WAGE DETERMINATION IN THE UNION AND NONUNION SECTORS: A SAMPLE SELECTIVITY APPROACH

GREGORY M. DUNCAN and DUANE E. LEIGH*

This paper re-examines the question, recently raised in this journal by Bloch and Kuskin, of whether wages are determined differently in the union and nonunion sectors. Whereas Bloch and Kuskin employed ordinary least squares to estimate separate wage equations for the two sectors, this study uses a methodology proposed by Heckman and by Lee to correct for the possibility that wage differences may determine the union status of workers as well as vice versa. The authors find that union status is strongly related to the predicted union-nonunion wage differential, but their evidence nevertheless reinforces Bloch and Kuskin's empirical finding that the union earnings function is less sensitive than the nonunion earnings function to changes in nearly every observable attribute of workers, such as education and experience. The authors also conclude that previous studies using separately estimated union and nonunion wage equations may have understated the success of unions in raising the relative wages of their members.

IN A recent article in this *Review*, Bloch and Kuskin (hereafter abbreviated B-K) raise the issue that wages may be determined differently in the union and nonunion sectors.[1] In particular, if marginal payoffs to individual characteristics, such as education and experience, are smaller in the union sector than in the nonunion sector, unionized employers have an incentive to hire more highly skilled workers because the marginal cost of a unit of skill to them is relatively low. This would be expected to cause a divergence in the quality of the work forces in the two sectors leading to an accentuation of existing union-nonunion wage differentials.[2] Using 1973 Current Population Survey data for white males employed in the private sector, B-K estimate separate wage equations for the union and nonunion sectors using ordinary least squares (OLS). As anticipated, they find that wage rates in the nonunion sector are more responsive to differences in personal

*Gregory Duncan is an associate professor and Duane Leigh a professor of economics at Washington State University. They gratefully acknowledge the helpful comments of Gary Fields and Ronald S. Warren.

[1]Farrell E. Bloch and Mark S. Kuskin, "Wage Determination in the Union and Nonunion Sectors," *Industrial and Labor Relations Review*, Vol. 31, No. 2 (January 1978), pp. 183 – 92.

[2]For additional discussion of this point, see Lawrence M. Kahn, "Unionism and Relative Wages: Direct and Indirect Effects," *Industrial and Labor Relations Review*, Vol. 32, No. 4 (July 1979), pp. 520 – 32.

Industrial and Labor Relations Review, Vol. 34, No. 1 (October 1980).
0019-7939/80/3401-0024$01.00

characteristics and price level variation than are union sector wages.

This paper investigates the same issue addressed by B-K but uses a methodology that accounts for the fact that the wages actually observed for union and nonunion workers are not random samples from the wage distribution for the entire population. If workers' union membership decisions are related to the differential between anticipated earnings in their best union and nonunion alternatives, observed union and nonunion wages will be nonrandomly selected samples from the population wage distribution. In such a case, estimates obtained using OLS on observed wage rates will be inconsistent if disturbance terms of the wage equations are correlated with workers' union status.[3] Technically, the statistical problem closely resembles the issues of sample selectivity investigated by Heckman and others. Heckman characterizes the problem as a specification error where an "omitted variable" measuring the expected value of disturbance terms conditional on sample selection rules is missing from the right-hand side of the wage equations.[4]

To assess the magnitude of sample selection bias in B-K's estimated wage relationships, the first section of this paper presents a simple three-equation model determining union status as well as union and nonunion wage rates. A second section outlines a procedure recently described by Lee that provides consistent estimates of union and nonunion wage relationships adjusted for sample selection.[5] Then estimates of the model are presented using a sample of middle-aged white men from the National Longitudinal Surveys (NLS).[6] These results are compared with unadjusted wage equations estimated by B-K and with Lee's estimates of adjusted wage equations and a unionism equation obtained for semi-skilled workers in the 1967 Survey of Economic Opportunity (SEO) data set.

The third section addresses the issue of the interpretation to be given union-nonunion wage differentials calculated from estimated wage equations that have been corrected for selectivity bias. A critique of the approach taken by B-K and others is provided and a new interpretation is presented.

