Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers

RONALD L. OAXACA and MICHAEL R. RANSOM
University of Arizona, Tucson, AZ 85721

I. *Introduction*

Traditionally, estimation of union/nonunion wage differentials has been something of a consolation prize. To analyze the effects of unionism on resource allocation naturally requires knowledge of the magnitude of the union/competitive wage differential. On an operational level, it is difficult to obtain reliable estimates of this magnitude. In a world of unions, how does one ascertain what the wage structure would be in a perfectly competitive labor market, that is, in the absence of unions? Accordingly, the interests of economists have moved to where the light is: union/nonunion wage differentials. The estimation of union/nonunion wage differentials has become very refined since Lewis's (1963) pathbreaking work. Recent interest in union/nonunion wage differentials has focused on the joint determination of union status and union wage effects (e.g., Duncan and Leigh, 1980, 1985; Farber, 1983). Unless spillover effects are absent, however, estimated union/nonunion wage differentials, no matter how refined they are econometrically, cannot be interpreted as estimates of the union/competitive wage differential. Killingsworth (1983) does distinguish between the union wage gain (the union/competitive wage differential) and the union wage gap (the union/nonunion wage differential), using a simultaneous equations procedure applied to aggregate industry data.

In this paper, we present a method for determining the effects of unionism on the wages of both union and nonunion workers relative to a plausible competitive wage structure. For illustrative purposes, we employ this procedure with sample data on individual workers and compare the results with those obtained from some standard approaches found in the literature.

II. *Conceptual Framework*

From the algebra of logarithmic wage differentials, we have

$$\ln(\delta_{un} + 1) = \ln(\delta_{uc} + 1) - \ln(\delta_{nc} + 1), \qquad (1)$$

where δ_{uc} is the union/competitive wage differential, δ_{un} is the union/nonunion wage differential, and δ_{nc} is the nonunion/competitive wage differential. Assume for the moment that one is only interested in δ_{un}. The most direct way to obtain an estimate of δ_{un} is to include a dummy variable for union status in a wage regression run with a pooled sample of union and nonunion workers. A simple extension of this procedure is to allow for variation in the union/nonunion differential by interacting the union status dummy variable with qualitative worker characteristics, such as industry, occupation, and regional location (see e.g., Ashenfelter, 1972; Oaxaca, 1975). Of course, this approach constrains the union and nonunion wage structures to be identical apart from shifts in the intercept term.

Allowance for different wage structures entails a full interaction between the union status dummy variable and all of the wage determining variables. Equivalently, one estimates separate wage equations for union and nonunion workers. In keeping with most of the literature, we adopt the semilog functional form for the wage equations. Evaluation of the wage equations at the mean yields

$$\ln(\tilde{W}_u) = \bar{X}_u' b_u \tag{2}$$

$$\ln(\tilde{W}_n) = \bar{X}_n' b_n, \tag{3}$$

where u denotes union and n denotes nonunion; $\tilde{W}$ is the geometric mean wage; $\bar{X}$ is a row vector of mean characteristics; and b is a column vector of estimated coefficients.[1] Proper estimation of δ_{un}, δ_{uc}, and δ_{nc} requires an adjustment for mean differences in worker characteristics between union and nonunion workers. In effect, this means that a suitable decomposition of the gross union/nonunion wage differential must be found that isolates the effects of differences in mean characteristics from the effects of differences in structural parameters.

The gross or unadjusted union/nonunion wage differential, G_{un}, is defined by

$$\ln(G_{un} + 1) = \ln(\tilde{W}_u / \tilde{W}_n). \tag{4}$$

There are any number of possible decompositions of G_{un}, each implying a different weighting scheme for differences in mean characteristics between union and nonunion workers. One possible decomposition is given by

$$\ln(G_{un} + 1) = (\bar{X}_u' - \bar{X}_n') b_u + \bar{X}_n' (b_u - b_n). \tag{5}$$

