Union Wage Setting and the Distribution of Employees' Earnings: Evidence from Certification Elections

Brigham R. Frandsen*

Massachusetts Institute of Technology

July 12, 2010

Abstract

This paper estimates the causal effect of unionization on the distribution of employee earnings using a regression discontinuity design based on union certification elections. The results suggest unions raise the lower end of the distribution by around 25 percent, with a much smaller, even negative effect on the upper tail, and little effect on average earnings. Results on worker retention suggest unions decrease turnover among lower-productivity workers, but increase it among higher-productivity workers. The empirical results are consistent with a model of the political economy of union wage setting in which unions pursue a wage schedule to maximize the probability of winning a certification election, subject to a minimum profit constraint for the employer. The optimal union wage schedule pays low-skilled workers above marginal product but reduces the return to skill. The estimates suggest that around 13.5 percent of the increase in the variance of log earnings from 1979 to 2009 can be accounted for by falling U.S. private sector unionization rates.

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

1 Introduction

How do unions affect the earnings distribution? This question is at the heart of the debate over the causes of increasing U.S. inequality over the past three decades. While market forces such as international trade and the supply of and demand for skilled labor have probably played a role (Juhn, Murphy, and Pierce, 1993; Katz and Murphy, 1992), institutional forces such as falling unionization rates may also have contributed (Freeman, 1993; Blau and Kahn, 1996; DiNardo, Fortin, and Lemieux, 1996; Card, 2001). From 1979 to 2009 the U.S. private sector unionization rate fell from about 25 percent to 8 percent (Bureau of Labor Statistics, 2010; DiNardo, Fortin, and Lemieux, 1996). To estimate the impact of this drop in the unionization rate, DiNardo, Fortin, and Lemieux (1996) construct counterfactual wage densities based on observed characteristics and show unionization is associated with substantial wage compression. If this compression reflects the causal effect of unionization, then deunionization accounts for a significant part of the increase in U.S. earnings inequality.

Comparisons between the earnings of unionized and non-unionized workers robustly show a positive union wage gap, especially in lower skill groups, but recent efforts to estimate the causal effect of unionization have generated mixed results.[1] Quasi-experimental evidence from DiNardo and Lee (2004), for example, shows little effect on employer outcomes, apparently at odds with regression-based comparisons. At the same time, this small average effect may mask significant, but offsetting effects on different features of the distribution.

The main objective of this paper is to identify the causal effect of unionization on the distribution of employee earnings. The target of estimation can be understood in terms of a hypothetical experiment where a set of establishments are randomly assigned to be unionized or not. The causal effect of interest is the difference between the subsequent distribution of earnings among employees at the unionized and non-unionized establishments. To approximate this hypothetical experiment, this paper adapts DiNardo and Lee's (2004) regression discontinuity (RD) design based on union certification elections, using administrative records on individual earnings matched to establishment-level election results. If establishments where the union barely won and barely lost

[1] Lewis (1986) surveys the large early literature on union wage gaps, while Blanchflower and Bryson (2003) offer an updated look using newer data. Quasi-experimental evidence on the causal effect of unionization includes DiNardo and Lee (2004) and Lee and Mas (2009).

are otherwise comparable, the resulting difference in the distribution of employees' earnings is due to the causal effect of unionization.

The paper motivates the possibility of distributional effects by analyzing the union wage-setting problem in the context of a certification election. The theoretical section of the paper highlights how the the need to garner the support of a majority of workers can lead to a wage schedule that raises the wages of lower-productivity workers but reduces the wages of higher-productivity workers, resulting in a wage-compressing effect of unionization.

Consistent with DiNardo and Lee (2004), the RD estimates reported here show little effect of unionization on average earnings. At the same time, my results provide clear evidence of a distributional effect. Specifically, unionization raises the lower tail of the earnings distribution by around 25 log points, while reducing earnings at the very high end. Further empirical results show unionization increases turnover for higher productivity workers, but not for lower productivity workers, consistent with the model's interpretation of union wage compression as reflecting a wage schedule that raises lower-productivity workers' wages but reduces the return to skill.

The remainder of the paper proceeds as follows. The next section describes the U.S. private sector unionization process and highlights the theoretical implications this process has for the distribution of earnings. Section 3 describes the data used in the empirical work. Section 4 lays out the research design and the econometric framework for identifying and estimating the effect of unionization on the distribution of earnings, and section 5 presents the estimation results. Section 6 summarizes the findings and concludes.

2 Background

2.1 NLRB election process

Since 1935, most U.S. private sector unionization has been governed by the National Labor Relations Act (NLRA), which specifies the rights of unionizing workers. While an employer may voluntarily bargain with the workers' chosen representative, or in some cases may be required to do so even without an election, the traditional process by which workers unionize is through a National Labor

Relations Board (NLRB) secret ballot election.[2] Although in practice an organizing drive is often fraught with disputes and delays, the following steps describe the stylized path a group of workers follows to form a union:[3]

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

2. Election: Eligible workers vote for or against the union, and the union wins if it receives a simple majority (50 percent + 1) of the votes cast.

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

4. Bargaining: The employer negotiates with union representatives over a collective bargaining agreement. If an agreement is reached, the contract becomes binding for all employees in the unit.

NLRB certification elections may include two or more competing unions on the ballot. In the case of multiple competing unions, a simple majority is still required for certification. Elections may also be held to remove union representation altogether (decertification) or to replace one union with another. These cases, however, occur relatively infrequently, and the analysis focuses on certification elections.

2.2 Theoretical framework

How do the incentives a union faces in a certification election affect wage setting? The model developed here takes as its starting place that the primary objectives of union leaders are survival and expansion of the organization, and retention of their offices (Atherton, 1973; Ross, 1948; Berkowitz,

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

[3] The simple process laid out here follows the procedures described in NLRB (2010). See Ferguson (2008) and DiNardo and Lee (2004) for a more complete description of the possible complications and objections that can be raised at each step.

1954; Ashenfelter and Johnson, 1969). Accordingly, a union facing a certification election will pursue a wage agreement that maximizes the probability of winning. The theoretical framework in this section is similar to Farber (1978), Booth (1995), Acemoglu, Aghion, and Violante (2002), and Lee and Mas (2008), who also consider the effects of majority-rules politics on union wage policies, but I focus on the implications for the distribution of workers' wages.

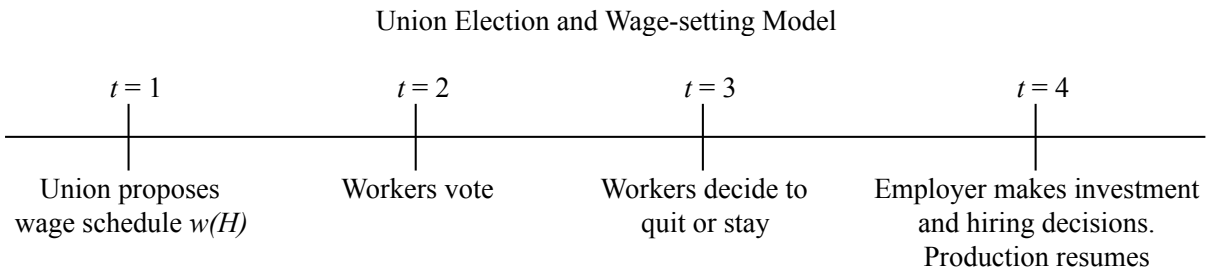

Figure 1: The model

The model casts the union election and wage-setting process as a four-stage game. The stages, shown in Figure 1, follow the stylized union certification process described above. In the first stage, a union petitions to represent the workers at a plant (currently producing in the competitive sector) and proposes a wage schedule, $w(H) = (v + r(H))H$, which gives a worker's wages as a function of his or her human capital, H. The outside (competitive) price of human capital is v, and $r(H)$ denotes the union rent earned by a worker with human capital H. The term "union rent" hightlights that $r(H)$ is the difference between the competitive and union price of skill. In the second stage, workers vote for the union if their union rent, $r(H)$, exceeds their individual cost of union representation, η, provided the wage schedule doesn't cause the plant to shut down. The voting rule can therefore be written $1(\eta \leq r(H))$, subject to the plant not shutting down. The cost of union representation η reflects any pecuniary (e.g., union dues) or nonpecuniary factors affecting workers' preferences for union representation outside of wage differences, and is assumed to be independent of H. In the third stage, workers decide whether to quit or stay, after observing the outcome of the election. Workers stay if their union rent exceeds their cost of unionization, net of an individual-specific switching cost, ε, also independent of H. The worker's decision rule after a union victory is therefore $1(\eta - \varepsilon \leq r(H))$, where the worker stays if the indicator is equal to one. Finally, the employer makes investment and hiring decisions to maximize profits.

The production technology combines H with another factor, K, in fixed proportions with con-

stant returns to scale, so the production function can be written $Y(H) = yH$. Normalizing the H/K ratio to unity, the competitive return to K is $y - v$. A fraction ϕ_K of K is sunk in the production relationship.