Determinants of Union Status and Union and Nonunion Wages

Following the general approach of Lee,[7] a comparison of the percentage union-nonunion wage differential with an individual's reservation wage is presumed to determine union status. If the union wage differential exceeds the reservation wage, the individual opts for a union job; if it does not, nonunion employment is selected. The reservation wage, in turn, depends on the following factors: a vector of personal characteristics that represents both workers' tastes for unionism and the greater selectivity in hiring by organized employers, the cost of providing union services, and an error term reflecting unobservable random influences.[8] Since direct measures of the costs of union services are not available,

[3]In their concluding section, B-K (p. 192) note that the probability of union membership may be positively related to what they term preexisting wage rates. They do not attempt to assess the impact of this relationship on the wage equation estimates presented in their paper.

[4]James J. Heckman, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, Vol. 5, No. 4 (1976), p. 478. See also Heckman, "Sample Selection as a Specification Error," *Econometrica*, Vol. 47, No. 1 (January 1979), pp. 153 – 61.

[5]Lung-Fei Lee, "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, Vol. 19, No. 2 (June 1978), pp. 415 – 33.

[6]Data for white males only are examined to ensure reasonable comparability with the sample investigated by B-K. Applying a similar statistical methodology to a similar model, results for middle-aged black men and young white and black men in addition to middle-aged whites are presented in Duane E. Leigh, "Racial Differentials in Union Relative Wage Effects: A Simultaneous Equations Approach," *Journal of Labor Research*, Vol. 1, No. 1 (Spring 1980), pp. 95 – 114.

[7]Lee, "Unionism and Wage Rates," pp. 416 – 18.

[8]A useful attempt to distinguish between the selectivity of unionized employers and individuals' choice of union status is found in John M. Abowd and Henry S. Farber, "Relative Wages, Union Membership, and Job Queues: Econometric Evidence Based on Panel Data," unpublished paper, Princeton University, July 1978.

these costs are assumed to depend on workers' personal characteristics and on industry of employment. As proxies for the costs of union services, personal characteristics represent factors such as employment stability and industry of employment capture factors such as firm or plant size and ease of entry into the industry.

Let UN_i be an unmeasurable union status variable that takes the following form: If $UN_i > 0$, individual i joins a union; otherwise he does not. Only the sign of UN_i is observable. For the i^{th} individual, the union threshold equation is defined as

$$(1) \qquad UN_i = a_0 + a_1 X_i + a_2 Y_i + a_3(\ln W_{ui} - \ln W_{ni}) - \varepsilon_i$$

where X_i is a vector of personal characteristics; Y_i is industry of employment; W_{ui} and W_{ni} are union and nonunion wage rates, respectively; and ε_i is a random disturbance term. The term ($\ln W_{ui} - \ln W_{ni}$) is used to approximate the percentage union wage differential.

Personal characteristics included in X_i in Equation 1 are region of residence, residence in a rural community, years of schooling completed, marital status, dependents other than wife, and occupational affiliation. The residence variables capture the relatively weak union preferences expected for southerners and residents of rural communities. As suggested by Lee,[9] the expected impact of schooling is uncertain because more highly educated individuals may prefer individual bargaining while unionized employers are likely to have a preference for workers with greater formal education. Measures of marital status and other dependents are included to test Freeman's hypothesis that workers with family responsibilities have a higher propensity toward unionization than persons with small or no family responsibilities.[10] A measure of white-collar occupational status is included in X_i to test Lewis's observation that white-collar workers tend to possess an above average relative preference for individual as opposed to collective bargaining.[11]

In addition to the unionism equation, the model includes two equations representing the wage determination process for union and nonunion workers. The wage equations take the form

$$(2) \qquad \ln W_{ui} = b_{u0} + b_{u1} X_{ui} + \varepsilon_{ui} \quad \text{and}$$

$$(3) \qquad \ln W_{ni} = b_{n0} + b_{n1} X_{ni} + \varepsilon_{ni}$$

where X_{ui} and X_{ni} are vectors of personal characteristics for union and nonunion workers, respectively; and ε_{ui} and ε_{ni} are random disturbance terms. Observed personal characteristics included in X_{ui} and X_{ni} are region of residence, rural residence, length of schooling completed, completion of a postschool training program, age, marital status, and occupation. Formal postschool training is measured by two dummy variables representing company-sponsored training of six or more weeks and completion of a noncompany training program.[12] Residential dummy variables included in the wage equations in lieu of B-K's regional price index account, in part, for regional differences in consumer market baskets and cost of living. Dummy variables for occupational groups allow union-nonunion wage differentials to depend upon skill level.

