The Impact of Unionization on Establishment Closure: A Regression Discontinuity Analysis of Representation Elections*

John DiNardo

University of Michigan, Ann Arbor and NBER

David S. Lee

UC Berkeley and NBER

September 2001

Abstract

Using data on more than 27,000 establishments (1983-1999) in the United States, this paper exploits an institutional feature of the union organizing process in order to generate an estimate of the causal effect of unionization on the probability of establishment closure. In the U.S., most establishments become "unionized" as a partial consequence of a secret ballot election among the workers. If employers where unions *barely* won the election are ex ante comparable in all other ways to employers where unions *barely* lost, differences in their subsequent outcomes should represent the true impact of union recognition. The regression discontinuity analysis finds a negligible effect on short- and long-run establishment survival rates – suggesting that distortions that are necessarily implied by the monopoly union model would more likely be found along the intensive margin of employment.

* Matthew Butler and Francisco Martorell provided outstanding research assistance. We thank David Card for helpful discussions, and are grateful to Hank Farber for providing election data.

1 Introduction

Arguably no single issue distinguishes "modern" labor economics research from its predecessors more than its concern with the question of whether or not "institutions" – especially those which ostensibly serve the interests of workers – have a "distortionary" impact on the allocation of resources and the level of economic activity. Central to this discussion has been the question of whether unions primarily act as a cartel for labor services (the "monopoly union or "right–to–manage" model) in which wage gains for union workers come at the expense of non–union workers. In the standard monopoly union model, the mechanism is straightforward: unions negotiate "above market" wages for their members, and profit-maximizing employers substitute away from unionized labor, and to the extent that they cannot, business establishments become more susceptible to closure; either way, the result is a decrease in output and a misallocation of resources in the economy.[1]

As is well understood, however, obtaining convincing evidence on the mechanism which creates this distortion or on the magnitude of this distortion is difficult. First, large-scale establishment-level data with the necessary information about unionization is typically not available. Second, and more importantly, even with such data, the identification of such an effect ultimately requires some comparison of unionized to non-unionized establishments. The confidence with which one can interpret such comparisons as *causal*, however, is generally limited by the extent to which one is ready to believe that the unionization of establishments is unrelated to the unobserved determinants of establishment survival. Addressing the concerns of selection and omitted variable biases - the hallmark of modern empirical analyses of the labor market - is a requirement of any conclusive analysis of the employer response to unions.

Using data on more than 27,000 U.S. establishments from 1983 to 1999, this paper utilizes a potential quasi-experiment inherent in union representation elections to generate an estimate of the causal effect of unionization on the probability of establishment closure. In the U.S., most establishments become

[1] Indeed, the presence of a "deadweight welfare loss" to unionization is a staple of textbook treatments of unionization. Even Freeman and Medoff (1984) - who suggest the possibility for allocation improvements under unionization - stipulate the existence of such a welfare loss, although they note that their estimate of this loss is small. They observe that their estimate is very close to the calculations in Rees (1963).

"unionized" as a partial consequence of a secret ballot election among the workers. By law, a simple majority vote in favor of the union requires the management to recognize and bargain "in good faith" with the victorious union in collective bargaining negotiations. In this paper, we attempt to minimize selection and omitted variable biases by comparing the probabilities of closure between business establishments that faced elections where the union *barely* won (and hence won legal recognition in the bargaining process) and those that faced elections where the union *barely* lost (and hence has no legal recognition).[2]

We argue that these "*close* election" comparisons are likely to be valid in identifying a causal effect if there is at least *some* component of randomness and unpredictableness in the determination of the *exact* vote tally in these NLRB elections. This particular comparison sharply contrasts to the comparison of winners and losers more generally (ignoring the vote count), where in general, establishments that were able to prevent a union victory are likely to be systematically different from those that could not. We present the conditions under which the differences that potentially confound union effects are likely to be negligible when examining elections decided by a very narrow margin. In focussing on close elections, the research design is readily recognized as an application of the regression discontinuity design, where the treatment of interest (union recognition) is a known deterministic function (simple majority voting) of an observable continuous variable with at least some random component (the final vote tally).[3]

Using data merged from three different sources – the National Labor Relations Board (NLRB), the Federal Mediation and Conciliation Service (FMCS) and a commercial database from InfoUSA, Inc. – we report the following findings.

First, as would be predicted by a valid regression discontinuity design, employers where the union barely won appear to be quite comparable to employers where the union barely lost – at least along observable dimensions such as bargaining unit size, industrial classification, and the pre-election presence of

[2] The approach of Lalonde, Marschke, and Troske (1996) comes closest to our analysis. They examine the impact of union victories in NLRB certification elections on wages, employment, and total value of shipments using the LRD. Using a "difference-in-difference" approach to address selectivity, they find that successful organization is associated with significant declines in subsequent employment and output. However, their results *also* imply that successful union organizing is associated with a decline in output and employment, *even before* the representation election.

[3] Regression discontinuity designs are described in Thistlethwaite and Campbell [1960] and Campbell [1969], and formally examined as an identification strategy recently in Hahn, Todd, and van der Klaauw [2001]. Recent examples include Angrist and Lavy [1998] and van der Klaauw [1996], and in the election context, Lee [2001].

a different union. Second, we find a striking effect of union certification on the probability that a union contract expiration notice is filed at the Federal Mediation and Conciliation Service (FMCS) subsequent to the election, suggesting the outcome of the election is somewhat "binding." Third, we estimate a negligible union recognition effect on survival (henceforth, the "extensive" margin of employment) – the point estimates range from -0.01 to -0.02 in probability over 1- to 15-year horizons. We conclude that extensive margin employer responses are far from overwhelming, and more likely to be relatively small or non-existent. Fourth, we also find that "close winner" and "close loser" establishments appear quite comparable along observable dimensions *conditional on survival*. Among other things, this suggests that sample selection bias in an analysis of employment conditional on survival (henceforth the "intensive" margin of employment") may be a second order issue. We conclude that the small estimated effects on establishment closure suggest that the distortions that are necessarily implied by the monopoly union model would more likely be found along the intensive margin of employment. Our own estimates of the *intensive margin* employment response, however, are not precise enough to rule out economically meaningful magnitudes, suggesting that future research utilizing this research design may require more detailed establishment-level characteristics in order to produce more informative *intensive margin* estimates.

The paper is organized as follows. Section 2 establishes some basic facts about establishment survival and the extent of unionization, and graphically illustrates our regression discontinuity estimates. Section 3 places our analysis in the context of industrial relations in the U.S., and describes some important aspects of our data. In Section 4 we illustrate the sufficient conditions for identifying union effects in a regression discontinuity framework. We present the main results in Section 5, discuss their economic implications in Section 6, and suggest directions for future research in Section 7.

2 Establishment Survival and Unionization: Basic Facts

2.1 Establishment Survival, by "Union Presence"

Establishing the first-order descriptive statistics regarding the short- and long-run economic outcomes of business establishments and how those outcomes differ between "unionized" and "non-unionized" establishments poses a significant challenge in itself.[4] For this study, we have compiled data that can provide a provide a preliminary, albeit incomplete, portrait of some economic outcomes of establishments that are at risk of becoming unionized. Three important patterns emerge from our sample of establishments that experience NLRB representation elections: 1) as might be expected, establishments' survival probabilities decline as one examines longer and longer intervals, 2) using our rough proxy for the presence of a union, it appears that "unionized" establishments in fact have a *higher* probability of surviving than non-unionized establishments both in the short- and long-term, and 3) establishments' death/exit rates appear to be the dominant contributor to the overall decline in the total employment they provide over time. Table I illustrates these basic patterns.

Table I and the subsequent analysis in this paper is essentially based on the universe of establishments that experienced NLRB representation elections between 1983 and 1999.[5] Establishment survival is determined by whether or not the employer in the NLRB data matches (on the basis of name and address) as of 2001 with an entry in an annually updated commercial database maintained by InfoUSA, Inc.; this database ostensibly contains the entire universe of all establishments with a telephone listing. Our proxy of whether the establishment was "exposed" to a union, is whether or not, in the period between the election and the year 2001, an employer or union located at the establishment's exact address filed notice with the Federal Mediation and Conciliation Service (FMCS), indicating an impending expiration of an existing collective bargaining agreement, as parties to agreements are legally required by to do.[6] The details and limitations of these data are deferred to Section 3.

The first column of Table I shows that in this sample, the probability of survival declines significantly as one examines longer and longer intervals. For example, the table shows that among the employers that experienced NLRB representation elections in 1984, roughly 28 percent of them were still in existence as of the year 2001. By contrast, about 58 percent of establishments that experienced an election in 1999

[4] First, representative survey data on business establishments is not readily available. More importantly, we know of no representative sample of establishments with information about whether or not the workers are unionized.

[5] The NLRB election data are representation election cases that are disposed within the fiscal years from 1984 to 1999; thus, most of the elections were held between the years of 1983 and 1999, with a few elections occuring before 1983.

[6] See the discussion of the data in Section 3.

had survived as of 2001. The survival probability grows monotonically as we examine more recent elections; this would be expected if the establishments that experienced elections were, on average, comparable over time. The implied exit rates are comparable to other estimates from existing research.[7]

The next three columns of Table I show that, if anything, establishments that were exposed to unions, are significantly *more* likely to survive both in the short- and long-run. For example, among elections that occurred in 1998, 67 percent of those employers for which we observe a contract expiration were still "alive" by the year 2001, compared to 56 percent for the "non-unionized" employers. The difference is also large for survival rates 13 years after the election (elections in 1988). One might be tempted to infer from this naive comparison that the causal effect of unionization on establishment survival is *positive*. However, it is also consistent with a *negative* causal effect of unionization if, for example, unions are sufficiently more likely to appear in inherently "healthy" establishments.

Taken together, the fifth and ninth columns in Table I suggest that in this sample, employer death or exit is a significant component in the decline of total employment among these establishments as longer intervals are examined.[8] Among employers facing elections in 1984, the average employment level - where "dead" employers are counted as having zero employment - is about 60, while the corresponding numbers in the late 90s are over 100.[9] By contrast, the average log employment (ninth column) *conditional on survival by 2001* appears to be relatively stable over time.

Finally, the differences in the 8th and 12th columns illustrate that the differences in total employment between "unionized" and "non-unionized" employers mostly stem from differences in employment at the extensive margin of employer survival/death. Conditional on survival, union and non-union establishments appear to be roughly the same size throughout the sample period, with a difference that is not statistically different from zero.

[7] For example, Dunne, Roberts, and Samuelson [1989] report a 5-year exit rate of about 40 percent, as calculated from Census of Manufacturing data (1967-1977). And an analysis of food-manufacturing plants by McGuckin, Nguyen, and Reznek [1998] imply a 10-year exit rate of about 60 percent in the LRD from 1977 to 1987. As seen in Table I, the corresponding implied exit rates in our data are 48 and 59 percent respectively.

[8] Employment data is from the InfoUSA, Inc. database.

[9] All "dead" establishments were assigned zero employment. There are 17622-16355=1267 missing values for employment among the surviving employers as of 2001.

Overall, Table I illustrates the empirical importance of the extensive margin of survival to employment patterns over time, and between ostensibly unionized and non-unionized employers; this motivates our focus on the impact of union recognition on employer survival. The evidence also suggests that significant positive selection is necessary – e.g. unions are attracted to inherently longer-lived establishments – in order to be consistent with negative union effects at the extensive margin of employment. The extent to which such a positive selection might obscure a small or large union effect is the focus of our study.

2.2 Establishment Outcomes and Characteristics, by Election Outcomes

In the U.S., union recognition, and the obligation of employers to bargain "in good faith" with the union is often determined through a simple-majority secret-ballot vote conducted by the National Labor Relations Board (NLRB). When we examine establishment outcomes and other pre-determined characteristics, by the *outcome of the NLRB representation election*, four broad patterns emerge: 1) employers that faced elections in which unions were victorious have a slightly smaller probability of surviving to the year 2001 than those establishments where the union lost, 2) establishments where unions won are significantly smaller, in terms of employment and sales volume as measured in 2001, than those where unions lost, 3) across establishments, the election outcome is quite significantly associated with our own proxy of union presence, and 4) the election outcome is also associated with several other pre-determined characteristics of the establishment (e.g. industry, size of the voting unit), providing reason to doubt the interpretation that differences in 1), 2), and 3) represent causal effects.

Table II provides the details of these findings. The first row reports that among establishments where the union won the representation election, 40 percent survive as of 2001, compared to 43 percent for the employers where the union lost. While the difference on this metric is somewhat modest, the differences in employment and sales volume are quite significant. The 2nd through 5th rows report that "union-win" establishments employ about 22 percent fewer workers and generate about 35 percent less in terms of sales by the year 2001.

Table II also shows that the outcome of the representation election is highly correlated with our proxy for union presence - whether or not we observe a contract expiration notice in the time between the election and the year 2001. Among "union-loss" establishments we observe a contract expiring a little less than 9 percent of the time. When the union wins the election, on the other hand, there is a 29 percent chance that we observe a union contract ending after the election.

The rest of Table II provides good reason for the analyst to resist interpreting these union-won/union-loss differences as the causal effect of union certification. For example, the establishments where the union won are about 15 to 20 percent smaller than the "union-loss" establishments, as measured by the number of eligible voters or the ultimate number of votes cast in the NLRB election. In light of these differences, it is thus not surprising that we observe differences in employment *after* the election, in the same direction, and of roughly the same magnitude. Undoubtedly, the post-election differences represents a combination of the true union effect and *ex ante* systematic differences between the "union-won" and "union-loss" establishments.

Similarly, row (7) of Table II reveals that establishments where we observe a union contract expiration were more likely to have been exposed to a union *before the election* than establishments for which we do not observe a contract expiration: the respective proportions of "union presence" are 0.136 and 0.077.

