SEARCHING FOR THE UNIONIZATION WAGE DIFFERENTIAL

Yalman Onaran

Follow this and additional works at: https://collected.jcu.edu/jep
SEARCHING FOR THE UNIONIZATION WAGE DIFFERENTIAL

Yalman Onaran
The College of Wooster

INTRODUCTION

A trade union is "a continuous association of wage earners for the purpose of maintaining or improving the conditions of their work lives" (Eatwell, Milgate & Newman, 1988). One reason for unions from the workers' point of view is to protect and raise wages. Although trade unions (or labor unions) aim to improve working conditions, the bargaining emphasis is on wages, which are a part of working conditions. It is claimed that unions have been strong restraints against the decline of wages during depressions when the employers desired to lower them. We see that union membership in England doubled during World War I, and wages stayed the same despite the war. During the great depression in the United States, union membership increased, and the small reduction of wages during that time period is striking (Smith, 1976).

Different theoretical models of unions all predict increased wages for workers covered by union contracts (Hamermesh & Rees, 1988). The spill-over effect states that some of this influence of unions on unionized workers' wages is carried over to the non-unionized workers' wages (Ibid). In other words, in a firm or industry with high level of unionization, the wages of unionized and nonunionized workers should be higher than those of their counterparts in a firm or industry with lower level of unionization.

This paper builds a mathematical model to test the influence of unionization on wages of unionized and nonunionized workers to empirically test the hypothesis that unionization increases wages of workers in a firm or industry, more so of the workers covered by union contracts than those who are not.

REVIEW OF LITERATURE

Earlier industry-level studies on the effect of unionization on wage levels have in general concluded that the average earnings differential between union and nonunion workers (manual, nonsupervisory) is 20% (Geroski & Stewart, 1986; Balkin, 1989). Some settle with a lower figure, usually between 10 and 20 percent (Lewis, 1963; Johnson, 1975; Mitchell, 1980; Ashenfelter and Johnson, 1972).

Balkin(1989) and Mitchell(1986) argue that this wage differential has varied over time. According to Balkin and Mitchell, the differential was reduced during the inflationary 1970's, and has increased in the 1980's because "union wages are locked into the terms of a labor contract" (Balkin, p.301) during inflationary periods. Thus, Mitchell concludes that "empirical investigations of wage determination have often produced autocorrelated residuals from time-series wage equations" (Mitchell, p.249). The reason for the change in the differential during any time period is given by economic and political forces at play. Thus, using 1970's and 1980's data together for a time-series analysis will create autocorrelation. 1970's experienced inflation, while the 1980's have been a period of real growth after the recession in early 1980's. The latter part of 1980's have been relatively similar in economic situation, and we can use these data in a cross-sectional study to get a better estimate of the wage differential.

The conventional model for estimating the wage differential using inter-industry data is commonly defined by the equation

$$\log W_i = f(X_i) + aC_i + e_i$$

where "W_i is the average earnings of manual workers in industry i, aC_i is the proportion of manual workers in the industry covered by collective agreements and f(X_i) is a function of a control vector of other characteristics" (Geroski and Stewart, 1986). This model simply looks at the direct relation between unionization and average wages.

Edwards and Swaim (1986) employ a model that looks at the earnings of the ith worker which is covered by collective bargaining or not.

$$\ln(W_i) = B_2X_i + B_1U_i + e_i$$

The X_i captures the control variables, including demography, education, region and occupation, while U_i is a dummy variable for the ith worker and takes on the value "0" if the worker is not covered by collective bargaining, "1" if the worker is covered. The authors use time series data in this model to see the changes of the wage differential during the end of 1970's; however, we could employ the same model to analyze cross-section data.

Control variables almost always include the demographic variables sex and race, while the form of the function is commonly linear (Mitchell, 1986; Geroski and Stewart, 1986; Edwards and Swaim, 1986).
Sometimes skill or educational attainment is also added to the set of control variables. Occupational differences are commonly controlled for by comparing manual, non-supervisory workers only, and exclude management.

Utilizing the findings of the earlier studies we have talked about, we conclude the following:
1. We should expect a wage differential for unionization;
2. The mathematical model can be estimated using a semi-log equation unless we make additional assumptions about the labor market;
3. We use 1980's data in our analysis as it is a more "normal" time than the 1970's;
4. Sex and race are the two most common control variables employed in unionization and wage analysis.

