DOES MEMBERSHIP PAY OFF FOR COVERED WORKERS? A DISTRIBUTIONAL ANALYSIS OF THE FREE RIDER PROBLEM

OZKAN EREN*

This paper examines the union membership wage premium among private sector employees covered by collective bargaining agreements. Using Current Population Survey data for 2000–2003, the author not only estimates the conditional mean wage premium—the metric on which most previous research has focused—but also employs recently developed (instrumental) quantile regression techniques to estimate the wage effect of membership across the wage distribution. Members enjoyed, on average, a wage premium of 9% over comparable covered nonmembers. Further analyses find no evidence that this mean premium is explained either by unobserved differences or by measurement error. The author also finds that a narrow focus on the mean impact partially masks heterogeneity in the impact across the distribution. Notably, membership wage effects were considerably more pronounced for low wage earners than for high wage earners.

There has been a vast amount of research analyzing the role of unions in the labor market, in particular the impact of unions on wages. Less attention, however, has been devoted to the relevant measure of unionization. In the United States, an employee can be covered by a collective bargaining agreement yet not be a union member. For example, in 2000, roughly 8% of private sector employees covered by bargaining agreements were not union members. Federal labor laws require unions, in their role as bargaining agents, to negotiate the same wage settlement for all employees in the bargaining unit regardless of their membership status. The major difference between members and covered nonmembers is the payment of union dues, which usually correspond to 1.25% of monthly wages (Budd and Na 2000). Moreover, covered nonmembers are not obliged to participate in strikes (Chaison and Dhavale 1992). Many employees, to avoid the monetary costs of membership, may choose to free ride, remaining covered nonmembers.

One way unions can try to discourage free riding is by offering member-only benefits. A relatively small strand of the union literature has attempted to examine the benefits exclusive to members in the form of a wage premium (a wage penalty for free riders). Early studies usually found small statistically significant membership effects. Jones (1982), using National Longitudinal Survey (NLS) data, obtained a wage premium of 2–3% for young member workers. Similar results were obtained by Christensen and Maki (1983). More recent studies have found larger differences. Blakemore et al. (1986),

*The author is Assistant Professor of Economics, University of Nevada, Las Vegas. He thanks Esfandiar Maasoumi, Daniel Millimet, Daniel Henderson, Ozan Sula, and Barry Hirsch for helpful comments that led to an improved version of this paper. The paper also benefited from the comments of participants at the Economics Seminar Series at the University of Nevada, Las Vegas.

The data used in this article can be obtained from the author upon request. Contact Ozkan Eren, Department of Economics, University of Nevada, Las Vegas, NV 89154; ozkan.eren@unlv.edu.

Industrial and Labor Relations Review, Vol. 62, No. 3 (April 2009). © by Cornell University.
0019-7939/00/6203 $01.00

using NLS data and employing longitudinal estimation techniques, found a membership premium of 13%. Using longitudinal estimation with Current Population Survey (CPS) data, Schumacher (1999) found a 6% wage penalty for free riders. Finally, Budd and Na (2000), using the CPS for a different time span, obtained a membership premium of 11–14% with ordinary least squares (OLS) and larger estimates with instrumental variable (IV) techniques.[1]

While the aforementioned studies provide important evidence on the membership wage premium, numerous gaps remain. First, recent studies, whether finding large or small effects, have used standard OLS and IV methods, and therefore have primarily focused on a single measure of central tendency, the conditional mean. The mean impact is an interesting and important measure, but its silence on the membership effect at specific points along the wage distribution limits its usefulness, especially if the effects are *heterogeneous.* The heterogeneity may arise because of the costs of membership, which consist mainly of union dues and costs incidental to obligatory participation in strikes called by the union. For instance, because the relative burden of union dues can be higher for low-wage earners, unions may try to provide additional benefits to these workers to encourage them to join. It is not possible to capture these kinds of potentially important variations with the single central tendency focus inherent in OLS and IV analyses.[2]

Second, researchers analyzing union wage effects usually overlook problems that may arise due to imputed (allocated) earnings within the available data sets (for example, the Current Population Survey). The prevailing view is that including or excluding imputed earnings does not make much difference. However, Hirsch and Schumacher (2004) found a substantial downward bias in the union wage gap due to earnings imputation. Thus, a reexamination of the free rider problem would be warranted even if the only purpose were to determine how, if at all, the results are affected when this potential bias is taken into account.

In this paper, in addition to the conditional mean (OLS and IV), I estimate the membership effect over the wage distribution. The econometric approach I employ is based on quantile regression, which was initially introduced by Koenker and Bassett (1978) as a robust estimation technique for use when the assumption of normality of the error term is not strictly satisfied. Among many others, Buchinsky (1994, 1998) and Powell (1986) extended the use of quantile regression to get information about the effects of exogenous explanatory variables on the dependent variable at different parts of the distribution. Most recently, Chernozhukov and Hansen (2004, 2005, 2006) formulated an instrumental quantile regression model from which the conditional quantiles of the wage distribution can be recovered through the use of instruments under a set of assumptions.

Background

In the United States, private sector workers are covered by the National Labor Relations Act, which was amended in 1947 to allow states to pass right-to-work (RTW) laws. In the "open shop" states adopting RTW, employees covered by a collective bargaining agreement cannot be required to join the union representing them. In states without RTW, on the other hand, covered employees can be required to join the union or pay association fees absent membership.

Recent explanations of member/covered nonmember wage differentials fall into two categories. The first category pertains to the discriminatory behavior of unions (or firms and unions). As already stated, a union, operating as the bargaining agent, must by

[1]In contrast to U.S. studies, Booth and Bryan (2004), using employer-employee linked survey data for Britain, did not find any membership effect. Similarly, Hildreth (2000), using the British Household Panel Survey and after accounting for endogeneity and measurement error, did not obtain any significant membership effect.