Employers differ by election outcome on a number of other characteristics; these differences give more reason to maintain some doubt in any causal interpretation of the comparisons in the first rows of Table II. For example, as row (11) of Table II indicates, establishments where the union won the election are much less likely (33 versus 42 percent) to be classified in the manufacturing sector. Moreover, as rows (12) and (13) of Table II indicate, establishments where the union won are more likely to be in the service sector (35 percent in "union-win" establishments versus 22 percent for "union-loss" establishments) and less likely to be classified by the NLRB as "truck drivers". On the other hand, measures of state economic conditions are not strongly related to the outcome of the election. The union won/loss differences in the levels and changes in the unemployment rate and the log(employment) level are statistically but not economically significant.

In sum, Table II provides evidence that *caution the analyst against* making inferences about the impacts of union certification on employer outcomes from simple differences in outcomes by election outcome.[10] The evidence is suggestive that union recognition by election may be *negatively selected* – that unions are more likely to prevail in a representation election in smaller, and potentially *less* robust establishments. This would be consistent with the notion that larger establishments with greater resources may be more able to resist organizing drives. However, this theory is at best speculative in the absence of a credible estimate of the causal effect of union certification.

2.3 Establishment Outcomes and Characteristics, by Union Margin of Victory

A more refined analysis of the comparison of employers where the union *barely won* the election to those where the union *barely lost* is the essence of the regression discontinuity analysis of our study. The identification relies on the proposition that these two groups of establishments are likely to be, on average, *ex ante* comparable in all other ways, except that only one group actually attains official union certification and recognition.

A graphical analysis of our data reveal four important patterns: 1) generally smooth empirical associations between the union margin of victory and the conditional expectations of the post-election outcomes and pre-election characteristics, with the exception of 2) a striking discontinuity in the conditional probability of observing a contract expiring after the election (our measure of union presence), 3) no visible discontinuity in the conditional probability of surviving as of the year 2001, and 4) no visible discontinuities in the pre-election characteristics of the establishment. Figures I-III illustrate these patterns.

Figure Ia plots local averages of the indicator variable which equals one when we observe a contract expiration after the election but before the year 2001 against a transformation of the vote margin of victory.[11] The dots to the right of the "0" threshold represent averages for establishments where the union won, whereas dots to the left represent those where the union lost. The most striking feature of the figure is the

[10] The researcher might be tempted to conduct the analysis conditional on the pre-determined characteristics such as industry, and size of voting unit, under the presumption that the election outcome is random conditional on those covariates. Besides being somehwat ad hoc, by "using up" the covariates, this approach has the drawback of eliminating any possibility of gauging the internal validity of the comparison.

[11] More specifically, the actual margin of victory was transformed according to $V^* = (1/\pi) \cdot \arctan\left(\frac{V}{17}\right)$ (the cumulative distribution function of a shifted Cauchy distribution), where V is the actual difference between votes for the strongest union (the one with the most votes), and the vote that would mean that the union would barely lose. Unions must attain 50 percent plus one vote in order to win. For example, if 25 votes were cast, and 15 went to the union, then V would be $15 - 12 = 3$. The monotonic transformation was applied to make the resulting V^* resemble a normal distribution. See Data Appendix for details. The local averages are thus for intervals that are 0.01 wide in units of V^*.

discontinuous jump precisely at the threshold between a union win and a union loss. So, for example, the figure shows that among elections where unions lost by less than 5 votes, the probability that we observe a contract expiration at that establishment sometime after the election is almost 0.10. The corresponding proportion for establishments where unions won by less than 5 percent is just over 0.20.

Under the assumption that these two groups of establishments are *ex ante* comparable (on average) in all other ways, the discontinuity jump reflects a causal effect of a union win on "union presence". However, it is important to note that, technically, our proxy for "union presence" is not independent of establishment survival (our outcome of interest). That is because, in principle, the exit of an establishment before the expiration of the first contract counts as a "zero", even if the union in fact secured an initial agreement. Thus, at best, Figure Ia, suggests that we can reject the hypothesis that the union victory had simultaneously no impact on the securing of an initial agreement, and no impact on establishment survival. Thus, the discontinuity implies at least one of the effects is strong.[12]

Instead of speculating about the degree to which the discontinuity represents a union effect on employer exit, we can evaluate that directly, since we observe whether or not the establishment remained in operation until the year 2001. Figure Ib illustrates the analogous plot for the estimate probability of establishment survival (as of 2001), conditional on the union vote margin of victory. The figure reveals that there is no visible discontinuity in establishment survival at the threshold of union victory. This is the main substantive finding of the paper – that there appears to be a negligible causal effect of union certification/recognition on survival probabilities for this sample of establishments. *How this finding can be reconciled with a standard monopoly union explanation is deferred to Section* 6.

These causal inferences are valid to the extent to which all other factors which determine employer survival – whether observed by the econometrician or not – are similar or "balanced" between *bare winners* and *bare losers*. As in any empirical analysis, assessing whether "unobservables" are balanced is impossible. However, we can at least assess whether or not the regression discontinuity design is succeeding in

[12] In other words, the discontinuity jump could, in principle, be consistent with the outcome of the election having no impact on the probability of securing an initial bargaining agreement, if there were a strong *positive* effect of union recognition on survival in the years immediately following the election, so that we do not observe a contract expiration for the bare losers, because those establishments die off before we can observe it.

balancing *observable* determinants of establishment survival. Since the conjecture of the design implies that it should, the magnitudes of any differences in other pre-determined factors different between bare winners and losers provides a gauge by which we can assess the internal validity of the design.

Figure IIa, IIb, IIIa, and IIIb provide evidence suggesting that observable pre-determined characteristics are balanced between bare winners and losers. For example, in Figure IIa, we see a relatively smooth empirical relationship between the probability of a "union presence" *before* the election and the union margin of victory. If establishments where the union barely won were otherwise inherently more susceptible to "unionization" than their union-loss counterparts, the figure should exhibit a discontinuity at the zero threshold. No discontinuity is apparent from the figure.

The remaining figures exhibit a similar pattern: a fairly smooth empirical relationship between the vote margin of victory and the log of the total vote cast in the election (Figure IIb), the probability that the employer belongs to the manufacturing sector (Figure IIIa), and the probability of belonging to the service sector (Figure IIIb). In all three figures, there is no significant discontinuity jump at the threshold, implying that the restrictions implied by the regression discontinuity design are not violated, lending credence to the causal interpretation of the discontinuity jump (and lack thereof) in Figure Ia and Ib.

A more formal discussion of the stochastic assumptions that are required for valid identification of the union effects in this context is deferred to Section 4.

3 Background

3.1 Institutional Background: the industrial relations climate and the NLRB Election Process

Our sample of establishments is limited to a particular "selection" of employers that are at risk for becoming "unionized." Thus, before proceeding any further, it will be instructive to place our analysis in the context of labor relations and the conduct of representation elections in the U.S.

The administration of fair, secret ballot elections to determine union recognition is one of the chief responsibilities of the NLRB, the most significant administrative agency to be a consequence of the National

Labor Relations Act (NLRA) – the Wagner Act – of the 1930s. The law has been changing continuously since its enactment, most notably with the passage of Taft–Hartley Acts in 1947 (which among other things, provided for temporary government seizure of struck facilities in the event of a strike that creates an "emergency") and the Landrum–Griffin Act of 1959 (which among other things, outlawed a number of successful union tactics including "secondary boycotts"). In principle, the NLRA provides a neutral setting in which the right for workers to bargain collectively is enforced.

It is important to note, however, that where U.S. law gives workers the right to unionize, an NLRB representation election is not required. In general, nothing prevents an employer from recognizing a union without the formalities of an election. Voluntary recognition of a union, however, is thought to be quite rare, and employers generally will attempt to resist an organizing drive. 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.
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.[13]
4. Alleging unfair labor practices, disputing the choice of bargaining unit, etc.[14]

Against this backdrop of employer opposition to unionization, it is perhaps not surprising that there is no single path to an NLRB election and eventual recognition of the union by the employer. Nonetheless, it is useful to describe a prototypical scenario that results in a establishment agreeing to bargain with its workers through a labor union:

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.

[13] For a colorful, albeit idiosyncratic discussion see Levitt (1993).

[14] In the case of graduate students at universities, for example, employers have often attempted to argue – sometimes successfully – that graduate student employees are not "employees" but "students receiving financial aid." Another example is employers arguing that its employees are not workers but "independent contractors" who are not covered by the provisions of NLRA.

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 in principle, only 30% is 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 dispute about the appropriate bargaining unit – employers generally prefer larger and more heterogenous groupings than do unions.
4. Next, typically within 30 days from the card submission, 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.[15]
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." Again, even at this state, however, there is no guarantee that the firm will recognize the union, or that a contract secured by collective bargaining is inevitable. Indeed, analysis by the "Dunlop Commission" found that only 55% of those unions who win elections eventually secure a first contract.

Two aspects of the industrial relations climate deem our sample of establishments particularly appropriate for an analysis of the impact of union recognition on employer outcomes.

First, that employers are thought to generally oppose organization drives suggests that both parties have "something at stake" in the outcome of the election. For example, we expect that both the union and management are expecting that a union win will generally lead to higher wages, more benefits, or better working conditions, potentially at the cost of the employer. If very little were at stake, and if the elections themselves were *pro forma* events, then we would not expect to see a significant employer response to a union election victory. Such a finding would say more about the small size of the "treatment" ("unionization") than the potential magnitude of distortionary effects of an aggressive union. This seems unlikely, however, since we analyze establishments which faced NLRB elections. Such a focus would seem likely to select establishments where union-management relations are contentious since in the overwhelming majority of cases the management of such establishments always have the option of *voluntarily* recognizing the union *without* a (costly) NLRB election. Thus, it would seem more reasonable to assume that the outcome

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

of the election is far from inconsequential to both parties.[16]

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. As shown in Section 4 a certain degree of uncertainty is important to generate the "near-random" assignment of the outcome in close decisions. For example, if all that was required for legal recognition was the petitioning of 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 (strange!) group of establishments/unions where the workers submitted signatures that totalled 49 percent. By contrast, it is very easy to imagine in a secret-ballot context that those unions that gained 26 out of 50 votes had virtually the same chance of winning as the unions that only gained 25 out of 50 votes.

3.2 Dataset Construction: the NLRB, FMCS, and InfoUSA, Inc.

The dataset used for the analysis was constructed as follows. First, electronic records on all representation election cases handled by the NLRB in the fiscal years from 1984 to 1999 were obtained. 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.

Importantly, these files contain the establishment name and exact address. The names and addresses alone were submitted to a commercial marketing database company called InfoUSA, Inc. 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. InfoUSA matched as many of the submitted records to their current database (as of May, 2001) and

16 Of course, whether union recognition ultimately has an effect on conditions, such as wages, is an empirical question.

appended information to the record. They were not given any information beyond the name and address.

This merged data was then additionally merged 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, in principle, parties to collective bargaining agreements are required to file so-called "30-day notices" with the FMCS. This was used to obtain our proxy for "unionization" or the "presence" of a union, under the presumption that contracts eventually expire, typically after two or three years.

There are a few important limitations to our data. First, our data do not constitute a comprehensive panel dataset. We can potentially observe "survival" or "death" as of one point in time - in the year 2001. We know very little about what happens between the time of the election and 2001, except the observation of contract expirations on that particular location. While we do observe a few "baseline" characteristics from the NLRB election file, since InfoUSA does not retain historical records, we do not have employment data for the establishments as of the time of the election. The lack of precise duration data are likely to have contributed to producing relatively noisier estimates of union effects.

Second, since we are measuring employer "survival" as a match (by name and address) in the InfoUSA database, there will undoubtedly be some measurement error. Consequently, we will inevitably treat some firms as having "died", when instead we have simply been unable to match them. However, while this may mean that estimates of the *level* of survival rates may be downward biased, it is highly unlikely that establishments with close union winners are systematically less or more likely to match to the InfoUSA database than counterpart close union losers, except if there is a true impact of union certification on survival probabilities.

Likewise, our measure of "union presence" will also likely be biased in *levels*, although this is unlikely to have important consequences for our comparison of close winners to close losers. For example, we understate the extent of unionization to the extent that our matching algorithm fails to locate a match in the FMCS data when such a match exists or to the extent that noncompliance with the law (regarding notifying the FMCS) is widespread. Alternatively, it could be upward biased to the extent that we are matching elections to contract expirations from *other* bargaining units at the same establishment, and to the extent that our matching algorithm produces "false positive" matches. Although the levels may be mismeasured, it seems reasonable to assume that these sources of measurement error are unlikely to be systematically different between close winners and close losers. On balance, we believe the benefits of being able to compare the bare winners and losers on the basis of some other measure of "union presence" other than the certification that results from winning the election outweighs the inability to obtain an accurate measure of the overall *level* of union presence.

4 Econometric Framework

4.1 Regression Discontinuity Framework

In our context of union representation elections, the internal validity of our regression discontinuity analysis primarily depends on two assumptions: 1) that the "treatment" (union recognition) is a known, discontinuous function of an observed variable (the votes for the union), and 2) that there is at least *some* random component of the exact vote tally. U.S. labor law ensures 1); the importance of 2) is discussed below within a reduced-form econometric framework. We show the sufficient stochastic assumptions for identifying union effects in our context.

