WORKING PAPER #452
PRINCETON UNIVERSITY
INDUSTRIAL RELATIONS SECTION
MAY 2001

Notes on the Economics of Labor Unions*

Henry S. Farber
Princeton University

Abstract

A common definition of a labor union is that it is an association of workers who bargain collectively with their employer regarding the terms and conditions of employment. Economic analysis of labor unions falls largely into four related categories: 1) unions as maximizing agents, 2) dispute resolution: strikes and arbitration, 3) the effect of unions on wages, and 4) the determination of union membership. This essay is organized around discussions of these topics.

*Industrial Relations Section, Firestone Library, Princeton University, Princeton, NJ 08544-2098. email: farber@princeton.edu. This is an expanded version of entries prepared for *The International Encyclopedia of the Social and Behavioral Sciences*. I thank John Pencavel for comments on an earlier draft.

Introduction

A common definition of a labor union is that it is an association of workers who bargain collectively with their employer regarding the terms and conditions of employment. Economic analysis of labor unions falls largely into four related categories:

1. Unions as Maximizing Agents.

2. Dispute Resolution: Strikes and Arbitration.

3. The Effect of Unions on Wages.

4. The Determination of Union Membership.

This essay is organized around discussions of these topics. The discussions rely on the (admittedly atypical) institutional arrangements that exist in the United States, but the key points are applicable more generally to unions in market economies. The presentation is necessarily selective, and it is meant to give an overview of a very substantial literature.

1 Unions as Maximizing Agents

The general approach to modeling union behavior is to assume that the union's goal is to maximize a well-defined utilty function of wages and employment, $U(W, L)$. The union is assumed to maximize this objective function subject to constraints imposed by the employer. Three classes of solutions to this maximization problem have been used. The first is the labor-demand-curve (LDC) model, also known as the right-to-manage model. In this model the union chooses the wage that maximizes $U(W, L)$ subject to the constraint that the firm selects the employment level corresponding to that wage on the labor demand curve. This model leads to an economically inefficient outcome in that the wage will generally be set above the competitive level so that employment will be lower than required for economic efficiency. The outcome is also inefficient in a private sense because both the union and the firm could be made better off by choosing a wage-employment combination that is off the labor demand curve. But the structure of the wage-employment determination process in the LDC model does not admit solutions off the labor demand curve.

The second class of model is what is called the weakly-efficient model. In this model, the union maximizes $U(W, L)$ with respect to W and L subject to a minimum profit constraint for the firm. The same outcome can be derived by solving the dual of this problem where the firm chooses W and L to maximize profits subject to a minimum utility constraint for the union. This outcome is privately efficient in that the outcome is Pareto-efficient from the standpoint of the the union and the firm, but it is not economically efficient in that it does not maximize the dollars available to be divided between workers and the firm.

Economic efficiency requires that employment be set at the level where the value of marginal product (VMP) of labor equals the opportunity cost of an hour for workers. The third class of model, the strongly-efficient model, has this property. In other words, the union and employer agree on wage and employment levels that depend only on the opportunity wage of labor and labor's VMP. The strongly efficient model (e.g., Hall and Lilien, 1979) is a special case of the weakly-efficient model where the union objective function has the form $U(W, L) = G((W - W_a)L)$, where $G(\cdot)$ is an increasing monotonic function and W_a represents the opportunity wage of workers.

Each of these classes of models has been investigated empirically. For example, the labor-demand-curve model was applied by by Farber (1978b) the United Mine Workers and by Dertouzos and Pencavel (1981) to the International Typographers Union (ITU). The weakly-efficient model was applied by MaCurdy and Pencavel (1986) to the ITU. Brown and Ashenfelter (1986) also use data from the ITU to examine the weakly-efficient and strongly-efficient models. They estimate an employment function of the form

$$lnL = X\beta + \gamma_1 W + \gamma_2 W_a + \epsilon \tag{1.1}$$

where $X\beta$ measures non-wage factors affecting employment. The central proposition of the strongly-efficient model is that employment is affected by the alternative wage of workers ($\gamma_2 < 0$) but not by the contract wage ($\gamma_1 = 0$). In contrast, the weakly-efficient model implies that employment will be related to both the contract wage and the alternative wage. However, a finding of a relationship of employment with both wage measures is necessary but not sufficient to conclude that the contract is weakly efficient. Neither MaCurdy and Pencavel (1981) nor Brown and Ashenfelter (1986) find strong evidence for strong efficiency, but they reach differing conclusions regarding weak efficiency. MaCurdy and Pencavel conclude that the weakly-efficient model may be consistent with the data, while Brown and Ashenfelter conclude that this model is not consistent with the data.

Abowd (1989) provides an interesting test of the strongly-efficient model that relies on the stock market response to unanticipated changes in the wage rate as a result of contract negootiations. Using a large sample of labor contracts for publicly traded firms, Abowd finds that the change in the market value of firms in response to the present discounted value of unanticipated wage changes is not significantly different from one for one. While this is consistent with the strongly-efficient model, Abowd's estimate of the tradeoff is not sufficiently precise to rule out divergence from the model.

It is important to recognize that rejection of the strongly-efficient model does not imply that the weakly-efficient model is appropriate. After all, most bargains are explicitly about wages alone and not about employment levels. While there are sometimes work-rules negotiated that attempt to control capital-labor ratios or output-labor ratios, they are unlikely to eliminate discretion in setting employment levels from the employer. Most bargaining takes place in a LDC model framework (bargaining over the wage with the employer free to set employment), and work rules are an attempt (partially successful, at best) to move toward efficiency by pushing the employer off the labor demand curve.