[2]There are numerous studies in the union-nonunion wage gap literature that examine the distributional effects of unions (for example, Chamberlain 1994; Card 1996; Hirsch and Schumacher 1998). These studies generally have found lower effects of unions for high wage earners.

federal law negotiate for all employees in the collective bargaining unit, regardless of their membership status. Thus, each individual in the bargaining unit benefits from the rents generated by unions. The major difference between members and covered nonmembers, particularly in the RTW states, is the payment of union dues, which equal roughly 1.25% of monthly wages. Hence the potential for free riding, which unions can discourage by offering some benefits exclusively to members. A further inducement to refrain from joining the union is to avoid the obligation to take part in strikes the union calls. The more free riders there are, the less effective the union will be. Eventually, a union whose membership is small as a percentage of the bargaining unit may be viewed as a candidate for decertification or for a challenge from a competing union.

As stated above, the most straightforward way to deter free riding—by negotiating different wage settlements for members and covered nonmembers—is disallowed by law. Any differential wage benefits in favor of members must therefore be effected by some end-run around the contractual wage settlement. For example, Blakemore et al. (1986) suggested the possibility of a cooperative union/firm arrangement whereby the union monitors the employment contract and disciplines workers without firm involvement, thus promoting work force stability and productivity, in return for which the firm abets or ignores discriminatory behavior by the union. For instance, firms may allow unions to participate in the decision process of training program allocations, job improvements, or promotions and thereby confer wage advantages for members. In other words, the union benefits the firm by promoting efficiency, and in return is given the privilege of differentially rewarding the members. In a variant of the same kind of *quid pro quo,* a firm, in return for union cooperation, may pay nonmembers from a point lower down the union wage scale (Booth and Bryan 2004), or may exploit employees' lack of access to wage information by paying covered nonmembers less than the wage to which they are contractually entitled (Budd and Na 2000). In all these cases, wages of members and nonmembers will differ.[3]

In the second category of hypotheses to explain observed member/covered nonmember wage differentials are selection bias and measurement error hypotheses. There can be unobservable differences between members and covered nonmembers. If unobservable differences are correlated with the membership decision, the resulting premium could be an artifact of this correlation. In addition, individuals are not required to join the union during the probationary period. If many workers refrain from joining during this period, members will, on average, have longer tenure at the firm, and since tenure is positively correlated with wages, omission of this variable may lead to upward bias. It may also be the case that members are systematically associated with firms in which union/bargaining power is strong (Schumacher 1999). Finally, Jones (1982) argued that some individuals who say they are covered nonmembers probably are mistaken, and are not covered by the bargaining agreement at all. This misclassification, too, could impart an upward bias to the estimate of the membership premium.[4]

Econometric Approach and Estimation

Traditional (Exogenous) Quantile Model

The basic quantile regression (QR) model, which specifies the conditional quantile as a linear function of explanatory variables, is given by

$$Y = X'\beta + \varepsilon,$$

$$(1) \qquad Q_\theta(Y|X = x) = x'\beta(\theta) \text{ and } 0 < \theta < 1,$$

where Y is the dependent variable, X is a vector of explanatory variables including the treatment indicator (membership), ε is the

[3]Discrimination against rather than in favor of members may also lead to wage differentials. However, there is no direct evidence of such an effect in the membership premium literature.

[4]The upward bias stems from the collective bargaining effect. On average, covered workers earn 17% more than comparable nonunion workers; see Lewis (1986).

error term, and $Q_\theta(Y|X = x)$ denotes the θ^{th} quantile of Y conditional on $X = x$. Unlike in OLS, the distribution of the error term ε is left unspecified; by the terms of equation (1), it is only assumed that ε satisfies the quantile restriction $Q_\theta(\varepsilon \mid X = x) = 0$. In addition, instead of minimizing the sum of squared residuals in order to obtain the OLS (mean) estimate $\hat{\beta}$, the θ^{th} regression quantile estimate, $\hat{\beta}(\theta)$, is the solution to the following minimization problem:

$$(2) \qquad \underset{\beta \in \Re}{Min} \sum_{Y \geq X'\beta} \theta |Y - X'\beta| + \sum_{Y < X'\beta} (1 - \theta)|Y - X'\beta|.$$

The left (right) term is a sum of positive (negative) residuals weighted by the factor θ. In the special case when $\theta = 0.5$, both terms are equally weighted and the procedure leads to minimizing the sum of absolute deviations, which constitutes the well-known median effect. In a similar manner, by minimizing a sum of asymmetrically weighted absolute residuals, one can estimate any other quantiles. For instance, for $\theta = 0.20$, the positive residuals have less weight than the negative ones, and equation (2) is minimized when 80% of the residuals are positive—20^{th} quantile estimate. Therefore, repeating the estimation for different values of θ between 0 and 1, we can trace the distribution of Y conditional on X and obtain a much more complete view of the effects of explanatory variables on the dependent variable.

Before proceeding, it is important to keep in mind the distinction between quantile regression and a within-quantile OLS model. The notion that something like quantile regression can be achieved by segmenting the dependent variable (unconditional distribution) and then applying OLS estimation to these subsets would be mistaken. As indicated in Koenker and Hallock (2001), this form of truncation on the dependent variable is generally doomed to failure for all the reasons laid out in Heckman's (1979) work on sample selection.

Instrumental Quantile Model

The traditional QR model, similar to OLS, relies on the assumption that the explanatory variables are exogenous. If the treatment indicator (membership) is endogenous, then the use of conventional quantile regression to infer the treatment effect over the distribution of Y will yield biased results. Chernozhukov and Hansen (2004, 2005, 2006) proposed an instrumental quantile regression (IQR) model that takes into account the possible endogeneity of the treatment indicator.