Suppose an employer outcome, such as the probability of survival by time $t+1$, is determined by the equation

$$y_{t+1}^{*} = WIN_t\beta + X\gamma + \varepsilon \tag{1}$$

where y_{t+1}^{*} is the probability of survival, and WIN_t is an indicator variable determining whether the union won the representation election at the time of the election t. X and ε are all other observable and unobservable determinants of establishment survival, respectively, and they can represent variables that are determined at $t+1$, t, or earlier. To simplify the exposition and notation of the discussion below, we abstract from the obvious fact that y_{t+1}^{*} must be bounded below and above by 0 and 1, respectively.[17]

The first key aspect of the regression discontinuity design is that we know something very specific about what determines WIN_t: a simple majority vote. We can represent this fact as

$$WIN_t = \begin{cases} 1 \text{ if } v_t > 0 \\ 0 \text{ if } v_t \leq 0 \end{cases} \tag{2}$$

where v_t is the realized the union vote minus half the total vote cast.[18]

It is straightforward to show that the expectation of y_{t+1}^{*} conditional on the observed vote v_t is thus

$$E\left[y_{t+1}^{*}|v_t\right] = \begin{cases} \beta + E\left[X|v_t\right]\gamma + E\left[\varepsilon|v_t\right] \text{ if } v_t > 0 \\ E\left[X|v_t\right]\gamma + E\left[\varepsilon|v_t\right] \text{ if } v_t \leq 0 \end{cases} \tag{3}$$

It is clear, then, that the comparison of survival probabilities between employers where the union won by Δ votes and those where the union lost by Δ votes will be a composition of the true impact β, the bias due to differences in observables $(E\left[X|v_t = \Delta\right] - E\left[X|v_t = -\Delta\right])\gamma$, and the bias due to differences in the unobservables $(E\left[\varepsilon|v_t = \Delta\right] - E\left[\varepsilon|v_t = -\Delta\right])$.[19] In general, even if v_t is stochastic in nature, it may be incidentally correlated with both ε and X. For example, it could be that inherently "healthy" employers (e.g. high ε) or larger employers (e.g. high X) have more resources to resist an organizing campaign, and hence lower v_t. Any such correlation will induce bias in this comparison, when Δ is large relative to v_t.

On the other hand, one can minimize the observable and unobservable components of the bias by choosing to examine very close elections – with Δ relatively small. This notion is the basis of the identification strategy employed in our analysis. Thus, a sufficient stochastic assumption for identification is that the distributions of unobservable and observable determinants of the survival probability – while potentially dependent on v_t – are continuous with respect to v_t, precisely at the threshold that divides union victory from defeat.

17 With a bit more notation, it is easy to generalize the discussion below to a proper latent variable framework.

18 For more details of an endogenous dummy variable setup, see Heckman [1978].

19 A latent variable analogue could be written as $\Pr\left(Survival|v_t\right) = \begin{cases} S_{(\varepsilon+X\gamma)|v_t}\left(-\alpha-\beta\right) \text{ if } V_t > 0 \\ S_{(\varepsilon+X\gamma)|v_t}\left(-\alpha\right) \text{ if } V_t \leq 0 \end{cases}$ where $S_{(\varepsilon+X\gamma)|v_t}$ denotes the survivor function of the distribution of $(\varepsilon + X\gamma)$ conditional on v_t and α is a constant. The same intuition holds, the larger Δ is, the more different the conditional distribution of $(\varepsilon + X\gamma)$ is likely to be.

When might this assumption be violated? As an example, if the vote tally v_t were perfectly predictable (say, at the time of filing the petition) by the union, then the establishments where the union barely wins recognition would probably systematically differ from the select group of establishments where the union – knowing that the vote tally would fall just short of victory – chose to petition for the election anyway (perhaps to signal dissatisfaction to the employer). On the other hand, if the eventual vote tally is less than perfectly predictable, and hence at least *some* stochastic component to the eventual vote count, bare winners and losers are likely to contain the roughly same proportion of unions who care about the outcome of the election and unions who are participating in the election as a symbolic gesture.[20]

As in any empirical investigation, it is impossible to rule out violation of the identifying assumption. However, in this case, if examining close elections fails to reduce bias along unobservable dimensions, it is likely to fail to reduce bias along observable dimensions as well. We can partially assess this bias by estimating $(E[X|v_t = \Delta] - E[X|v_t = -\Delta])$. Put another way, while it is impossible to assess whether "unobservable" characteristics are roughly balanced between the union-win and union-loss groups, we *can* test the restriction of the research design that *observable* characteristics are roughly balanced between the union-win and union-loss groups of employers.

4.2 Empirical Evidence of the Regression Discontinuity Design

Table III shows that the pre-determined characteristics (X) between the union-won and union-loss establishments *do* become more balanced as we examine closer and closer elections. For example, in elections where the union lost, the average number of eligible voters is about 113, compared to about 92 where the union eventually won. That difference falls dramatically to about 2, when we focus on the comparison among elections decided by 8 votes (about half the sample).[21] The same holds true in percentage terms. The difference in terms of the log of the vote cast falls from -0.19 to about 0.01 when we move from the

[20] To see this, suppose that $V_t = \hat{V}_t + u_t$ where $\hat{V}_t$ is akin to a "forecast" (say, at the time of the filing of the petition) and u_t is a prediction error. Then the bias in unobservables would be $\left(E\left[\varepsilon|u_t = \Delta - \hat{V}_t\right] - E\left[\varepsilon|u_t = -\Delta - \hat{V}_t\right]\right)$. Thus, even if there is a discontinuity in $E\left[\varepsilon|\hat{V}_t\right]$ at the threshold and correlation between u_t and ε, the vias will vanish as long as $E[\varepsilon|u_t]$ is continuous at the threshold and Δ becomes smaller and smaller relative to the variance in u_t.

[21] Margin is defined as the number of votes attained by the union (with the most votes) minus the vote that would mean that the union barely lost (by one vote).

first to second set of columns.

While most of the initial differences are reduced by examining that set of narrower victories and defeats, the small differences that remain can be reduced even more by examining even narrower margins. In the third set of columns, we report the means and differences for the elections decided by less than 2 votes. Comparison of the sixth and ninth columns reveal that the differences in the pre-determined characteristics are even smaller when examining elections decided by 2 or fewer votes. For example, the win/loss difference in the probability that the employer is a service sector establishment falls from 0.130 to 0.072 to 0.048 when moving from all elections, elections decided by 8 or fewer votes, to elections decided by 2 or fewer votes.

The final set of columns in Table III demonstrate that these small differences remain when we calculate alternative estimates of the pre-determined characteristics for close union winners and losers. Given the apparent "smoothness" in the relation between each of the variables on each side of the threshold – as revealed in Figures I, II, and III – it seems natural to incorporate this restriction. In this way, we attempt to use observations "away from the threshold" to estimate averages at the left- and right-limits of the threshold. The estimating equation is simply a flexible-form parametrization of the expectation of X conditional on v_t where we let $E[X|v_t] = g_0(v_t)$ if $v_t \leq 0$, and $E[X|v_t] = g_1(v_t)$ if $v_t > 0$, where $g_0(v_t)$ and $g_1(v_t)$ are cubic splines in v_t with two knot points each. We estimate each function by restricted least squares, and the estimates $\widehat{g_0}(0)$, $\widehat{g_1}(0)$, and $\widehat{g_1}(0) - \widehat{g_0}(0)$ are reported in the last three columns of Table III, respectively, and the estimated functions using this procedure are plotted in Figures I, II, and III.[22] For the most part, the estimated standard errors fall, relative to the third set of columns. Compared to the third set of columns, roughly half of the estimated differences are larger and half are smaller in absolute magnitude.

[22] This amounts to the following (data to the "left" and "right" side of the threshold were used in seperate regressions): regressing X on dummy variables indicating to which of three "segments" - intervals along the v_t dimension - the data belong, as well as their interactions with a linear term in v_t, and an interaction of the middle "segment" dummy with quadratic and cubic terms in v_t, while restricting the coefficients such that the overall pieced-toether function is continuous and has continuous first derivatives at the two knot points that define the three segments. Thus, the function is a smooth piece-wise linear-cubic-linear function. The knot points were chosen by iterating and minimizing the root mean-squred error from the regression through a grid search. We reiterate here that before estimation, we actually use $v_t = \left(\frac{1}{\pi}\right)\arctan\left(\frac{\text{Margin}}{17}\right)$ (a translated Cauchy cdf) because the actual margin of victory has a distribution with extremely "thin" tails. In order to spread the data to look more "normal", we apply this monotonic transformation. See Appendix Figure Ib.

Overall, Table III provides evidence strongly consistent with the observable implications of the regression discontinuity design: win/loss differences in pre-determined characteristics of the establishments vanish as one compares closer and closer margins of victory. To the extent this provides evidence of the validity of the regression discontinuity design, differences in survival rates between close winners and losers should represent a *causal* effect. The survival rate differences are reported in the first row of Table III. It shows that the initial union win-loss difference in survival probability falls slightly from -0.03 to an estimate about -0.02, within sampling error of zero. Given the estimated sampling error, the estimate is consistent with both a negative effect of up to -0.048 as well as a positive impact of up to 0.008 at conventional levels of significance.

Finally, note that the one estimate that remains strongly significant, both statistically and meaningfully in a probability sense, is the impact of a union victory on the probability that we observe a contract expiration notice sometime after the election (our "union presence" proxy) – an effect of 0.133. As with the impact on survival probability, we interpret this as a causal effect and not the artifact of selection bias or omitted variable bias.

We examine the robustness of these final two results to various specifications in the next section.

5 Estimates of the Impact of Union Recognition

5.1 Estimates from Alternative Specifications

If the assumptions of the regression discontinuity design hold, then the estimates of the effect of union recognition on establishment survival should be robust to various specifications that include any combinations of pre-determined characteristics in the estimation procedure. Table IV provides evidence that the estimates are indeed robust to various specifications.

To reiterate, in Table III, the "cubic spline" estimate of β from the specification

$$E\left[y_{t+1}^{*}|v_t\right] = WIN_t\beta + E\left[X|v_t\right]\gamma + E\left[\varepsilon|v_t\right] \equiv WIN_t\beta + h\left(v_t\right) \quad (4)$$

was implemented by regressing the survival indicator on the WIN_t indicator and an approximation of h –

two cubic splines with two knot points each, with one cubic spline approximating $h(v_t)$ when $v_t \leq 0$ and the other approximating $h(v_t)$ when $v_t > 0$.[23] The estimate is reported again in Column (1) (top panel) of Table IV.

If the assumptions of the research design hold, then the inclusion of any set of pre-determined characteristics into the regression should not significantly affect the estimate of β. This is because Equation 3 can be manipulated to imply

$$E\left[y_{t+1}^{*}|v_t, x_1\right] = WIN_t\beta + \widetilde{h}(v_t) + x_1\gamma_1 \tag{5}$$

where $x_1$is any particular realization of a subset of the pre-determined characteristics X, where we again approximate $\widetilde{h}(v_t)$ with the cubic spline specification.[24] If the estimate of β depends significantly on which variables are included, then it would suggest that either the cubic spline approximation is poor and/or the identifying assumption is violated.

The first row of Table IV demonstrates that the estimate of the impact of union recognition on survival probabilities is robust to the inclusion of any combination of pre-determined characteristics, with the estimates ranging from -0.013 to -0.020. In Column (2), x_1 is a set of election-year dummy variables. In Column(3) state dummies are additionally included, and in Column (4) industry dummies and unit dummies (indicated how the NLRB classified the primary type of worker in the bargaining unit) are included. In Columns (5) and (6) our proxy for the presence of a union *before* the election is included (as it is pre-determined as of the election date), as is the log of the number of eligible voters in the bargaining unit. It is important to note that the coefficient on the pre-election union presence dummy should not be interpreted as a *causal effect*, but simply indicative that this proxy appears to absorb some of the "residual" variance in the dependent variable.

These robustness tests are analogous to the tests one could perform in a classical randomized experiment. In the true experiment, the randomization has ostensibly made the distribution of observables and

---

[23] Again, the identifying assumption was the continuity $h(\cdot)$ at $v_t = 0$.

[24] If $X\gamma = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \cdot \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$, then $\widetilde{h}(v_t)$ would simply be $h(v_t) + \widehat{u}$ where $\widehat{u} \equiv E[X_2|v_t, x_1]\gamma_2 + E[\varepsilon|v_t, x_1] - h(v_t)$, where x_1 is a particular realization of X_1. If the joint distribution of ε and X are continuous with respect to v_t at $v_t = 0$, then $\widetilde{h}(v_t)$ is also.

unobservables the same in both "treatment" and "control" groups. Thus, the inclusion of *any* set of "baseline" characteristics in a regression of the dependent variable on the treatment indicator should yield similar estimates.[25] The similarity of the estimates across specifications is an *implication* of the randomization; analogously, the similarity of the estimates across the specifications here, are also an *implication* of a valid regression discontinuity design.

Another alternative specification that should give a similar estimate is to regress the survival indicator on X, and use the resulting residuals as the dependent variable in estimating β from Equation 4.[26] Column (7) shows that the estimate is -0.013, similar in magnitude to the other estimates in the first row.

The preceding discussion holds equivalently for examining the effect of union recognition on our *post-election* "union presence" proxy. Table IV shows that estimates from repeating the exercise – using our post-election union presence variable as the dependent variable – are quite stable across specifications. This would be expected if the assumptions of the regression discontinuity design were valid. The estimates of the effect of union recognition on the probability that we observe a contract expiration notice at the establishment are precisely estimated and range from 0.119 to 0.133. Note from the comparison of Columns (3) and (5) or (4) and (6) that the inclusion of the *pre-election* union proxy does not meaningfully affect the union recognition effect, despite its own independent predictive power (t-stat over 40).[27]

5.2 Estimates of Heterogeneous Effects

Table V presents that the data cannot reject the hypothesis that the estimated "treatment" effect of union recognition, as presented above, are constant along three dimensions by which the union certification effect

---

[25] For an example of this in practice, see LaLonde [1986].