Estimation of the Unionism and Wage Equations

B-K use OLS to estimate wage relationships similar to Equation 2 and 3. As noted in the introductory section, this procedure ignores the dependence of the observed wage rate on the worker's union status; indeed, only W_{ui} or W_{ni} can be observed for a particular individual. Correlation between the error terms in the union threshold equation and the wage equation means, in general, that expected values of the errors

[9] Lee, "Unionism and Wage Rates," pp. 418–19.

[10] Richard B. Freeman, "Individual Mobility and Union Voice in the Labor Market," *American Economic Review*, Vol. 66, No. 2 (May 1976), pp. 361–68.

[11] H. Gregg Lewis, "Competitive and Monopoly Unionism," in Philip D. Bradley, ed., *The Public Stake in Union Power* (Charlottesville: University of Virginia Press, 1959), p. 191.

[12] Noncompany sources of training include business college or technical institute training, vocational training in the armed forces, and vocational training other than on-the-job training.

in the observed union and nonunion wage equations are nonzero. More precisely,

$$E(\varepsilon_{ui} \mid UN_i > 0) \neq 0 \text{ and}$$
$$E(\varepsilon_{ni} \mid UN_i \leq 0) \neq 0$$

so that OLS estimates of the wage equations are inconsistent. Moreover, the errors will be heteroscedastic.

The underlying idea of the estimation procedure proposed by Heckman[13] and by Lee[14] and used here is first to correct the wage equation error terms to remove their nonzero expected values and then to eliminate the heteroscedasticity. The first correction requires calculating an expression for the means $E(\varepsilon_{ui} \mid UN_i > 0)$ and $E(\varepsilon_{ni} \mid UN_i \leq 0)$. This correction is accomplished by using these expressions to form "selectivity variables" that measure the truncation effect associated with sample selectivity. The selectivity variables are then appended to the sets of explanatory variables in the wage equations and an OLS estimation performed. Using the OLS estimates, the second correction involves estimating variances of the corrected wage equation error terms and then performing weighted least squares. The first part of this section describes the calculation of selectivity variables as part of a consistent two-stage procedure for estimating wage equation coefficients and standard errors using observed data. This is followed by a discussion of empirical results yielded by the procedure.

Estimation procedure. In the first stage of the procedure, Equations 2 and 3 are substituted into Equation 1 to yield the reduced form of the union status equation. The resulting reduced-form equation is a threshold function determining sample selection into union and nonunion employment. Using a weighted nonlinear least squares probit (WNLSP) method,[15] we fit for each individual a union threshold equation of the form $\hat{UN}_i = \theta_0 + \theta_1 X_i' + \theta_2 Y_i$, where X_i' is the vector of all personal characteristics in the model. Then for the union and nonunion sectors, selectivity variables are calculated as $-f(\hat{UN}_i)/F(\hat{UN}_i)$ and $f(\hat{UN}_i)/[1 - F(\hat{UN}_i)]$, respectively, where $F(\bullet)$ is the cumulative distribution of a standard normal variable, and $f(\bullet)$ is its density function. With the inclusion of the selectivity variables, the wage relationships may be rewritten as

$$(2') \quad \ln W_{ui} = b_{u0} + b_{u1} X_{ui} + b_{u2}\left[-\frac{f(\hat{UN}_i)}{F(\hat{UN}_i)}\right] + \delta_{ui} \quad \text{and}$$

$$(3') \quad \ln W_{ni} = b_{n0} + b_{n1} X_{ni} + b_{n2}\left[\frac{f(\hat{UN}_i)}{1 - F(\hat{UN}_i)}\right] + \delta_{ni}$$

where b_{u2} and b_{n2} are covariances between the error term of the reduced-form unionism equation and the wage equation error terms ε_{ui} and ε_{ni}, respectively; and the expected value of the adjusted error terms δ_{ui} and δ_{ni}, conditional on union status, is zero.[16]

The second step of the estimation procedure is to obtain OLS estimates of the coefficients in Equations 2′ and 3′. These estimates will be consistent but inefficient because the conditional variances of the error terms in 2′ and 3′ are heteroscedastic; moreover, OLS standard error formulas will be incorrect. Thus, a generalized least squares (GLS) procedure is utilized, which yields appropriate standard error estimates and more precise coefficient estimates.

After adjusting for selectivity bias and heteroscedasticity, second-stage estimates of Equations 2′ and 3′ can be used to predict a union-nonunion wage differential for each sample member. The predicted series $(\ln \hat{W}_{ui} - \ln \hat{W}_{ni})$ is substituted for $(\ln W_{ui} - \ln W_{ni})$ in Equation 1, and estimates of the structural parameters of the unionism equation are then calculated using WNLSP.