This decomposition was implicitly adopted in Duncan and Leigh (1980), which estimates δ_{un} from the second term in (5). That is,

$$\hat{\delta}_{un} = \exp [\bar{X}_n' (b_u - b_n)] - 1. \tag{6}$$

An interesting interpretation of the decomposition given by equation (5) arises if one believes that the current wage structure in the *union* sector represents what the competitive wage structure would be in the absence of unions. First of all, this would imply that the value of δ_{uc} is zero. From equation (1) this implies

$$\ln(\delta_{un} + 1) = -\ln(\delta_{nc} + 1). \tag{7}$$

It is clear from equation (7) that the (adjusted) union/nonunion wage differential is entirely the result of the effects of unionism on the *nonunion* wage (structure).

[1] Technically, $\tilde{W}$ is the geometric mean wage when the wage equation has been estimated by a procedure that forces the regression hyperplane through the mean, e.g., least squares.

If we denote the average wage of nonunion workers in a competitive labor market by $\tilde{W}_n^c$, then equation (5) and our assumptions about the competitive wage structure imply that

$$\ln(\hat{\delta}_{un}+1) = -\ln(\hat{\delta}_{nc}+1) = -\ln(\tilde{W}_n/\tilde{W}_n^c) = \bar{X}_n'(b_u - b_n),$$

where in this case it is clear that

$$\ln(\tilde{W}_n^c) = \bar{X}_n' b_u. \tag{8}$$

Furthermore, the first term in equation (5) is interpreted as an estimate of the productivity wage differential (in logs) between union and nonunion workers in a competitive labor market with the currently observed *union* wage structure. That is,

$$\ln(\tilde{W}_u^c/\tilde{W}_n^c) = \ln(\tilde{W}_u/\tilde{W}_n^c) = (\bar{X}_u' - \bar{X}_n')b_u.$$

An alternative decomposition to equation (5) is given by

$$\ln(G_{un}+1) = (\bar{X}_u' - \bar{X}_n')\, b_n + \bar{X}_u'(b_u - b_n). \tag{9}$$

From this decomposition, one could estimate $\hat{\delta}_{un}$ from the second term in equation (9). That is,

$$\hat{\delta}_{un} = \exp\,[\bar{X}_u'(b_u - b_n)] - 1. \tag{10}$$

Analogous to the previous case an interesting interpretation of the decomposition specified by equation (9) arises if one believes that the current wage structure in the *nonunion* sector represents what the competitive wage structure would be in the absence of unions. First, this would imply that the value of δ_{nc} is zero (which is theoretically plausible if the union sector is small). From equation (1), it follows that

$$\ln(\delta_{un}+1) = \ln(\delta_{uc}+1). \tag{11}$$

It is clear from equation (11) that the union/nonunion wage differential is entirely the result of the effects of unionism on the *union* wage (structure). In other words, this is the case that corresponds to the absence of spillover effects on the nonunion sector. Under the circumstances, δ_{un} is interpreted as an estimate of δ_{uc}. If we let $\tilde{W}_u^c$ denote the average wage of union workers in a competitive labor market, then equation (9) and our current assumption about the competitive wage structure imply

$$\ln(\hat{\delta}_{un}+1) = \ln(\hat{\delta}_{uc}+1) = \ln(\tilde{W}_u/\tilde{W}_u^c) = \bar{X}_u'(b_u - b_n),$$

where in this case it is clear that

$$\ln(\tilde{W}_u^c) = \bar{X}_u' b_n. \tag{12}$$

The first term in equation (9) is interpreted as an estimate of the productivity wage differential (in logs) between union and nonunion workers in a competitive labor market with the currently observed *nonunion* wage structure. That is,

$$\ln(\tilde{W}_u^c/\tilde{W}_n^c) = \ln(\tilde{W}_u^c/\tilde{W}_n) = (\bar{X}_u' - \bar{X}_n')\, b_n.$$