[26] Equation 3 is equivalent to

$$y^*_{t+1} - X\widehat{\delta} = WIN_t\beta + X\gamma - X\widehat{\delta} + \varepsilon$$

where $\widehat{\delta}$ is the OLS coefficient of y^*_{t+1} on X (using the whole sample), which leads to

$$E\left[y^*_{t+1} - X\widehat{\delta}|v_t\right] = WIN_t\beta + \widetilde{\widetilde{h}}(v_t)$$

where $\widetilde{\widetilde{h}}(v_t) = E[X|v_t]\gamma - E[X|v_t]E\left[\widehat{\delta}\right] + E[\varepsilon|v_t]$. This follows because, denoting x as a particular realization of X, $E\left[X\widehat{\delta}|v_t, x\right] = E\left[X\widehat{\delta}|x\right] = xE\left[\widehat{\delta}|x\right]$ and $xE\left[\widehat{\delta}|x\right] = xE\left[\widehat{\delta}\right]$ (since $\widehat{\delta}$ is estimated using the whole sample) and the law of iterated expectations. Again, as long as X and ε are distributed continuously with respect to v_t at $v_t = 0$, $\widetilde{\widetilde{h}}$ will be continuous there also.

[27] Again, the coefficient on the pre-election union presence variable is not to be interpreted as a causal effect; rather it is more properly thought of as a partial correlation; its inclusion "absorbs" residual variation.

could potentially differ.

First, since we observe survival at one point in time (the year 2001), the "overall" estimate we obtained is an average of 2- through 18-year survival rates. In principle, a *positive* effect on survival in the short-run could be canceling out a long-run *negative* effect, it is instructive to stratify the analysis by groups of years. The 2nd row of the first column presents the estimate from Column (6) of Table IV. The 3rd row reports that among elections that were held before 1988, the corresponding effect of union recognition on the probability of survival is about -0.009 with a standard error of 0.024. The following three rows report the interaction effects for the periods 1988-1991, 1992-1995, and after 1995, respectively. Two of the interaction effects are negligible in magnitude, and the joint hypothesis that all interaction terms are zero cannot be rejected at conventional levels of significance.[28]

Second, the effect could potentially vary significantly by industry. The next three rows in the first column of Table V show modestly larger negative effects for service sector establishments, but again the interactions are not jointly statistically significant. Third, the effects could vary by the size of the voting unit (a rough proxy for initial size of the establishment). The final three rows of Table V show that the estimates are positive (0.016 and 0.016-0.012=0.004) for voting unit sizes between 20 and 40 workers, and for units with greater than 100 individuals. For units between 40 and 100, the estimated effect is 0.016-0.052=-.036. Again all estimates are not statistically meaningful and the null hypothesis of equality among the 3 groups cannot be rejected at conventional levels of significance. On the other hand, it should be noted that the sampling error on the 40-100 interaction is not trivial, so the test of equality of the estimates have relatively low power.

The second column reports that the estimates of the overall effect and the various interactions do not change significantly, when we use an alternative measure of establishment survival. As mentioned in Section 3, we consider that an establishment has survived as of 2001 if the company name and address matches with an entry in the InfoUSA database. However, in principle, if bare losers are much more likely

[28] Specifically, the following specification was estimated: a regression of the survival indicator on year-period dummies and their interactions with the WIN_t indicator and their interactions with v_t, as well as state dummies, industry dummies, unit dummies, log of the number of eligible voters, and the pre-election union presence dummy, as well as two cubic splines in v_t with two knots each (one spline for the $v_t \leq 0$, and the other for $v_t > 0$).

than bare winners to go through an ownership change – and hence change their name – then our primary measure of establishment survival may mask a true effect on establishment closure. A comparatively robust way to address this issue is to consider that an establishment has survived if *any* establishment is present at that exact address as of the year 2001 – irrespective of whether the company name changes. The second column of Table V reports the estimates from using this measure (Survival (2)), and shows that the estimates mirror that of the first column.[29]

On balance, due to the magnitude of our sampling error, while we cannot rule out small heterogeneous effects by these three observable dimensions, we interpret the estimates as indicating that our main estimate is not being wholly driven by a particular subsample, as defined along these observable dimensions.

5.3 Evidence on "Intensive Margin" Effects

Table V also reports regression discontinuity estimates of union impacts on four other measures of employer outcomes: the levels of employment and estimated sales volume, as well as the logs of both variables. Overall, the results are mixed; the employment and sales responses are small relative to the overall variability in outcomes across establishments. However, the sampling errors are too large to rule out large, economically meaningful negative effects on employment (as well as large positive effects on sales volume). The results suggest that more detailed information is needed to absorb the cross-sectional variation in the scale of operation. A more comprehensive set of covariates may allow one to generate informative estimates about the magnitude of union effects on the intensive margin of employment.

The third column reports regression discontinuity estimates for the level of employment, where we have assigned "0" to those establishments that have closed by the year 2001. The point estimate is positive 2.33 with a standard error of 5.51. While the null hypothesis of a zero effect cannot be rejected at conventional levels of significance, neither can a negative response of about 8 employees – about 10 percent of the mean (83.4).

The fifth column reports the estimates for estimated sales volume, where again we have assigned

[29] Apparent from the first row, the obvious exception, as expected, is that the proportion of establishments that have *any* business (regardless of name) as of the year 2001, is significantly higher, at about 0.643.

"0" to establishments no longer in existence. The estimates imply a positive impact on sales volume on the order of 1.5 million dollars, but it is not statistically different from zero. Furthermore, a moderate negative effect could not be ruled out at conventional levels of significance. As with employment, the cross-sectional variability in sales is quite significant, and suggests that even large, economically meaningful effects on sales are small, relative to the cross-sectional heterogeneity in output across establishments, even while including industry dummies.

The fourth and sixth columns report the discontinuity estimates for the log of employment and sales volume, respectively; hence, "dead" or "0 employment" establishments are necessarily dropped from the sample. Again, while the point estimates are not statistically different from zero, we are only able to statistically rule out large effects: for example, a 25 percent negative effect on employment and a 20 percent positive effect on sales volume.

This last set of results, however, should be interpreted with more caution, compared to the other findings of the paper. Although the small estimated effects of unionization survival suggest otherwise, it remains possible that these last set of comparisons are contaminated by sample selection bias.[30] To investigate this point further, we compared the pre-determined characteristics between the bare union winners and the bare union losers – *this time, conditional on the establishments surviving as of 2001*. If the sample selection process were strongly related to unobservable determinants of employment or output, it is plausible to expect the bare winners and losers – conditional on surviving as of 2001 – to be systematically different on observable dimensions. Indeed, our analysis of this selected sample suggests that the *surviving* bare winners and losers are quite similar on the basis of a number of pre-determined characteristics, including the number of eligible voters, and industrial composition. The interested reader is referred to Appendix Table III.

6 Economic Implications

In this paper, we have primarily focussed on assessing the internal validity of using NLRB rep-

[30] See Heckman [1976, 1979].

resentation elections to generate regression discontinuity estimates of the impact of union certification on employer survival. It is impossible to rule out with certainty that strategic voting induces "sorting" around the voting cutoff in such a way to invalidate the quasi-experiment. However, we believe, against the backdrop of a contentious industrial relations atmosphere in the U.S., and given the time between the filing of the petition and the actual election, and given that the election is a secret-ballot vote, it is more plausible that, conditional on petitioning for an election, there is at least some (potentially large) unpredictable component to the exact vote count. In this case, we would expect the establishments where the union barely won to be quite comparable to those where the union barely lost, along both unobservable and observable dimensions. Most importantly, the empirical evidence presented here corroborates this *pre-specified* prediction of the regression discontinuity design.

We believe that our estimates constitute an important first step in empirically assessing the magnitude of economic distortions caused by unions – while seriously considering the potential biases induced by self-selection and omitted variables. Considering it as a first step, we are hesitant to draw any strong inferences regarding the magnitude of potential welfare losses or normative implications. However, it is instructive to consider our findings' implications in light of four issues of economic interpretation.

6.1 Is the effect "economically" small?

Our analysis obtains estimates in the range of -0.01 to -0.02, and the null hypothesis of "no effect" cannot be rejected at conventional levels of statistical significance. However, it is also instructive to consider whether or not the estimates are consistent with an alternative null hypothesis – *that the decline in the union sector in the U.S. in recent years is entirely attributable to the union impact on employer survival*.

Such "back-of-the-envelope" calculations require a great deal of abstraction. However, suppose we consider an economy initially made up of N_0 establishments with identical, constant hazard rate of closure of d per year, and a constant inflow of bN_0 establishments per year. Normalizing $N_0 = 1$, it is straightforward to show that the number of establishments at time t is represented by

$$n(t) = \lambda + (1 - \lambda) e^{-dt} \tag{6}$$

where $\lambda = \frac{b}{d}$, so that if $b = d$, the number of establishments remains constant over time.[31]

Over the period from 1983 to 1998, union density among private sector workers fell by approximately 40 percent. The implied baseline hazard rate from our data of establishments implies a constant hazard of $d = 0.10$.[32] This implies $\lambda = \left(0.6 - e^{-0.10(15)}\right) \frac{1}{1-e^{-0.10(15)}} = 0.485$. Thus, if the decline of the union sector was entirely due to the union effect on the ability of establishments to survive (i.e. if the establishments were not unionized, λ would be 1), then the causal effect of union recognition would have to be a doubling of the hazard rate d. With a base hazard rate of 0.10, this implies that we should expect causal estimates of the 1-15 year survival rates to be on the order of -0.20. To the contrary, however, our causal estimates are centered on -0.01 to -0.02 and we are able to reject magnitudes larger than -0.05. Rather, our point estimates are consistent with unions causing a 0.005 increase in the *probability of dying* (that is, a 5 percent increase in the *rate*) and given our standard errors we can reject anything larger than a 0.02 increase in this probability (i.e. a 20 percent increase in the *rate*).

Thus, our estimates cast considerable doubt on the proposition that the primary mechanism of recent union decline is through union effects on establishment closure. Of course, these rough calculations should be viewed with caution, as they are based on a number of compositional, aggregation, and behavioral assumptions – the empirical relevance of which can only be assessed with large micro-data on establishments in the U.S.

6.2 Is the "treatment" small?

One interpretation of the seemingly small survival rate effects is that the extensive margin distortions induced by unions are, in fact, negligible. *An alternative explanation* is that the impact of union *certification* on survival rates is small because the impact of union *certification* on affecting employees' compensation and working conditions is small – even if an employer's response to a binding collective bargaining agreement is large.[33]

---

[31] This follows from a the differential equation $n'(t) = dn(t) + b$.

[32] Actually, our data seem to fit a Weibull baseline better; however, as usual, it is difficult to know whether the empirical duration dependence is "real" or an artifact of unobserved heterogeneity. The 10 percent hazard rate is also roughly consistent with exit rate estimates from Dunne, Roberts, and Samuelson [1989] and McGuckin, Nguyen, and Reznek [1998].

[33] Another obvious point is that the effect is only identified precisely at the 0-vote margin threshold. In principle, the effect could

This could be occurring in at least two ways. One possibility is that, on the whole, U.S. workers who are attempting to organize ultimately are not demanding much in the way of wages, benefits, or working conditions anyway: the "treatment" itself is small. If this is the case, then – even if the marginal response is large – the net result would be a negligible distortionary response on the part of employers.

Another possibility is that only a tiny fraction of establishments are induced to substantially raise wages and benefits as a consequence of having to recognize the union. In this case, even large responses to higher compensation could be diluted by the many establishments for which the provisions of the NLRA are "non-binding" as a matter of practice.[34]

In this paper, we argue that these possibilities seem incongruent with the notion that employers are strongly opposed to organization drives, and that it seems more plausible that something substantial is at stake with the representation elections. However, whether or not the outcome of a representation election has any measurable impact on compensation and the working conditions is ultimately an empirical question. Should future research show that the outcomes of elections *are* inconsequential in terms of altering compensation and working conditions, then it suggests that, on average, workers who attempt to organize are not particularly aggressive or demanding and/or that the provisions for union recognition and the duty to "bargain in good faith" in the NLRA are simply not effectively enforced as a matter of practice.

6.3 Is "no effect" in survival rates consistent with monopoly unionism?

Put simply, yes. There are many margins on which unions could cause distortionary effects. First, even if employers' probabilities of closure are entirely unaffected, they could cut their level of employment. This could be occurring *within* every establishment in the union sector. Second, even if employment is not affected, if the employer's production techniques are constrained by terms of a contract or union threat,

be small at the 50/50-threshold, but very large when the vote is 90 percent in favor of the union. However, given that the distribution of vote shares look roughly normal, centered around the 50/50-threshold, it is not clear whether an effect – even if we could identify it – at the 90 percent vote share is any more interesting than the effect in the "middle" of the distribution of vote shares.

[34] In principle, one could attempt to estimate the causal effect of "obtaining a first contract". This would involve dividing our estimate by an estimate of the fraction of newly certified bargaining units who secure an initial agreement. An analysis by the Dunlop Commission suggests about 55 percent of newly certified units obtain contracts, suggesting that the point estimates (and standard errors) could be multiplied by 2. On the other hand, it is not entirely clear that union certification would affect an employer only through its effect on securing an initial agreement. It is quite plausible that the simple threat of a union contract may induce the employer to capitulate to worker demands to some degree.

productivity and hence output could fall. Third, the threat of unions could prevent establishments from appearing in the first place. In this mechanism, union threat distorts the allocation of resources even before there is a representation election.

Thus, although the small estimated effects cannot provide us with a full picture of the extent of potential distortionary re-allocation of resources, our findings do suggest that if we are to find large effects, we are likely to find them on the intensive margin.

A similar set of calculations provides an instructive benchmark for what one might expect to observe in teh analogous case where adjustments occur primarily through the "intensive" margin of employment. Interpret $\lambda = \frac{(1-\Delta)b}{a} = 1 - \Delta$ in terms of the proportion of total employees at an establishment, so that Δ is now the "within-employer" response. If we assume that the inflow of new establishments equals the hazard rate, under the null hypothesis that the sole reason for the union decline was due to an intensive margin effect, we would expect the measured causal impact of union certification on employment to be about a 50 percent decline in the level of employment at the establishment. [35]Even given the imprecision of our intensive margin estimates, however, we can easily reject an effect of such magnitude. On the other hand, we cannot rule out a smaller negative intensive margine employment effect of unionization of, say, 15 percent. Given the often cited estimate of the union wage effect of 15 percent [Lewis, 1963] this wuold imply a labor demand elasticity of -1.