Two assumptions in the model setup are important for the results. The first is that workers' decisions to vote for a union or quit their job depend on more than simply a comparison of wages. These other factors are modeled as union costs η and switching costs ε. The presence of the union cost in the model prevents the union from being able to assure 100 percent vote share with only an infinitesimal wage increase to all workers. The presence of the switching cost prevents the unrealistic scenario that all workers who vote against the union quit after a union victory, and thus widens the scope for the union to cater to specific groups of workers. The union and switching costs are assumed to reduce the effective union and outside return to human capital, respectively. While the specific functional form is a simplification, the substantive assumption that the "equivalence premium"—the dollar amount by which a union would have to raise a worker's earnings to make him indifferent between a union and no union—is on average increasing in a worker's outside wage is important. A concrete motivating example is that union dues are commonly collected as a percentage of wages. The consequence of this assumption for the model is that it makes it more efficient for the union to shift resources to attract the votes of workers with lower outside options.

The second key assumption is that employers face short term rigidities in adjusting inputs. The model captures this in a simple way following Caballero and Hammour (1998) by assuming the chosen technology takes fixed ratios of inputs. The assumption is that while technology can adjust over time, it is essentially fixed over the period relevant for an initial collective bargaining agreement. Consequently, employers cannot undo the effects of a union wage schedule by immediately adjusting the production inputs. An alternative way of capturing this that leaves the results substantively unchanged is to allow the employer to fully adjust K, but have the union impose firing costs.

The union's optimal wage schedule is shown in the theory appendix to maximize the expected vote share, subject to a minimum profit constraint for the employer:

$$\max_{r(h)} \int F_\eta\left(r\left(h\right)\right) dF_H\left(h\right)$$
$$\text{s.t.} \int \left(\phi_K\left(y - v\right) - r\left(h\right)\right) hF_{\eta-\varepsilon}\left(r\left(h\right)\right) dF_H\left(h\right) \geq 0.$$

Concrete functional forms illustrate the main implications of the model. Assume H is log-normally distributed and η and ε are exponentially distributed with parameters λ_η and λ_ε, where the mean switching cost, $1/\lambda_\varepsilon$, exceeds the average rent, $\phi_K(y-v)$. In this case the optimal rent schedule is

$$r(H) = \begin{cases} r^+(H,\lambda^*) & , \quad H \leq h_1(\lambda^*) \\ -\left(\frac{1}{\lambda_\varepsilon} - \phi_K(y-v)\right) & , \quad H > h_1(\lambda^*) \end{cases}, \tag{1a}$$

where $r^+(H,\lambda)$ satisfies

$$\exp(\lambda_\eta r) = \frac{\lambda_\eta}{\lambda H} + \frac{\lambda_\varepsilon}{\lambda_\eta + \lambda_\varepsilon}\left(1 + (\phi_K(y-v) - r)\lambda_\eta\right). \tag{1b}$$

The threshold above which rents are negative is

$$h_1(\lambda) = \frac{1}{\lambda}\frac{\lambda_\eta + \lambda_\varepsilon}{1 - \lambda_\varepsilon \phi_K(y-v)}, \tag{1c}$$

and λ^* is the value of λ that satisfies the profit constraint with equality.

The optimal union wage schedule, (1), has two features with stark implications for the union effect on the distribution of earnings. First, the union rent is positive for lower-productivity workers, and negative for higher-productivity workers. Interestingly, the less the union can extract from the employer (i.e., the smaller is ϕ_K), the lower the threshold above which rents are negative. The second feature is that even where rents are positive, they are decreasing in the level of human capital:

$$\left.\frac{dr}{dH}\right|_{H=h\leq h_1} = -\left(\lambda h^2\left(\frac{\lambda_\varepsilon}{\lambda_\eta + \lambda_\varepsilon} + \exp(\lambda_\eta r)\right)\right)^{-1} < 0.$$

Thus compared to the competitive equilibrium, the union wage schedule compresses the distribution of potential wages and shifts it to the right.

To see the effects on the distribution of earnings directly, I solve a numerical example based on the distribution of wages observed in the the sample of full-time, nonunion workers from the 1998-2000 Current Population Survey (CPS). Taking the hourly wage as a measure of human capital, H, I assume the sunk fraction of K is $\phi_K = .3$. Normalizing the competitive return to human capital to be $v = 1$, and assuming labor's share in income is about .7, I set the production function parameter to be $y = 10/7$. Finally, I assume unionization and switching costs both have means of

$1.5 \times \phi_K (y - v)$.

The optimal wage schedule in this example distributes rents disproportionately to low-wage workers and substantially compresses the distribution of earnings. Figure 2 plots the union premium (in dollars), $r(H) \times H$, implied by the optimal union wage schedule as a function of human capital. The union premium is positive for lower levels of human capital, and decreasing and eventually negative for higher levels of human capital. Figure 3 compares the distribution of potential union

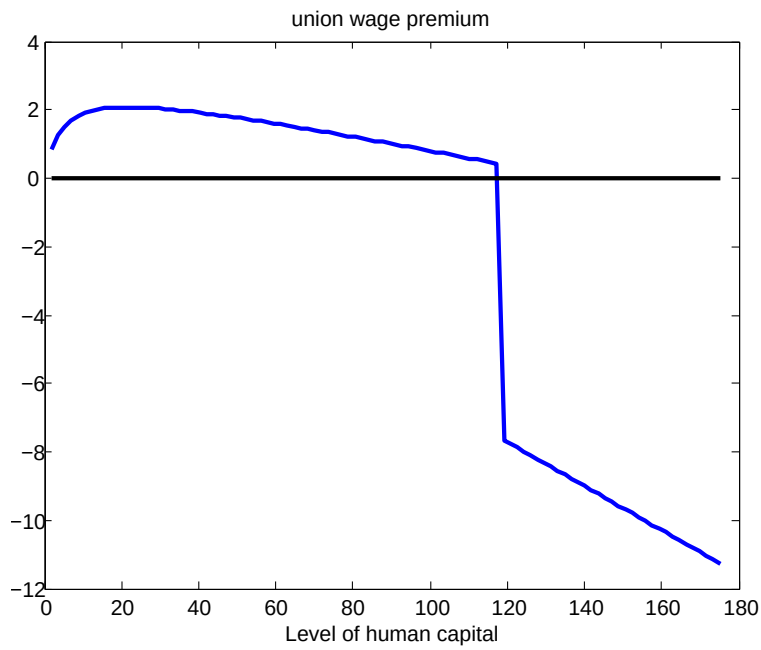

Figure 2: Union wage premium as a function of human capital for parameter values $\phi_K = .3$, $v = 1$, $y = 10/7$, $\lambda_\eta = \lambda_\varepsilon = (1.5\phi_K (y - v))^{-1}$.

and non-union log wages implied by this example. The union distribution is compressed and shifted to the right relative to the non-union distribution.

Given the assumption of fixed K, these implications are likely to apply to the short term. In the long term, however, K can adjust, and the employer may be able to terminate lower-productivity workers, who are being paid above marginal product. In the longer term, therefore, the degree to which the union wage schedule inflates the wages at the bottom of the distribution relative to the top will be attenuated.

The theoretical discussion shows that unions facing certification elections have an incentive to commit to a wage schedule that favors lower-productivity workers at the expense of higher-productivity workers. In the example, the union wage schedule raises the wages of low-productivity

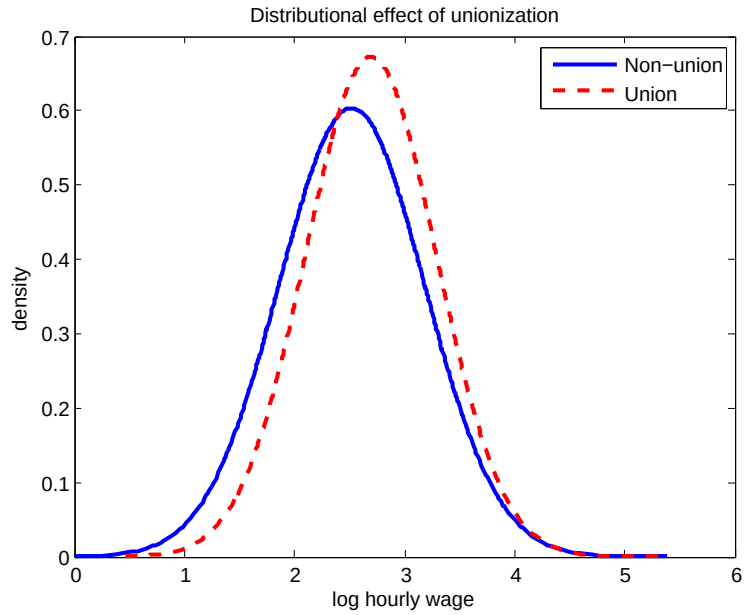

Figure 3: Densities of potential union and non-union log wages implied by the optimal wage schedule from the example in the text.

workers and lowers the wages of high-productivity workers, compressing the distribution of wages.