Equations (5) and (9) represent polar cases in terms of assigning weights to the mean differences in characteristics of union and nonunion workers. This is especially true if one wishes to go beyond the measurement of δ_{un} and to attempt to measure δ_{uc}. In the second case, expression (5) represents the competitive wage structure by the current union wage structure and expression (9) represents the competitive wage structure by the current nonunion wage structure. A third alternative decomposition of the unadjusted union/nonunion wage differential is implicit in the specification of $\hat{\delta}_{un}$ adopted in Duncan and Leigh (1985):

$$\hat{\delta}_{un} = \exp\,[\bar{X}'\,(b_u - b_n)] - 1, \tag{13}$$

where $\bar{X}'$ is a row vector of mean characteristics for the combined sample of union workers and nonunion workers.[2] The decomposition implicit in equation (13) is given by

$$\ln(G_{un} + 1) = (\bar{X}_u' - \bar{X}_n')\,\bar{b} + \bar{X}'\,(b_u - b_n), \tag{14}$$

where $\bar{b}$ is a weighted average of b_u and b_n. Let $\bar{U}$ be the sample proportion of union workers. It can be shown that

$$\bar{b} = b_u\,(1 - \bar{U}) + b_n\bar{U}. \tag{15}$$

If one were to treat $\bar{b}$ as the estimated competitive wage structure in the absence of unionism, then equation (14) generates yet another means for determining the effects of unionism on the union wage and the nonunion wage. The second term in equation (14) is the adjusted union/nonunion wage differential in logs and is decomposable according to

$$\ln(\hat{\delta}_{un} + 1) = \bar{X}'\,(b_u - \bar{b}) - \bar{X}'\,(b_n - \bar{b}) = \ln(\hat{\delta}_{uc} + 1) - \ln(\hat{\delta}_{nc} + 1).$$

It follows that δ_{uc} and δ_{nc} are estimated from

$$\hat{\delta}_{uc} = \exp\,[\bar{X}'\,(b_u - \bar{b})] - 1 \tag{16}$$

and

$$\hat{\delta}_{nc} = \exp\,[\bar{X}'\,(b_n - \bar{b})] - 1. \tag{17}$$

In this case, it is clear that

$$\ln(\tilde{W}_u^c) = \bar{X}_u'\,\bar{b} \tag{18}$$

and

$$\ln(\tilde{W}_n^c) = \bar{X}_n'\,\bar{b}. \tag{19}$$

Subdividing the adjusted union/nonunion wage differential into the separate effects of unionism on the wages of union and nonunion workers permits a finer decomposition of the unadjusted union/nonunion differential:

$$\ln(G_{un} + 1) = (\bar{X}_u' - \bar{X}_n')\,\bar{b} + \bar{X}'\,(b_u - \bar{b}) - \bar{X}'\,(b_n - \bar{b}). \tag{20}$$

Analogous to the previous decompositions, the first term in equation (14) is interpreted as an estimate of the productivity wage differential (in logs) between union and nonunion workers in a competitive labor market with the wage structure

[2] Note that this method evaluates the adjusted union/nonunion wage differential at the mean characteristics of the combined sample of union and nonunion workers.

implied by $\tilde{b}$; that is, $\ln(\tilde{W}_u^c / \tilde{W}_n^c)$. One drawback to approximating the competitive wage structure by $\tilde{b}$ is evident from the reverse weighting scheme of equation (15). Weighting the estimated union (nonunion) parameter vector by the sample proportion of nonunion (union) workers is not a very intuitive procedure. Paradoxically, the larger the sample proportion of union workers the greater the weight given to the estimated nonunion parameter vector in determining the competitive wage structure.