Again, whether or not these effects – including those on the intensive margin – are in fact substantial is an empirical question. But examining the extensive margin is a pre-requisite to examining the intensive margin. Even in the ideal scenario (from a purely scientific standpoint) where union recognition was randomly assigned, it would not be straightforward to estimate intensive margin effects on employment if the "treatment" group were more likely to "die" and hence "drop out of the sample". Our findings that there are negligible survival effects and that the survivors in both the bare winners and losers look similar along observable characteristics suggest that the sample selection issue may well be a second order issue.

[35] Alternatively, if the effects were on the margin of affecting the appearance of establishments in the union sector, it is easy to show that if it entirely explained recent union decline in the U.S., the effect would be a quite large effect of 0.50 in probability.

6.4 Is "no effect" unsurprising?

Finally, it may be tempting to conjecture that a small or no effect on survival probabilities is exactly what we might expect if workers or the union leadership were acting rationally. After all, presumably, unionized workers also lose if an employer is forced to close its operations. This is certainly true in the unrealistic case that a collective bargaining agreement led to closure with complete certainty. But if the union effect works through a negative impact on the *probability* of survival, a large effect (say a doubling of a hazard rate of 0.10) could easily be consistent with rationally behaving workers and unions.[36] Thus, arguments purely based on the presumed rationality of workers are not informative in bounding the certification effects on employer survival.

Nonetheless, the notion that the organizing workers and union leadership may be behaving rationally – and have a direct stake in maximizing the viability of the employer – has important implications. Indeed, not only could the union be interested in reducing the risk of closure, but they could be interested in maximizing the profitability of the business – insofar as it would lead, for example, to a larger wage bill for the workers. Indeed, this possibility has been the focus of the "efficient contracts" literature (McDonald and Solow (1981), Brown and Ashenfelter (1986), MacCurdy and Pencavel (1986), Card (1986), Abowd (1989)). The efficient contracts model departs conceptually from the monopoly union/"right–to–manage" model by observing that union members could value both the level of employment and the wage. And if the firm enjoys some economic rents – through monopoly power in the product market, for example – the monopoly union outcome is "inefficient" in the sense that at least one of the two parties can be made better off without making the other party worse off. In some situations – when there is a "strongly efficient" contract – there is a possibility that unions do not lower employment, but instead act to redistribute rents from firms to workers.[37]

[36] If the alternative to the union wage is the market wage, then it is trivial that any worker will prefer the collective bargaining agreement, as long as there is some probability that the establishment will still exist so that they can collect an above-market wage. If the worker's alternative was "above-market" and the union doubled the hazard rate, for the indifferent worker (under risk neutrality) we would have $\frac{W_A - W_M}{1-\delta(1-d)} = \frac{W_u - W_M}{1-\delta(1-2d)}$, where W_U, W_A, and W_M denote the union, alternative, and market wages respectively, and δ is the discount factor, and d is the constant hazard rate for the establishment. If $W_U - W_A$ were 0.10 in logs and δ was 0.95, and d were 0.10, then $W_A - W_M$ would have to be about 0.06 (in logs) – a plausible number – in order for the worker to be indifferent

[37] Indeed, the notion that unions have an interest in the businesses ultimate success is explicitly mentioned by union activists.

7 Conclusion: Directions for Future Research

We believe that a prerequisite to deepening our understanding of both the magnitude and mechanism of potential distortionary effects of unions on employer outcomes, is to learn about the impact on the extensive margin of survival, given that turnover among establishments is an important phenomenon. We also argue that any conclusive empirical analysis must 1) use representative establishment-level micro-data, and 2) seriously address the potential correlation between union status and unobserved determinants of employee outcomes.

In this paper we have attempted to meet these challenges, by exploiting a peculiar feature of the determination of union recognition in the U.S within a regression discontinuity design. Using a dataset of over 27,000 establishments who faced NLRB representation elections, we examine comparisons between bare winners and losers of the elections; the analysis reveals that the impact of union recognition on employer survival is relatively small, implying that any large distortionary effects on employment would likely occur on the "intensive" margin. The analysis also implies that the potential sample selection biases that arise when examining employer outcomes *conditional on survival* may be a second order issue.

Our findings suggest three important areas for future research. First, since the regression discontinuity design generates testable predictions about the comparability of close winners and losers, the collection of even more pre-election "baseline" characteristics will prove useful to subject the design to further specification tests. Our confidence in the internal validity of the design rests on the ability of the design to pass those tests. Second, more detailed longitudinal information about the employer can help absorb the heterogeneity across employers, and should prove useful to generate more informative estimates of magnitude of employment responses on the intensive margin. Finally, there is potential to use the regression

---

Commenting on the first industry-wide contract signed by the United Mine Workers of America (UMWA), president of the UMWA John L. Lewis once noted:

> [with this contract] the industry can apply itself – both management and labor, to the problem of producing coal in quantity *at the lowest cost possible by modern techniques.* The Mine Workers stand for the investors in the industry and for a return on capital. They stand for the public to have coal at the lowest possible price consistent with the Mine Workers having a decent life

discontinuity framework to study the impacts of unionism on wages, the primary mechanism by which a monopoly union could lead to distortionary employment effects.

References

[1] Abowd, John M. “The effect of wage bargains on the stock market value of the firm” *American Economic Review* 79:774–800, September 1989.

[2] Angrist, Joshua D., and Victor Lavy. “Using Maimondies’ Rule to Estimate the Effect of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics* 114 (1998):533-75.

[3] Boal, William M., and John Pencavel, “The Effects of Labor Unions on Employment, Wages, and Days of Operation: Coal Mining in West Virginia,” *Quarterly Journal of Economics* 109 (1994) 267-298.

[4] Bronars, Stephen G., and Donald R. Deere, “Union Organizing Activity, Firm Growth, and the Business Cycle,” *American Economic Review*, 83 (1993) 203-220.

[5] Bronfenbrenner, Kate. “Employer behavior in certification elections and first contracts: Implications for labor law reform” In Sheldon Friedman, Richard Hurd, Rudy Oswald, and Ronald Seeber, editors, *Restoring the Promise of American Labor Law*, pages 75–89. ILR Press, Ithaca, New York, 1994.

[6] Brown, James N. and Orley Ashenfelter. “Testing the efficiency of employment contracts” *Journal of Political Economy* 94:S40–S87, 1986.

[7] Campbell, D. T. “Reforms as Experiments.” *American Psychologist* 24 (1969): 409-29.

[8] Card, David. “Efficient contracts with costly adjustment: Short run employment determination for airline mechanics” *American Economic Review* 76:1045–1071, December 1986.

[9] Dubofsky, Melyvn and WarrenVan Tine *John L. Lewis, A Biography* Quadrangle/The New York Times Book Company, Inc., New York, 1977.

[10] Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. “The Growth and Failure of U.S. Manufacturing Plants.” *Quarterly Journal of Economics* 104 (1989): 671-698.

[11] Freeman, Richard B. and James L. Medoff. *What Do Unions Do?* Basic Books, New York, 1984.

[12] Freeman, Richard, and Morris Kleiner, “The Impact of New Unionization on Wages and Working Conditions,” *Journal of Labor Economics*, 8 (1990) S8-S25.

[13] Gronau, R. “Leisure, home production and work – the theory of the allocation of time revisited.” *Journal of Political Economy* 85 (1977): 1099-1124.

[14] Freeman, Richard B. “Why are unions faring poorly in NLRB representation elections?” In Thomas Kochan, editor, *Challenges And Choices Facing American Labor*. MIT Press, Cambridge, 1985.

[15] Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica* 69 (2001): 201-209.

[16] Heckman, James. “The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and A Simple Estimator for Such Models.” *Annals of Economic and Social Measurement* 5 (1976): 475-492.

[17] Heckman, James J. “Dummy Endogenous Variables in a Simultaneous Equations System.” *Econometrica* 46 (1978): 931-59.

[18] Heckman, James J. “Sample selection bias as a specification error.” *Econometrica* 47 (1979): 153-62.

[19] Lalonde, Robert J. “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *American Economic Review* 76 (1986): 604-620.

[20] Lalonde, Robert J., G. Marschke, and K. Troske, “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 (1996) 155-185.

[21] Lee, David S. “The Electoral Advantage to Incumbency and Voters’ Valuation of Politicians’ Experience: A Regression Discontinuity Analysis of Close Elections” *Center for Labor Economics Working Paper #31.* UC Berkeley, April, 2001.

[22] Lewis, H. Gregg. *Unionism and relative wages in the United States*. University of Chicago Press,

Chicago, 1963.

[23] Macdonald, Ian and Robert M. Solow. "Wage bargaining and employment". *American Econoimc Review* 71:886–908, 1981.

[24] MaCurdy, Thomas and John Pencavel. "Testing the efficiency of employment contracts" *Journal of Political Economy* 94:S3–S39, 1986.

[25] McGuckin, Robert H., Sang V. Nguyen, and Arnold P. Reznek. "On Measuring the Impact of Ownership Change on Labor: Evidence from U.S. Food-Manufacturing Plant-Level Data." in *Labor Statistics Measurement Issues*, J. Haltiwanger, M. Manser, and R. Topel, eds. Chicago, Illinois: University of Chicago Press, 1998.

[26] Thistlethwaite, D., and D. Campbell. "Regression -Discontinuity Analysis: An alternative to the ex post facto experiment." *Journal of Educational Psychology* 51 (1960): 309-17.

[27] Van der Klaauw, Wilbert. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach." *Unpublished manuscript* (1996).