3 Data

3.1 Union Elections

The analysis uses a dataset on the universe of NLRB union representation election results from 1963 to 2006, which was compiled and analyzed by Ferguson (2008). Each record in this dataset represents a union certification election held at an establishment, and includes the number of votes cast for and against union representation, the date of the election, and the employer's name and address. The data appendix explains in detail how the dataset was constructed.

The main sample used in the analysis covers the years 1992-2001, the period covered by the earnings data described below. This sample contains data on 37,354 representation elections, involving over 1.7 million votes cast. Unions received 50.4 percent of the votes cast, and won 54.5 percent of the elections.[4] Figure 4 shows the distribution of the union vote share in the sample. The mode is around 40 percent, with a significant number of elections in which the union received

[4]The votes-weighted union success rate is only 41.5 percent, indicating that unions fare considerably worse at large establishments, as noted by Farber (1999).

all votes. The figure corresponds to Figure II in DiNardo and Lee (2004), who use similar data. Figure 5 shows the distribution of the union margin of victory in terms of number of votes, close to the threshold. This figure shows that close elections represent the typical case, a fact that is important for the interpretation of the estimation results.

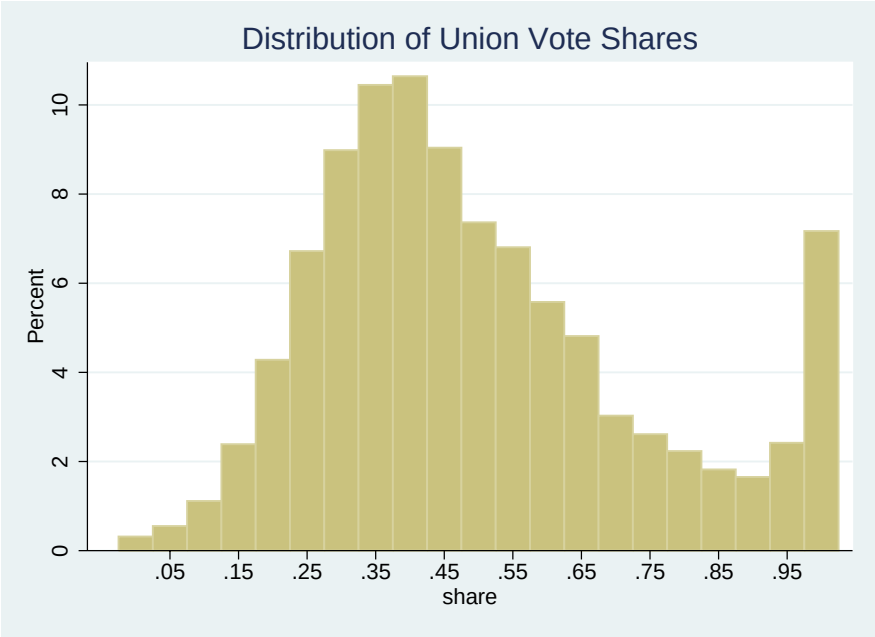

Figure 4: Votes-weighted histogram of the union vote share in representation elections from 1992 to 2001. Data are from NLRB election records, restricted to elections where 10 or more votes were cast.

3.2 LEHD

The second data component contains individual-level earnings from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) database. The LEHD combines data from a wide variety of state and federal administrative records and surveys. In particular, the LEHD integrates the universe of unemployment insurance-covered (UI) earnings records held by participating state agencies into a cohesive data structure using person and employer identifiers, allowing linkages to other sources of data.[5]

The Employment History Files (EHF) within the LEHD contain quarterly records of individuals' UI-covered earnings. The EHF for each of the 23 covered states contains a record for each employee-

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

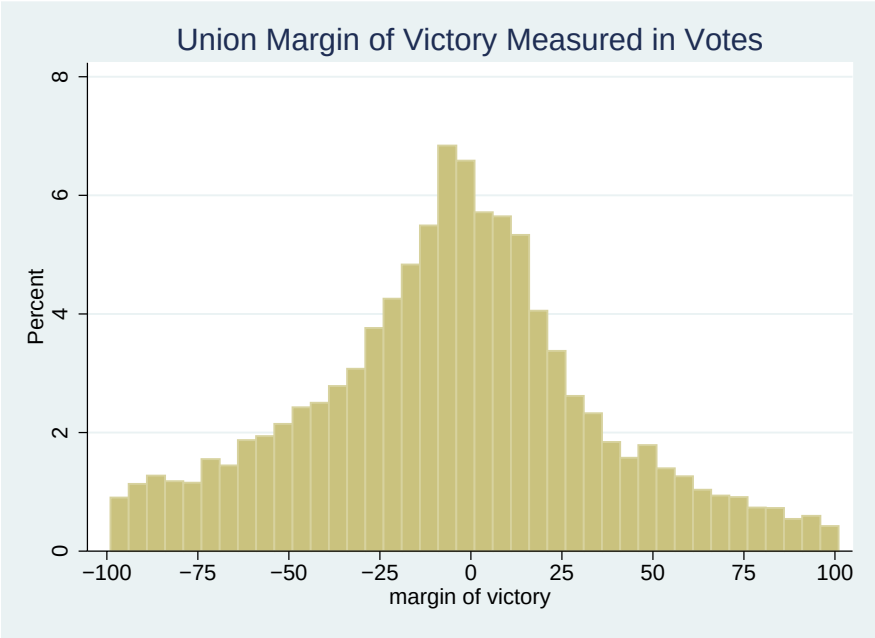

Figure 5: Votes-weighted histogram of the union margin of victory (in terms of number of votes) in representation elections from 1992 to 2001. Only elections decided by 100 or fewer votes are included in the histogram. Data are from NLRB election records, restricted to elections where 10 or more votes were cast.

employer combination—a job—that produced at least one dollar of wages in that state in each year. The data cover a period as wide as 1985 to 2004, although for most states the data only go back to the early 1990s. The EHF contains more than 2.8 billion records, although I focus on full-time, full-year workers with high labor force attachment. This is the sample for whom earnings most closely approximates the hourly wage, on which the theory is based. Restricting to this sample also allows us to separate wage effects from labor supply effects. See the data appendix for details on the sample selection.

Crucially for this study, individual-level earnings records in the LEHD can be matched to establishments. For each union election record in the NLRB election dataset, employees at the time of the election can be identified by matching employer name and address information from the election record with employer information in the LEHD. The data appendix describes in detail the procedure used for matching the two datasets. The matching procedure identified over 1.5 million individuals who were employed at establishments at the time a union election was held from 1992-2001. The subsequent earnings and employment histories of these individuals constitute the main outcomes of interest.

Table 1 reports summary statistics on pre-election earnings, post-election earnings, and retention by union status for full-time, full-year workers in my sample. Average (post-election) annual earnings—defined as the sum of the four quarterly earnings beginning six months after the election—in this sample is just under $30,000. Unionized workers earn on average nearly $3,300 more than non-unionized workers, consistent with the large literature on union wage gaps finding a union premium of about 10 percent. This wage gap, however, reflects both the causal effect of unionization, as well as a selection effect. The nearly $5,000 pre-election earnings gap suggests the selection effect is large. The table also reports statistics for the sample restricted to close elections (+/- 5 percent). Most of the earnings differences disappear in the restricted sample, consistent with bias in the full-sample comparisons by union status, and with DiNardo and Lee's (2004) findings of a small average effect of unionization. However, this small average effect could be masking significant, but offsetting effects elsewhere in the distribution. Average earnings also appear to be lower in the sample restricted to close elections, which reflects that the largest (and highest-paying) employers are less likely to be involved in close union elections.

Table 1: Pre-election Earnings and Outcomes by Unionization Status

	Full Sample			Discontinuity Sample		
	All	Non-unionized	Unionized	All	Non-unionized	Unionized
Annual earnings	29,299	27,319	30,614	22,559	22,306	22,871
	(41,966)	(46,565)	(38,556)	(42,546)	(50,493)	(29,911)
Pre-election earnings	43,202	40,129	45,065	34,474	34,884	34,028
	(47,082)	(46,173)	(47,528)	(38,862)	(47,543)	(26,351)
Stayed employed	0.523	0.505	0.535	0.48	0.492	0.465
at plant	(.499)	(.500)	(.499)	(.500)	(.500)	(.499)
N	1,539,325	614,159	925,166	145,739	80,618	65,121

Notes: the table reports sample means and standard deviations of the dependent variables at left by union status for the sample of full-time, full-year workers employed at plants in covered states where a union election was held from 1992-2001. Earnings variables are measured in 2000 dollars. Annual earnings are defined as the sum of the four quarterly earnings starting two quarters after the union elections. Pre-election earnings are the sum of the four quarterly earnings prior to the union election closing date. Stayed employed at plant is as of 10 quarters after the union election closed. The discontinuity sample restricts to elections that were decided by 5 percent or less.