The existing work on wage and employment determination largely ignores the bargaining problem. In fact, the wage, and employment where relevant, are set in a bargaining context. In the labor-demand-curve model, bargaining changes the wage but leaves the wage-employment combination on the labor demand curve. In the weakly-efficient model, bargaining can be characterized as determining the minimum profit level that the union must offer the firm or, in the dual, the minimum utility level the firm must offer the union. In the strongly-efficient model, bargaining determines the wage rate paid to the worker but does not affect employment levels.

2 Dispute Resolution: Strikes and Arbitration

Negotiation is common in many settings. A common feature of all negotiation is an understanding by the parties to the negotiation about the process that will be used to resolve a dispute in the event the parties fail to reach agreement. Disputes between labor unions and employers can end in a strike or can be resolved by arbitration. Disagreement over contracts or tort damages can result in civil litigation and a decision by a judge or jury. Disagreement between nations can result in military disputes. The dispute settlement process defines the expectations that the parties have about what will happen in the event that no agreement is reached. And this process, in turn, determines negotiator behavior.

There are two fundamental types of dispute settlement mechanisms. The first is the systematic imposition of costs on each side. Strikes and military disputes are examples of this type of mechanism. No determinate end to the dispute in implied, but presumably one or both parties concedes sufficiently to make an agreement possible and end the dispute. The second mechanism is the mandating of a settlement by a third party. Arbitration of labor and commercial disputes and litigation of civil disputes are examples of this type of mechanism.

Both types of mechanisms (strikes and arbitration) are used in the resolution of labor disputes and the discussion here focuses on that area. The discussion here is necessarily selective and is meant to provide a brief introduction to important concepts and issues. See Hirsch and Addison (1986) for a more extensive survey.

2.1 Strikes

A difficult problem in the economics of labor unions is the modeling of strikes. The difficulty arises because strikes appear to be non-equilibrium phenomena. Both sides would have been better off if they had been able to agree on whatever the outcome was without bearing the costs of a strike. In other words, strikes are Pareto-dominated by the same bargaining outcome without a strike. While research on strikes in the United States has diminished with the level of strike activity in the U.S. over the last 15 years, there was an active empirical literature in the 1970s and 1980s.

Ashenfelter and Johnson (1969) developed a non-equilibrium model of strike activity by attributing to the union an arbitrary concession schedule positing an inverse relationship between the minimum acceptable wage and the length of strike. The firm selects the point on this schedule that maximizes profits, the tradeoff being longer strikes involved larger up-front strike losses but lower labor costs into the future. Ashenfelter and Johnson used this model to motivate an empirical analysis of the aggregate level of strike activity in the United States. Farber (1978a) used the same theoretical framework to estimate a structural model of the joint determination of wage changes and strike length using data on the outcomes of a time-series of cross-sections of negotiations for large U.S. corporations.

Tracy (1986) developed an equilibrium model of strikes based on learning in a model of asymmetric information (Fudenberg and Tirole, 1981; Sobel and Takahashi, 1983). The model is based on the idea that the union does not know the profitability of the firm. The union would like to extract a higher wage from high-profit firms and a lower wage from low-profit firms. The key is that high-profit firms have a high cost of bearing strikes relative to low-profit firms. In this case the union can make a high initial wage demand. High-profit firms will prefer to pay this wage immediately rather than to lose profits during a strike. Low-profit firms will prefer to lose their lower profits during the strike in exchange for a lower wage into the future. Tracy used this model to generate the empirical implication that there will be more strikes where there is more uncertainty about the firm's profitability. He then implemented this model using data on a large number of negotiations and using the volatility of daily stock returns as his measure of uncertainty about profitability. He finds evidence that uncertainty is positively correlated with strike activity. A final point is that the declining concession schedule in the Ashenfelter-Johnson model can be interpreted as a reduced-form expression of the union learning function. As the strike goes on, the union is willing to accept a lower wage as they learn that the firm is lower-profit.

It is quite possible that strikes are not equilibrium phenomena and that the insistence in the economics literature, both theoretical and empirical, on equilibrium explanations is inappropriate. Strikes can result from miscalculation. In particular, the parties may be relatively optimistic about the outcome after a strike. For example, the union may expect a wage settlement of $20/hour after a strike while the firm expects to pay only $18/hour after the same strike. If the cost of a strike is \$0.50/hour to each side, then the union will accept no less than \$19.50 to avert the strike while the firm will be willing to offer no more than \$18.50. The result will be a strike. Of course, both sides expectations cannot be accurate, *ex post*. This relative optimism can have a variety of sources. One possibility, with a long history in the industrial relations literature, is that the parties selectively overweight arguments that are favorable to their side (Walton and McKersie, 1965; Neale and Bazerman, 1991). However, relative optimism can be generated even with appropriate weighting of available information. If each side is getting independent (and accurate in expectation) information on the likely uncertain outcome, some fraction of the time the information will lead to relatively optimistic expectations sufficient to outweigh the costs of a strike. Strikes will occur in these cases.

2.2 Arbitration

In the United States, arbitration is commonly used to determine the terms of labor contracts in the public sector in cases where the parties cannot agree on their own. In conventional arbitration, the arbitrator is free to impose whatever settlement he or she deems appropriate. A common alternative is final-offer arbitration, where the arbitrator must select one of the final offers submitted by the parties and cannot compromise.