Figure Ia: Probability of Post-Election Union Presence, by Margin of Victory

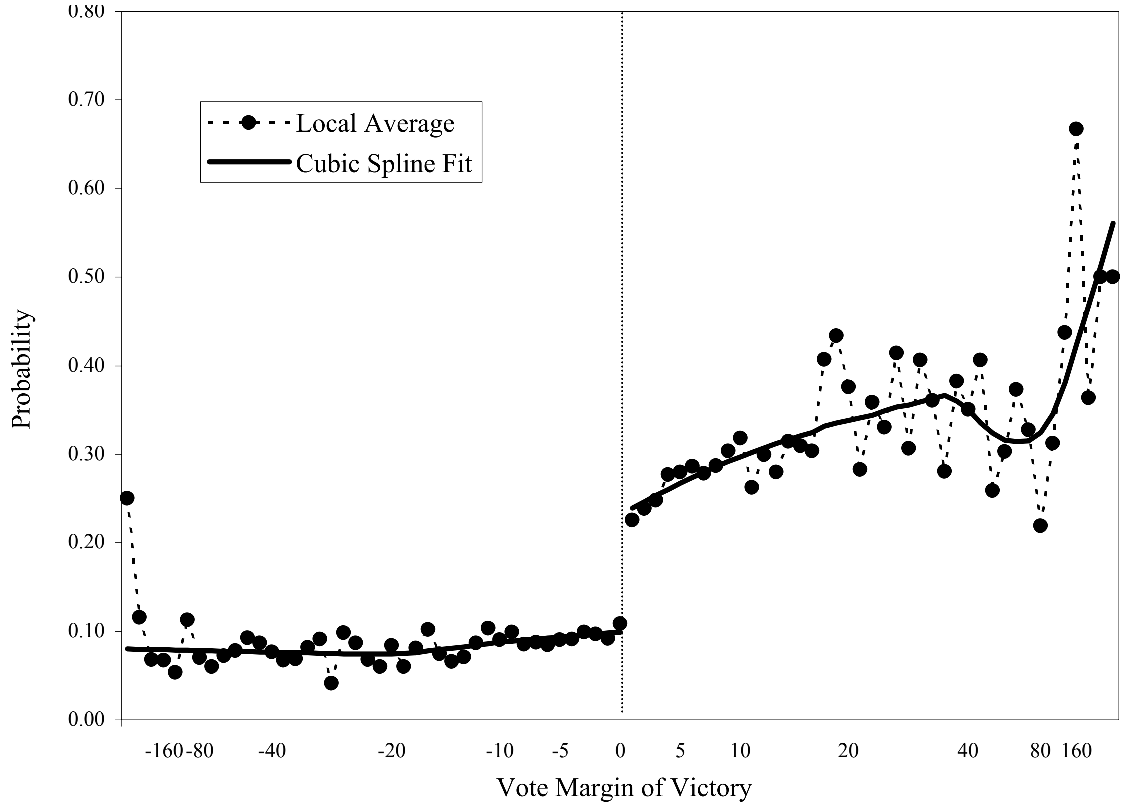


Figure Ib: Probability of Establishment Survival by 2001, by Margin of Victory

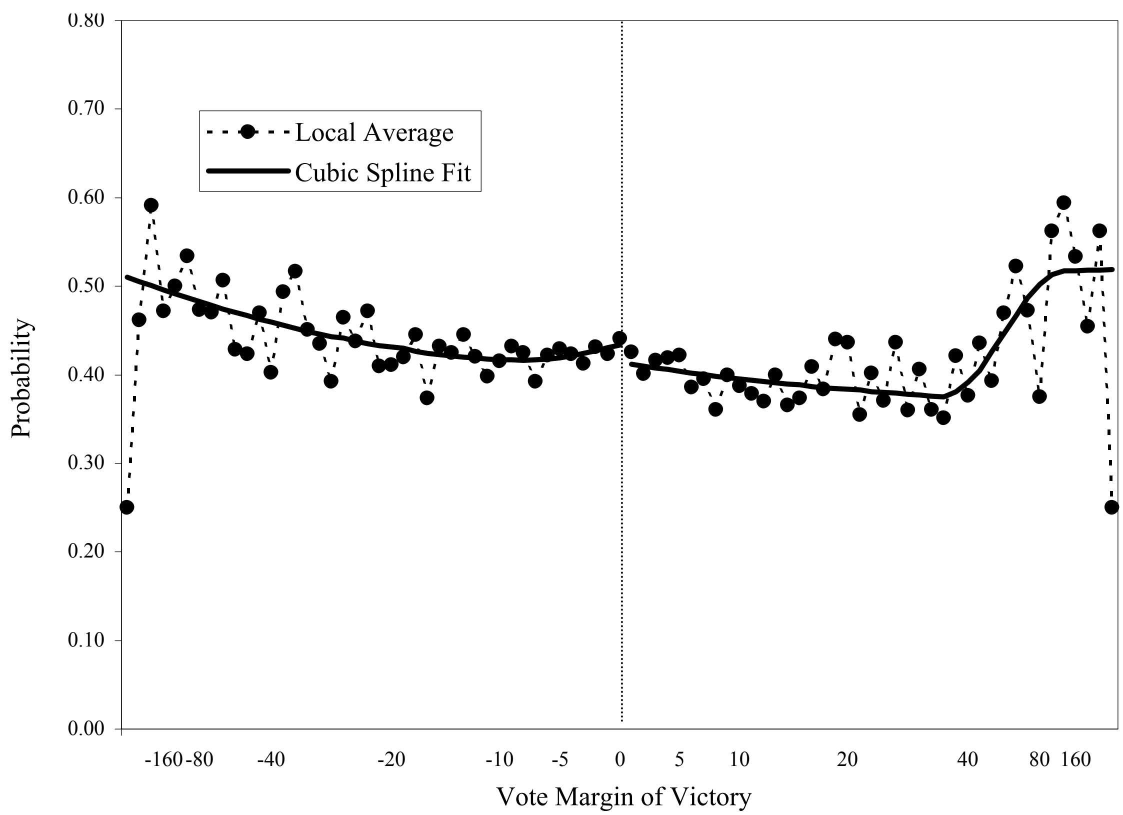


Figure IIa: Probability of Pre-Election Union Presence, by Margin of Victory

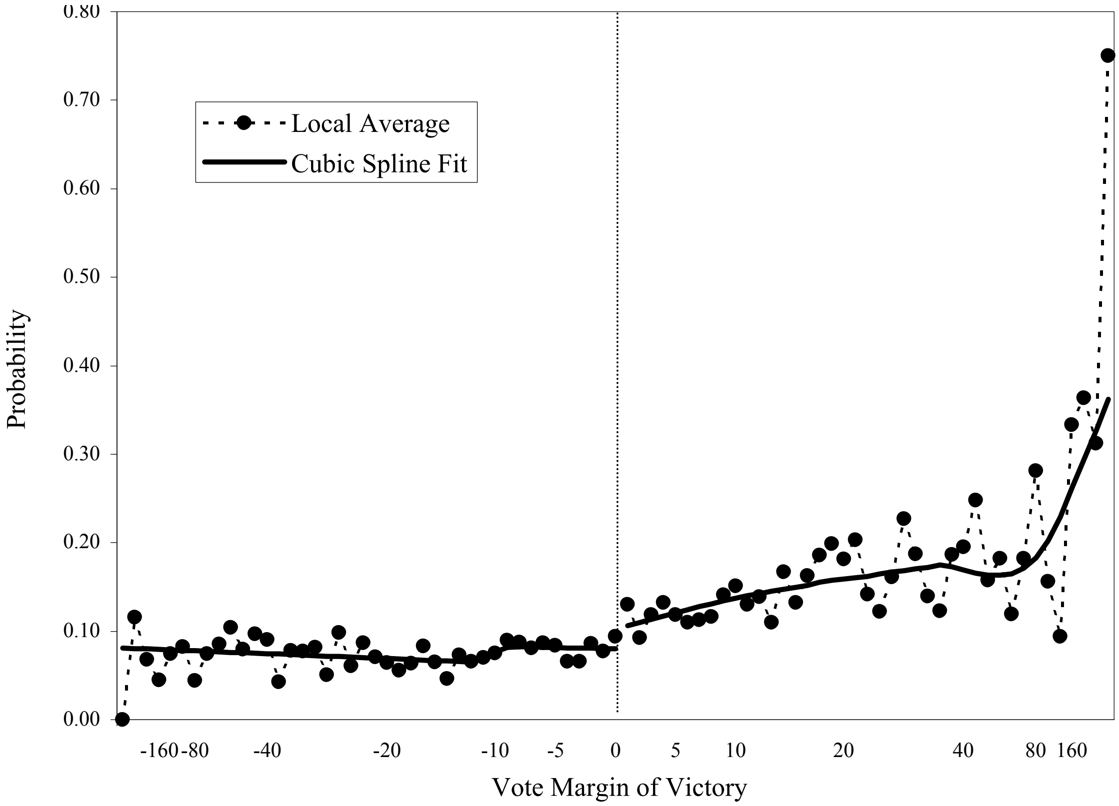


Figure IIb: Log of Total Votes Cast in NLRB Election, by Margin of Victory

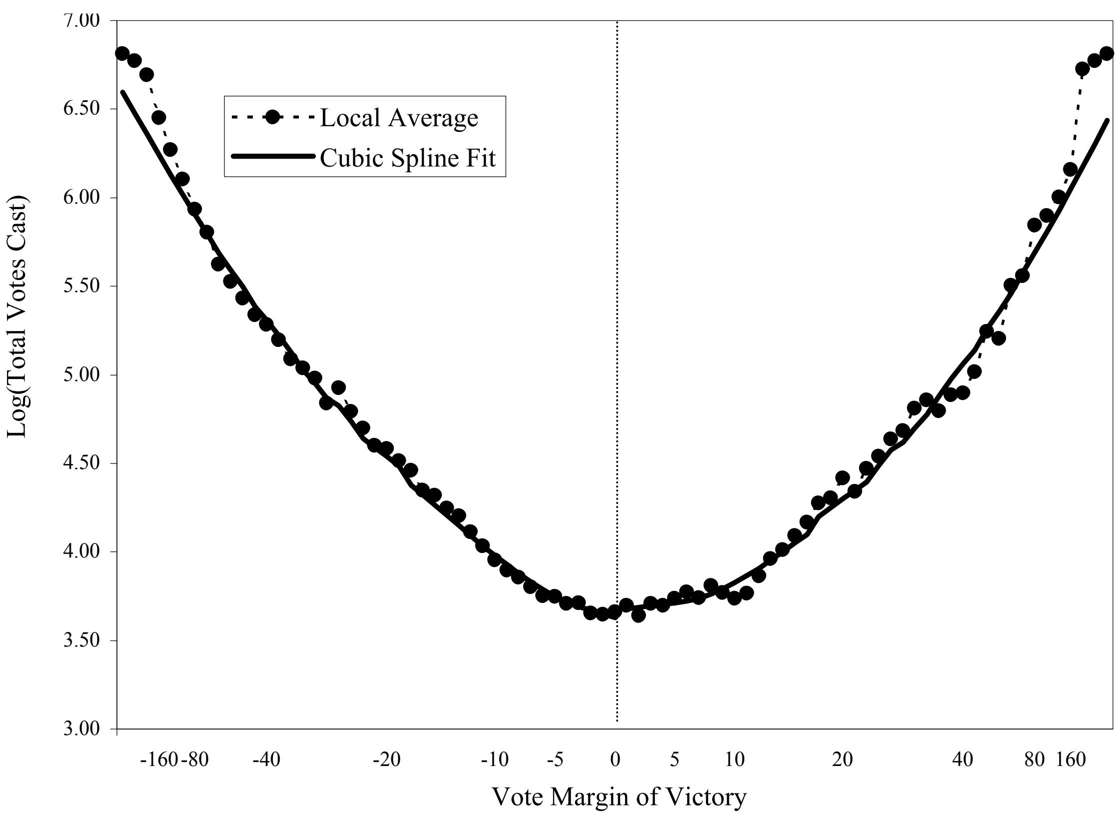


Figure IIIa: Probability Establishment in Manufacturing, by Margin of Victory

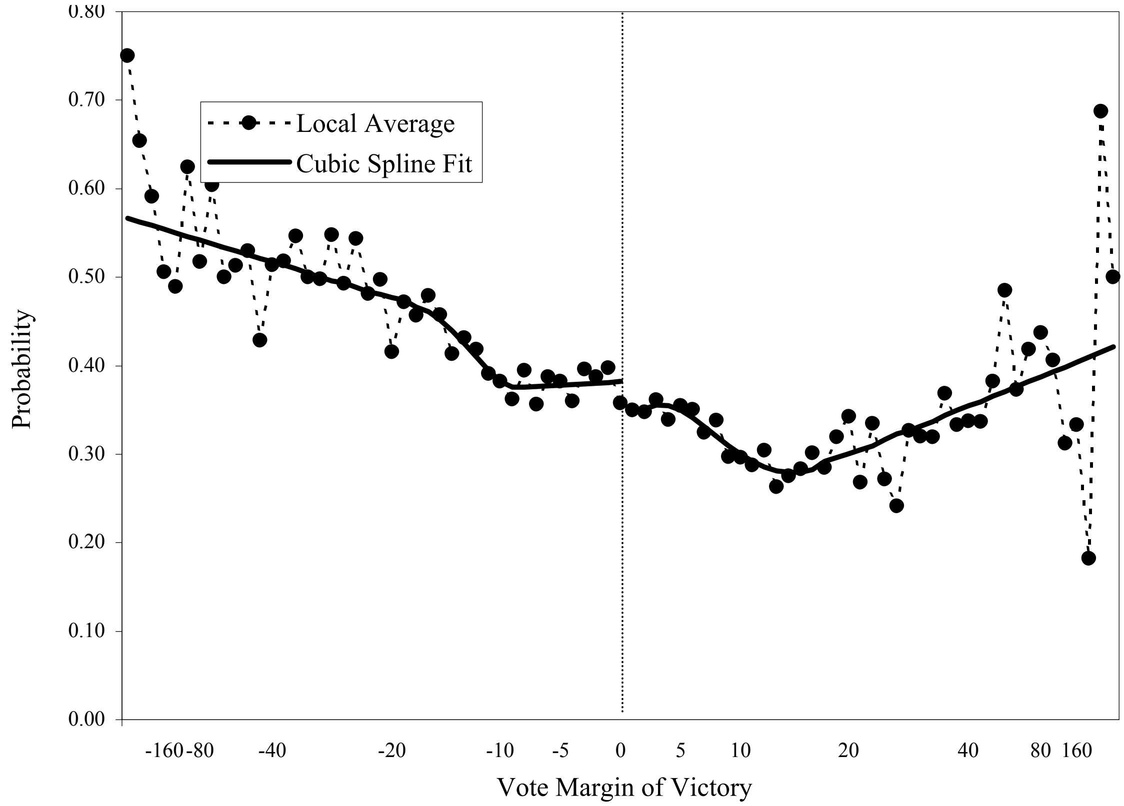
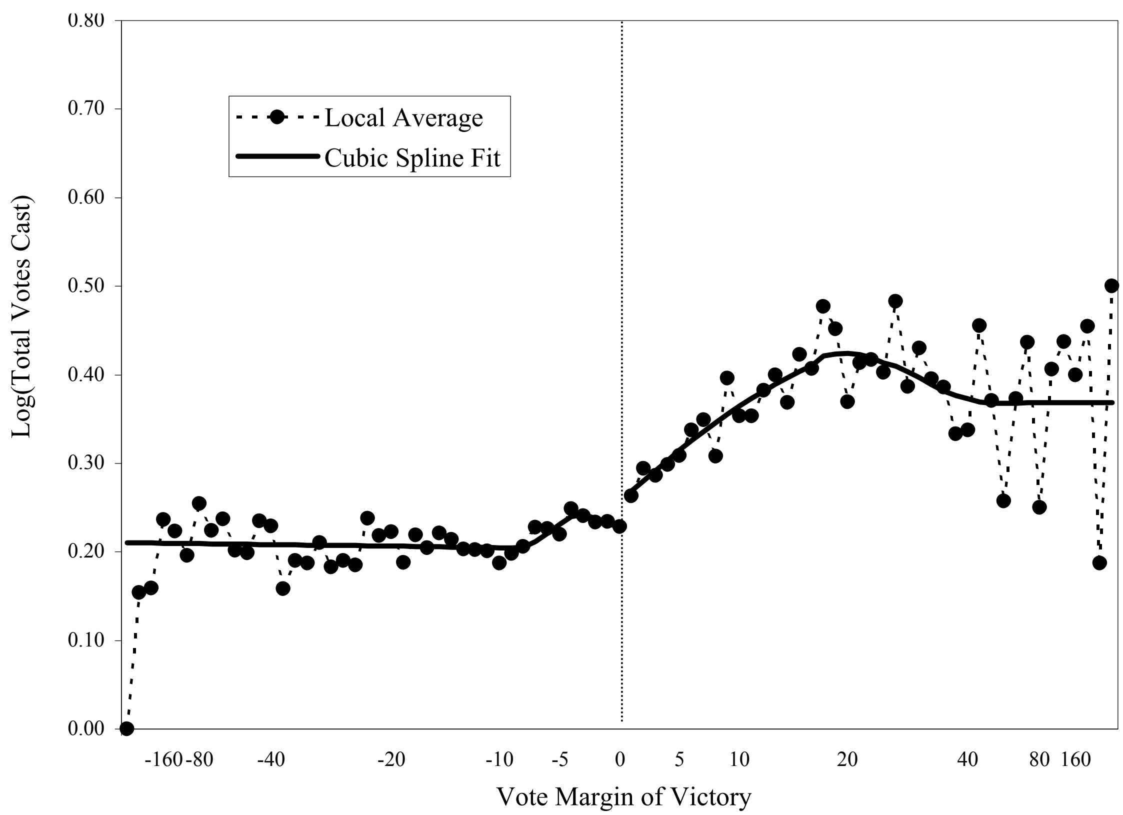
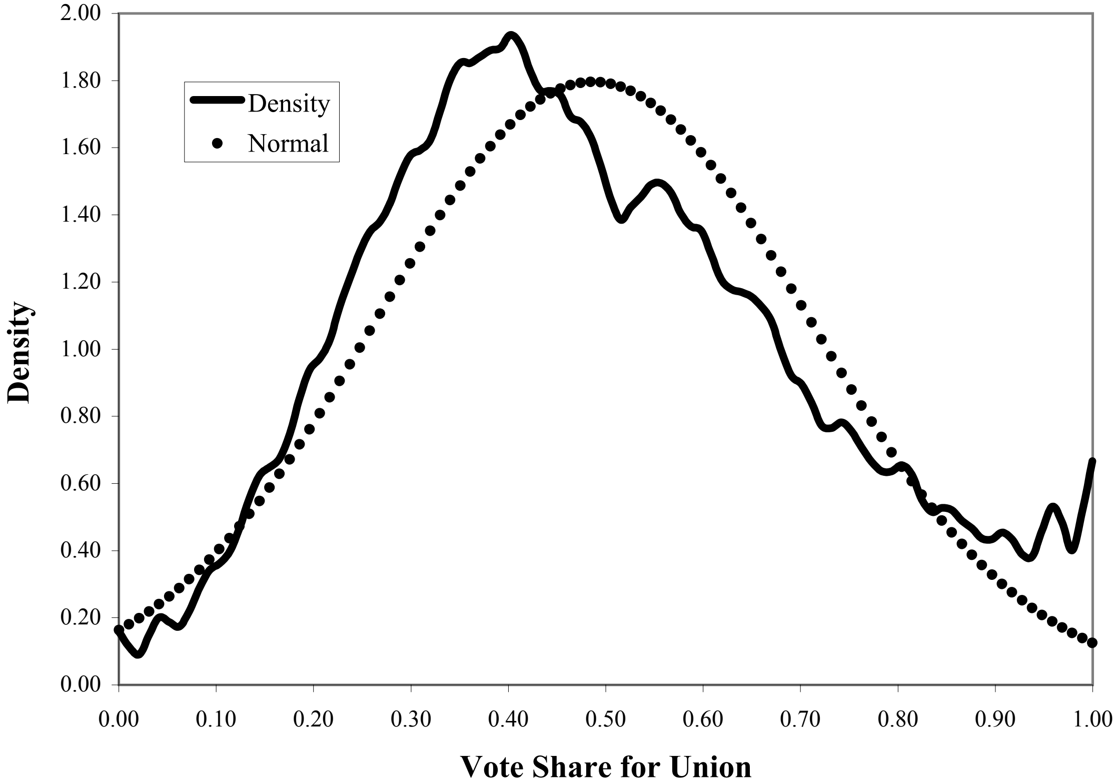