**A MODEL OF UNIONIZATION AND WAGES**

If we look at the average wage in a firm or industry, it has to be made up of the average union-wage and the average nonunion-wage. By the average union-wage, we mean the average wage of workers who are covered by collective bargaining, and by nonunion wage, those who are not covered. Thus, the total average wage in the firm or industry can be written as:

\[ W = w_u f_u + w_n (1 - f_u) \]  

(1)

where \( W \) is average non-supervisory wage in the industry; \( w_u \) is average wage of unionized workers in the same industry; \( w_n \) of nonunionized workers; \( f_u \) is the ratio of unionized workers to the total. \( W, w_u \) and \( w_n \) are in dollars, while \( f_u \) is a fraction.

Theory tells us that unionized workers and nonunionized workers are affected differently by the amount of unionization in the industry (Hamermesh & Rees, 1988). We can write two separate equations based on this theory, the first of which depicts \( w_u \) and the second one \( w_n \).

\[ w_u = w_c + w_1 f_u \]  

(2)

where \( w_c \) is the competitive wage (hypothetical), which all the workers at that industry would be paid if there were no unions. \( w_1 \) is the influence the unionization ratio will have on the wages of workers who are covered by collective bargaining. We can write the equation for the average nonunionized worker by

\[ w_n = (w_c - a) + b f_u \]  

(3)

where \( b \) is the influence unionization exerts on nonunion workers, and \( a \) is the difference of the starting point between unionized and nonunionized. \( a \) is expected to be negative, but may turn out to be zero. In other words, the intercepts of the two equations can be equal to the competitive wage; however, that doesn't mean that unions have no influence on wages. The different influence is measured by \( w_1 \) and \( b \).

We should note here that we are building our model on a direct linear relation between unionization and wages, in contrast to the semi-log models employed in previous research. One of our goals, then, is to prove this linear relation.

When we combine equations (1), (2) & (3) above, we come up with

\[ W = (w_1 - b) f_u^2 + (a + b) f_u + (w_c - a) \]  

(4)

We can rewrite equation (4) in the following way by substituting a single letter for the coefficients:

\[ W = p + k f_u^2 + m f_u \]  

(5)

where \( p = w_c - a \), \( k = w_1 - b \), and \( m = a + b \).

When we run equation (5) in a regression, the coefficient \( k \) will give us an estimate of the differential of unionization effect on unionized and nonunionized workers' wages, and coefficient \( m \) will be (the difference between the y-intercepts for union and non-union wage curves) + (slope of the non-union wage curve).

**METHODOLOGY**

In the previous section, we have created a quadratic equation to use in testing our hypothesis. To this equation we add two control variables, race and sex, to measure the effect of the proportion of female workers and/or minority workers in the industry.

\[ W = p + k f_u^2 + m f_u - n \text{RACE} - q \text{SEX} \]  

(6)

where \( \text{RACE} \) is the percentage of black and Hispanic workers in the industry; \( \text{SEX} \) is the percentage of female workers. The expected signs of \( \text{RACE} \) and \( \text{SEX} \) are negative because of the theoretical reasons we have discussed in the preceding section.

The sample is made up from industries in the United States. Data for 1984 and 1989 are included in the analysis to increase the number of observations. The five-year gap is due to the need for allowing enough time for different values in the variables to occur. Data from the 1970s are not used because other studies suggest that the impact of unions on wages was different in the 1970s.
To account for the wage differences between 1984 and 1989 that is due to economic factors irrelevant to unionization, a dummy variable YEAR is introduced at this stage.

\[ W = p + k_f u^2 + m_f u - n \text{RACE} - q \text{SEX} - r \text{YEAR} \]  

(7)

YEAR takes on the value "1" when the observations are from 1984 and "0" when they are from 1989. Equation (7) is the final equation that we use in our regression analysis.

**RESULTS**

**First OLS Results**

Table 1 presents the OLS estimates of equation (7) in our model. This equation includes race and sex control variables as well as a dummy variable for the year. The significance of the estimated coefficient for SQUARE rejects the null hypothesis, that is unionization has no second-order influence on wage levels. The coefficient of the variable SQUARE was \( w_1 - b \), where \( w_1 \) is the effect of unionization on the workers covered by collective bargaining and \( b \) is its effect on workers not covered. SQUARE, UNION, SEX and RACE are measured as fractions of the total; YEAR is a dummy variable that takes on the value 1 if the data are from 1984; WAGE is the dollar amount of hourly earnings.