Consider the potential outcomes as Y_d, where $d \in \{0,1\}$ denotes the endogenous treatment and takes a value of one (member) or zero (covered nonmember). We call Y_d potential outcomes because we can observe only one state for any individual at a given time. Y_d can be represented as

$$(3) \qquad Y_d = q(d, X_1, U_d),\ U_d \sim U(0,1).$$

Similar to the basic quantile regression, the IQR specifies the conditional quantile as a linear function of explanatory variables, as well as the endogenous treatment, and is given by

$$Q_\theta(Y|d, X_1) = \alpha(\theta)d + X_1'\beta(\theta),$$

where X_1 is the vector of explanatory variables excluding the treatment and $U_d \sim U(0,1)$ is the unobserved heterogeneous random variable that determines the relative ranking of individuals in terms of potential outcomes. Chernozhukov and Hansen (2005) called U_d the "rank variable," which may be determined by many unobserved factors, including skills and motivation. Letting $q(d, X_1, \theta)$ denote the conditional θ^{th} quantile of potential outcome Y_d for given $X_1 = x_1$, the quantile treatment effect (QTE) that summarizes the difference between the quantiles of two states (member versus covered nonmember) is defined as

$$q(1, x_1, \theta) - q(0, x_1, \theta),$$

which is equivalent to $\alpha(\theta)$. In order to obtain the QTEs, Chernozhukov and Hansen (2005, 2006) derived an estimation equation of the form

$$(4) \qquad P[Y \leq \alpha(\theta)d + X_1'\beta(\theta)|X_1, Z] = \theta$$

under a set of assumptions with Z being the instrument(s). Although I will not repeat the discussion of the underlying assumptions of the IQR, two points are noteworthy.[5] First, analogous to the traditional instrumental variable approach, the IQR model requires that (i) conditional on $X_1 = x_1$, the error term $\{U_d\}$ is independent of the instrument Z, and (ii) the instrument is not independent of the treatment d. Second, unlike the traditional instrumental variable estimation, the IQR model to estimate the QTE ($\alpha(\theta)$) imposes the "rank similarity" assumption. This assumption states that given the information (X_1, Z), the expectation of U_d does not vary across treatment states—that is, individuals who are strong earners as members expect to remain strong earners as covered nonmembers for the same set of X_1 and Z. Similarity allows QTE to be interpreted as the effect of treatment, holding the level of unobserved heterogeneity constant across the treatment states.[6]

Equation (4) provides a moment restriction, which can be used to obtain the IQR estimates $\alpha(\theta)$ and $\beta(\theta)$. Specifically, for a given value of α, we run the conventional QR of $w - \alpha(\theta) d$ on X_1 and Z to estimate $\hat{\beta}(\alpha,\theta)$ and $\hat{\gamma}(\alpha,\theta)$, where $\hat{\gamma}(\alpha,\theta)$ is the estimated coefficient on the instrument. The moment equation in (4) is equivalent to the statement that zero is the quantile solution of $w - \alpha(\theta) d - X_1'\beta(\theta)$ conditional on (X_1, Z). Hence, to find an estimate for $\alpha(\theta)$, we will search for a value α that makes the coefficient on the instrumental variable $\hat{\gamma}(\alpha,\theta)$ as close to zero as possible.

In practice, with one endogenous variable (d) and one instrument (Z), the estimation strategy is as follows:

(i) Run a series of traditional quantile regressions of $Y - d\alpha_j$ on X_1 and Z where α_j is a grid over α.

(ii) Take the α_j that minimizes the absolute value of the coefficient on Z as the estimate of α. Estimates of β are then the corresponding coefficients on X_1.

Data, Sample Statistics, and Evaluation of the Instrument

Data and Sample Statistics

The data for this study are drawn from the Current Population Survey (CPS), a monthly Census Bureau survey of about 60,000 households. Each household entering the CPS is administered 4 monthly surveys, then ignored for 8 months, then surveyed again for 4 months and thereby included in CPS for 8 monthly surveys (or 8 rotation groups). Beginning in January 1979, CPS earnings supplement questions (usual weekly earnings, usual weekly hours, and so on for the primary job in the previous week) were asked in every month rather than just in May, but only the outgoing rotation groups (ORG) 4 and 8 (individuals in the fourth month of a 4-month survey period) were administered these questions. The National Bureau of Economic Research has compiled extracts of the public use ORG files, also known as the Annual Earnings Files.

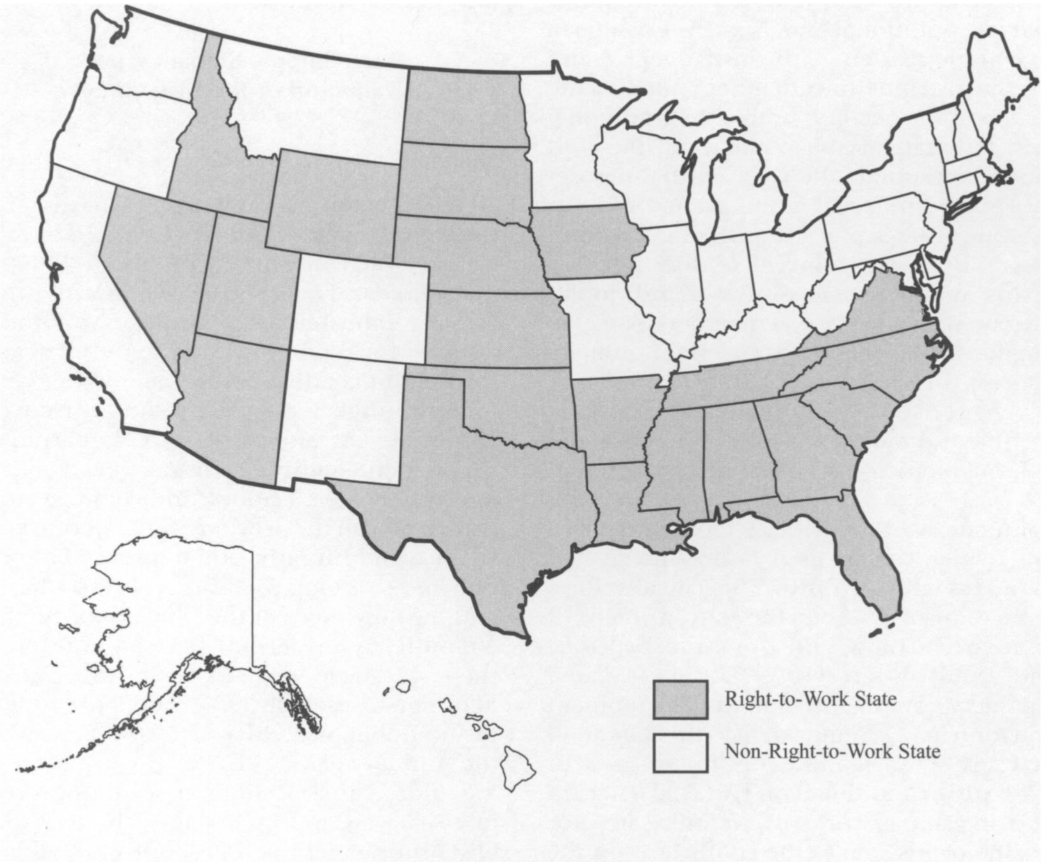
Union status questions, first administered in 1977, were not included in the monthly ORG supplement in 1979, but were added beginning in January 1983. Individuals were first asked, "On this job is .... a member of a labor union or an employee association similar to a union?" Those who answered no were then asked, "On this job is .... covered by a union or employee association contract?" For purposes of the present study, I define those answering yes to the first question as members and those answering yes to the second question as covered nonmembers. Note that the second question was not put to respondents who answered the first question in the affirmative; those individuals are assumed to have been covered by a labor union or employee association contract.