4 Research Design and Econometric Framework

A fundamental obstacle to measuring the effect of unionization on earnings is selection bias: earnings within unionized plants may differ for reasons other than union representation. This study seeks to overcome selection issues by using a regression discontinuity (RD) research design, originally

developed by Thistlethwaite and Campbell (1960). The motivation for the design, first used in this context by DiNardo and Lee (2004), is that new unions arise through a majority-rule election. If plants and workers where the union barely won and barely lost are comparable, then close union elections approximate a randomized experiment, and the resulting difference in the distribution of earnings provides a reliable estimate of the causal effect of unionization.[6] To represent this idea formally, let $D = 1(R > 0)$ be an indicator for union representation, where R is the union margin of victory (negative for losses). Let Y_1 be an individual's earnings under union representation, and let Y_0 be the earnings otherwise, so that observed earnings is $Y = Y_0 + (Y_1 - Y_0) D$.

Since a worker is never observed simultaneously with and without union representation, we cannot measure the individual specific treatment effect, $Y_1 - Y_0$, but we can estimate the treatment effect on the distribution of outcomes, that is, the difference between the distributions of Y_1 and Y_0 at the margin of union victory. The distributional effects of unionization are captured by the quantile treatment effect, or the difference between the quantiles of potential earnings:

$$\delta(\tau) \equiv Q_{Y_1|R=0}(\tau) - Q_{Y_0|R=0}(\tau). \tag{2}$$

Note that this is the treatment on the distribution of earnings as a whole, rather than the effect of treatment on any particular individual. While (2) does not capture the effect of unions that win or lose by large margins, Figure 5 implies that the effect conditional on a close election reflects the typical case.

The key identifying assumption is that the conditional distribution of potential earnings as a function of the union vote share is smooth near the threshold of union victory, and thus any jumps in the observed distribution of earnings at the threshold is due to the treatment. Formally:

Assumption 1: Local Smoothness $F_{Y_d|R}(y|r)$ is continuous in r over an ε-neighborhood of zero, and is strictly increasing in y over the same neighborhood, for $d \in \{0, 1\}$.

This assumption is satisfied if, for example, unions, workers, and firms are *a priori* uncertain about the outcome of the election when it is close (see Lee, 2008 for a formal proof). The condition that the distribution be increasing in y ensures that quantiles are uniquely defined at the threshold.

[6] See Lee (2008) for further discussion on the conditions under which a close election provides as-good-as-randomized variation.

Given the sharp RD design setup and this local smoothness assumption, the identifying conditions given in Frandsen (2008) are satisfied.[7]

I focus here on the effect of union representation, which leads to the sharp RD design (Campbell, 1969). Another possible treatment of interest is an indicator for a collective bargaining agreement, leading to a "fuzzy" design since an agreement does not always follow from a union victory (Ferguson, 2008). Data on whether an agreement was reached and a contract signed can be inferred from Federal Mediation and Conciliation Service (FMCS) records on contract expiry and renewal. While the econometric framework described in the text applies equally well to the sharp and fuzzy designs, studying the effect of a collective bargaining agreement in this context suffers from two problems. The first is that FMCS records severely undercount union agreements (DiNardo and Lee, 2004), introducing a potential source of bias. Second, there is typically a time lag of several months between a certification election and a collective bargaining agreement being reached. Any responses by employees or employers to the outcome of the election, but prior to an agreement being reached, contaminates the design and would lead to a discontinuity in the distribution of Y_0 (potential earnings under no collective bargaining agreement) at the margin of union victory. For these reasons, the current research design is ill-suited for studying the effect of collective bargaining agreements, and I focus on the effect of union representation.

Frandsen (2008) outlines a tractable procedure for estimating quantile treatment effects in this framework. In the sharp RD design considered here, the estimator becomes particularly simple: it is the difference between kernel-smoothed local linear estimates of the quantiles of earnings approaching the threshold of union victory from the right and from the left. Formally, the estimator can be written:

$$\hat{\delta}_{LQTE}(\tau) = \hat{F}^{-1}_{Y_1|R=0}(\tau) - \hat{F}^{-1}_{Y_0|R=0}(\tau),$$

where

$$\begin{aligned} \hat{F}^{-1}_{Y_1|R=0}(\tau) &= \inf\left\{a : \hat{F}_{Y_1|R=0}(a) = \tau\right\}, \\ \hat{F}^{-1}_{Y_0|R=0}(\tau) &= \inf\left\{b : \hat{F}_{Y_0|R=0}(b) = \tau\right\}, \end{aligned}$$

[7] The other identification conditions in Frandsen (2008)—that the probability of treatment jumps discretely at the threshold, and that crossing the threshold has a montonic effect on treatment status—are automatically satisfied in the sharp RD setup considered here.

and $\hat{F}_{Y_1|R=0}(y)$, $\hat{F}_{Y_0|R=0}(y)$ are local linear, consistent estimates of the conditional distribution functions of potential earnings at the threshold,

$$\begin{aligned}
\hat{F}_{Y_1|R=0}(y) &= \frac{1}{\sum_{j:D=1} w_j(h_1)} \sum_{j:D=1} w_j(h_1)\,\Omega\left(\frac{y-Y_j}{h_2}\right), \\
\hat{F}_{Y_0|R=0}(y) &= \frac{1}{\sum_{j:D=0} w_j(h_1)} \sum_{j:D=0} w_j(h_1)\,\Omega\left(\frac{y-Y_j}{h_2}\right).
\end{aligned}$$

The weighting function associated with local linear fitting is given by:

$$w_j(h_1) = K\left(\frac{R_j}{h_1}\right)\left[S_{n,2} - R_j S_{n,1}\right],$$

with

$$S_{n,l} = \sum_{i:D=0,Z=0} K\left(\frac{R_i}{h_1}\right) R_i^l, \quad l = 1, 2.$$

The bandwidths h_1 and h_2 are chosen to minimize the approximate mean squared error; $K(\cdot)$ is a kernel density fucntion; and $\Omega(\cdot)$ is a kernel distribution function.

The estimator for the local quantile treatment effect is asymptotically normally distributed with the following limiting distribution:

$$n^{\frac{1-b}{2}}\left(\hat{\delta}_{LQTE}(\tau) - \delta_{LQTE}(\tau)\right) \xrightarrow{d} N\left[0, \frac{\omega^+\tau(1-\tau)}{f_{Y_1|R=0}\left(Q_{Y_1|R=0}(\tau)\right)} + \frac{\omega^-\tau(1-\tau)}{f_{Y_0|R=0}\left(Q_{Y_0|R=0}(\tau)\right)}\right],$$

where the bandwidths h_1 and h_2 are proportional to n^{-b}, with $b \in (1/5, 1)$, and $f_{Y_1|R=0}$ and $f_{Y_0|R=0}$ are the densities of Y above and below the threshold, respectively. The other constants are given by:

$$\begin{aligned}
\omega^+ &= \frac{\int_0^\infty \left(s_2^+ - s_1^+ u\right)^2 K(u)^2\, du}{f_R(0)\,\gamma_{h_1} \cdot \left[s_2^+ s_0^+ - \left(s_1^+\right)^2\right]^2}, \\
\omega^+ &= \frac{\int_{-\infty}^0 \left(s_2^- - s_1^- u\right)^2 K(u)^2\, du}{f_R(0)\,\gamma_{h_1} \cdot \left[s_2^- s_0^- - \left(s_1^-\right)^2\right]^2},
\end{aligned}$$

where $s_l^+ = \int_0^\infty K(u)\,u^l du$ and $s_l^- = \int_{-\infty}^0 K(u)\,u^l du$. In the empirical results, I estimate the distribution of the estimator via the nonparametric bootstrap. The validity of the bootstrap in this

setting is discussed in Frandsen (2008).

A regression discontinuity estimator for the average treatment effect at the threshold is given by:

$$\hat{\delta}_{ATE} = \hat{E}\left[Y_1|R=0\right] - \hat{E}\left[Y_0|R=0\right], \tag{3}$$

where

$$\begin{aligned}
\hat{E}\left[Y_1|R=0\right] &= \frac{1}{\sum_{j:D=1} w_j\left(h_1\right)} \sum_{j:D=1} w_j\left(h_1\right) Y_j, \\
\hat{E}\left[Y_0|R=0\right] &= \frac{1}{\sum_{j:D=0} w_j\left(h_1\right)} \sum_{j:D=0} w_j\left(h_1\right) Y_j,
\end{aligned}$$

and the weighting functions and bandwidths are as described above.

5 Results

Comparisons of earnings by union status suggest that unionized workers' earnings are higher on average and more compressed than non-unionized workers' earnings. These findings can be seen in Table 2, panel A, which reports OLS and quantile regression coefficients from a regression of post-election log earnings on an indicator for union representation status for the sample of full-time, full-year workers at establishments where 10 or more votes were cast in a union election. The first column in panel A shows the estimated difference in log earnings is on average .146 with a standard error of .0024. The remaining columns in panel A report differences in the sample quantiles of log earnings. The tenth percentile of unionized earnings is .1931 (s.e.=.0068) log points higher than the tenth percentile of non-unionized earnings. The difference in median earnings is .1667 (s.e.=.0023), and the difference in the 90th percentile is .0765 (s.e.=.0017). These estimates are consistent with regression-based comparisons, such as those surveyed by Lewis (1986), which robustly find a significant positive union wage gap, and quantile regression estimates (e.g., Chamberlain, 1994), which show the union-nonunion wage differential is monotonically declining in wage percentile.