2.2.1 Conventional Arbitration

Consider the following simple model of the negotiation process where conventional arbitration is the dispute settlement mechanism (Farber and Katz, 1979). Suppose the parties are negotiating over the split of a pie of size one. The share of the pie going to the union is Y, and the share going to the firm is $(1-Y)$. The arbitrator will impose a settlement of Y_a, and each party has a prior distribution on Y_a. Let Y_a^u and Y_a^f represent the prior expected value to the union and firm respectively of the arbitration award. Finally, let C_u and C_f represent the cost to the union and the firm of using arbitration (denominated in pie shares). Assume further that the parties are risk neutral so that the union will prefer any negotiated share for itself (Y_n) such that $Y_n > Y_a^u - C_u$ and the firm will prefer any negotiated share for itself such that $1-Y_n > 1-Y_a^u-C_f$. These are the reservation values of the parties in negotiation. The result is a *contract zone* of settlements preferred by both sides to arbitration that consists of all shares that fall in the range

$$Y_a^u - C_u < Y_n < Y_a^f + C_f. \tag{2.1}$$

The size of the contract zone is

$$CZ = Y_a^f - Y_a^u + C_u + C_f. \tag{2.2}$$

In the simplest model, there will be a negotiated settlement whenever $CZ \geq 0$, and there will be an arbitrated outcome whenever $CZ < 0$.

This model provides important insight into when one might expect a dispute. Relative optimism about the likely arbitration award $(Y_a^u > Y_a^f)$ can be a reason for a dispute. If the relative optimism is sufficiently great to outweigh the total costs of arbitration $(C_u + C_f)$, then there will be no contract zone and arbitration will be required. Thus, relative optimism is a cause of disputes. Clearly, higher total costs of arbitration imply a larger contract zone. Larger costs make it less likely that any degree of relative optimism will cause there to be no contract zone. Thus, larger total costs of disputing reduce the likelihood of disputes. This latter property is implied by any reasonable model of disputes.

The model also demonstrates how the dispute settlement mechanism fundamentally determines negotiated outcomes. Consider the case where the parties have identical expectations about the arbitration award of $Y_e = Y_a^u = Y_a^f$. The size of the contract zone in this case is determined entirely by the costs of arbitration (equation 2.2), and there will be no dispute. The negotiated settlement will lie in the range defined by

$$Y_e - C_u < Y_n < Y_e + C_f. \tag{2.3}$$

In other words, the negotiated settlement will lie in some range around the common expectation of the arbitration award.

One criticism of conventional arbitration as a dispute settlement mechanism in labor disputes is that arbitrators appear to split the difference between the demands of the parties, resulting in an arbitration award that is close to the average of the demands. However, equation 2.3 makes clear that the opposite is likely to be true: The negotiated settlement is likely to lie between the reservation values of the parties. Mnookin and Kornhauser (1979) and Farber (1981) present analyses that make this point.

2.2.2 Final-Offer Arbitration

The perception that arbitrators split the difference in conventional arbitration led the belief that conventional arbitration "chilled" bargaining by providing the parties with an incentive to maintain extreme demands. This led to the development of final-offer arbitration, which prevented the arbitrator from compromising (Stevens, 1966). Farber (1980) developed a simple model of final-offer arbitration where the arbitrator selects the offer closest to what he or she would have imposed in conventional arbitration (Y_a). In this model, the probability that the union's offer is selected is

$$P_u = Pr(Y_a > \frac{Y_u + Y_f}{2}) = 1 - F(\frac{Y_u + Y_f}{2}) \tag{2.4}$$

where Y_u and Y_f represent the final offers of the union and the firm respectively and $F(\cdot)$ represents the cumulative distribution function of the parties' common prior on Y_a.

Assuming risk-neutrality, the expected payoff to the union from final-offer arbitration is

$$V_u = P_u Y_u + (1 - P_u) Y_f, \tag{2.5}$$

and the expected payoff to the firm is

$$V_f = 1 - [P_u Y_u + (1 - P_u) Y_f]. \tag{2.6}$$

The Nash equilibrium final offers of the parties are the offers such that 1) Y_u maximizes V_u conditional on Y_f and 2) Y_f maximizes V_f conditional on Y_u. These offers have the property that neither party will desire to change its offer.

Since the parties are uncertain about Y_a, the Nash equilibrium final offers reflect a tradeoff between the value to the party of its offer if its offer is selected and the probability that its offer is selected. A more extreme offer yields a higher payoff to the the party if its offer is selected by the arbitrator, but it lowers the probability that the offer will be selected.

A key implication of this model, is that final offer arbitration may not encourage more convergence in negotiation than conventional arbitration because of the desire to present optimal final offers to the arbitrator that may be extreme.

2.2.3 Empirical Evidence

Ashenfelter and Bloom (1984) present evidence from arbitration of contract disputes for public sector employees in New Jersey, where both conventional and final-offer arbitration are used, that is consistent with the general model of arbitrator behavior outlined here. They find that arbitrators use a common underlying notion of Y_a to make awards in both conventional and final-offer arbitration. Their evidence is consistent with the arbitrator's decision in final-offer arbitration being to select the offer closest to Y_a. Bazerman and Farber (1986) reach the same conclusions based on evidence from arbitration awards, both conventional and final-offer, made by professional arbitrators in hypothetical cases.