In some sense, a philosophical question is being posed when one asks what the competitive wage structure is in the absence of unionism. If by the absence of unionism one means what would have been the case had unionism never existed, then there is no practical answer. If, however, we mean the complete cessation of existing union activity, then it is possible to generate a plausible estimate of the competitive wage structure. Under this scenario, it is reasonable to suppose that the resulting competitive wage structure would be a blend of the currently observed wage structures in the union and nonunion sectors. A natural weighting scheme for approximating the competitive wage structure that would emerge can be derived from the parameter vector estimated with the pooled sample of union and nonunion workers. This is most easily demonstrated in the case of ordinary least squares estimation of the appropriate wage equation.[3] The resulting weights involve the cross-product matrices formed from the observation matrices of the union, nonunion, and pooled samples. Let $\hat{b}$ be the column vector of the estimated parameters from the pooled sample and let Z_u be a square matrix, such that $Z_u = (X'X)^{-1}(X_u'X_u)$, where X, X_u, and X_n are the observation matrices for the pooled sample, the union sample, and the nonunion sample. The interpretation of Z_u as a weighting factor is easily seen by noting that $X'X = X_u'X_u + X_n'X_n$. It is straightforward to show that

$$\hat{b} = Z_u b_u + (I - Z_u) b_n. \tag{21}$$

Adoption of $\hat{b}$ as our estimate of the competitive wage structure yields the decomposition

$$\ln(G_{un} + 1) = (\bar{X}_u' - \bar{X}_n')\hat{b} + \bar{X}_u'(b_u - \hat{b}) - \bar{X}_n'(b_n - \hat{b}). \tag{22}$$

The first term in equation (22) is interpreted as an estimate of $\ln(\tilde{W}_u^c / \tilde{W}_n^c)$, because

$$\ln(\tilde{W}_u^c) = \bar{X}_u'\hat{b}$$

and

$$\ln(\tilde{W}_n^c) = \bar{X}_n'\hat{b}.$$

[3]For analytical convenience, we abstract from the potential issues of sample selection bias and simultaneous equations bias.

$\ln(\delta_{uc}+1)$ and $-\ln(\delta_{nc}+1)$ are estimated by the second and third terms in equation (22). Thus, δ_{uc} and δ_{un} are estimated according to

$$\hat{\delta}_{uc} = \exp\,[\bar{X}_u'(b_u - \hat{b})] - 1 \quad (23)$$

and

$$\hat{\delta}_{nc} = \exp\,[\bar{X}_n'(b_n - \hat{b})] - 1. \quad (24)$$

Finally, the last two terms in (22) yield an estimate of $\ln(\delta_{un}+1)$. Therefore, the adjusted union/nonunion wage differential is determined by

$$\hat{\delta}_{un} = \exp\,[\bar{X}_u'(b_u - \hat{b}) - \bar{X}_n'(b_n - \hat{b})] - 1. \quad (25)$$

III. *Data*

The data that we use in the empirical example are taken from the 1981 cross-section of the Panel Study of Income Dynamics. Of the 6,742 households available in the sample, we have selected a subsample of 933 white, male heads of households who currently have a job, are not self-employed, and who are paid on an hourly basis. Appendix Table 1 presents descriptive statistics for our subsample, broken down according to whether or not the individual's job is covered by a union contract (the concept of unionism maintained in this paper). It is apparent that we do not have a random sample of the U.S. labor force. For example, over 45 percent of the sample works under a union contract, a much larger proportion than in the general population. The union sample is more urban, more likely to live in the northeastern and northcentral regions, and has markedly higher levels of experience and tenure than the nonunion sample.

IV. *Empirical Findings*

For the sample used in this study, the value of the gross or unadjusted union/nonunion wage differential in logs was 0.3745. This implies a value of G_{un} of 0.4543; that is, the average hourly wage of union workers exceeded that of nonunion workers by 45 percent. In Table 1, we report the values of δ_{un}, δ_{uc}, δ_{nc}, and $(W_u^c/W_n^c) - 1$, estimated according to the interpretations given the decompositions (5), (9), (20), and (22). Estimates of the (adjusted) union/nonunion wage differential range from about 0.24 to 0.36. Coincidentally, approximations of the competitive wage structure by both the estimated nonunion wage structure and the wage structure estimated from the pooled sample yield a union/nonunion wage differential of 0.24. Nevertheless, these two procedures imply quite different estimates of the effects of unionism on union and nonunion wages. Adoption of the estimated nonunion wage structure as the competitive norm implies the absence of union effects on the nonunion wage. Therefore, this procedure constrains the 24 percent union wage advantage over the nonunion sector to equal the effect of unionism relative to the competitive wage. On the other hand, adoption of the estimated wage structure from the pooled sample as the competitive norm implies that the 24 percent union wage advantage over the nonunion sector arises