Estimation results. The NLS sample of middle-aged men consists of about 5,000

[13]Heckman, "The Common Structure of Statistical Models of Truncation," pp. 475–80.

[14]Lee, "Unionism and Wage Rates," pp. 420–23.

[15]The WNLSP program is written in the 1976 version of SAS. Copies of a sample program are available from the authors.

[16]The terms $b_{u2}\left[-\frac{f(\hat{UN}_i)}{F(\hat{UN}_i)}\right]$ and $b_{n2}\left[\frac{f(\hat{UN}_i)}{1-F(\hat{UN}_i)}\right]$ are, in fact, the means $E(\varepsilon_{ui} \mid UN_i > 0)$ and $E(\varepsilon_{ni} \mid UN_i \leq 0)$, respectively.

individuals aged 45 to 59 in 1966. Predominantly black neighborhoods were oversampled relative to predominantly white neighborhoods so that the sample contains roughly 3,500 whites. Included in the empirical analysis are white respondents who were employed by a private firm or the government. In addition, respondents must have reported their wage and educational attainment to be included in the sample.

The version of the NLS tapes used in this study reports unionism variables for 1969 and 1971. Hourly wage rates are also available for these two years. Table 1 reports estimated coefficients for the structural equation determining union status and for adjusted and unadjusted union and nonunion wage equations measuring all variables, except age, in 1971.[17] Union status is measured by a dummy variable that takes the value one if the respondent's wage is determined by a collective bargaining agreement between his employer and a union or employee association. Size of the sample in 1971 is 1,879 individuals, with about 40 percent of the sample employed in jobs covered by a union contract. With the exception of schooling all exogenous variables are measured as categorical variables. Omitted categories of the exogenous variables are residence in the South, residence in a SMSA, no postschool formal training, age 45 – 49 in 1966, married with spouse present, no dependents other than wife, service worker, and employment in manufacturing.[18]

The first column of Table 1 displays probit estimates obtained for Equation 1, and these results can be briefly summarized. (Coefficient estimates on the industry dummy variables are not reported to conserve space but are available upon request.) Residence outside the South and residence in a rural community have the expected positive and negative effects, respectively, on union coverage. Similarly, family responsibilities associated with having a wife and other dependents increase the propensity of workers to affiliate with unions. Employment in a white-collar occupation seems to decrease workers' preferences for collective bargaining. Although not shown in the table, results for the industry dummy variables indicate that employment in the transportation industry increases the likelihood of union coverage (relative to the effect of manufacturing), while employment in trade, finance, services, and public administration has a negative impact.

The schooling coefficient in the unionism equation captures the net effect of greater formal education on union status. As suggested in the previous section, more schooling may lead to a preference for individual bargaining as opposed to collective bargaining. But this effect may be offset in whole or in part since a positive union-nonunion wage differential could cause a queue of workers for jobs in the union sector necessitating the use of education as a rationing device. As seen in Table 1, the estimated schooling coefficient is positive and larger than its asymptotic standard error. Although the coefficient is not statistically significant at customary levels, it does provide some evidence that the selectivity of employers dominates workers' tastes. (A significantly positive schooling coefficient is obtained for 1969 data.) Similar findings are reported by Lee using SEO data for operatives and by Farber using the NLS sample of young men.[19]

The final coefficient shown for the unionism equation indicates that the predicted union wage differential has a positive and highly significant effect on union coverage. Lee likewise reports a positive effect (2.455) for the union-nonunion wage differential in his union membership equation esti-

[17]Similar results for 1969 are presented in an appendix available from the authors. Sample size for 1969 is 2,075 individuals, with 38 percent of the sample covered by union contracts.

[18]Other industry categories are agriculture, forestry, and fisheries; mining; construction; transportation, communications, and other public utilities; wholesale and retail trade; finance, insurance, and real estate; services; and public administration.

[19]Lee, "Unionism and Wage Rates," pp. 428 – 30 and Farber, "Unionism, Labor Turnover, and Wages of Young Men," *Research in Labor Economics*, forthcoming. For a sample of male workers from the Michigan Panel Study of Income Dynamics, Abowd and Farber report that individuals with more schooling are more likely to be selected by employers for union jobs if they desire one; but they are less likely to seek employment in the union sector. See Abowd and Farber, "Relative Wages, Union Membership, and Job Queues."