Convincing evidence is scarce on whether final-offer arbitration leads to more settlements and reduces reliance on arbitration. Ashenfelter, Currie, Farber, and Spiegel (1992) present experimental evidence that does not show higher rates of agreement in final-offer arbitration than in conventional arbitration

3 The Effect of Unions on Wages

3.1 Unions and Wages: What is Measurable?

Lewis's (1963) classic monograph set the standard for empirical analysis of the effect of unions on wages using aggregate (e.g., industry level) data. As individual level data become available beginning in the late 1960's, a substantial literature, surveyed comprehensively by Lewis (1986), developed addressing this same question using the new data. Other useful surveys of this literature are contained in Freeman and Medoff (1984), Hirsch and Addison (1986), and Booth (1995).

There are at least two levels on which unions can affect wages in an economy. The first level is the direct effect of unions on the wages of workers in jobs where wages are set through collective bargaining (for non-members as well as for members). The second and less well understood level is that the presence of unions in an economy can change the level and distribution of wages generally. For example, unions can raise the wages of their members, but the wages of nonunion workers can fall or rise as well. A standard competitive market model would imply a fall in the wage of nonunion workers. The mechanism is that unions, in raising wages of their members, cause a decline in employment in the union sector. The result is an increase in supply of workers to the nonunion sector as displaced union workers move to that sector, implying a drop in the nonunion wage rate, called the "spillover effect" of unions. A force working to raise nonunion wages is the so-called "threat effect" of unions. The threat effect results from nonunion employers raising the wages of their workers in order to avoid becoming unionized (Rosen, 1969; Freeman and Medoff, 1981). Thus, based on theoretical considerations, the effect of unions on the level of aggregate wages in an economy is not clear.

In order to be more precise, define the effect of unions on the wages of union and nonunion workers respectively as

$$\Delta_u = \frac{W_u - W_0}{W_0}, \tag{3.1}$$

and

$$\Delta_n = \frac{W_n - W_0}{W_0}, \tag{3.2}$$

where W_u and W_n represent the union and nonunion wage respectively and W_0 represents the wage that would have prevailed in the absence of unions in the economy. It is not possible to estimate either of these wage effects because W_0 is not observable.

The union-nonunion wage differential, called the union wage gap and defined as

$$\Delta = \frac{W_u - W_n}{W_n}, \tag{3.3}$$

is estimable and is related to the underlying union and nonunion wage effects by

$$1 + \Delta = \frac{1 + \Delta_u}{1 + \Delta_n}. \tag{3.4}$$

As long as all differentials are small (less than 0.25 or so), this expression is usefully approximated by

$$\Delta \approx \Delta_u - \Delta_n. \tag{3.5}$$

This relationship states that the measured union wage gap is approximately equal to the difference in the proportional effects of unions on the union and nonunion wage.

The union wage gap, defined in equation 3.3, can be usefully approximated by the difference in log wages. This implies that

$$\Delta \approx ln(W_u) - ln(W_n), \tag{3.6}$$

and estimation of this quantity is the goal of much of the empirical literature on the effects of labor unions.

The union wage gap reflects a combination of effects. Clearly it includes the direct effect of unions on the wage of unionized workers, and it includes the offsetting effects, described above, of unions on the wage of nonunion workers. It is not possible to parse estimates of the union wage gap into its component effects, and estimates of Δ have been broadly interpreted as the effect of unions on wages.

A further complication is that workers are heterogeneous and only a single wage is observed for each worker: the union wage for union workers and the nonunion wage for nonunion workers. If both wages were observed for workers then calculation of Δ would be straightforward. But workers make choices regarding whether to seek union or nonunion employment in part based on the wage they would receive in each sector. Thus, union status is endogenous to wages, and the result is a classic selection bias problem that has been treated extensively in the literature. In the presence of this selection bias, the value of Δ estimated from a standard cross-section of workers can be interpreted as the average difference in wages between union and nonunion workers, but it cannot be interpreted as the effect of union membership on the wage of a particular worker.

3.2 Estimating the Union Wage Gap

Labor unions have been variously characterized as organizations whose operational goals have been to "take wages out of competition", to impose a "standard rate", and to "attach wages to jobs rather than to workers." This characterization has important implications for who will be union members and, hence, for estimation of the effect of union membership on wages.

The standard approach to estimation of the union-nonunion wage gap using individual-level data is to estimate by OLS

$$lnW_i = X_i\beta + \delta U_i + \epsilon_i, \tag{3.7}$$

where i indexes individuals, X_i is a vector of worker characteristics including such skill measures as education and labor market experience, U_i is a dummy variable indicating union membership, and ϵ_i is a random component. The parameter δ represents the average proportional difference in wages between union and nonunion workers adjusted for measured differences in worker characteristics, and it is the regression adjusted analogue of Δ. Most estimates of δ are in the 0.1 - 0.25 range.

A central problem with this approach is that it constrains the marginal effect of skill on earnings to be the same in both union and nonunion jobs. Given that unions tend to "standardize wages" and "attach wages to jobs", it is likely that the function characterizing the earnings of union workers will be flatter in skill dimensions than the function for nonunion workers. Specifying separate earnings functions for the two sectors yields

$$lnW_{ni} = X_i\beta_n + \epsilon_{ni}, \tag{3.8}$$

for nonunion workers and

$$lnW_{ui} = X_i\beta_u + \epsilon_{ui} \tag{3.9}$$

for union workers. The usual *modus operandi* of unions implies that $\beta_u < \beta_n$ in skill dimensions such as education and experience. Since wages are tied less closely to individual characteristics in the union sector and since the X_i vector largely consists of individual characteristics, the R^2 from cross-sectional estimation is generally smaller for the union earnings function than for the nonunion earnings function. It is also the case that the residual variation in earnings is smaller in the union sector than in the nonunion sector, again likely the result of standardization. Bloch and Kuskin (1978) were among the first to estimate separate union and nonunion earnings functions and to note the systematic differences. Separate earnings functions imply that the estimated union-nonunion wage gap varies with worker characteristics such that the estimated union wage gap is declining in skill.