Are these differences due to the causal effect of unionization? Regression discontinuity estimates support the notion that unionization has little effect on the average, but significantly compresses the distribution of employee earnings. These findings can be seen in Panel B of Table 2, which

Table 2: Union Log Earnings Effect

	Quantiles				
Average	0.1	0.25	0.5	0.75	0.9
		A. OLS / Quantile regression			
.1464	.1931	.1679	.1667	.1341	.0765
(.0024)	(.0068)	(.0032)	(.0023)	(.0016)	(.0017)
		B. Regression Discontinuity			
.0015	.2472	.0701	.0749	.0564	.0447
(.0124)	(.0855)	(.0267)	(.0113)	(.0234)	(.0243)

Notes: The table reports estimates of the union effect on post-election log earnings for full-time, full-year workers in covered states from 1992-2001. Panel A shows full sample OLS and quantile regression estimates, and Panel B shows regression discontinuity estimates. The bandwidth (in terms of union vote share) is .025. Earnings (in 2000 dollars) are defined as the sum of the four quarterly earnings starting two quarters after the union election closed. Only workers at plants where more than 10 votes were cast in a union election are included.

reports RD estimates of the average treatment effect, (3), in the first column, and estimates of the quantile treatment effect, (2), in the remaining columns. The first column shows the average effect of unionization conditional on a close election is small: .0015 with a standard error of .0124. This result is consistent with DiNardo and Lee's (2004) finding of little union effect on average wages. This small average effect masks larger effects elsewhere in the distribution, as the remaining columns in Panel B show. The effect on the 10th percentile is large and positive, .2472 with a standard error of .0855, while the effect on the median is a more modest .0564 (s.e.=.0234), and the effect on the 90th percentile is smaller still at .0447 (s.e.=.0243). The distributional effects of unionization are summarized graphically in Figure 6, which plots estimates and pointwise confidence intervals for quantile treatment effects for quantile indices from .1 to .9. The figure shows unionization significantly raised the lower end of the distribution, but had more modest effects through the middle and upper end of the distribution.

Motivated by the model's prediction that unions may reduce the highest productivity workers' wages, I take a closer look at the union effect on the upper tail of the earnings distribution. The results suggest that unionization significantly reduces the highest quantiles of employee earnings. These findings are summarized in Figure 7, which plots estimates and confidence intervals for the effect of unionization on the 90th through the 99th percentile of earnings. While the effect on the

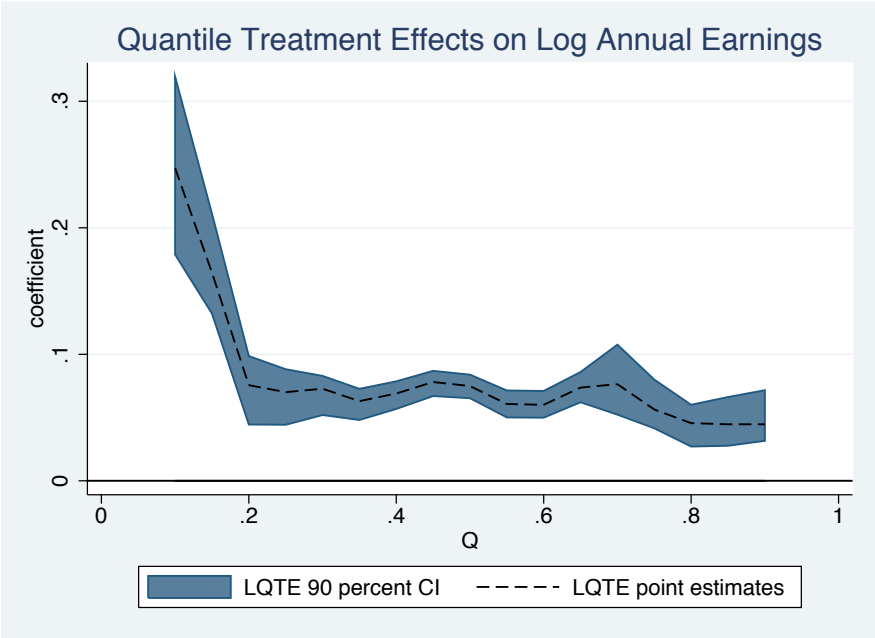

Figure 6: Estimates and 90-percent confidence intervals for the effect of unionization on the quantiles of employees' log annual earnings.

90th percentile is small but positive, the confidence interval begins to include negative values around the 95th percentile, and the estimated effect becomes large and significantly negative starting with the 97th percentile, reaching -.54 at the 99th percentile. In terms of magnitude, unionization's largest impact on earnings appears to be to cut off the upper tail.

The results thus far imply that unionization compresses and shifts the earnings distribution slightly to the right. This effect can be seen directly in Figure 8, which plots estimates of the counterfactual earnings densities by union status, conditional on a close election. The solid curve represents the density of potential earnings without unionization, and the dashed curve represents the density of potential earnings under unionization. The union density is lower in the tails, reflecting compression, and shifted slightly to the right, reflecting the modest positive effect throughout much of the distribution seen in Figure 6.

It is tempting to interpret the estimates plotted in Figures 6 and 7 as giving the effect of unionization on an individual of a given rank in the earnings distribution. In order to make this interpretation, one would have to invoke the rank invariance assumption that a worker with rank τ in the non-unionized potential earnings distribution also has rank τ in the unionized distribution (Heckman, Smith, and Clements, 1997; Chernozhukov and Hansen, 2005). In this case, the τ-th

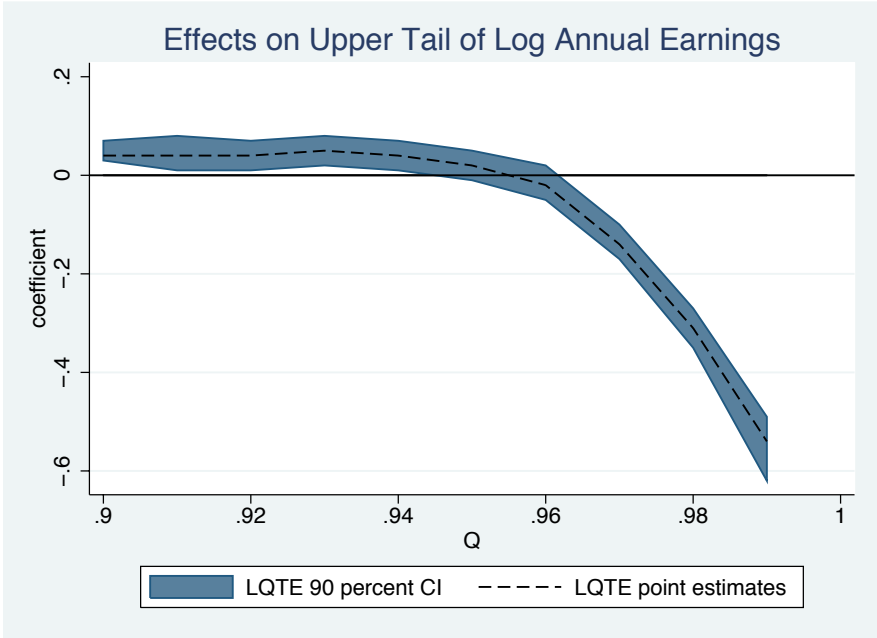

Figure 7: Estimates and 90-percent confidence intervals for the effect of unionization on the upper tail of employees' log annual earnings.

quantile treatment effect—the difference between the τ-quantile of the unionized and non-unionized potential outcome distributions—corresponds to the effect of unionization on a worker of rank τ, since that workers rank is unchanged by unionization. Formally, this assumption could be written: $F_{Y_1|R=0}(Y_1) = F_{Y_0|R=0}(Y_0)$ almost surely. However, given the model's suggestion that unions may increase the pay of some workers at the expense of others, and the empirical results that union effects on quantiles vary drastically, this assumption may be unrealistic in this setting. For example, if unionization led to the termination of some higher-paid workers (e.g., management), ranks would almost certainly be affected.

An alternative assumption that may be more plausible, and allows for inference on the union effect on individuals across the distribution, is that an individual's rank in the non-unionized potential earnings distribution is equal to his or her rank in the pre-election earnings distribution. This would be true if, for example, a worker's rank remains unchanged from one year to the next, barring changes in union status. Denoting the distribution of pre-election earnings by $F_{Y_{-1}}$, formally this assumption can be written: $F_{Y_0|R=0}(Y_0) = F_{Y_{-1}|R=0}(Y_{-1})$ almost surely. Making this assumption, I turn to the union effect on average earnings by pre-election earnings quintile to get a more direct measure of how unionization affects individuals at different points in the distribution.