When we check for econometric problems in this OLS output, we run into a few. First of all, the estimated coefficients of SQUARE, UNION, and RACE are not statistically significant at any level. Thus, we cannot arrive at the conclusions about the unionization-wage differential as we have just tried to do above. Moreover, when we look at the correlation matrix, we see that there is a high linear correlation between SQUARE and UNION, SEX and every other independent variable (except YEAR), RACE and every other variable (except YEAR). These econometric problems force us to rethink our model and come up with a new equation that captures what our hypothesis is out to test without endangering the theoretical justification behind it.

<table>
<thead>
<tr>
<th>Table 1 Variables and Coefficients (First OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Variables</strong></td>
</tr>
<tr>
<td>SQUARE: unionization ratio squared</td>
</tr>
<tr>
<td>UNION: unionization ratio</td>
</tr>
<tr>
<td>SEX: female workers to total ratio</td>
</tr>
<tr>
<td>RACE: black workers to total ratio</td>
</tr>
<tr>
<td>YEAR: 1 if 1984, 0 if 1989</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td><strong>Estimated coefficient</strong></td>
</tr>
<tr>
<td>17.345</td>
</tr>
<tr>
<td>-4.871</td>
</tr>
<tr>
<td>-9.681</td>
</tr>
<tr>
<td>3.998</td>
</tr>
<tr>
<td>-1.561</td>
</tr>
<tr>
<td>13.907</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.7930 \text{ (adjusted=}0.7068) \]

The high linear correlation between SQUARE and UNION requires us to drop one of them from the equation. The coefficient of UNION would give us \( a + b \), where \( a \) is the difference between the intercepts of the wage curves of unionized and nonunionized workers and \( b \) is the effect of unionization on nonunionized workers' wages. This coefficient is not necessary for the explanatory function of our model, while as we have seen above, the coefficient of SQUARE, which gives \( w_1 - b \), is absolutely necessary. Because there is a very high linear correlation between SQUARE and UNION (0.963), and because the two are conceptually linked as well, the impact of the two variables on the dependent variable should be captured by one of them. Thus, we can drop UNION from our equation without hurting the explanatory power of the model too much.

SEX has high linear correlation with most of the other independent variables; however, it also has a statistically significant coefficient and a large coefficient, the second biggest after SQUARE. Given the historical and current context of American society, it is not hard to understand the high input gender has on wage differences. The literature reveals women's wages to be lower than men in average. Thus, deleting SEX variable from the equation might hurt the explanatory power of the system deeply. When we keep SEX, and drop UNION and RACE, the only linear correlation that is bothersome is between SEX and SQUARE, and it doesn't constitute too big a problem because the correlation between SQUARE and WAGE (0.498) is almost equal to the above correlation (0.500).

The reason we included RACE as a control variable was because of historical evidence that suggested lower wages for blacks in the U.S. However, our data analysis showed us that racial composition of the industry is not an important determinant of wages in that industry. Moreover, the high linear relations RACE has with SEX and SQUARE should be adequate for us to capture its influence on wages without including it as a control variable.

**Second OLS Results**

After we drop RACE and UNION from our equation for the reasons we discussed above, we get the results on Table 2. The coefficient of SQUARE is still proving our hypothesis of unionization's influence on unionized and nonunionized workers' wages. The corrected differential is now -9.576, much lower than the 17.345 value the first OLS had predicted, but we can understand the reason for this lower value when we think about the high
multicorrelations that existed between variables in the first equation, especially the one between SQUARE and UNION.

### Table 2 Variables and Coefficients (Second OLS)

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Estimated coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQUARE: unionization ratio squared</td>
<td>9.576</td>
</tr>
<tr>
<td>SEX: female workers to total ratio</td>
<td>-8.502</td>
</tr>
<tr>
<td>YEAR: 1 if 1984, 0 if 1989</td>
<td>-1.608</td>
</tr>
<tr>
<td>Constant</td>
<td>13.331</td>
</tr>
</tbody>
</table>

$R^2 = 0.7911$ (adjusted=0.7464)

The coefficient of SQUARE is now statistically significant at the 0.0005 level, constituting a big jump from no significance in the previous OLS. The significance of SEX has also increased from 0.1 level to 0.0005 in the new output. The high value of $R^2$ (0.7911; adjusted-0.7464) and the significance of the F-statistic (at 0.001 level) are also positive indications for the explanatory power of the new equation.