Ordinary least squares estimates of nonunion and union earnings functions like equations 3.8 and 3.9 provide evidence consistent with the standardization-of-rates view of union wage setting. Table 1 contains estimates of nonunion and union log-earnings functions based on data from the outgoing rotation groups of the 1999 U.S. Current Population Survey. These regressions control for age, education, sex, marital status, the interaction of sex and marital status, and sector of employment (private vs. public). The general character of these estimates is consistent with the large number of similar estimates that have appeared in the literature over the last thirty years.

TABLE 1: OLS Estimates of Nonunion and Union Earnings Functions.
1999 U.S. Current Population Survey, Merged Outgoing Rotation Groups.
(standard errors)

Variable	Nonunion	Union	U - NU
Constant	1.3829	1.7927	0.4099
	(0.0082)	(0.0157)	(0.0197)
Age	0.0094	0.0088	-0.0006
	(0.0001)	(0.0003)	(0.0004)
ED<12	-0.2535	-0.2194	0.0341
	(0.0055)	(0.0131)	(0.0159)
ED 13-15	0.1662	0.1148	-0.0514
	(0.0041)	(0.0080)	(0.0100)
ED≥16	0.5834	0.3616	-0.2218
	(0.0042)	(0.0085)	(0.0106)
Female	-0.1250	-0.1637	-0.0388
	(0.0049)	(0.0108)	(0.0132)
Married	0.2114	0.0958	-0.1157
	(0.0048)	(0.0091)	(0.0114)
Married*	-0.1650	-0.0840	0.0809
Female	(0.0065)	(0.0133)	(0.0164)
Nonwhite	-0.0817	-0.0609	0.0208
	(0.0045)	(0.0083)	(0.0104)
Public	-0.0677	-0.0177	0.0500
	(0.0051)	(0.0071)	(0.0095)
N	88556	17223	
R-Squared	0.350	0.217	
Root MSE	0.475	0.409	

Note: The base group consists of single white males with 12 years of education who are employed in the private sector. Estimates are weighted by the CPS final sampling weights. The third column is the difference between the second and the first columns.

With respect to education, the most important skill measure, the college-high school wage differential is 22 log points (25 percentage points) larger in the nonunion sector than in the union sector. Thus, skill is rewarded more in the nonunion sector, and the union wage gap is considerably larger for less-skilled workers. This is a consistent finding in the literature on the union wage gap.

Additionally, other personal characteristics are of lesser importance in the union sector. Nonwhites in nonunion jobs earn about 8.2 percent less than whites while nonwhites in union jobs earn about 6.1 percent less than whites. This confirms the earlier work of Ashenfelter (1972, 1978) that the union wage gap is larger for nonwhites than for whites. Married males in the nonunion sector earn 21 log points (23.5 percent) more than single males while the marital status differential is only 9.5 log points (9.9 percent) in the union sector.

One apparently anomalous result is that the sex differential in wages is *larger* in the union sector than in the nonunion sector. This is due to the fact that females are substantially segregated in particular occupations, both within and outside the union sector.

The union wage gap varies with other measures of job and worker characteristics that are not included in the estimates in table 1. The union wage gap differs substantially across occupations, with relatively large wage gaps in the less-skilled occupations. This is a manifestation of the organization of unions in the United States along industrial lines and the standardization of wage rates in the union sector. It also appears that the union wage gap varies inversely with firm size. Workers in small firms receive a large wage advantage from unionization while workers in larger firms receive a much smaller advantage. Given that unions are more common in large firms than in small firms, this might reflect the well-known firm-size wage differential where large firms pay higher wages than small firms (Brown and Medoff, 1989).

Overall, existing estimates suggest a substantial positive union wage gap for most workers. The raw (not regression adjusted) union wage gap in the data used in table 1 is 28.5 log points (33 percentage points). The wage gap for a single 40 year old white male high school graduate predicted by the estimates in table 1 is 38.6 log points (47 percentage points). This declines sharply as education increases, falling to 16.4 log points (17.8 percentage points) for an otherwise-similar college graduate, but it remains substantial and positive. The wage gaps are smaller for otherwise-similar married males, falling to 27 log points (31 percentage points) for married high-school graduates and to 4.8 log points (4.9 percentage points) for married college graduates.

Other evidence consistent with the standardization-of-rates view is that the regression mean squared error is larger in the nonunion sector than in the union sector. This suggests that unobservable characteristics of workers are more important in wage determination in the nonunion sector than in the union sector. It is also the case that the R^2 of the estimated earnings function is considerably larger in the nonunion sector than in the union sector which is also consistent with the view that observable individual characteristics are a relatively more important determinant of wages in the nonunion sector than in the union sector.

The standardization of wages in the union sector and the concomitant lower variation in wages in the union sector implies that unions reduce the overall dispersion of wages. A force for increased dispersion as a result of unions is the potential decline in the nonunion wage as the union wage increases and the supply of labor to nonunion jobs increases. Freeman (1980a) analyzed the effect of unions on the dispersion of wages, and he concludes that, on balance, unions reduce the aggregate level of wage dispersion. Card (2001) finds that 15 to 20 percent of the increase in male wage inequality in the United States between 1973 and 1993 can be accounted for by the decline in union membership.