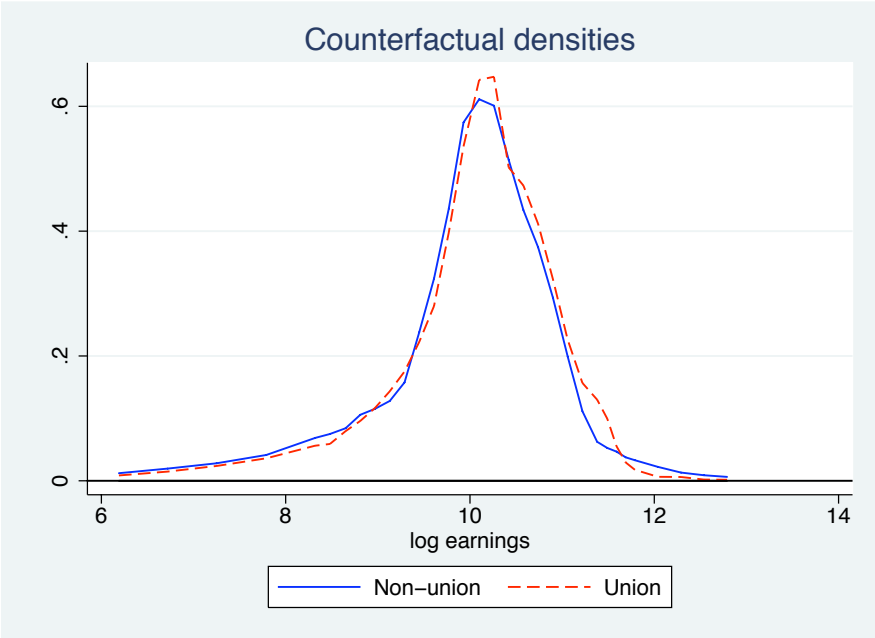

Figure 8: Estimated counterfactual densities of log annual earnings conditional on a close union election.

The results suggest that union representation raises the earnings of those at the lower end of the pre-election earnings distribution, and lowers the earnings of those at the upper end. These findings are shown in Figure 9, which plots RD estimates of the average treatment effect of unionization on log annual earnings by quintile of pre-election earnings. The estimated effects are between .07 and .09 for the bottom three quintiles, but fall to -.05 for the highest quintile of pre-election earnings. These effects mirror the quantile treatment effects reported above, and suggest that individuals at the lower end of the the earnings distribution prior to a close election can expect to see their earnings increase, while individuals at the upper end may see their earnings decrease as a result of union representation. The similarity of the pattern of effects in Figures 6 and 9 also suggests that, as a first approximation, the assumption that ranks are preserved across union status may not be unreasonable, and the quantile treatment effects roughly correspond to the expected effect on an individual at a given point in the earnings distribution.

The earnings compression implied by the estimates is consistent with the theoretical predictions in Section 2, but does it reflect higher wages for lower-skilled workers, and a reduced return to skill, as the model suggests? Unions may compress earnings for other reasons, including shifting risk

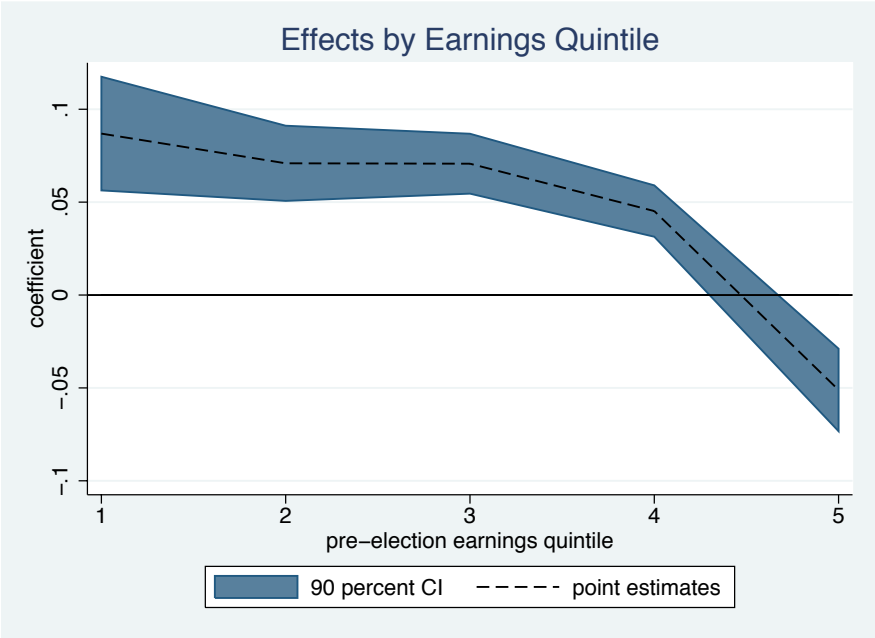

Figure 9: Estimates and 90-percent confidence intervals for the average effect of unionization on log annual earnings, by quintile of pre-election earnings.

to the employer, which could be pareto-improving for workers (Burda, 1995).[8] To get a fuller picture of the welfare consequences of unionization I turn to estimates of the effect on worker retention. The results suggest unionization increases retention among workers at the lower end of the pre-election earnings distribution, while increasing turnover among workers at the higher end. These findings can be seen in Figure 10, which plots estimates and confidence intervals for the effect of unionization on an indicator for retention 10 quarters after the election, by quintile of pre-election earnings. The figure shows unionization has a significantly positive effect on retention for the bottom two quintiles, essentially no effect for the middle quintile, and a significantly negative effect for the top two quintiles. The effect ranges from around 5 percent for the bottom quintile to negative 10 percent for the top quintile, with an overall effect of -.043 (s.e. = .004). The pattern of effects on retention supports the view that unionization makes employment differentially more attractive for lower earners relative to higher earners, consistent with the model's interpretation that union wage compression reflects higher pay for lower-skilled workers and a reduced return to skill.

[8] The literature identifies several other possible sources of union wage compression. Freeman and Medoff (1984) argue that unions provide more efficient levels of public goods in the workplace. Acemoglu and Pischke (1999) show union wage compression can encourage firms to invest in employees' general human capital. Freeman (2005) and Budd (2004) discuss non-wage benefits.

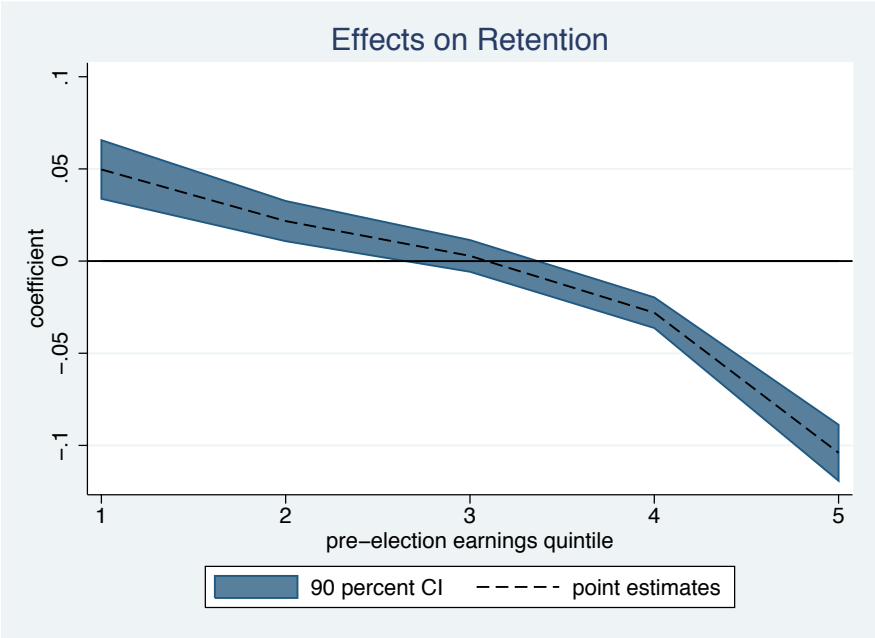

Figure 10: Point estimates and 90-percent confidence intervals for the average effect of unionization on retention 10 quarters after the union election, by quintile of pre-election earnings.

The effect on retention in Figure 10 also has implications for selection into employment at a unionized establishment. Card (1996) developed a two-sided selection model incorporating both employer and employee behavior. If unions compress the distribution of wages, employers are more likely to want to retain (or hire) high-skilled workers, while lower-skilled workers are more likely to want to stay. Figure 10 suggests that on net, selection on the part of employees dominates.

The results in this section provide evidence for the union wage compression found in previous regression-based studies, but against a large average effect of unionization. Thus the principal substantive conclusion of Freeman (1993), Card (1996), DiNardo, Fortin, and Lemieux (1996) that unions reduce dispersion holds up even in a quasi-experimental setting, but the finding in these studies and many others that union earnings are on average significantly higher than non-union earnings appears to be largely a selection effect, consistent with DiNardo and Lee (2004), at least for the typical case of a close union election. When unions win by a wider margin, Lee and Mas (2008) find large negative effects on employer profitability, suggesting that the effect on average earnings may increase with the union's vote share. Interpreted in the light of the model in Section 2, plants where the union barely wins may have fewer sunk costs (that is, lower ϕ_K), and thus unions extract less from the employer and garner fewer votes compared to plants where ϕ_K is large.