I use data for the years 2000–2003 for those individuals who were surveyed the last time (ORG 8) in the given period to avoid duplication. Furthermore, following Budd and Na (2000) and Schumacher (1999), I restrict the

[5]See Chernozhukov and Hansen (2004, 2005, 2006) for an extensive discussion of the IQR model, its assumptions, and its identification.

[6]It is important to note that the QTEs do not correspond to quantiles of the distribution of the treatment effect unless the assumption of rank similarity holds. Absent this assumption, the QTE simply reflects differences in the quantiles of two marginal distributions. In the Results section, I will discuss the plausibility of this assumption in the context of this study.

Figure 1. U.S. States by Right-to-Work (RTW) Laws.



Note: Oklahoma's RTW legislation was enacted in 2001; Idaho's, in 1985; and all others', in the late 1940s or early 1950s.

analysis solely to RTW states, where individuals are free in their membership decision. As mentioned in the Background section, in non-RTW states, unions can negotiate union shop provisions requiring membership after a short probationary period. Although union shop clauses cannot be used to force membership, many workers do not have full information, and some may join the union under the misapprehension that they must do so. Figure 1 depicts the RTW states.[7]

Researchers analyzing union wage effects usually overlook problems that could arise due to imputed earnings in the CPS data. As widely recognized, many individuals surveyed in the CPS refuse to report their earnings. The Census, rather than compiling official statistics based on incomplete records, imputes or allocates earnings for those with missing values. The prevailing view among researchers is that including or excluding

[7]Twenty-two states have RTW laws: Alabama, Arizona, Arkansas, Florida, Georgia, Idaho, Iowa, Kansas, Louisiana, Mississippi, Nebraska, Nevada, North Carolina, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, and Wyoming. Oklahoma's RTW legislation was enacted in 2001; Idaho's, in 1985; and all the other states' laws, in the late 1940s or early 1950s. In the effective sample, Oklahoma is included beginning with 2001.

allocated earners does not make much difference. However, after comparing regression estimates with and without allocated earners included, Hirsch and Schumacher (2004:691) stated, "Failure to account for earnings imputation causes a substantial understatement in the union wage gap. This bias is particularly severe since 1994." The downward bias arises because the Census does not use "union status" as a matching criterion when imputing earnings for nonrespondents based on the known earnings of their counterparts. Most union nonrespondents are assigned the earnings of nonunion respondents, while some nonunion nonrespondents are assigned the earnings of union respondents, and as a consequence, the union-nonunion wage differential among the imputed earners is close to zero. Hirsch and Schumacher calculated that including imputed earnings resulted in a downward bias, on average, of 20% in 1996–98 and 30% in 1999–2001. Given this potential bias, I exclude individuals with imputed earnings from the analysis. This restriction reduces the sample by 27.7%.

The effective sample consists only of private sector member and covered nonmember wage and salary workers in right-to-work states. Respondents who were below age 18 or above age 65, and those enrolled in school in the week prior to the survey, are dropped. I use the straight-time wage rate (exclusive of tips, overtime, and commission) as the hourly wage for hourly workers. For all other workers, the wage is obtained as the ratio of usual weekly earnings (inclusive of tips, overtime, and commission) to usual hours worked per week.[8] For respondents whose usual weekly earnings are top-coded, I assign the estimated mean earnings above the cap based on the assumption that the upper tail is characterized by a Pareto distribution.[9] I further restrict the sample to individuals whose hourly wage was between $1 and $100.[10] The dependent variable is log hourly wages. In addition to the parameter of interest (member/covered nonmember), I include the following explanatory variables: years of education, potential experience and its square, and the state unemployment rate; and dummies for sex, marital status, full-time, metropolitan status, region, race, industry, occupation, and year effects.

Table 1 displays the sample statistics by membership status for several variables used in the analysis. The mean log hourly wages were 2.821 and 2.689 for members and covered nonmembers, respectively, indicating a raw gap of 0.132 log points. On average, compared to nonmembers, members had less education and more experience, worked in states with lower unemployment rates, and were more likely to be male, married, and residing in metropolitan areas. Appendix Table A1 reports summary statistics for a sample that includes individuals with imputed earnings. The finding that this group's inclusion leads to a reduction of 0.045 log points (from 0.132 to 0.087) in the mean log hourly member/nonmember gap is raw evidence of downward bias.