The substantial variation in the union wage gap, particularly by worker skill level, has important implications for worker preferences for union representation and for employer hiring decisions. A reasonable inference, made explicit in the next section, is that the union status of workers is related to skill, both observable and unobservable. This implies that union status is endogenous to the determination of wages so that OLS estimates of the union wage gap are likely to suffer from selection bias.

3.3 The Non-wage Effects of Unions

In addition to the work on the effect of unions on wages, there has also been substantial research on the effect of unions on non-wage aspects of employment. This literature is surveyed by Freeman and Medoff (1984) and by Hirsch and Addison (1986). Some of the findings are

- union members receive a larger fraction of their compensation in the form of fringe benefits such as health insurance and pensions (Freeman, 1981).

- workers in union jobs are less likely to quit, reflecting the higher level of wages and benefits. Freeman (1980b) suggests that this is due to union provision of a mechanism for "voice" as an alternative to "exit" in addressing workplace concerns.
- workers in union jobs are more likely to be laid off, reflecting the fact that union employers cannot adjust wages in the short run in response to demand shocks (Medoff, 1979).

4 The Determination of Union Status of Workers

There are two components to the determination of union status of workers. First, workplaces must become organized by labor unions. In the United States, the presumption is that workplaces are not unionized, and it generally takes a substantial effort on the part of workers and unions to organize the workplace. However, once the workplace is organized and a contract is negotiated, the terms and conditions of employment of all workers, present and future, are determined through negotiation between the union and the employer. This organization process effectively sets the stock of union jobs in the economy, and it is largely determined by past organization activities.

The second component of the determination of a worker's union status is the individual decision regarding whether to seek a job in a union or a nonunion workplace. Taking a job in a unionized workplace is essentially costless (other than the payment of a modest initiation fee and the regular payment of dues) and does not involve any investment in organization.

4.1 The Worker-Choice Model

Early empirical analyses of the union status of individual workers (e.g., Lee, 1978) relied on a model (the worker-choice model) where the stock of union jobs was not taken as a binding constraint and where individual workers compared their utilities in union and nonunion employment and selected the sector that yielded the highest utility. In order to focus the discussion, consider a worker-choice model where unions affect only the wage received by workers but where the effect on the wage can differ across workers. Let each worker i have a nonunion wage of W_{ni} and a union wage of W_{ui} so that the wage advantage to workers i

of being in the union is $V_i = W_{ui} - W_{ni}$. Given a cost of union membership of C_i, worker i is unionized if $V_i > C_i$ and is nonunion otherwise.

This simple model has direct implications for the composition of the union sector. Abstracting from differences across workers other than in skill, suppose that the earnings of worker i in a nonunion job are defined by

$$lnW_{ni} = \beta_0 + \beta_1 S_i + \beta_2 \mu_i + \epsilon_{ni}, \tag{4.1}$$

and that the earnings of worker i in a union job are defined by

$$lnW_{ui} = \gamma_0 + \gamma_1 S_i + \gamma_2 \mu_i + \epsilon_{ui} \tag{4.2}$$

where S_i is an observable measure of skill, μ_i is an unobservable measure of skill with mean zero, and the ϵ's are random components. The β's and γ's are parameters. The worker-choice model of union status determination implies, using equations 4.1 and 4.2, that workers are unionized if and only if

$$V_i = (\gamma_0 - \beta_0) + (\gamma_1 - \beta_1)S_i + (\gamma_2 - \beta_2)\mu_i + (\epsilon_{ui} - \epsilon_{ni}) > C_i. \tag{4.3}$$

To the extent that unions raise the wages of at least low-skilled workers, it is the case that $\gamma_0 > \beta_0$. The characterization of unions as standardizing pay implies that the marginal effect of both observed and unobserved skill on wages is likely to be greater in the nonunion sector than in the union sector ($\beta_1 > \gamma_1$ and $\beta_2 > \gamma_2$). The clear implication of the worker-choice model is that the union sector will be composed disproportionately of less-skilled workers. Consistent with this, Farber and Saks (1980) find that the likelihood that a nonunion worker will desire union representation is inversely related to the worker's position in the earnings distribution of the potential bargaining unit being organized.

The worker-choice model of union status determination implies a standard binomial discrete choice specification, such as a linear probability, probit, or logit model. While Lee (1978) made the structural formulation explicit, most others who have estimated binomial discrete choice models of union status determination are not clear about the foundations and estimate a reduced from specification. Hirsch and Addison (1986) attempt to interpret a variety of estimates of models such as this in the context of the worker-choice model.

4.2 The Queuing Model

A more realistic model of the determination of the union status of workers takes account of the fact that not all workers who desire union employment can find union jobs. This is because, as noted above, workplaces are costly to organize, but, once organized, remain so into the indefinite future. The result is that there will be workers who would desire employment in an already-unionized workplace but who would not be interested in undertaking organization themselves.

Denote the cost of organizing an unorganized workplace by C_o and the cost of taking a job in an existing unionized establishment by C_e so that $C_o > C_e$. Then a worker will desire a job in a unionized establishment when $V_i > C_e$, but a worker will be willing to invest in organization activity only when $V_i > C_o$. Since the organization decision is what determines the stock of union jobs, there will be workers who would like an existing union job but who are unwilling to invest in organization themselves. These are the workers for whom $C_e < V_i < C_o$. The result is that there will be more workers who desire an existing union job than there are union jobs available. Thus, there is likely to be excess demand by workers for existing union jobs, resulting in an implicit queue. Aside from some unions of skilled trades workers, union employers in the United States are free to hire workers of their choosing to fill vacancies. Thus, the union status of a given worker is the result of decisions by both the worker and union employers (Abowd and Farber, 1982; Farber, 1983).