### Implications

Using the second OLS results as a basis for our analysis, we can create a "unionization wage differential" by looking at the coefficient of the variable SQUARE. The value of the coefficient, 9.576, represents $w_1 - b$, the difference between the influence of unionization on unionized workers' wages and nonunionized workers' wages. In other words, there is a difference of 9.576 between the slopes of the wage curves of unionized workers and nonunionized workers. For example, if the slope of $w_n$ was 3.5, the slope of $W_u$ would be 13.076. We can illustrate this graphically in Figure 1.

![Figure 1 Wage and Unionization](image)

The hypothesis that unionization has positive effect on wages is supported by our findings. We have a strong association between union contract existence and higher wages. In addition, we have seen that the demographic factor, the ratio of women to men in an industry, has significant influence on unionization as well as affecting wage levels directly. The higher the ratio of women to men in an industry, the lower the amount of unionization and average hourly wages.

Another implication from the results of our study is the surprising finding that racial composition of the industry has very little impact on wage levels. More surprising is that, although we were expecting a negative sign for the RACE variable, it turned out to be positive in our first OLS output.

In sum, unionization increases wage levels, more of the unionized workers than the nonunionized; the ratio of women in an industry is negatively correlated to wages and unionization; and racial composition has smaller impact on wages than we have expected.

A methodological implication arising from our results is the strength of the quadratic model employed in our analysis. Contrary to the well-accepted assumption that unionization has a semi-log correlation with wages, we
have built and tested a quadratic model which revealed higher $R^2$ than most previous models. The strength of the quadratic model employed can be attributed to the strong emphasis it places on the spill-over effect, which, although prominent in theory, is often overlooked in empirical models.

**LIMITS & CONCLUSIONS**

As we have mentioned in the previous sections, the model has some weaknesses and we have to consider these fully before reaching any final conclusions about unionization and wages. Because the variable SEX has impact both on wages and unionization, we have to realize that the unionization wage differential we have found above is not a perfect measure of unionization impact on wages. The influence of SEX on unionization, and thus, the indirect influence of SEX on wages, is not captured by our model. Although we have eliminated the control variable RACE from our equation due to its low coefficient, multicollinearity with other variables and lack of significance, we should realize that a model testing a different hypothesis could find a significant correlation between racial composition and wages. We have a positive coefficient for RACE in our results, contrary to our expectation of a negative coefficient. We should not jump to the conclusion that RACE does have a positive impact on wages because of the lack of significance the coefficient of this variable has in this model. Also, we have only looked at the ration of blacks in the industries, and the omission of Hispanics and other minorities in our study might be a weakness causing some of the problems we have encountered with this variable.

Another important factor that might have a negative impact on our findings is our data sample. We have compared the wages, unionization levels, female to male ratios of the 9 major industrial divisions in the United States. As we have discussed in the review of literature, some economists are skeptical of the predictory value of inter-industry studies. Our model does not attempt to control for the inter-industry differences in wages due to the economic cycles each industry was going through. The manufacturing industry might be in a better situation than the retail industry based on demand and other economic factors, and that might have a positive impact on the wages in the manufacturing industry. Thus, the fact that we have looked at industries instead of individual firms in one industry could have reduced the explanatory power of our results.

A similar study employing inter-firm data would be appropriate for further analysis and improving consistency. Also, a separate study looking at both the impact of racial composition and female to male ratio in American industries could be of high interest to the reader and could help clarify the understanding of our model.

A structural problem might have arisen due to the dropping of UNION from the original equation. The final equation (7) derived was quadratic in structure with $f_u^2$ and $f_u$. When we drop $f_u$, we run the risk of injuring the model structurally. We justified this omission by pointing out to the high linear correlation between UNION ($f_u$) and SQUARE ($f_u^2$), and that the effect of UNION on wages should be captured by SQUARE. Thus, we can hope that this assumption is true, but should keep our reservations about the explanatory power of the model.

Given the weaknesses above, we still conclude that unionization has positive effect on wages, with greater impact on the wages of unionized workers than nonunionized workers. Even if the assumptions we have made above to justify the omissions above are not true, the model is still strong enough to indicate that $w_1 - b$ is positive and $w_1$ is greater than $b$. Thus, we conclude by saying that our basic finding about the effect of unionization on wage levels holds out in the time period (1984-89) we have selected for analysis.

**BIBLIOGRAPHY**