Evaluation of the Instrument

In this paper, I use the public sector unionization rate (expressed as a percentage) in the individual's state of residence as instrument.[11] To the extent that money is the ruling consideration when workers decide whether to become union members, a covered worker is unlikely to have joined a union if he or she does not expect any pecuniary gains from doing so. However, even in the absence of any pecuniary gains, a covered worker may still join a union if membership has psychological value. This is more likely to be the case if the individual's community is supportive of unionization.

[8]I also run the regressions in the paper by using the ratio of usual weekly earnings to usual hours per week as the measure of hourly wage for all workers. The results are similar.

[9]Usual weekly earnings are top-coded at $2,885 for the years 2000–2003. The estimated mean above the earnings cap values is taken from the Union Membership and Coverage Database, available at http://www.unionstats.com/; see Hirsch and Macpherson (2003) for a general description of the Database.

[10]All wages are deflated by the year's average consumer price index (2003 = 100).

[11]Public union density rates for states are obtained from the Union Membership and Coverage Database.

Table 1. Descriptive Statistics for Members/Covered Nonmembers.

Independent Variable	*Members* *Mean* *(Std. Error)*	*Covered Nonmembers* *Mean* *(Std. Error)*
Log Hourly Wage	2.821 (0.436)	2.689 (0.460)
Years of Education	12.598 (2.082)	13.214 (2.281)
Experience	22.897 (10.764)	20.290 (11.551)
State Unemployment Rate (%)	4.659 (1.024)	4.721 (1.083)
Female	0.269 (0.444)	0.383 (0.486)
Marital Status (1 = yes)	0.662 (0.472)	0.616 (0.486)
Full-Time	0.869 (0.336)	0.844 (0.362)
Metropolitan (1 = yes)	0.725 (0.446)	0.703 (0.457)
Sample Size	3,145	610

Notes: The variables are only a subset of those used in the analysis. The remainder are excluded in the interest of brevity. The full set of sample statistics is available upon request to the author.

Since public sector unionization is greatly affected by the collective bargaining laws adopted in each state, it may be a reasonable indicator of the community's social support for unionization. In other words, I use the state's public sector union percentage as a proxy for the community's attitude toward unions. I expect union density in the public sector to be positively correlated with private sector membership among covered workers.

For union density in the state's public sector to be a valid instrument, (i) it must be a determinant of the membership decision, but (ii) it must not be a determinant of wages; that is, it must be uncorrelated with the error term in the wage equation. To see whether this variable is a determinant of the membership decision, I first present the F-statistic for a test of the hypothesis that the coefficient on the instrument is zero in a regression of membership on the instrument. Staiger and Stock (1997) proposed 10 as the threshold value defining a strong instrument versus a weak one. The first row of Table 2 shows an F-statistic (29.23) well above that threshold. The second row presents the F-test from a regression of residualized membership on the residualized instrument; the F-statistic (16.73) is still above 10.[12] Finally, to test whether union density in the state's public sector is correlated with the membership decision, I run a probit model of membership, controlling for all other explanatory variables, on public sector union density. Shown in the third row is the probit coefficient estimate, which is a positive value of 0.016 (0.004). This coefficient implies that a 1% increase in public sector union density significantly raises the probability of membership among private sector covered workers.

Overall, simple checks indicate that the correlation of interest is not weak, and thus the validity of the instrument depends on the second condition. The most straightforward way to address this issue is to include state public sector union density in the wage equation. I recognize that this is not a formal test, but it does offer a clear sense of the pattern in the data. When public union density (as a percentage) is included in the wage equation and estimated by OLS, the coefficient is 0.0002 (0.0009) and does not even approach statistical significance. Moreover, the estimated effect is very small in magnitude—almost zero. This informal test may suggest that public sector union density is not correlated with the wages of private sector covered workers.

Empirical Results

Mean Results

Ordinary least square estimations. The research on the union-nonunion wage differential typically treats the coverage versus membership distinction as a measurement issue. In his seminal study, Lewis (1986) concluded that the difference between these two measures is negligible and that the union wage gap literature usually does not distinguish between them. Indeed, for

[12]Residualized values are obtained by purging out the effects of the remaining explanatory variables.

right-to-work states in the years 2000–2003, the estimated union effect is 0.169 (0.006) if union membership is used to measure union status and 0.154 (0.006) if union coverage is used.

That being said, I first present the benchmark OLS estimate for the effective sample (the sample restricted to those under a collective bargaining agreement) by including a single membership status dummy in Table 3. The standard error is corrected for any forms of arbitrary heteroskedasticity. The first row of Table 3 shows a statistically significant positive estimate of 0.086 (0.016), a 9% membership premium.[13] If individuals with imputed earnings were not excluded, the OLS estimate would be 0.053 (0.015). This indicates a downward bias of 40% (3.5 percentage points), which is consistent with the findings of Hirsch and Schumacher (2004).

The benchmark model constrains all the right-hand-side variables except membership to be equal for members and covered nonmembers. Earnings profiles, however, may differ between these two groups due to the compression of wage structures (flatter or lower estimated coefficients) in the union sector (see, for example, Budd and Na 2000; Eren 2007). If this is the case, then the benchmark model may be misleading. To address this possibility, I provide two additional OLS estimates. First, I run separate member and covered nonmember wage equations. Relative to covered nonmembers, the positive effects of education, being white, and being male are less pronounced for member workers. The estimated premium, assuming that the covered nonmember wage structure applies to all workers, is 0.113.[14] Second, rather than restricting the sample to those under the bargaining agreement and using a single dummy, I specify a log wage of the form

$$\ln W = \gamma Covered + \delta Member + X'\beta + \varepsilon,$$

where *Covered* is a dummy variable equal to one if the individual is covered by a collective bargaining agreement and zero otherwise, and *Member* is similarly defined.[15] The difference between δ and γ, which is the membership effect, is 0.118. Even though our benchmark estimate implies some differential due to differences in the earnings profiles, this is not a substantial disparity. I therefore rely on this benchmark estimate for the balance of the paper, since it greatly simplifies the presentation of the distributional results.