Table 1. Probit Estimates of Union Status Equation and Adjusted GLS and Unadjusted OLS Estimates of Union and Nonunion Wage Equations, 1971.
(Standard errors in parentheses[a])

Explanatory Variables	*Union Status*[b]	*Adjusted Wage Eqs.* Union	*Adjusted Wage Eqs.* Nonunion	*Unadjusted Wage Eqs.* Union	*Unadjusted Wage Eqs.* Nonunion
Constant	– 1.243* (.389)	1.125 (.084)	.409* (.086)	1.079* (.069)	.516* (.083)
Region:					
Northeast	.771* (.111)	.021 (.037)	.061 (.046)	.033 (.034)	.148* (.043)
North Central	.635* (.107)	.017 (.036)	.023 (.043)	.028 (.035)	.096* (.041)
West	.768* (.129)	.085* (.042)	.060 (.053)	.098* (.039)	.147* (.051)
Rural residence	– .553* (.110)	– .022 (.026)	– .158* (.035)	– .028 (.025)	– .188* (.034)
Schooling	.028 (.021)	.021* (.004)	.048* (.005)	.020* (.004)	.047* (.005)
Training:					
Company		.043 (.031)	.167* (.039)	.041 (.031)	.146* (.040)
Noncompany		.003 (.018)	.048* (.023)	.001 (.018)	.036 (.023)
Age: 50 – 54		.006 (.025)	.032 (.035)	.005 (.025)	.031 (.036)
55 – 59		.016 (.030)	– .030 (.042)	.016 (.030)	– .039 (.042)
Not married	– .306* (.150)	– .053 (.040)	– .232* (.056)	– .055 (.040)	– .223* (.056)
1+ dependents	.117 (.071)				
Occupation:					
White-collar	– .514* (.175)	.226* (.053)	.516* (.064)	.229* (.053)	.442* (.063)
Craftsmen		.254* (.051)	.270* (.066)	.269* (.049)	.362* (.065)
Operatives		.084 (.054)	.026 (.075)	.105* (.049)	.138 (.071)
Laborers		– .019 (.064)	– .126 (.081)	– .005 (.063)	– .151 (.082)
Selectivity variable		.047 (.049)	.331* (.076)		
$\ln \hat{W}_u - \ln \hat{W}_n$	1.718* (.479)				

[a]Asymptotic standard errors are shown for union status and adjusted wage equations.
[b]Also included in the equation are nine industry dummy variables.
*Significant at .05 level using a two-tailed test.

mated for operatives. The probit coefficient estimate on ($\ln \hat{W}_u - \ln \hat{W}_n$) can be readily transformed to give the more familiar interpretation of the partial derivative of the probability of union coverage with respect to a change in the union-nonunion wage differential.[20] Evaluating all explanatory variables at their mean values, the probit estimate 1.718 corresponds to a partial derivative of 0.62, which implies that a one percentage point increase in the union wage differential above its mean of 34.1 percent yields an increase of .62 percentage points in the proportion of workers in the sample covered by a union contract.

The second through fifth columns of Table 1 present a comparison of Equations 2′ and 3′ estimated by GLS with OLS estimates of Equations 2 and 3. The unadjusted results shown in the fourth and fifth columns, if they were reliable, would suggest that estimated impacts of personal characteristics, including formal schooling, company and noncompany training, and marital status, are considerably larger in the nonunion sector than in the union sector. In addition, region and type of residence are seen to be more important in determining nonunion sector wages than union sector wages. B-K also report that their regional price index has a larger impact in the nonunion sector.

Wage equation estimates adjusted for selectivity bias reported in the second and third columns of Table 1 show little change from the unadjusted results in both sectors for schooling, training, age, and marital status.[21] Our adjusted estimates as well as the unadjusted results thus echo B-K's finding that the union sector earnings function is less responsive than the earnings function for the nonunion sector to changes in nearly every observable attribute. An interpretation consistent with this finding is that the long-term concentration of organized labor on standardizing rates of pay under collective bargaining agreements has resulted in a substantial reduction in the dispersion of earnings within the union sector.[22] Among the occupational categories appearing in the table, the only differences between the adjusted and unadjusted results worth noting are the increase in the coefficient on white-collar occupations and the decrease in coefficients estimated for craftsmen and operatives in the nonunion sector. Thus, estimated wage differentials between white-collar and blue-collar occupations show a slight increase in the nonunion sector going from the unadjusted to the adjusted results.