Empirical studies looking directly at sunk costs and quasi-rents have also found that unions have a larger effect on employers and industries where sunk costs and rents are higher (see, e.g., Freeman, 1983).

6 Summary and Conclusions

Quasi-experimental estimates based on close union certification elections show unionization substantially compresses the distribution of employee earnings. Union representation raised the tenth percentile of earnings by about 25 log points with a much smaller effect in the middle of the distribution, a large negative effect on the upper tail of earnings, and little effect on the average. Estimates of the union effect on employee retention by quantile of pre-election earnings showed a similar pattern: among workers in the bottom two quintiles, unionization significantly increased retention, it had little effect on retention in the middle quantile, and significantly decreased retention in the top two quantiles.

The pattern of effects on the distribution of earnings and worker retention is consistent with a model where unions pursue a wage schedule to achieve political objectives. A union whose growth as an institution depends on new unionization has incentives to set wage schedules to maximize the probability of winning certification elections. The theoretical model in the paper showed unions will raise the wages of lower-skilled workers, but reduce the return to skill, resulting in a compressing effect on the distribution of workers' earnings. The empirical results on worker turnover by earnings quantile also support this interpretation. Further estimation and testing of the model is a subject for future research.

The results imply that unions close to the margin of victory unambiguously reduce dispersion in the overall earnings distribution, since they compress earnings within the unionized sector, but have little effect on the average union earnings gap. Deunionization therefore explains part of the increased inequality in the U.S. income distribution since the 1970s, as Freeman (1993), Card (1996), and DiNardo, Fortin, and Lemieux (1996) also found. A crude estimate of how much of the increased inequality the fall in unionization rates can explain may be obtained using the sample of full-time, full-year private sector wage and salary workers from the 1979 and 2009 Current Population Survey (King, Ruggles, Alexander, Leicach, and Sobek, 2009). The variance of log

earnings in this sample increased by about 26 percent from 1979 to 2009, while the private sector unionization rate fell from about 25 percent to 8 percent (Bureau of Labor Statistics, 2010; DiNardo, Fortin, and Lemieux, 1996). Assuming the deunionization occurred among workers at marginally unionized plants, deunionization accounts for about 13.5 percent of the increase in the variance of log earnings.[9] This estimate is close to the 15-20 percent found by Card (2001) and the 6-21 percent found by DiNardo, Fortin, and Lemieux (1996).

The estimates apply only to workers at private sector establishments where a close union election was held. While the estimates thus reflect the causal effect of typical private sector unionization in recent years, they miss the effect of public sector unionization, which now accounts for the majority of U.S. union membership (Bureau of Labor Statistics, 2010). More research is needed on the effects of public sector unionization and on the mechanisms driving those effects.

Theory Appendix

The optimal wage schedule is part of a subgame perfect equilibrium: it maximizes the union's probability of winning the certification election, given that workers vote sincerely conditional on the wage schedule, and given that the employer invests and hires to maximize profits conditional on the wage schedule and the outcome of the election. In this setting maximizing the probability of winning is equivalent to maximizing the vote share. To see this, consider an election at a plant with N workers voting. Let $V_i(\theta)$ be the i-th worker's voting rule, as a function of the union's choice parameter, θ. The union solves $\max_{\theta} \Pr\left(\sum_{i=1}^{N} V_i(\theta) > N/2\right)$. The number of votes is a Binomial random variable with parameters $(P(\theta), N)$, where $P(\theta)$ is the expected vote share. Denoting the cdf of this binomial random variable by F_B, the union's problem can be rewritten as $\max_{\theta} 1 - F_B(N/2; (P(\theta), N))$. The first order condition is $\frac{\partial F_B}{\partial P}\frac{\partial P}{\partial \theta} = 0$. Since for the Binomial distribution $\frac{\partial F_B}{\partial P} \neq 0$, this reduces to $\frac{\partial P}{\partial \theta} = 0$, which is the first order condition for maximizing the expected vote share.

Intuitively, for a proposed wage schedule to garner a positive vote share in equilibrium, it must

[9] 13.5% was arrived at as follows. The distribution of unionized potential log earnings in Figure 8 has a variance .087 less than the non-unionized distribution. The difference in unionization rates from 1979 to 2009 is $25\% - 8\% = 17\%$. Since the means of the unionized and non-unionized potential log earnings distributions are essentially equal, the increase in overall variance due to deunionization is therefore $17\% \times .087 = .0149$, which is 13.5% of the total increase in variance.

result in K earning at least $(1-\phi_K)(y-v)$. Otherwise, the employer would simply shut down, and all workers would incur the switching cost to find a job elsewhere, and thus would prefer no union in the first place.[10] Thus, although the union would like to set the wage schedule so as to garner the most votes possible, it must also take into consideration the direct effect of the wage schedule on the firm's profits via the payroll, as well as the indirect effect via the wage schedule's effect on the distribution of workers' human capital employed at the firm. In consequence, any incentive to redistribute rents among workers of different levels of human capital is tempered by the tendency of workers who are losers under the union wage schedule to quit, further tightening the firm's profit constraint. The following proposition formally characterizes the optimal union wage schedule.

Proposition 1 *A subgame perfect Nash equilibrium union wage schedule is* $w^U(H) = (v + r(H, \lambda^*))H$, *where the union rent schedule* $r(H, \lambda)$ *satisfies*

$$\frac{f_\eta(r)}{F_{\eta-\varepsilon}(r)} = \lambda H\left(1 - (\phi_K(y-v) - r)\frac{f_{\eta-\varepsilon}(r)}{F_{\eta-\varepsilon}(r)}\right),$$

and λ^* *satisfies*

$$\int (\phi_K(y-v) - r(h, \lambda^*))\, hF_{\eta-\varepsilon}(r(h, \lambda^*))\, dF_H(h) = 0,$$

and $F_{\eta-\varepsilon}$ *denotes the distribution function of the random variable* $\eta - \varepsilon$, *the unionization cost net of the switching cost.*

Proof. Working backwards, consider the firm's investment and hiring decision given $r(H) = w(H)/H - v$, the outcome of the election, and workers' quitting decisions. First take the case where the union loses. Then the price of H hasn't changed, and the competitive equilibrium remains optimal for the employer. If any workers quit, the employer hires from the pool of applicants (in this case identical in distribution to the population of workers) to replace them and production continues. If no workers quit, no additional hiring or investment takes place and production resumes as before. Now take the case where the union wins. Since in the short run the production technology takes fixed proportions of H and K, if no workers quit, the available actions are hiring additional H

[10] A recent paper by Kremer and Olken (2009) highlights unions' incentives to take into account employers' profitability.

and K, or releasing currently employed H and K. The employer would hire additional H and K only if the return to the additional K is greater than $y - v$ (the purchase price of K). However, the employer must pay at least v in order to hire more H, so the equilibrium cannot involve the employer hiring additional H or investing in more K. Still considering the case where the union wins but no workers quit, the employer releases currently employed H and K if the return to K under the union wage schedule is less than $(1 - \phi_K)(y - v)$. In consequence of constant returns to scale and fixed proportions, if it's optimal to release any H and K, it is optimal to shut down completely. Finally, in the case where the union wins and some workers quit, it is never optimal for the employer to replace the lost workers, for the same reason the employer doesn't hire additional workers. Instead the employer will divest the freed up K (recouping $(1 - \phi_K)(y - v)$ per unit). The employer will either resume production with the remaining H and K, or shut down, again depending on whether the return to the remaining K under the union wage schedule is greater than or less than $(1 - \phi_K)(y - v)$. In summary, the employer's equilibrium strategy is to divest any K freed up by quitting workers, and continue production with what remains if the return to K is at least $(1 - \phi_K)(y - v)$, and shut down otherwise.

Turn now to the worker's decision to stay or leave conditional on $r(H)$ and given the employer's equilibrium strategy. A worker of human capital H chooses to stay if her union rent, $r(H)$, exceeds her cost of unionization net of her switching cost. Thus the equilibrium decision rule is $1(\eta - \varepsilon \leq r(H))$. The density of human capital conditional on staying at the firm is therefore

$$f_{H|\text{stay}}(h) = \frac{F_{\eta-\varepsilon}(r(h)) f_H(h)}{\int F_{\eta-\varepsilon}(r(s)) dF_H(s)},$$

where $F_{\eta-\varepsilon}$ denotes the distribution function of the random variable $\eta - \varepsilon$, the cost of unionization net of the switching cost.

Next, consider the workers' voting choice conditional on $r(H)$ and given the employer's equilibrium strategy. No worker will vote for the union if $r(H)$ is such that the employer shuts down production, since all workers would then incur the switching cost ε, and earn the same wage, v, elsewhere. Conditional on $r(H)$ not resulting in a shut-down, a worker of human capital H votes for the union if her rent under the union schedule exceeds her cost of unionization, η. The equilibrium voting function is therefore $1(\eta \leq r(H))$, again subject to the condition that the plant stays

open under the wage schedule.