The queuing model also has important implications for estimating models of the union status of workers. A worker will be in a union job only if the worker prefers union employment to nonunion employment *and* a union employer hires the worker. As in the worker-choice model, the most skilled workers are less likely to want a union job. But the standardization of wages reduces the marginal cost of skill to employers so that union employers will want to hire the most skilled workers from among those workers who desire a union job. Thus, union workers will be those in the middle of the skill distribution. The most skilled will not want union jobs and the least skilled will not be hired by union employers. This selection process could account for at least part of Freeman's (1980a) finding that there is less dispersion of earnings in the union sector.

Empirical implementation of the queuing model requires a bivariate discrete choice model in order to capture both the employer and worker decisions. Identification is a potential problem because it is not known *a priori* whether nonunion workers are nonunion because they did not desire a union job or because they were not hired by a union employer. Farber (1983), Farber (1990), and Farber and Krueger (1993) use information available in a few U.S. surveys on preferences of nonunion workers for union representation to provide this information.

4.3 Implications for Estimation of the Union Wage Gap

Regardless of which model of union status determination is appropriate, applying standard OLS techniques to the estimation of the union wage gap using either equation 3.7 or equations 3.8 and 3.9 will yield biased estimates of the key parameters. Both models imply that workers' union status is determined at least partly on the basis of unobservable skills (μ_i), so that there will be classic "selection bias".

In the context of the worker-choice model, less-skilled workers are more likely to be union members so that workers in union jobs are more likely to be those with low values of μ_i relative to workers in nonunion jobs. The result is that OLS will yield a downward-biased estimate of the effect of unions on the wage of a particular worker.

The effect of the double selection implied by the queuing model on OLS estimates of the union wage gap is not clear. The two selection processes appear to have offsetting effects on the estimated wage gap, with the worker selection implying a negative bias and the employer selection implying a positive bias.

Two general approaches have been used to address the problem of selection bias in estimating the union wage gap. The first approach is to model union status determination simultaneously with earnings and to estimate an econometric model that takes account of the simultaneity However, this requires strong assumptions regarding functional form of the distribution of the errors in the earnings and union status determination functions. Virtually all existing work proceeds using a two-step estimator where the earnings functions and the union status determination function are assumed to have errors that are jointly normal. Lewis (1986) surveyed most of the studies that used this general approach, and he concluded that the specific results are unreliable and depend heavily on arbitrary assumptions regarding functional form and untestable exclusion restrictions (variables assumed to affect union status but having no direct effect on earnings).

The second approach relies on repeated observations on wages for individuals who change union status (Freeman, 1984). Thus, both union and nonunion wages are observed for these workers, and the difference in wages for these individuals can be thought of as their union wage gap. However, this approach relies on the strong assumption that the probability of switching sectors is not related to unobservable skill.

Another problem with the use of longitudinal data in the context of difference or fixed effects models is that there is error in the classification of workers as union or nonunion. Card (1996), based on data from the January 1977 Current Population Survey (Mellow and Sider, 1983) estimates that the misclassification rate of union status in the CPS is approximately 2.7 percent. Even this relatively low rate of misclassification implies that number of changes in union status will be substantially overstated. It is well known that measurement error is a more serious problem in estimating fixed effect or difference models than in estimating standard levels models. The result is that estimates derived using longitudinal data can be subject to substantial attenuation bias (bias toward zero). Even relatively low rates of misclassification of the key union status change variable implies that estimates of the union wage gap derived using fixed effect or difference models are likely to be substantially biased downward. Lewis (1986) presents a full discussion and analysis of this problem. He finds that the estimates derived using these techniques are substantially smaller than those derived using cross-section data, and he attributes this to measurement error bias.

References

[1] Abowd, J 1989 The Effect of Wage Bargains on the Stock Market Value of the Firm.' *American Economic Review* **90**: 774-809.

[2] Abowd J M, Farber H S 1982 Job Queues and the Union Status of Workers. *Ind. Lab. Rel. Rev.* **36**: 354–367

[3] Ashenfelter O 1972 Racial Discrimination and Trade Unions. *J. Pol. Econ.* **80**: 435–64

[4] Ashenfelter O 1978 Union Relative Wage Effects: New Evidence and Survey of their Implications for Wage Inflation. In: Stone R and Peterson W (eds.) *Econometric Contributions to Public Policy.* St. Martin's Press, New York

[5] Ashenfelter O, Bloom D E 1984 Models of Arbitrator Behavior. *Amer. Econ. Rev.* **74**: 111–124

[6] Ashenfelter 0, Currie J, Farber H S, Spiegel M 1992 An Experimental Comparison of Dispute Rates in Alternative Arbitration Systems. *Econometrica* **60**: 1407–1433

[7] Ashenfelter O, Johnson G 1969 Trade Unions, Bargaining Theory, and Industrial Strike Activity. *Amer. Econ. Rev.* **59**: 35–94

[8] Bazerman M H, Farber H S 1986 The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration. *Econometrica* **54**: 1503–1528

[9] Bloch F E and M S Kuskin 1978 Wage Determination in the Union and Nonunion Sectors. *Ind. Lab. Rel. Rev.* **31**: 183–192

[10] Booth A L 1995 *The Economics of the Trade Union.* Cambridge University Press, Cambridge UK

[11] Brown J N and Ashenfelter O 1986 Testing the Efficiency of Employment Contracts. *J. Pol. Econ.* **94**: S40—S87.