The most important difference between the adjusted and unadjusted wage equation estimates is the decrease in size of adjusted coefficient estimates on the residence variables in the nonunion sector. The sizable union-nonunion differences in the effect of region of residence on earnings in the unadjusted results largely disappear in the adjusted wage equations. Indeed, region of residence is found to have a small effect in both adjusted wage equations. Similar results are obtained using 1969 data. Since region is an important component of the selectivity variables, inclusion of the selectivity variables in the adjusted wage equations eliminates the positive impact of a non-South residence on wages for both sectors. In other words, the estimated effect on earnings of region of residence is substantially reduced once account is taken of regional differences in preferences toward unionism.

[20]The partial derivative of the probability of union coverage with respect to a particular explanatory variable Z_k is written as

$$\frac{\partial Pr(UN=1)}{\partial Z_k} = a_k f(Za)$$

where a_k is the coefficient on Z_k, and $f(Za)$ is the value of the normal density function at the point Za.

[21]It should also be noted that while GLS estimates of the adjusted union wage equation differ hardly at all from corresponding OLS estimates, the GLS procedure produces a noticeable difference (in the second place to the right of the decimal point) between GLS and OLS coefficient estimates of the adjusted nonunion sector equation.

[22]To the extent that unions raise the average wage and reduce the dispersion of intrafirm earnings, workers near the bottom of the earnings distribution would be expected to show the greatest interest in unionization. Recent evidence provided by Farber and Saks shows that the probability of voting for a union in NLRB representation elections is strongly and inversely related to the individual's position in his employer's earnings distribution. See Henry S. Farber and Daniel H. Saks, "Why Workers Want Unions: The Role of Relative Wages and Job Characteristics," *Journal of Political Economy*, Vol. 88, No. 2 (April 1980), pp. 349–69.

Turning to the estimated coefficients on the selectivity variables, the coefficient in the nonunion equation is positive and precisely estimated but the corresponding coefficient in the union equation is imprecisely estimated. Since the selectivity variable in the nonunion equation is positive, its positive coefficient estimate (.331) indicates a positive selection bias in the determination of nonunion sector employment (or, more formally, that $E(\varepsilon_{ni} | UN_i \leq 0)$ is positive). To interpret this result, consider the subsample of union and nonunion workers with the same personal characteristics as those of nonunion workers. Positive selectivity (or positive truncation) implies that the wage distribution actually observed for nonunion employees is higher than the distribution that would be observed for the average individual in the subsample had he found nonunion sector employment. That is, average wages of workers with given personal characteristics actually selected into nonunion jobs exceed what nonunion sector wages would have been for workers with the same characteristics selected instead into unionized employment. Using data specific to operatives, Lee also reports evidence of positive selection for workers actually found in the nonunion sector.[23]

Given the negative sign of the union sector selectivity variable, failure to obtain a negative estimate of b_{u2} indicates that positive selection does not occur in the union sector wage determination process. That is, average wages of workers with given personal characteristics actually found in union jobs are no more likely to exceed than to fall below what wages would have been for similar workers employed in the nonunion sector. In contrast, Lee reports positive selectivity for union sector observations. Apparently, moving from Lee's relatively homogeneous sample of operatives to the full range of occupations examined here raises the distribution of wages available in both sectors to individuals with the same average characteristics of union workers. Hence, while unionized operatives may enjoy superior earnings opportunities as union members than would nonunion operatives with similar personal characteristics, the same statement cannot be made with respect to middle-aged union members vis-à-vis nonunion workers across occupational groups. It is also important to recognize that the absence of positive selectivity in the union sector is not inconsistent with evidence of a strong direct impact on union status of the predicted union-nonunion wage differential. Lee points out that the signs of coefficients on the selectivity variables depend on the second moments of the disturbances ε, ε_u, and ε_n since the error term in the reduced-form unionism equation is a composite of the error terms in the structural model.[24] If there are important but unmeasured factors common to all equations, a nonnegative estimate of b_{u2} may be obtained even though the coefficient on the predicted union wage differential is positive.

Estimated Union-Nonunion Wage Differentials

In a model containing a single wage equation, the estimated union-nonunion wage differential can be calculated immediately from the coefficient estimate on the union dummy variable. When separate union and nonunion wage equations are estimated, however, coefficient estimates on the right-hand-side variables must be weighted in some manner to arrive at an estimated union wage differential. Since there are several weighting schemes that might be used, a number of different estimates of the relative wage effect of unions are possible. One approach to assessing the validity of alternative estimates is to attempt to describe the conceptual experiments to which they correspond.

