Chapter 13
Unions and the Labor Market

Summary

Chapter 12 introduced an objective framework for analyzing the earnings differences that accompany gender, race, and ethnicity. This framework was then used as the basis for a systematic inquiry into labor market discrimination, a frequently misunderstood topic where passions tend to run high. The topic of labor unions also stirs up strong feelings in many people. The purpose of this chapter is to provide a framework for objective and critical thinking about the role labor unions play in the labor market. This framework draws heavily on the discussion of the elasticity of labor demand presented in Chapter 4.

Labor unions are organized as either industrial or craft unions. An industrial union, like the United Steel Workers, represents most or all of the workers in a particular firm or industry regardless of the job they perform. A craft union, like the carpenters union, represents only workers performing a certain type of job. Bargaining on behalf of the collective membership can take place at the national or local level and often involves more than one employer. Subjects raised in bargaining sessions include wages and employee benefits, job security, safety, and workplace rules. Unions also play a role in enforcing the eventual agreement and in helping workers communicate with management.

The importance of unions varies a great deal internationally. The United States and Japan, for example, have relatively small percentages of workers that are covered by collective bargaining agreements. Some countries, such as France, have very low union membership but almost all workers are covered by a collective bargaining agreement. In Sweden, nearly all workers are union members and are covered by a collective bargaining agreement, but some unions are far stronger and have better bargaining power than others.

In the United States, unions faced a hostile legal environment throughout most of the 19th and early 20th centuries, and this helped to hinder their growth and development. Beginning in 1932, with the Norris-LaGuardia Act, the legal climate turned in organized labor’s favor by banning the use of labor injunctions. In 1935 the Wagner Act established the National Labor Relations Board to conduct union elections and to stop employers from interfering with unions through unfair labor practices. Later laws focused on keeping unions from becoming too powerful. Under the 1947 Taft-Hartley Act, states were allowed to pass right-to-work laws. These laws prohibit agreements requiring employees to become union members. The 1959 Landrum-Griffin Act was directed towards increasing union democracy. Since 1962, federal government workers have also been allowed to organize, although laws for state workers vary.

Union membership in the U.S. peaked at about one-third of the labor force just after World War II. Since then it has steadily declined, standing at 13.7% in 2002. Unionized workers usually belong to a local union, which bargains at the firm level, but the local unions are generally part of a national or international union. In cases where wage agreements are industry-wide, the national union will bargain. Most of the national unions are affiliated with the AFL-CIO, a national association of unions that acts to coordinate initiatives between member unions, act as a unified political voice, and provide information and research. More men than women are unionized, and the highest rates of unionization are in manufacturing, construction, transportation, and public utility industries.
What do unions want, and how do unions and employers behave in their negotiations over wages and benefits? Unions clearly seek higher compensation for their members, a package that includes both wages and other benefits. But they face constraints. As unions raise the price of union workers, firms will be induced to substitute capital for labor, and thus the demand for labor decreases (as does the ability of unions to achieve higher wages). Increasing wages also reduces the quantity of workers demanded, all other things equal. In industries that are stable or declining and where the elasticity of demand for labor is high, unions will be weak because any increase in the wage will cause a sharp decrease in employment. In industries where demand for labor is inelastic and in industries that are growing, unions may be able to obtain higher wages without employment effects.

The traditional monopoly view of unions has been that unions bargain for higher wages and then the employer is free to determine the employment level that will maximize the firm’s profits. This occurs at the point on the firm’s demand curve consistent with the negotiated wage rate. The problem with this scenario is that if the union values both the wage and employment of its members, such an outcome is not economically efficient; in other words, there is a whole range of wage and employment combinations that still exist where at least one party can be made better off without hurting the other. Using indifference curves to represent union preferences and isoprofit curves to represent the firm’s constraints, the efficient contracts model shows the set of combinations preferable to the monopoly outcome. Within that set, those combinations where it is impossible to make one party better off without making the other worse off form the contract curve or locus of efficient contracts. These wage and employment combinations are shown to lie to the right of the firm’s demand curve for labor.

If the level of union membership is thought of as the result of the demand for, and supply of, union activity, what accounts for the decline in the percentage of the United States labor force that is unionized? Three factors have combined to shift the demand for union services to the left. The first is the substantial increase in the percentage of the labor force that is female. Seniority provisions, job security arrangements, and retirement plans are of little interest to any worker who expects to be with the firm for only a short period of time. However, these are some of the main issues for which unions typically fight. Historically, women had shorter attachments to the labor force than men, and were consequently less likely to demand union representation. More recently, however, women’s attachment to the labor force has come to more closely parallel that of men, and it is not likely that increased labor force participation among women has been much of a factor in declining union membership for the past two decades.