Table 2. Evaluation of the State's Public Sector Unionization Density as an Instrument.

Tests	*Public Sector Union Density*
F-Statistic	29.23
Residualized F-Stat.	16.73
Probit Coefficient (Std. Error)	0.016 (0.004)

Notes: F-statistics below the value of 10 indicate a weak instrument. The residualized F-statistic is obtained by running an OLS regression of residualized union membership on the residualized instrument. Residualized values are obtained by purging out the effects of the remaining explanatory variables.

The finding of a large and statistically significant membership premium suggests that more is going on than just a measurement issue. I now investigate the econometric problems that may explain the membership premium.

Measurement error. An important concern regarding the membership estimate is the possibility of measurement error arising due to misclassification of union status. Jones (1982) stressed that individuals who report being covered nonmembers are likely to be mistaken, and some of them may not be covered by a bargaining agreement at all. If that is the case, the resulting misclassification is likely to cause an upward bias in the membership premium estimate. To check this, I follow the approach in Budd and Na (2000) and use the fact that any individual in rotation 4 in year t is resurveyed as rotation 8 in year $t+1$. The nature of ORG files allows

[13] The wage differential is obtained as $(\exp(\hat{\beta}) - 1) * 100$; see Halvorsen and Palmquist (1980) for a discussion of log wage and percent wage differentials.

[14] The estimated premium, assuming that the member wage structure applies to all workers, is 0.080.

[15] Note that I include not only members and covered nonmembers, but also nonunion workers.

Table 3. Mean Estimates of Membership Effects.

Union	*Dependent Variable: Log Hourly Wages* *Coefficients (Std. Error)*
OLS	0.086 (0.016)
IV	0.104 (0.254)

Notes: Standard errors are corrected for any forms of arbitrary heteroskedasticity. Estimations control for years of education, experience, experience squared, state unemployment rate, and dummies for sex, marital and metropolitan status, full-time, region (5), race (4), industry (11), and occupation (5) and year effects (4).

matching of individuals across two years.[16] Therefore, I match individuals in ORG 8/(t+1) with those in ORG 4/t and keep those who report the same membership status, sex, race, and an age difference of +2 in two successive surveys (since surveys occur on different days of the month, the age difference can be greater than one year). This yields a sample of 1,225 observations, with 1,120 (91.4%) members and 105 (8.6%) covered nonmembers. Note that, by matching, I exclude not only the individuals with possible measurement error but also the union (collective bargaining) switchers, as well as those whose household has changed (due to a move out of a given household, for example) across the observation period. The estimated effect from the matched sample is statistically significant, with a value of 0.090 (0.032), which translates to a 9.4% wage differential for member workers. The membership coefficient for the matched sample is larger in magnitude than the one presented in Table 3. Therefore, if we assume that consistent responses in two successive surveys remove the problem of measurement error, then there is no evidence of the full sample membership premium being upwardly biased.

Instrumental variable estimation. In the OLS estimates described above, a major econometric concern is ignored. Specifically, some recent studies have addressed the possibility that membership is endogenous with respect to wages (see, for example, Booth and Bryan 2004; Budd and Na 2000; Hildreth 2000). This endogeneity may significantly bias the OLS estimate. For instance, individuals with low unobserved skills may require membership association to maintain their wages. Under this circumstance, the OLS estimate will understate the membership effect. Alternatively, a bias in the opposite direction might arise if, say, members are systematically associated with firms where unions are strong and are able to negotiate higher wages. Given these possibilities, it would seem most prudent to attempt to control for the endogeneity of membership.

The second row of Table 3 presents the IV estimate using the public sector unionization rate in the individual's state of residence as instrument. Similar to OLS, the standard error is corrected for any forms of arbitrary heteroskedasticity. The membership effect is positive, although imprecisely estimated, with a value of 0.104 (0.254). Precision set aside, the IV estimate is larger than the OLS and indicates a very slight downward bias.

In sum, the mean results demonstrate a non-negligible wage advantage for member workers over comparable covered nonmembers. Including those with imputed earnings leads to a downward bias of 40%. Alternative estimates uncover no evidence that the membership premium (the wage penalty for covered nonmembers) can be explained either by unobserved differences or by a greater degree of misclassification among covered nonmembers. On the contrary, these trials generate coefficients slightly *larger* than the benchmark coefficient, though the IV estimate is not statistically significant. The mean findings reported here are consonant with results of previous studies using U.S. data. For instance, Budd and Na (2000), using CPS data for 1983–93, obtained a

[16]The ORG files only contain the household identifiers, so a matching algorithm is necessary. I constructed separate files for ORG groups 4 and 8 for the years 2000, 2001, 2002, and 2003. Individuals in 8/(t+1) were then matched with 4/t according to the household identification number, the person line number in the household, and the survey interview month.

Table 4. Quantile Estimates of Union Membership Effects.

	Dependent Variable: Log Hourly Wages *Coefficient* *(Standard Error)*				
	$\theta = 0.20$	$\theta = 0.40$	$\theta = 0.50$	$\theta = 0.60$	$\theta = 0.80$
Quantile Regression	0.089	0.088	0.082	0.082	0.054
	(0.026)	(0.021)	(0.022)	(0.020)	(0.020)
Instrumental Quantile Regression	0.098	-0.036	0.085	0.137	-0.021
	(0.258)	(0.217)	(0.210)	(0.215)	(0.212)

Notes: Bootstrap standard errors, based on 500 replications, are reported. Estimations control for years of education, experience, experience squared, state unemployment rate, and dummies for sex, marital and metropolitan status, full-time, region (5), race (4), industry (11), and occupation (5) and year effects (4).

membership premium of 11–14% with OLS and larger estimates of around 16% with IV techniques. Similarly, using longitudinal estimation with the CPS data for a different time span, Schumacher (1999) found a 6% wage penalty for free riders.