B-K suggest that alternative measures of the union wage differential might be calculated as $\hat{b}_u(\overline{X}_n - \overline{X}_u)$ and $\hat{b}_n(\overline{X}_n - \overline{X}_u)$, where $\hat{b}_u$ and $\hat{b}_n$ are vectors of estimated coefficients in the union and nonunion wage equations, respectively; and $\overline{X}_u$ and $\overline{X}_n$ are vectors of mean values of personal characteristics of union and nonunion workers, respectively.[25] B-K's first measure,

[23]Lee, "Unionism and Wage Rates," pp. 425–26.

[24]Ibid., p. 426.

[25]Bloch and Kuskin, "Wage Determination in the Union and Nonunion Sectors," p. 188.

for example, weights differences in means of personal characteristics of nonunion and union workers by estimated market returns to these characteristics in the union sector. Since their model does not describe the mechanism by which some workers are employed in the union sector while others hold nonunion jobs, it is not possible to formulate a conceptual experiment that corresponds to this measure. It is clear, nevertheless, that by weighting intersector differences in personal characteristics, the alternative measures presented by B-K do not capture the customary definition of the relative wage effect of unions that is the presence of differences in pay for union and nonunion workers of the same productivity. Instead of measuring the relative wage effect of unions, B-K's formulas partly measure the effect of unions in generating a differential in personal characteristics between union and nonunion workers.[26]

The approach taken here is to weight differences in estimated coefficients in the union and nonunion wage equations by mean values of workers' characteristics in the nonunion sector. Using notation previously introduced, the logarithmic differential in wages is

$$d = \overline{X}_n(\hat{b}_u - \hat{b}_n),$$

which implies a corresponding percentage union-nonunion wage differential D of the form

$$D = e^d - 1.$$

Table 2 compares Ds calculated using adjusted GLS and unadjusted OLS wage equation estimates for the five occupational groups specified in the wage relationships. Separate sets of estimates are shown for 1969 and 1971.

The first two columns of Table 2 display D estimates calculated from adjusted wage equation coefficients displayed in the sec-

Table 2. Union-Nonunion Wage Differentials Estimated Using Adjusted and Unadjusted Wage Coefficients, 1969 and 1971, by Occupation.[a]

	Using Adjusted Coefficients		
Year and Occupation	*Omitting Selectivity Variables*	*Including Selectivity Variables*	*Using Unadjusted Coefficients*
1969: White-collar	.217	.112	.092
Craftsmen	.429	.306	.165
Operatives	.425	.302	.164
Laborers	.416	.294	.175
Service workers	.402	.282	.161
1971: White-collar	.331	.195	.156
Craftsmen	.534	.377	.260
Operatives	.551	.392	.282
Laborers	.553	.394	.335
Service workers	.540	.383	.287

[a]Nonunion means used as weights.

[26]Lee's approach to measuring the union wage differential involves first cross-classifying his sample by race, sex, market experience, education, industry, and percent of industry organized. For each subgroup, estimated union and nonunion wage equations are used to predict union and nonunion wage rates from which the percentage union relative wage effect is calculated. Finally, percentage union wage differentials are summed across all subgroups to yield an overall estimate of the union-nonunion wage differential. Lee does not discuss his procedure in any detail, but he seems to assume homogeneity of sample members within each subgroup. If union and nonunion workers differ with respect to personal characteristics on which the sample is not stratified (region of residence, size of residence, and health, for example), union wage differentials calculated for each subgroup capture the effect of intersector differences in personal characteristics as well as intersector differences in wages. In this respect, Lee's procedure is subject to the same criticism levied against B-K. (Lee, "Unionism and Wage Rates," p. 426.)

ond and third columns of Table 1. The two columns of Table 2 differ from each other by omitting (in the first column) and including (in the second column) in the calculations means and estimated coefficients of the selectivity variables. *D*s shown in the first column correspond to the following conceptual experiment: From a sample of both union and nonunion workers in a particular occupation, pick at random an individual with the average characteristics of nonunion workers. Then predict a union and a nonunion wage rate for this individual. The percentage union wage differential calculated from predicted wages measures how much greater the wage predicted for a given set of personal characteristics is in union than in nonunion employment. In other words, estimated *D*s appearing in the first column represent the outcome of an unconditional experiment in which the union selection process is modeled in addition to specifying how wages are determined in the union and nonunion sectors.