Table 1

Estimated Wage Effects of Unionism

Competitive Wage Structure	$\hat{\delta}_{un}$	$\hat{\delta}_{uc}$	$\hat{\delta}_{nc}$	$(\bar{W}_u^c/\bar{W}_n^c) - 1$
Union	0.3642	0	−0.3642	0.0658
Nonunion	0.2412	0.2412	0	0.1710
Reverse weighted	0.3067	0.1299	−0.1353	0.1129
Pooled	0.2369	0.1220	−0.0929	0.1752

$\hat{\delta}_{un}$, $\hat{\delta}_{uc}$, and $\hat{\delta}_{nc}$ are the estimated union/nonunion, union/competitive, and nonunion/competitive wage differentials. $(\bar{W}_u^c/\bar{W}_n^c) - 1$ is the productivity wage differential between union and nonunion workers.

from unionism simultaneously raising the wages of union workers by 12 percent above and depressing the wages of nonunion workers by 9 percent below their respective competitive levels. When the competitive wage structure is approximated by the estimated union wage structure, the large estimated union/nonunion wage differential of 36 percent is entirely imputed to the depressing effects of unionism on the wages of nonunion workers. This imputation method constrains the effects of unionism on the wages of union workers to equal zero. The reverse weighting procedure estimates the union/nonunion wage differential to be about 0.31. This estimated union wage advantage over nonunion workers arises from unionism simultaneously raising the wages of union workers by 13 percent above and depressing the wages of nonunion workers by 14 percent below their respective competitive levels.

All four methods for measuring the wage effects of unionism indicate a productivity advantage of union workers. Based on measured personal productivity characteristics, it is estimated that even in the absence of union wage effects union workers would earn from 7 to 18 percent more than nonunion workers. Approximation of the competitive wage structure by either the nonunion wage structure or the pooled wage structure yields a union productivity wage advantage of 17 to 18 percent. Assuming that the estimated union wage structure is the competitive norm results in a modest union productivity wage advantage of 7 percent. Finally, the reverse weighting scheme produces an estimated union productivity wage advantage of 11 percent. All of these results emphasize the fact that union members *are* different from nonunion members. Some of this difference is due to lower separation rates due to higher union wages. The difference might also be attributed to greater union success in organizing higher productivity firms, competition (via job changing) by individuals seeking rents.

V. *Concluding Remarks*

We have sought to take the estimation of union wage effects beyond merely estimating the union/nonunion wage differential. It is probably safe to assume that there is widespread agreement that knowledge of the effects of unionism relative to some competitive norm is fundamentally the more interesting objective. Once the competitive wage structure is known, one can always deduce the magnitude of the union/nonunion wage differential. The catch, of course, is to first define the absence of unionism and then estimate the wage structure that would exist in that absence of unionism. Conventional methods for estimating adjusted union/nonunion wage differentials through decompositions of gross (unadjusted) differentials can be twisted to yield estimated wage effects of unionism relative to assumed competitive wage structures. This procedure assumes that the competitive wage structure is approximated by either the current union wage structure or the current nonunion wage structure. Either case is admittedly extreme and restrictive. It is more palatable to approximate the competitive wage structure by some combination of the current union and nonunion wage structures. Both the reverse and the pooled sample weighting methods factor in the estimated union and nonunion current wage structures. Both methods yield results that are more similar to one another than to those from either of the other two methods. Our pooled sample procedure, however, does have an advantage over the reverse weighting method: it offers a readily understood and intuitive interpretation of the implicit weighting factors.