Figure IIIb: Probability Establishsment in Service Sector, by Margin of Victory



Appendix Figure Ia: Estimated Density of Union Vote Share



Appendix Figure Ib: Estimated Density of Union Margin of Victory

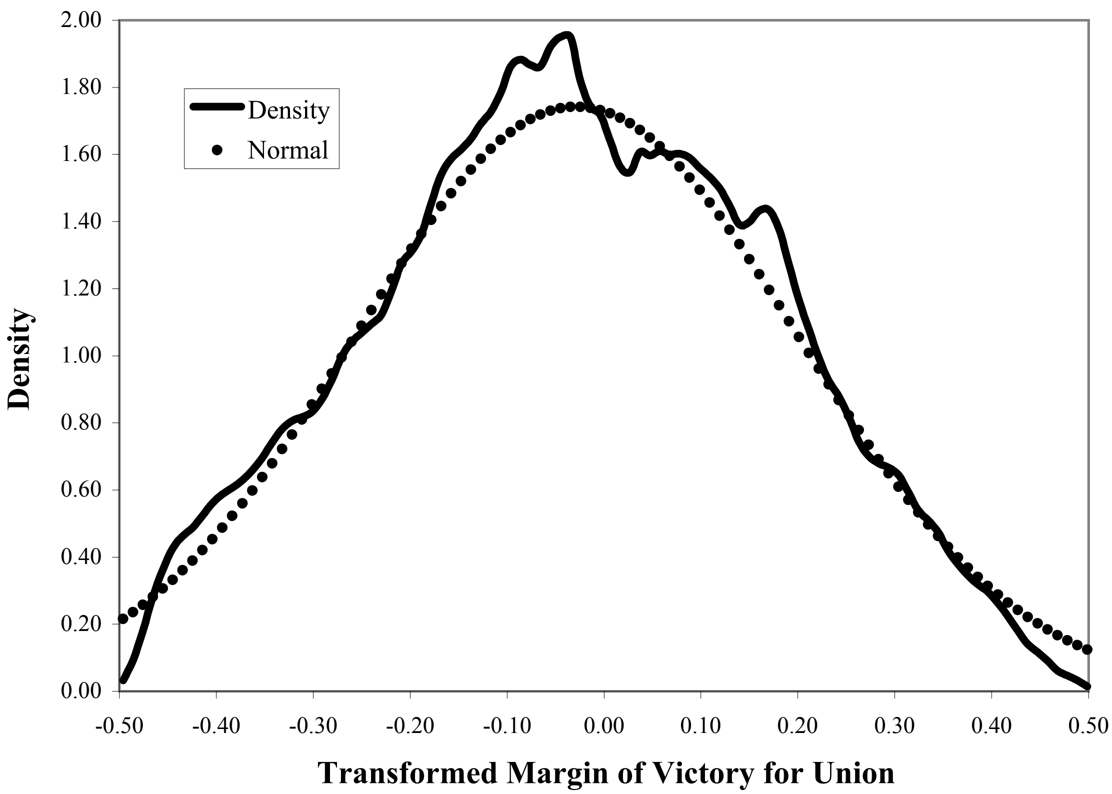


Table I: Establishment Outcomes by Presence of Union, as of 2001: NLRB Elections 1983-1999

Year of	Proportion Survived as of 2001				Mean Employment as of 2001				Mean Log of Employment as of 2001			
Election	All	No Union	Union	Diff.	All	No Union	Union	Diff.	All	No Union	Union	Diff.
<=1983	0.304	0.303	0.308	0.006	65.4	66.8	59.8	-7.0	4.35	4.45	3.95	-0.51
1984	0.277	0.261	0.337	0.076	61.4	48.7	106.1	57.3	4.38	4.34	4.47	0.13
1985	0.312	0.290	0.403	0.112	57.6	42.4	123.5	81.1	4.35	4.29	4.53	0.24
1986	0.309	0.290	0.383	0.093	53.2	45.5	83.2	37.7	4.33	4.31	4.37	0.05
1987	0.327	0.323	0.343	0.021	57.4	49.3	85.8	36.5	4.39	4.41	4.33	-0.08
1988	0.360	0.341	0.425	0.084	68.6	61.9	92.3	30.4	4.43	4.45	4.38	-0.06
1989	0.372	0.340	0.510	0.170	75.8	64.3	126.8	62.5	4.40	4.44	4.28	-0.16
1990	0.393	0.365	0.512	0.146	66.2	58.2	100.9	42.6	4.25	4.31	4.08	-0.24
1991	0.411	0.392	0.498	0.107	84.9	74.7	130.0	55.2	4.37	4.38	4.35	-0.03
1992	0.432	0.397	0.576	0.179	80.0	69.5	124.2	54.8	4.40	4.39	4.43	0.05
1993	0.416	0.390	0.537	0.147	75.1	57.6	154.2	96.6	4.37	4.35	4.41	0.06
1994	0.465	0.433	0.621	0.188	105.4	83.2	212.4	129.2	4.53	4.49	4.66	0.17
1995	0.509	0.486	0.623	0.137	107.1	93.4	177.7	84.3	4.59	4.56	4.70	0.15
1996	0.522	0.495	0.680	0.186	107.5	86.6	226.4	139.8	4.47	4.43	4.64	0.22
1997	0.572	0.563	0.643	0.080	133.7	111.6	296.6	185.1	4.57	4.56	4.61	0.05
1998	0.572	0.564	0.672	0.108	119.8	115.6	172.3	56.8	4.43	4.47	4.02	-0.45
1999	0.578	0.567	0.730	0.163	115.5	99.3	344.4	245.2	4.42	4.36	5.03	0.67
All	0.417	0.402	0.491	0.089	83.4	72.0	137.2	65.2	4.42	4.42	4.43	0.01
	(0.003)	(0.003)	(0.007)	(0.008)	(1.7)	(1.6)	(6.1)	(6.3)	(0.01)	(0.02)	(0.04)	(0.04)
Obs.	27622	22779	4843		26355	21755	4600		10265	8130	2135	

Note: Standard errors in parentheses. Entries are outcomes, as of 2001, of establishments that experienced an NLRB certification election in a given year. Zero is assigned to the employment of "dead" establishments. "Union/No Union" Indicates whether or not the establishment's location appeared in the FMCS contract expiration notice, implying the presence of a union. Details of the merged data from the NLRB, FMCS, and InfoUSA are in the appendix.

Table II: Means of Establishment and Election Outcomes and Characteristics, by Representation Election Outcome, 1983-1999

		N	Full Sample	Union Loss	Union Win	Difference
1	Survival (Indicator Variable), 2001	27622	0.417	0.430	0.400	-0.030
			(0.003)	(0.004)	(0.005)	(0.006)
2	Employment, 2001	26355	83.4	88.3	76.8	-11.5
			(1.7)	(2.2)	(2.7)	(3.5)
3	Log of Employemt, 2001	10265	4.42	4.51	4.30	-0.22
			(0.01)	(0.02)	(0.02)	(0.03)
4	Sales Volume, 2001	25719	14225.3	16250.5	11500.6	-4749.9
			(321.4)	(453.7)	(441.0)	(632.7)
5	Log of Sales Volume, 2001	9629	9.34	9.48	9.14	-0.35
			(0.02)	(0.02)	(0.03)	(0.04)
6	Presence of Union Post-Election	27622	0.175	0.087	0.293	0.206
	(Indicator Variable)		(0.002)	(0.002)	(0.004)	(0.005)
7	Presence of Union Pre-Election	27622	0.102	0.077	0.136	0.060
	(Indicator Variable)		(0.002)	(0.002)	(0.003)	(0.004)
8	Number of Eligible Voters	27622	104.1	113.4	91.6	-21.8
			(0.8)	(1.2)	(1.1)	(1.6)
9	Log of Eligible Voters	27622	4.22	4.29	4.14	-0.15
			(0.01)	(0.01)	(0.01)	(0.01)
10	Number of Votes Cast	27622	91.7	101.9	78.0	-23.9
			(0.7)	(1.0)	(0.9)	(1.4)
11	Log of Votes Cast	27622	4.10	4.18	3.99	-0.19
			(0.01)	(0.01)	(0.01)	(0.01)
12	Manufacturing Sector (Indicator Variable)	27622	0.380	0.421	0.326	-0.094
			(0.003)	(0.004)	(0.004)	(0.006)
13	Service Sector (Indicator Variable)	27622	0.273	0.218	0.348	0.130
			(0.003)	(0.003)	(0.004)	(0.005)
14	Trucking Voting Unit (Indicator Variable)	27622	0.150	0.174	0.119	-0.055
			(0.002)	(0.003)	(0.003)	(0.004)
15	Log of State Employment, Election Year	27622	15.08	15.06	15.10	0.04
			(0.01)	(0.01)	(0.00)	(0.01)
16	Log of State Employment, 2000	27622	15.19	15.18	15.21	0.04
			(0.01)	(0.01)	(0.02)	(0.02)
17	Change In Log Emp. (2000 - Election Year)	27622	0.115	0.117	0.113	-0.005
			(0.001)	(0.001)	(0.008)	(0.008)
18	State Unemployment Rate, Election Year	27622	6.23	6.24	6.22	-0.02
			(0.01)	(0.02)	(0.02)	(0.02)
19	State Unemployment Rate, 2000	27622	4.14	4.11	4.17	0.06
			(0.01)	(0.01)	(0.01)	(0.01)
20	Change in UR (2000 - Election Year)	27622	-2.09	-2.13	-2.04	0.08
			(0.01)	(0.01)	(0.02)	(0.02)

Note: Standard errors in parentheses. Details of the merged data from the NLRB, FMCS, and InfoUSA are in the appendix. Presence of Union post-election (pre-election) indicates whether or not a union at the location of the establishment filed a contract expiration between the election date and 2001 (between the beginning of the FMCS data and the date of the election).