[12] Brown C and Medoff J L 1989 The Employer Size-Wage Effect. *J. Pol. Econ.* **97**: 1027-1059

[13] Card D 1996 The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica* **64**: 957–999.

[14] Card D 2001 The Effects of Unions on Wage Inequality in the United States. *Ind. Lab. Rel. Rev.* **54**: 329–348

[15] Dertouzos J N and Pencavel J H 1981 Wage and Employment Determination Under Trade Unionism: The International Typographical Union. *J. Pol. Econ.* **89**: 1162–1181

[16] Farber H S 1978 Bargaining Theory, Wage Outcomes, and the Occurrence of Strikes: An Econometric Analysis. *Amer. Econ. Rev.* **68**: 262–271 (a)

[17] Farber H S 1978 Individual Preferences and Union Wage Determination: The Case of the United Mine Workers. *J. Pol. Econ.* **86**: 923–942 (b)

[18] Farber H S 1980 An Analysis of Final-Offer Arbitration. *J. Conflict Resolution* **24**: 683–705

[19] Farber H S 1981 Splitting-the-Difference in Interest Arbitration. *Ind. Lab. Rel. Rev.* **34**: 70-77

[20] Farber H S 1983 The Determination of the Union Status of Workers. *Econometrica* **51**: 1417–1438

[21] Farber H S 1990 The Decline of Unionization in the United States: What Can be Learned from Recent Experience? *J. Lab. Econ.* **8**:S75-S105

[22] Farber H S, Katz H C 1979 Interest Arbitration, Outcomes, and the Incentive to Bargain. *Ind. Lab. Rel. Rev.* **32**: 55–63

[23] Farber H S and Krueger A B 1993 Union Membership in the United States: The Decline Continues. In: Kaufman B and Kleiner M (eds.) *Employee Representation: Alternatives and Future Directions* Industrial Relations Research Association, Madison, WI USA

[24] Farber H S and Saks D H 1980 Why Workers Want Unions: The Role of Relative Wages and Job Characteristics. *J. Pol. Econ.* **88**:349-369

[25] Freeman R B 1980 Unionism and the Dispersion of Wages. *Ind. Lab. Rel. Rev.* **34**: 3–23 (a)

[26] Freeman R B 1980 The Exit Voice Tradeoff in the Labor Market: Unionism, Job Tenure, Quits, and Separations. *Q. J. Econ.* **94**: 643–74 (b)

[27] Freeman R B 1981 The Effect of Unionism on Fringe Benefits. *Ind. Lab. Rel. Rev.* **34**: 489–509

[28] Freeman R B 1984 Longitudinal Analyses of the Effects of Trade Unions. *J. Lab. Econ.* **2**: 1-26

[29] Freeman and Medoff 1981 The Impact of the Percent Organized on Union and Nonunion Wages. *Rev. Econ. Stat.* **63**: 561–572

[30] Freeman and Medoff 1984 *What Do Unions Do?*. Basic Books, New York

[31] Fudenberg D, Tirole J 1981 Sequential Bargaining with Incomplete Information. *Rev. Econ. Stud.* **50**: 221-48

[32] Hall R E and Lilien D 1979 Efficient Wage Bargains Under Uncertain Supply and Demand. *Amer. Econ. Rev.* **69**: 868–879.

[33] Hirsch B T and Addison J T 1986 *The Economic Analysis of Unions: New Approaches and Evidence.* Allen & Unwin, Boston, MA USA

[34] Lee, L 1978 Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables. *Intl. Econ. Rev.* **19**: 415–434

[35] Lewis H G 1963 *Unionism and Relative Wages in the United States: An Empirical Inquiry.* University of Chicago Press, Chicago, IL USA

[36] Lewis H G 1986 *Union Relative Wage Effects: A Survey.* University of Chicago Press, Chicago, IL USA

[37] MaCurdy T Pencavel J 1986 Testing Between Competing Models of Wage and Employment Determination in Unionized Labor Markets. *J. Pol. Econ.* **94**: S3–S39.

[38] Medoff, J L 1979 Layoffs and Alternatives Under Trade Unions in U.S. Manufacturing. *Amer. Econ. Rev.* **69**: 380–395

[39] Mellow W and Sider H 1983. Accuracy of Response in Labor Market Surveys: Evidence and Implications. *J. Lab. Econ.* **1**: 331–344.

[40] Mnookin R, Kornhauser L 1979 Bargaining in the Shadow of the Law – The Case of Divorce. Yale Law J. **88**: 950-997

[41] Neale M A, Bazerman M H 1991 *Cognition and Rationality in Negotiation.* The Free Press, New York USA

[42] Rosen S 1969 Trade Union Power Threat Effects and the Extent of Organization. *Rev. Econ. Stud.* **36**: 185–196

[43] Sobel J Takahashi I 1983 A Multi-stage Model of Bargaining. *Rev. Econ. Stud.* **50**: 411-426

[44] Stevens C M 1966 Is Compulsory Arbitration Compatible with Bargaining? *Ind. Rel.* **5**: 38–52

[45] Tracy J S 1986 An Investigation into the Determinants of U.S. Strike Activity. *Amer. Econ. Rev.* **76**: 423–436.

[46] Walton R E, McKersie R B *A Behavioral Theory of Labor Negotiations; An Analysis of a Social Interaction System.* McGraw-Hill, New York USA