Finally, given the workers' and the employer's equilibrium strategies, the union chooses $r(H)$ to maximize the expected vote share, which is equivalent to maximizing the expectation of $1(\eta \leq r(H))$, subject to the firm earning an ex post return of at least $(1-\phi_K)(y-v)$. The problem the union solves can be written:

$$\begin{aligned} &\max_{r(h)} \int F_\eta(r(h))\, dF_H(h) \\ &\text{s.t.} \int (\phi_K(y-v) - r(h))\, hF_{\eta-\varepsilon}(r(h))\, dF_H(h) \geq 0. \end{aligned} \tag{4}$$

This is a straightforward calculus of variations problem of the type treated by, say, Theorem 1 in Gelfand and Fomin (1963, p. 43). The optimal rent schedule therefore satisfies

$$\frac{f_\eta(r)}{F_{\eta-\varepsilon}(r)} = \lambda H \left(1 - (\phi_K(y-v) - r)\frac{f_{\eta-\varepsilon}(r)}{F_{\eta-\varepsilon}(r)}\right),$$

where λ is the Lagrange multiplier on the minimum profit constraint for the firm. This condition can be solved for $r(H,\lambda)$. Substituting this into the profit constraint, (4), and solving for λ gives λ^*, and thus the optimal wage schedule can be written

$$w^U(H) = (r(H,\lambda^*) + v)\, H.$$

■

Data Appendix

Construction of the dataset

As described in the text, the dataset used in this paper consists of NLRB certification election results matched to employer-employee wage data from the Census Bureau's LEHD program.

The union certification election records were collected by the NLRB, and in large part maintained by the AFL-CIO. John-Paul Ferguson obtained the data by filing Freedom of Information Act requests with the NLRB, and has made them available for this research. The complete data

set covers the period 1963-2006, and contains records from about 250,000 union elections, although the main sample used in the analysis covers the years 1992-2001, including 37,354 elections. The raw data contains results from elections stemming from several different type of petitions, including cases where a union seeks to be certified (RC), an employer seeks an election to remove an existing union (RM), or employees seek to remove a union (RD). I restrict to RC-cases, where a union seeks certification. The dataset contains many duplicate records. In some cases they are true duplicates: one election generated multiple records in the database. In these cases I simply delete the redundant entries. In other cases, multiple entries arise from more than one union being on the ballot. In these cases the relevant union vote share is the largest one; I therefore retain the entry with the largest vote share, and delete the others. Finally, in some cases multiple elections were held at the same establishment because, for example, different groups of workers constituted different bargaining units. Since I can't distinguish between workers in different bargaining units, the relevant vote share is the largest, so again I keep only the entry corresponding to the election where the union received the highest vote share.

The second data component consists of the Employment History Files (EHF) within the LEHD database. As described in the text, the EHF contains employee, employer, and earnings data for each employment relationship that generated at least one dollar of wages. The EHF includes a state employer identification number (SEIN) with each record, and in some cases an identifer for the establishment within the employer, which is important for multi-unit employers. For the cases where there is no establishment identifier, the LEHD provides a Unit-to-Worker (U2W) imputation to assign workers to establishments. The employer name and address of these establishments—obtained from the Business Register's Standard Statistical Establishment List (SSEL)—are then used to link to the union election dataset to determine union coverage status.

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

matched BR records were then linked to employers in the LEHD's Employer Characteristics File (ECF) by the Business Register Bridge (BRB) via state, year, county, Employer Identification Number (EIN) and two-digit industry code. Finally the work histories (including earnings) of all individuals employed at the matched employers during the quarter of the certification election were drawn from the Personal History File (PHF), using the Unit-2-Worker imputation to complete the match in the case of multi-unit employers.

Defining the Running Variable

A critical feature of the regression discontinuity design is the running variable (in this case the amount by which the union's vote share exceeds 50 percent). As DiNardo and Lee (2004) point out, care must be taken when defining this variable to avoid biasing results toward the smallest employers. I follow their procedure of first subtracting .5/(# votes cast) from the vote shares where an even number of votes were cast, and then binning the resulting modified shares so that all elections with a share between .50 and .55 are assigned .525, and so forth. Finally, only elections where the number of total votes cast exceeded 10 were kept in the analysis.

Sample Selection

The sample included in the main analysis consists of those workers who have non-missing wage data for the one-year period beginning two quarters after the closing date of the election. I also condition (approximately) on full-time/full-year workers by keeping only those workers whose wages in the year prior to the election exceeded (in 2000 dollars) 20 hours/week $\times$ 40 weeks/year $\times$ \$5/hour = \$4000/year. Although crude, this conservative approximation to full-time, full-year status is based on pre-determined wages and so does not affect the validity of the estimation, and aids in interpretation.

References

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

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

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

Acemoglu, Daron, and Jörn-Steffen Pischke (1999): "The Structure of Wages and Investment in General Training," *The Journal of Political Economy*, 107, 539–572.

Ashenfelter, Orley, and George E. Johnson (1969): "Bargaining Theory, Trade Unions, and Industrial Strike Activity," *The American Economic Review*, 59, 35–49.

Atherton, Wallace N. (1973): *Theory of Union Bargaining Goals*. Princeton University Press, Princeton, New Jersey.

Berkowitz, Monroe (1954): "The Economics of Trade Union Organization and Administration," *Industrial and Labor Relations Review*, 7, 575–592.

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

Blau, Francine D., and Lawrence M. Kahn (1996): "International Differences in Male Wage Inequality: Institutions versus Market Forces," *The Journal of Political Economy*, 104, 791–837.

Booth, Alison L. (1995): *The Economics of the Trade Unions*. Cambridge University Press.

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

Budd, John W. (2004): "Non-Wage Forms of Compensation," *Journal of Labor Research*, 25.

Burda, Michael C. (1995): "Unions and Wage Insurance," CEPR Discussion Paper No. 1232. London, Centre for Economic Policy Research. Available at http://www.cepr.org/pubs/dps/DP1232.asp.

Bureau of Labor Statistics (2010): "Union Members Summary," in BLS annual union membership report, 2009.

Caballero, Ricardo J., and Mohamad L. Hammour (1998): "The Macroeconomics of Specificity," *The Journal of Political Economy*, 106, 724–767.

Campbell, Donald Thomas (1969): "Reforms as Experiments," *American Psychologist*, 24, 409–429.

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

——— (2001): "The Effect of Unions on Wage Inequality in the U.S. Labor Market," *Industrial and Labor Relations Review*, 54, 296–315.

Chamberlain, Gary (1994): "Quantile Regression, Censoring and the Structure of Wages," in *Proceedings of the Sixth World Cogress of the Econometrics Society, Barcelona, Spain*, ed. by Christopher A. Sims, and Jean-Jacques Laffont, pp. 179–209. Cambridge University Press, New York.

Chernozhukov, Victor, and Christian Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

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

Farber, Henry S. (1978): "Individual Preferences and Union Wage Determination: The Case of the United Mine Workers," *The Journal of Political Economy*, 86, 923–942.

——— (1999): "Union Success in Representation Elections: Why Does Unit Size Matter?," Working Paper No. 420, Industrial Relations Section, Princeton University.

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

Frandsen, Brigham R. (2008): "Nonparametric Estimation of Local Quantile Treatment Effects in the Regression Discontinuity Design," Available at SSRN: http://ssrn.com/abstract=1216062.

Freeman, Richard B. (1983): "Unionism, Price-Cost Margins, and the Return to Capital," NBER Working Paper No. 1164.

——— (1993): "How Much Has De-unionisation Contributed to the Rise in Male Earnings Inequality?," in *Uneven Tides: Rising Inequality in America*, ed. by Sheldon Danziger, and Peter Gottschalk, pp. 133–163. Russell Sage Foundation, New York.

——— (2005): "What Do Unions Do? The 2004 M-Brane Stringtwister Edition," NBER Working Paper No. 11410.

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

Gelfand, I. M., and S. V. Fomin (1963): *Calculus of Variations*. Prentice-Hall, New Jersey, Edited by Richard A. Silverman.

Heckman, James J., Jeffrey Smith, and Nancy Clements (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *The Journal of Political Economy*, 101, 410–442.

Katz, Lawrence F., and Kevin M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107, 35–78.

King, Miriam, Steven Ruggles, Trent Alexander, Donna Leicach, and Matthew Sobek (2009): "Integrated Public Use Microdata Series, Current Population Survey: Version

2.0," Machine-readable database, Minneapolis, MN: Minnesota Population Center [producer and distributor].

Kremer, Michael, and Benjamin A. Olken (2009): "A Biological Model of Unions," *American Economic Journal: Applied Economics*, 1, 150–175.

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

Lee, David, and Alexandre Mas (2008): "The Long-Run Costs of Unionization: New Evidence from Financial Markets, 1961-1999," Unpublished manuscript.

——— (2009): "Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961-1999," Unpublished manuscript.

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

Lewis, H. Gregg (1986): *Union Relative Wage Effects A Survey*. University of Chicago Press.

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

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

Ross, Arthur Max (1948): *Trade Union Wage Policy*. University of California Press, Berkeley.

Thistlethwaite, Donald L., and Donald T. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51, 309–317.