Quantile Regression Results

Quantile regression estimation. Turning to the distributional results, I first present traditional QR results in the first row of Table 4 for the 20th, 40th, 50th, 60th, and 80th quantiles, respectively. Bootstrapped standard errors are shown beneath each estimate. The membership coefficients are positive and statistically significant at conventional levels for all quantiles. The QR estimates range from 5.5% to 9.3% across quantiles. Even though the membership wage premium is monotonically decreasing in the quantile index, little heterogeneity occurs across the first four quantiles. However, the coefficient for the upper quantile is much smaller in magnitude. In other words, the wage penalty for free riding is lowest for high wage earners.[17] Interestingly, this finding is consistent with the conclusion of distributional studies of union-nonunion wage differentials; union wage effects sharply decrease as one moves from lower to upper quantiles (see, for example, Chamberlain 1994).

Assuming exogeneity of membership status, one interpretation of the traditional quantile regression estimates is as follows. As noted at the outset of the paper, the major difference between members and covered nonmembers is the payment of union dues, which usually correspond to 1.25% of monthly wages. The relative burden of the union dues is likely to be larger for low wage earners than for other workers. Unions, fully aware of this, may offer additional benefits to low earners in order to entice them to join. Similarly, if low wage-earners pay a greater relative price for strike participation than other workers do, this inequality, too, can be mitigated if the union awards a larger membership premium to these low-wage workers.

Instrumental quantile regression estimation. As stated in the Econometric Approach and Estimation section, the IQR model is identified and has a QTE interpretation if, among other assumptions, rank similarity is satisfied. Because rank similarity is an untestable assumption, I can only discuss its plausibility. Conditional on X_1 and Z, rank similarity entails that the expectation of the rank in the potential distributions does not vary with the choice of treatment (membership). That is, a worker at the θ^{th} quantile of the potential membership wage distribution conditional on X_1 and Z is expected to be at the same quantile of the potential covered nonmember wage distribution given the same set of X_1 and Z. I believe that rank similarity holds in the present context because it is unlikely that the qualities required of workers would differ

[17]In addition to the analysis on the full sample, I estimate separate quantile regressions for men versus women, as well for manufacturing versus nonmanufacturing industries. Identical patterns within subsamples are observed. All additional estimates are available upon request.

between member and covered nonmember groups. For instance, rank similarity would be violated if skills, motivation, or commitment to the organization were rewarded for members but not for covered nonmembers, which seems implausible.

One subject merits discussion before the IQR estimates are presented. In research on a question similar to that investigated here, the union-nonunion wage differential, Card (1996) and Hirsch and Schumacher (1998) explained the patterns of selection biases across the wage distribution with a *two-sided selection model* that incorporates both firm and employee behavior in the union selection process. The authors claimed that conditional on a high (low) level of observed skill, the worker's (firm's) selection criteria are more likely to be binding than the firm's (worker's) selection criteria, and thus for workers of higher (lower) levels of observed skill, those in the union sector are more likely to have negative (positive) values of unobserved skills. The opposite selection is a by-product of the "flattened" wage structure in the union sector; highly productive workers are less likely to want to work in the union sector, whereas unionized firms are less likely to want to hire a low-productivity worker. That being said, we may potentially describe the patterns of selection biases, if any, with a *one-sided selection model* in the present context since all the individuals are hired by the unionized firms and are already under the collective bargaining agreement. In other words, the firm selection criteria are not binding.

Of course, a natural question is why we should expect different patterns of biases for those under the bargaining agreement. It may be that individuals differ in the incentives driving their membership decision. For instance, a low-skilled worker may choose membership to maintain his or her wage, which will probably lead to a negative selection (downward) bias. Conversely, high wage earners may decide to join a union because of its reputation effect (Booth 1985) or because they are more motivated or more prepared than other workers to stay with the firm and invest in firm-specific human capital (Budd and Na 2000); if so, the result will be positive selection (upward) bias. Apart from or along with individual differences, union strength or unobserved establishment characteristics may cause different biases throughout the distribution.

In order to examine the one-sided selection model and also to highlight any endogenous heterogeneity, the IQR estimates of QTE are presented in the second row of Table 4 for the 20th, 40th, 50th, 60th, and 80th quantiles, respectively.[18] Bootstrapped standard errors are given beneath each estimate.[19] QTE estimates of the membership effect do not show a monotonic increasing (or decreasing) pattern in the quantile index and are not statistically significant. For the first quantile, we observe a wage premium of 10.3% for members. For the 40th quantile, the membership effect turns out to be negative. For the median and 60th percentile of the conditional distribution of log wages, the membership effects are 8.9% and 14.7%, respectively. Finally, at the 80th quantile, we observe a negative effect on wages. Unfortunately, the coefficients are so imprecisely estimated that we can draw no firm conclusions from them regarding patterns of biases or endogenous heterogeneity within the QTE estimates.[20]

Conclusion

In this paper I have examined the membership wage premium among private sector employees covered by collective bargaining agreements. In addition to an analysis of the mean premium, I have estimated the member-

[18]By including public sector union density in the OLS equation, I have informally shown that, on average, this variable is not a determinant of private sector covered worker wages. To see whether public sector union density affects wages in different parts of the distribution, I undertake a similar approach and test the validity of the instrument. Specifically, I include public sector union density in the traditional QR model. The coefficients on this variable fall far short of statistical significance and are close to zero across all quantiles.

[19]The analytical standard errors, based on kernel estimation, are similar in magnitude and are available upon request. For each bootstrap replication, I re-estimate steps (i) and (ii) of IQR.

[20]Estimations based on ORG groups 4 rather than ORG groups 8 yield qualitatively similar inferences. These estimates are available on request.

ship effect over the wage distribution using recently developed (instrumental) quantile regression techniques.