The second reason is the change in the composition of private sector employment. Workers employed in the growing areas of wholesale and resale trade, finance, insurance, and other services have traditionally not joined unions because in these competitive areas labor demand tends to be very elastic. When labor demand is elastic, the ability of unions to bargain for higher wages is constrained by the large employment losses such wage increases bring. Historically, female employment has been much stronger in the service than in the production industries. This probably served to reduce the independent impact of growing female participation, since women were not entering (in very large numbers) markets where unions were strong.

The third reason has been overall increase in product market competition due to deregulation and foreign competition. As product markets become more competitive, labor demand becomes more elastic and the expected benefits from union membership decline.

Two factors have also combined to shift the supply of union services to the left. Regional shifts in employment opportunities to the South, due in part to increased product market competition, have made union organizing more difficult since most of these states have passed right-to-work laws. Faced with more competition and a rising union to non-union wage ratio, many employers have also increased their resistance to union organizing campaigns. Some forms of employer resistance are perfectly legal, but some have taken the form of unfair labor practices.
As noted previously, given the limits that large employment losses place on a union’s ability to raise wages, unions are more likely to be successful when labor demand is inelastic. A growing demand for labor may also help to mask the reduction in employment opportunities brought about by higher wages. But what factors lead to an inelastic demand for labor? According to the Hicks-Marshall laws discussed in Chapter 4, labor demand elasticity decreases as the difficulty of substituting other inputs for labor in the production process increases, as the supply of other inputs becomes less elastic, as the demand for the product becomes less elastic, and as the share of total costs that are labor costs decreases. When these factors are not favorable to union success, unions can be expected to take steps to change them. For example, union support for a higher minimum wage, restrictions on immigration, staffing and subcontracting restrictions, strict specification of a worker’s job function, and limits on the supply of skilled workers to a craft can be traced to the union’s desire to limit the substitution possibilities for union labor. Similarly, support for import quotas or tariffs, domestic content legislation, and programs to “buy American” can be traced to the desire to stimulate the demand for union labor.

Given the elasticity and position of the labor demand curve, unions are able to secure higher wages for their members mainly because of their ability to impose costs on management through work slowdowns or strikes. While no single model of strike activity is completely satisfactory, a number of models exists to highlight different aspects of the process. According to Hicks’ model of the bargaining process, the optimal strike length occurs at the intersection of the firm’s upward-sloping employer concession curve and the downward-sloping union resistance curve. The employer concession curve slopes upward because longer strikes impose additional costs on the firm, creating an incentive to raise the wage offer. The union resistance curve slopes downward because strikes also impose costs on union members through lost income. As the strike lengthens, the income loss increases and union members reduce their wage demands. One problem with this framework is that if each side is aware of the other’s schedule, there is no need for a strike since agreeing to the predicted wage increase before the strike would make each side better off. Hence, the credibility of this model depends on the existence of imperfect information or the ability of a strike to influence future negotiations.

For public sector unions, the collective bargaining situation may be substantially different since public sector unions are not allowed to strike in most states, and instead must deal with the possibility of some form of third-party arbitration. Chapter 13 discusses the incentives faced by each party in a collective bargaining situation when there is the possibility of having an arbitrator impose a settlement, and a formal model is presented in Appendix 13A. The model shows that the wider the range of uncertainty about the arbitrator’s decisions and the greater the risk aversion of the parties, the greater is the range of acceptable settlements to both parties. Assuming the probability of reaching a negotiated settlement is positively related to the size of the acceptable range, increased uncertainty and risk aversion increase the likelihood that arbitration will not be necessary.

What effects do unions have on the labor market? The absolute effect that unions have on wages is impossible to determine since union wages can only be measured relative to non-union wages, and non-union wages are affected by union activities. Unions may have a positive or negative effect on non-union wages, so it is not always clear whether the relative wage advantages unions secure for their members overstate or understate the absolute wage effects of unions.

**Example**

Consider a labor market with union and non-union sectors. The demand and supply curves for union (U) labor are given by the equations

\[
\text{Demand: } D_U = 20 - 0.5W, \\
\text{Supply: } S_U = 1.5W.
\]
These curves are shown as lines $D$ and $S_0$ in Figure 13-1a. The demand and supply curves in the non-union ($N$) sector are given by

- **Demand**: $L_N^D = 40 - 2W$,
- **Supply**: $L_N^S = 2W$.