Including means and estimated coefficients of the selectivity variables in the calculations reported in the second column changes the conceptual experiment to a conditional experiment in which the union selection process is taken as given and market returns to personal characteristics in the union and nonunion sectors are predicted. That is, these *D* estimates correspond to the experiment in which union and nonunion wage rates are predicted for an individual with mean personal characteristics of nonunion workers who is chosen at random from the subsample of workers already selected into the nonunion sector. Estimates in the second column are seen to be uniformly smaller than those in the first column. This is because for both 1969 and 1971, the coefficient estimate on the selectivity variable for the nonunion sector is larger than that for the union sector and the nonunion selectivity variable is positive. The intuition underlying this result is that including the selectivity variables in the calculations and using nonunion means imply that union wage differentials are estimated for individuals who have already opted for nonunion employment. The model represented by Equation 1 presumes that nonunion workers are nonunion because, holding their tastes for unionism constant, they and their employers believe that they are relatively more productive in nonunion jobs. If their beliefs are correct, conditional union wage differentials should be smaller than unconditional estimates representing the situation in which choice of union status has not already been made. If union means were used in the calculations, the impact of including the selectivity variables should be to increase conditional estimates of *D* relative to unconditional estimates. Since the selectivity variable appearing in the union sector equation is negative, this conjecture is supported.

For completeness, the third column of Table 2 reports *D*s calculated using wage equation coefficients estimated without correction for sample selection (see the fourth and fifth columns of Table 1). These differentials are useful for measuring the impact of unionism on relative wages because the inclusion of union status results in an improved fit of estimated relationships. However, the differentials have no behavioral interpretation in the sense of providing information on the relative earnings opportunities facing any individual in the union and nonunion sectors. Despite this limitation of *D*s calculated using unadjusted wage equation estimates, it is interesting to note that the positive sign of the estimates in the third column is maintained in the estimated *D*s appearing in the first two columns. In fact, estimates obtained on the basis of the adjusted wage equations are uniformly larger, often substantially larger, than those appearing in the third column. This suggests that unions may be more important in raising the relative wages of their members than has been implied by past studies that ignore the selection of workers into union and nonunion jobs.

Conclusion

Recent evidence presented by Bloch and Kuskin (B-K) indicates that wages are determined differently in the union and nonunion sectors of the labor market. This paper corrects the statistical methodology used by B-K and provides additional empirical results using white respondents from the NLS sample of middle-aged men. Although the focus of the paper is on the determination of union and nonunion wages, the discussion is applicable to many situations in which market outcomes depend on a selection rule that sorts individuals into distinct groups. Examples of empirical studies that utilize the adjustment for selectivity bias include Heckman's estimation of a wage determination model for married women whether or not they work, Willis and Rosen's investigation of the impact of high school and college schooling on earnings allowing expected lifetime earnings to influence the decision to attend college, and Borjas's treatment of the impact of unionism on individuals' supply price to alternative jobs, taking account of self-selection in the proportion of the sample that provided a numerical answer to the reservation wage question.[27]

The first issue addressed is B-K's use of OLS to estimate separate union and nonunion wage equations without taking into account the fact that observed wage rates depend on workers' union status. Using a consistent two-stage estimation procedure, the union status of middle-aged whites is found to be strongly affected by the predicted union-nonunion wage differential. Nevertheless, consistent estimates obtained for union and nonunion wage equations support B-K's finding that the marginal impact on earnings of individual characteristics is smaller in the union sector than in the nonunion sector. The major difference between our results and those of B-K is that our study shows region of residence (representing regional price differences, among other factors) does not have a significant effect on wages in either sector once correction has been made for sample selectivity.

The other issue that is not adequately discussed by B-K or elsewhere in the literature is how to calculate and interpret union-nonunion wage differentials obtained from separately estimated union and nonunion wage equations. With respect to B-K's approach, it is pointed out that their calculations do not measure the relative wage effect of unions and, more generally, that union wage differentials based on wage equations uncorrected for sample selection do not provide information on the earnings opportunities in union and nonunion employment facing any individual. Using wage equation estimates adjusted for sample selection, the evidence for middle-aged white men indicates that unions have a sizable positive impact on relative wages measured across major occupations.

[27] James J. Heckman, "Shadow Prices, Market Wages, and Labor Supply," *Econometrica,* Vol. 42, No. 4 (July 1974), pp. 679 – 94; Robert J. Willis and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy,* Vol. 87, No. 5, Part 2 (October 1979), pp. S7 – S36; and George J. Borjas, "Job Satisfaction, Wages, and Unions," *Journal of Human Resources,* Vol. 14, No. 1 (Winter 1979), pp. 21 – 40.