The main OLS estimate indicates that members did indeed earn more than comparable covered nonmembers, with an effect of 9%. Including the imputed earnings in the sample imparts a downward bias of 40% to this estimate. Assuming that two successive surveys remove the problem of measurement error, there is no evidence that the membership premium is driven by a higher degree of misclassification of covered nonmembers than of members. Actually, a higher estimated wage premium is generated by the matched sample. Moreover, the analysis that accounts for possible unobserved differences supports a similar conclusion, although this evidence falls short of statistical significance. Therefore, my mean estimates support the hypothesis that unions (or firms and unions) discriminate in favor of members. When the analysis is extended to a *distributional* framework using traditional quantile regression, we observe that the discrimination in favor of members is lowest for high wage earners, which may reflect a lower relative cost of membership for this group than for more poorly paid workers. Finally, using an instrumental quantile regression approach, I attempted to identify membership premium heterogeneity and patterns of selection biases. Unfortunately, even though the one-sided selection model is a plausible one, the imprecision of the estimates militates against making any inferences. Perhaps future work can re-approach this issue in more detail.

Appendix Table A1
Descriptive Statistics Obtained by Including Imputed Weekly Earnings

Independent Variable	*Member* *Mean* *(Std. Error)*	*Covered Nonmember* *Mean* *(Std. Error)*
Log Hourly Wage	2.793 (0.460)	2.706 (0.478)
Years of Education	12.623 (2.070)	13.282 (2.268)
Experience	23.329 (10.812)	20.941 (11.602)
State Unemployment Rate (%)	4.685 (1.301)	4.757 (1.093)
Female	0.264 (0.441)	0.379 (0.485)
Marital Status (1 = yes)	0.671 (0.469)	0.624 (0.484)
Full-Time	0.882 (0.321)	0.851 (0.355)
Metropolitan (1 = yes)	0.730 (0.443)	0.704 (0.456)
Sample Size	4,424	769

Notes: The variables are only a subset of those used in the analysis. The remainder are excluded in the interest of brevity. The full set of sample statistics is available upon request to the author.

REFERENCES

Blakemore, Arthur E., Janet C. Hunt, and B. F. Kiker. 1986. "Collective Bargaining and Union Membership Effects on the Wages of Male Youths." *Journal of Labor Economics*, Vol. 4, No. 2, pp. 193-211.

Booth, Alison L. 1985. "The Free Rider Problem and a Social Custom Model of Trade Union Membership." *Quarterly Journal of Economics*, Vol. 100, No. 1, pp. 253-61.

Booth, Alison L., and Mark L. Bryan. 2004. "The Union Membership Wage Premium Puzzle: Is There a Free Rider Problem?" *Industrial and Labor Relations Review*, Vol. 57, No. 3, pp. 402-21.

Buchinsky, Moshe. 1994. "Changes in the U.S. Wage Structure, 1963-1987: Application of Quantile Regression." *Econometrica*, Vol. 62, No. 2, pp. 405-58.

____. 1998. "Recent Advances in Quantile Regression Models." *Journal of Human Resources*, Vol. 33, No. 1, pp. 88-126.

Budd, John W., and In G. Na. 2000. "The Union Membership Wage Premium for Employees Covered by Collective Bargaining Agreements." *Journal of Labor Economics*, Vol. 18, No. 4, pp. 783-807.

Card, David. 1996. "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis." *Econometrica*, Vol. 64, No. 4, pp. 957-79.

Chaison, Gary N., and Dileep G. Dhavale. 1992 "The Choice between Union Membership and Free Rider Status." *Journal of Labor Research*, Vol. 13, No. 4, pp. 355-69.

Chamberlain, Gary. 1994. "Quantile Regression, Censoring, and the Structure of Wages." *Advances in Econometrics, Sixth World Congress*, Vol. 1, pp. 171-209.

Chernozhukov, Victor, and Christian Hansen. 2006. "Inference on the Instrumental Quantile Regression Process for Structural and Treatment Effect Models." *Journal of Econometrics*, Vol. 132, No. 2, pp. 491-525.

____. 2004. "The Impact of 401K Participation on Savings: An IV-QR Analysis." *Review of Economics and Statistics*, Vol. 86, No. 3, pp. 735-51.

____. 2005. "An IV Model of Quantile Treatment Effects." *Econometrica*, Vol. 73, No. 1, pp. 245-61.

Christensen, Sandra, and Dennis Maki. 1983. "The Wage Effect of Compulsory Union Membership." *Industrial and Labor Relations Review*, Vol. 36, No. 2, pp. 230-38.

Eren, Ozkan. 2007. "Measuring the Union-Nonunion Wage Gap Using Propensity Score Matching." *Industrial Relations*, Vol. 46, No. 4, pp. 766-80.

Halvorsen, Robert, and Raymond Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review*, Vol. 70, No. 3, pp. 474-75.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, Vol. 47, No. 1, pp. 153-62.

Hildreth, Andrew K. G. 2000. "Union Wage Differentials for Covered Members and Nonmembers in Great Britain." *Journal of Labor Research*, Vol. 21, No. 1, pp. 133-47.

Hirsch, Barry T., and David A. Macpherson. 2003. "Union Membership and Coverage Database from the Current Population Survey: Note." *Industrial and Labor Relations Review*, Vol. 56, No. 2, pp. 349-54.

Hirsch, Barry T., and Edward J. Schumacher. 1998. "Unions, Wages, and Skills." *Journal of Human Resources*, Vol. 33, No. 1, pp. 201-19.

____. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics*, Vol. 22, No. 3, pp. 689-722.

Jones, Ethel B. 1982. "Union/Nonunion Differentials: Membership or Coverage?" *Journal of Human Resources*, Vol. 17, No. 2, pp. 276-85.

Koenker, Roger W., and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica*, Vol. 46, No. 1, pp. 33-50.

Koenker, Roger W., and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143-56.

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

Powell, James L. 1986. "Censored Regression Quantiles." *Journal of Econometrics*, Vol. 32, No. 1, pp. 143-55.

Schumacher, Edward J. 1999. "What Explains Wage Differences between Union Members and Covered Nonmembers?" *Southern Economic Journal*, Vol. 65, No. 3, pp. 493-512.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variable Regression with Weak Instruments." *Econometrica*, Vol. 65, No. 3, pp. 557-86.