0.0394 | 0.048 | 0.0684 | 0.0015 | 0.0051 | 0.0053 | 0.0058 | 0.0442 | 0.0523 | 0.0723 |

Note: Sample size is 1085. Some elections match to same establishment, so standard errors are clustered on the establishment level (998 independent establishments). Includes all elections that occurred before 1997 that has the same name and address as a pre-election-year entry in the LRD. Dependent variable is whether establishment appears in the 1997 Census of Manufactures. Least squares point estimates reported.

Appendix Table 6: Least-squares Regression-Discontinuity Estimates of Union Effects, Decertifications

| Dependent Variable | Coefficient on Won Election | | | | | | | | | |
|---------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Contract Expiration | 0.257 (0.060) [1098] | 0.328 (0.026) [5954] | 0.280 (0.042) [5954] | 0.261 (0.045) [5954] | 0.242 (0.056) [5954] | 0.226 (0.058) [5954] | 0.217 (0.056) [5954] | 0.220 (0.055) [5954] | 0.218 (0.056) [5954] | 0.232 (0.054) [5954] |
| Log(Hours) | -0.071 (0.054) [1098] | 0.016 (0.024) [5954] | 0.002 (0.038) [5954] | -0.005 (0.039) [5954] | -0.021 (0.049) [5954] | -0.038 (0.050) [5954] | 0.006 (0.037) [5954] | -0.006 (0.035) [5954] | -0.004 (0.035) [5954] | -0.005 (0.031) [5954] |
| Log(Output) | -0.243 (0.190) [1098] | 0.088 (0.081) [5954] | -0.115 (0.142) [5954] | -0.104 (0.143) [5954] | -0.414 (0.176) [5954] | -0.381 (0.174) [5954] | -0.160 (0.105) [5954] | -0.193 (0.100) [5954] | -0.196 (0.099) [5954] | -0.205 (0.094) [5954] |
| Log(Output/worker) | -0.233 (0.184) [1098] | 0.153 (0.088) [5951] | -0.014 (0.153) [5951] | 0.002 (0.150) [5951] | -0.484 (0.186) [5951] | -0.473 (0.183) [5951] | -0.221 (0.096) [5951] | -0.249 (0.092) [5951] | -0.252 (0.090) [5951] | -0.236 (0.087) [5951] |
| Log(Assets/worker) | 0.011 (0.153) [1098] | 0.065 (0.065) [5951] | 0.101 (0.110) [5951] | 0.106 (0.111) [5951] | -0.070 (0.144) [5951] | -0.092 (0.142) [5951] | -0.065 (0.063) [5951] | -0.067 (0.063) [5951] | -0.067 (0.063) [5951] | -0.043 (0.062) [5951] |
| Log(Wage) | -0.342 (0.226) [770] | -0.048 (0.098) [4205] | 0.077 (0.163) [4205] | 0.015 (0.162) [4205] | -0.128 (0.207) [4205] | -0.132 (0.207) [4205] | -0.156 (0.189) [4205] | -0.146 (0.176) [4205] | 0.001 (0.141) [4205] | -0.033 (0.128) [4205] |
| Sample | +/- 5% | All | All | All | All | All | All | All | All | All |
| Polynomial Terms | 0 | 0 | 1 | 2 | 3 | 4 | 4 | 4 | 4 | 4 |
| Dependent Variable | Level | Level | Level | Level | Level | Level | De-meaned | De-meaned | De-meaned | De-meaned |
| Include Base Mean? | No | No | No | No | No | No | No | Yes | Yes | Yes |
| Year dummies | No | No | No | No | No | No | No | No | Yes | Yes |
| Industry dummies | No | No | No | No | No | No | No | No | No | Yes |

Note: Within-election clustered standard errors in parentheses. Number of observations in brackets. Each entry is the estimated coefficient on the "Union won" indicator in a least squares regression. "Base Mean" is the average of the dependent variable for years strictly before the election year. "De-meaned" denotes that the dependent variable is the outcome minus the "base mean". "+/- 5%" sample are elections where the union vote share is between 45 and 55 percent.

Appendix Table 7: Least-squares Regression-Discontinuity
Estimates of Union Effects, Decertifications

| Dependent Variable | (1) | (2) | (3) |
|---------------------|-----------------------------|-----------------------------|-----------------------------|
| Contract Expiration | 0.106 (0.063) [2620] | 0.380 (0.068) [1839] | 0.192 (0.092) [1495] |
| Log(Hours) | -0.026 (0.032) [2620] | 0.021 (0.041) [1839] | 0.018 (0.067) [1495] |
| Log(Output) | -0.155 (0.082) [2620] | -0.138 (0.119) [1839] | -0.372 (0.174) [1495] |
| Log(Output/worker) | -0.139 (0.076) [2619] | -0.193 (0.112) [1838] | -0.450 (0.157) [1494] |
| Log(Assets/worker) | 0.012 (0.060) [2619] | -0.064 (0.085) [1838] | -0.108 (0.118) [1494] |
| Log(Wage) | -0.091 (0.133) [2071] | -0.075 (0.199) [1284] | -0.037 (0.211) [850] |
| Sample | 0-3 Years post-El. | 4-7 Years post-El. | 8+ Years post-El. |

Note: Within-election clustered standard errors in parentheses. Number of observations in brackets. Each entry is the estimated coefficient on the "Union won" indicator in a least squares regression. Specification is the same as Col. (10) in Table 2. Each column restricts the sample to the observations that are within the first three years, between four to seven years, and eight years or later, relative to the election year.