Table III: Establishment Survival, Union Presence, and Pre-determined Characteristics, by margin of victory (loss) in NLRB election

	Full Sample			-7 <= Margin <= 8			-1 <= Margin <= 2			Cubic Spline Fit		
	Loss	Won	Diff.	Loss	Won	Diff.	Loss	Won	Diff.	Loss	Won	Diff.
Survival, 2001	0.430	0.400	-0.030	0.423	0.405	-0.017	0.432	0.413	-0.019	0.434	0.414	-0.020
	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)	(0.009)	(0.012)	(0.012)	(0.017)	(0.011)	(0.009)	(0.014)
Union Present	0.087	0.293	0.206	0.094	0.263	0.169	0.100	0.232	0.132	0.099	0.232	0.133
Post-Election	(0.002)	(0.004)	(0.004)	(0.003)	(0.006)	(0.006)	(0.007)	(0.011)	(0.013)	(0.006)	(0.008)	(0.010)
Union Present	0.077	0.136	0.060	0.080	0.116	0.036	0.086	0.111	0.025	0.080	0.102	0.022
Pre-Election	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.007)	(0.008)	(0.010)	(0.005)	(0.006)	(0.008)
Eligible Voters	113.431	91.602	-21.829	56.602	58.679	2.077	53.455	54.150	0.695	33.225	33.055	-0.170
	(1.156)	(1.088)	(1.636)	(0.672)	(0.778)	(1.023)	(1.301)	(1.385)	(1.900)	(0.973)	(1.352)	(1.666)
Log(Elig. Voters)	4.289	4.138	-0.152	3.811	3.837	0.026	3.748	3.763	0.015	3.720	3.752	0.032
	(0.007)	(0.007)	(0.010)	(0.007)	(0.008)	(0.011)	(0.014)	(0.015)	(0.021)	(0.013)	(0.016)	(0.020)
Votes Cast	101.935	78.033	-23.902	50.752	51.930	1.178	48.477	48.794	0.317	30.026	31.297	1.271
	(1.032)	(0.925)	(1.437)	(0.584)	(0.683)	(0.894)	(1.186)	(1.209)	(1.696)	(0.864)	(1.178)	(1.461)
Log(Votes Cast)	4.184	3.990	-0.194	3.707	3.721	0.013	3.654	3.668	0.013	3.630	3.669	0.039
	(0.007)	(0.007)	(0.010)	(0.007)	(0.008)	(0.010)	(0.014)	(0.015)	(0.020)	(0.013)	(0.016)	(0.020)
Manufacturing	0.421	0.326	-0.094	0.379	0.346	-0.032	0.378	0.349	-0.029	0.382	0.342	-0.040
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.008)	(0.011)	(0.012)	(0.017)	(0.009)	(0.018)	(0.020)
Service Sector	0.218	0.348	0.130	0.233	0.304	0.072	0.231	0.279	0.048	0.227	0.257	0.029
	(0.003)	(0.004)	(0.005)	(0.005)	(0.006)	(0.008)	(0.010)	(0.011)	(0.015)	(0.011)	(0.009)	(0.014)
Trucking	0.174	0.119	-0.055	0.188	0.143	-0.045	0.174	0.145	-0.029	0.174	0.148	-0.025
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)	(0.009)	(0.009)	(0.013)	(0.008)	(0.009)	(0.012)
Number of Obs.	15818	11804		7143	5881		1785	1598				

Note: Robust standard errors in parentheses. Entries are means and differences by election outcome, in three different samples. The fourth set of columns includes all observations (N=27622), and are the estimates based on a cubic spline regression of the variable on the vote margin of victory, fit for winners and losers separately. Details of estimation in text, and of the data in the Data Appendix.

Table IV: Reduced Form Specification: Effect of Union Victory on Probability of Survival (2001), and Post-Election Presence of Union

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Probability of Survival							
Union Victory	-0.020	-0.018	-0.015	-0.015	-0.012	-0.014	-0.013
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
Presence of Union	---	---	---	---	0.093	0.091	---
(Pre-Election)	---	---	---	---	(0.010)	(0.010)	---
Log(Eligible Vote)	---	---	---	---	0.025	0.014	---
	---	---	---	---	(0.005)	(0.005)	---
Year Dummies	No	Yes	Yes	Yes	Yes	Yes	---
State Dummies	No	No	Yes	Yes	Yes	Yes	---
Industry Dummies	No	No	No	Yes	No	Yes	---
Unit Dummies	No	No	No	Yes	No	Yes	---
Probability of Post-Election Union Present							
Union Victory	0.133	0.132	0.132	0.129	0.126	0.123	0.119
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Presence of Union	---	---	---	---	0.421	0.417	---
(Pre-Election)	---	---	---	---	(0.009)	(0.009)	---
Log(Eligible Vote)	---	---	---	---	0.020	0.011	---
	---	---	---	---	(0.004)	(0.004)	---
Year Dummies	No	Yes	Yes	Yes	Yes	Yes	---
State Dummies	No	No	Yes	Yes	Yes	Yes	---
Industry Dummies	No	No	No	Yes	No	Yes	---
Unit Dummies	No	No	No	Yes	No	Yes	---

Note: N=27622. Robust standard errors in parentheses. Upper panel refers to survival, as of 2001 as the dependent variable; lower panel refers to the observation of a contract expiration after the election. All specifications are least squares regressions, for the upper (lower) panel include a natural cubic spline in the transformed vote margin of victory, separately for union wins and losses, with knot points at -.45, -.05, 0.35, and -.45 (-.3, -.15, .35, and .45). Column (7) regresses the residuals - from an initial regresssion of survival (presence of union) on all the covariates - on a natural cubic spline of the transformed vote margin of victory.

Table V: Reduced-Form Results: Impact of Certification on Establishment Outcomes, Overall, and by Year of Election, Industry, and Voting Unit Size

Dependent Variable	Survival	Survival (2)	Empl.	Log(Empl.)	Sales	Log(Sales)
Mean	0.417	0.643	83.400	4.424	14225.250	9.342
(Std. Dev)	(0.493)	(0.479)	(276.486)	(1.461)	(51541.690)	(1.672)
Overall Effect	-0.014	-0.004	2.327	-0.082	1571.021	0.029
	(0.014)	(0.013)	(5.509)	(0.071)	(1041.089)	(0.075)
Year: before 1988	-0.009	-0.006	13.187	0.035	4184.944	0.075
(Main)	(0.024)	(0.021)	(14.131)	(0.099)	(2679.523)	(0.113)
Year: 1988-1991	-0.036	-0.023	-19.719	-0.218	-3942.986	-0.130
(Interaction)	(0.029)	(0.028)	(17.594)	(0.138)	(3178.689)	(0.154)
Year: 1992-1995	0.017	0.017	-11.878	-0.103	-4084.175	0.005
(Interaction)	(0.030)	(0.029)	(18.567)	(0.132)	(3459.451)	(0.151)
Year: after 1995	0.008	0.021	-6.259	-0.170	-1266.068	-0.067
(Interaction)	(0.031)	(0.029)	(20.248)	(0.130)	(3853.649)	(0.154)
Other Industry	-0.001	0.012	12.846	0.023	3161.412	0.106
(Main)	(0.020)	(0.019)	(6.938)	(0.092)	(1721.428)	(0.113)
Manufacturing	-0.004	-0.018	-16.522	-0.235	-3512.184	-0.184
(Interaction)	(0.025)	(0.024)	(11.377)	(0.097)	(2648.038)	(0.129)
Service	-0.036	-0.029	-12.986	-0.112	-2854.243	-0.081
(Interaction)	(0.027)	(0.026)	(16.800)	(0.141)	(2477.028)	(0.159)
El. Vote: < 40	0.016	0.020	8.155	-0.099	2576.833	0.002
	(0.021)	(0.019)	(8.764)	(0.089)	(1589.781)	(0.107)
El. Vote: 40-100	-0.052	-0.028	-9.435	0.018	-997.231	0.015
	(0.028)	(0.027)	(11.530)	(0.121)	(2092.356)	(0.662)
El. Vote: > 100	-0.012	-0.065	13.455	0.060	-5700.273	0.002
	(0.039)	(0.038)	(21.150)	(0.187)	(4117.818)	(0.075)
Number of Obs.	27622	27622	26355	10265	25719	9629

Note: Standard Deviations in first row, robust standard errors otherwise. Survival(2) denotes whether any establishment was present at the exact street address as of 2001, third and fifth columns assign "0" to the dead establishments. Estimated Annual Sales Volume is in thousands of dollars. Specifications: Base specification in second row is a least squares regression on union win indicator, interacted with a cubic spline (see text for knot selection) in transformed vote margin of victory. See text for details on interaction specifications.

Appendix Table I: Summary Statistics for Merged NLRB-FMCS-InfoUSA data, full sample and sample for estimation

	All Certification Elections			Sample for Estimation		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Survival (Indicator Variable), 2001	51132	0.394	0.489	27622	0.417	0.493
Survival (2) (Indicator Variable), 2001	51132	0.623	0.485	27622	0.643	0.479
Employment, 2001	48836	66.0	290.1	26355	83.4	276.5
Log of Employemt, 2001	17863	4.021	1.536	10265	4.424	1.461
Sales Volume, 2001	29367	28602.6	103667.4	16349	32864.9	74883.7
Log of Sales Volume, 2001	29367	8.679	1.831	16349	8.988	1.837
Union Present, Post-Election	51132	0.184	0.388	27622	0.175	0.380
Union Present, Pre-Election	51132	0.117	0.321	27622	0.102	0.303
Eligible Voters	51132	62.604	110.943	27622	104.103	134.922
Log(Elig. Voters)	51132	3.327	1.252	27622	4.225	0.828
Votes Cast	51132	54.986	97.701	27622	91.721	118.772
Log(Votes Cast)	51083	3.195	1.254	27622	4.101	0.822
Manufacturing	51132	0.306	0.461	27622	0.380	0.485
Service Sector	51132	0.260	0.439	27622	0.273	0.446
Trucking	51132	0.177	0.382	27622	0.150	0.357
Log(State Emp.), Year of Election	50284	15.078	0.871	27622	15.079	0.853
Log(State Emp.), Year 2000	50284	15.193	0.867	27622	15.194	0.850
Change in Log(Emp.)	50284	0.115	0.097	27622	0.115	0.096
State UR, Year of Election	50284	6.227	1.807	27622	6.230	1.811
State UR, Year 2000	50284	4.155	0.870	27622	4.137	0.845
Change in UR	50284	-2.072	1.696	27622	-2.093	1.704

Note: Sample restriction is that the number of votes cast be greater than or equal to 20, and the establishment is within the 50 U.S. states and District of Columbia.

Appendix Table II: NLRB Elections: Win rates and Average Union Vote Share, 1983 - 1999

Year of Election	Obs.	Winning Percentage	Vote Share for Union Mean	Vote Share for Union Std Dev.
<=1983	721	0.376	0.463	0.199
1984	1802	0.416	0.474	0.220
1985	1810	0.403	0.471	0.217
1986	1793	0.408	0.476	0.222
1987	1768	0.445	0.494	0.229
1988	1809	0.425	0.493	0.227
1989	1890	0.442	0.496	0.229
1990	1846	0.413	0.488	0.224
1991	1648	0.397	0.474	0.221
1992	1446	0.431	0.488	0.223
1993	1710	0.447	0.499	0.222
1994	1625	0.431	0.485	0.226
1995	1508	0.417	0.477	0.215
1996	1629	0.427	0.487	0.218
1997	1760	0.445	0.492	0.222
1998	1741	0.455	0.505	0.218
1999	1116	0.470	0.511	0.226
All	27622	0.427	0.487	0.222

Note: Restricted sample of elections with 20 or more valid votes cast. See Appendix Table I.

Appendix Table III: Union Presence, Pre-determined Characteristics for Surviving Establishments, by margin of victory (loss) in NLRB election

	Full Sample			-7 <= Margin <= 8			-1 <= Margin <= 2			Cubic Spline Fit		
	Loss	Won	Diff.	Loss	Won	Diff.	Loss	Won	Diff.	Loss	Won	Diff.
Union Present	0.097	0.371	0.274	0.109	0.332	0.223	0.114	0.280	0.166	0.118	0.288	0.170
Post-Election	(0.004)	(0.007)	(0.008)	(0.006)	(0.010)	(0.011)	(0.012)	(0.019)	(0.021)	(0.008)	(0.015)	(0.016)
Union Present	0.092	0.180	0.087	0.100	0.148	0.048	0.117	0.125	0.008	0.110	0.118	0.008
Pre-Election	(0.004)	(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.012)	(0.014)	(0.018)	(0.007)	(0.011)	(0.013)
Eligible Voters	119.913	99.238	-20.675	57.973	61.084	3.111	56.046	55.619	-0.427	33.771	32.573	-1.198
	(1.995)	(2.068)	(2.956)	(1.147)	(1.364)	(1.772)	(2.095)	(2.188)	(3.044)	(1.626)	(2.382)	(2.884)
Log(Elig. Voters)	4.326	4.179	-0.147	3.829	3.863	0.034	3.785	3.780	-0.005	3.753	3.768	0.015
	(0.011)	(0.013)	(0.017)	(0.012)	(0.014)	(0.018)	(0.023)	(0.025)	(0.035)	(0.021)	(0.027)	(0.034)
Votes Cast	107.914	84.597	-23.317	51.880	54.488	2.608	51.027	50.383	-0.644	30.449	31.132	0.683
	(1.773)	(1.750)	(2.586)	(0.952)	(1.202)	(1.513)	(1.903)	(1.958)	(2.747)	(1.423)	(2.106)	(2.541)
Log(Votes Cast)	4.224	4.037	-0.188	3.729	3.755	0.026	3.695	3.690	-0.005	3.667	3.689	0.023
	(0.011)	(0.012)	(0.017)	(0.011)	(0.014)	(0.018)	(0.023)	(0.025)	(0.034)	(0.021)	(0.027)	(0.034)
Manufacturing	0.450	0.335	-0.115	0.411	0.368	-0.043	0.401	0.378	-0.023	0.416	0.387	-0.029
	(0.006)	(0.007)	(0.010)	(0.009)	(0.011)	(0.014)	(0.019)	(0.020)	(0.027)	(0.015)	(0.021)	(0.026)
Service Sector	0.234	0.380	0.147	0.251	0.330	0.079	0.248	0.284	0.035	0.247	0.263	0.016
	(0.005)	(0.008)	(0.009)	(0.008)	(0.010)	(0.013)	(0.016)	(0.019)	(0.025)	(0.017)	(0.029)	(0.034)
Trucking	0.168	0.112	-0.055	0.177	0.134	-0.043	0.155	0.139	-0.016	0.168	0.156	-0.012
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.011)	(0.014)	(0.014)	(0.020)	(0.012)	(0.011)	(0.016)
Number of Obs.	6092	4173		2722	2099		701	582				

Note: Robust standard errors in parentheses. Entries are means and differences by election outcome, in three different sub-samples of those establishments that survive by the year 2001 with non-mssing employment. The fourth set of columns includes all observations (N=10265), and are the estimates based on a cubic spline regression of the variable on the vote margin of victory, fit for winners and losers separately. Details of estimation in text, and of the data in the Data Appendix.