A relatively recent refinement in the estimation of union/nonunion wage differentials is the correction for sample selectivity bias. The importance of sample selectivity bias in estimating the wage effects of unionism can vary across data sets and with methods used to test for its presence. For example, Duncan and Leigh (1985) report a case where the Hausman specification test (used in conjunction with an instrumental variables estimator) rejects the null hypothesis of exogeneity of union status while the inverse Mills ratio procedure fails to find evidence of sample selectivity. In any event, if union status is endogenous, the consequent sample selectivity bias will somewhat diminish but not eliminate the attractiveness of the pooled sample methodology for determining the separate wage effects of unionism on the wages of union and nonunion workers. One would still estimate the competitive wage structure under the restriction that the wage equation parameters are the same for union and nonunion workers in the absence of unionism, except now the parameters estimated from the pooled sample will no longer be linearly related to the consistently estimated parameters of the separate union and nonunion wage equations.

Appendix Table 1

Descriptive Statistics of Data Used in Analysis of Union Wage Differentials

Variable	Full Sample		Union Sample		Nonunion Sample	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
ln wage	2.12	0.47	2.32	0.35	1.94	0.49
Years school	11.76	2.09	11.62	2.01	11.88	2.14
Months tenure with current employer	88.07	96.84	125.45	109.57	56.69	70.90
Years labor market experience	11.02	10.91	13.17	10.86	9.20	10.62
Lives in NE	0.20	0.40	0.23	0.42	0.17	0.38
Lives in NC	0.29	0.45	0.35	0.48	0.23	0.42
Lives in South	0.33	0.47	0.23	0.42	0.40	0.49
Lives in urban areas	0.64	0.48	0.69	0.46	0.60	0.49
Sample size	918		419		499	

Appendix Table 2

Determinates of log (Wage) Estimated Coefficients

Variable	Full Sample		Union Sample		Nonunion Sample	
	Estimated Coefficient	Standard Error	Estimated Coefficient	Standard Error	Estimated Coefficient	Standard Error
SCHOOL	.0847	.0382	0.0043	0.0463	0.1582	0.0547
$SCHOOL^2$	−.0019	.0016	0.0014	0.0021	−0.0046	0.0023
EXPERIENCE	.0262	.0045	0.0110	0.0065	0.0311	0.0060
$EXPERIENCE^2$	−.0006	.0001	−0.0003	−0.0002	−0.0007	0.0001
TENURE	.0032	.0004	0.0008	0.0005	0.0039	0.0007
$TENURE^2$	5.75×10^{-6}	1.17×10^{-6}	9.56×10^{-7}	1.32×10^{-6}	8.23×10^{-6}	2.32×10^{-6}
NE	−0.1858	0.0443	−0.1370	0.0521	−0.2276	0.0644
NC	−0.1137	0.0414	−0.0998	0.0488	−0.1480	0.0621
SOUTH	−0.1603	0.0407	−0.1673	0.0524	−0.1052	0.0557
URBAN	0.1521	0.0296	0.0906	0.0370	0.1509	0.0409
R^2	.2259		.1048		.2274	
Residual Sum of Squares	156.26		46.59		91.05	

REFERENCES

Ashenfelter, Orley. "Racial Discrimination and Trade Unionism." *Journal of Political Economy* 80 (May/June 1972): 435-64.

Duncan, Gregory M., and Duane E. Leigh, "Wage Determination in the Union and Nonunion Sectors: A Sample Selectivity Approach." *Industrial and Labor Relations Review* 34 (October 1980): 24-34.

____________. "The Endogeneity of Union Status: An Empirical Test." *Journal of Labor Economics* 3 (July 1985): 385-402.

Farber, Henry S. "The Determination of the Union Status of Workers." *Econometrica* 51 (September 1983): 1417-37.

Killingsworth, Mark R. "Union-Nonunion Wage Gaps and Wage Gains: New Estimates from an Industry Cross-Section." *Review of Economics and Statistics* 65 (May 1983): 332-36.

Lewis, H. Gregg. *Unionism and Relative Wages in the United States.* Chicago: University of Chicago Press, 1963.

Oaxaca, Ronald L. "Estimation of Union/nonunion Wage Differentials Within Occupational/Regional Subgroups." *Journal of Human Resources* 10 (Fall 1975): 529-37.