These curves are shown as lines $D$ and $S_0$ in Figure 13-1b. Note that the initial market-clearing wage in each sector is $10$. The optimal employment level is 15 in the union sector and 20 in the non-union sector.

Now suppose that the union succeeds in raising the wage of its members to $12$. This represents an **absolute wage effect** of 20%. What effect will this wage change have on the relative wages between the two sectors?

Note that the increase in the union wage reduces the quantity of union labor demanded to

$$L_U^D = 20 - 0.5(12) = 14,$$

while the quantity of labor supplied increases to

$$L_U^S = 1.5(12) = 18.$$

Together this creates a labor surplus in the union sector of 4 units.

![Figure 13-1a](image1)

**Figure 13-1a**

![Figure 13-1b](image2)

**Figure 13-1b**
How will these workers respond to the unemployment in the union sector? One possibility is that the surplus workers may seek employment in the non-union sector. Suppose the workers spill over from the union to the non-union sector in such a way that the supply curve in the union sector becomes

\[ L^s_U = 1.5W - 4, \]

while the supply curve in the non-union sector becomes

\[ L^s_N = 2W + 4. \]

These curves are denoted by the lines labeled \( S_1 \) in Figures 13-1a and 13-1b. The supply shift in the union sector is just enough to eliminate the unemployment in that sector. The increase in supply in the non-union sector creates a surplus in that sector at the original wage of $10. However, since the wage is free to adjust in the non-union sector, the market will clear where

\[
\begin{align*}
L^D_N = L^s_N & \Rightarrow 40 - 2W = 2W + 4 \\
& \Rightarrow W^*_N = $9 \quad \text{and} \quad L^*_N = 22.
\end{align*}
\]

Note that after the spillover adjustment is complete, the union wage ($12) exceeds the non-union wage ($9) by $3 for a relative wage advantage of 33%. Therefore, if labor market spillovers occur in response to union wages, the relative wage advantage will overstate the absolute effect of the union on its members’ wages. However, such a spillover effect may not be the most likely response to the increase in the union wage.

Another possibility is that a wage increase in the union sector may actually trigger a wage increase in the non-union sector as employers move to head off union attempts to organize non-union workers and secure an even larger wage increase. The ability of a union wage increase to bring about a wage increase in the non-union sector is called a threat effect. In this example, the threat effect may result in the non-union employer raising the wage somewhere between the original wage of $10 and the union wage of $12. If, for example, the wage was raised to $11, note that this would create unemployment in the non-union sector. Given the supply curve \( S_1 \), the quantity of labor supplied would increase to

\[ L^s_N = 2(11) + 4 = 26, \]

while the quantity of labor demanded would decrease to

\[ L^D_N = 40 - 2(11) = 18, \]

creating a surplus of 8 units in the non-union sector. Note however, that the threat effect would lower the relative union wage advantage to one-eleventh or 9.1%. In this case the relative wage advantage understates the absolute effect of the union on member wages.

Another problem with the standard spillover scenario is that expected earnings are not equal at the new market-clearing wage and employment levels. Letting \( F \) stand for the fraction of each period that a person expects to be employed, expected earnings \( (Y) \) are given by the equation

\[ Y = (W) (F). \]

Since there is no unemployment after the spillover is complete, \( F \) equals 1 in both sectors. However, since the union wage exceeds the non-union wage, expected earnings are greater in the union sector. This suggests that some workers may find it attractive to remain in the union sector even though they will be unemployed for at least part of the time. Such decisions lead to what is called wait unemployment.
For example, suppose one of the workers who spilled over to the non-union sector now returns to the union sector. Also suppose we adopt the simple rule that the fraction of time any individual will be employed is related to the unemployment rate ($UR$) in that sector according to the equation

$$F = 1 - UR.$$ 

Letting $E$ be the number employed and $U$ the number unemployed, the unemployment rate is given by the formula

$$UR = \frac{U}{U + E}.$$ 

If 3 workers return to the union sector, expected earnings in that sector would equal

$$Y_U = W_U F_U \Rightarrow Y_U = 12 \left(1 - \frac{3}{3+14}\right) \Rightarrow Y_U = (12)(0.823) = 9.88.$$ 

The flow of workers out of the non-union sector would reduce the supply of non-union workers to 

$$L^{N}_s = 2W + 1,$$

which means that the non-union sector clears where

$$40 - 2W = 2W + 1 \Rightarrow W^*_N = 9.75 \text{ and } L^*_N = 20.5.$$ 

Since there is no unemployment in the non-union sector, $Y_N$ also equals $9.75$. Note that the migration of 3 workers back to the union sector has significantly reduced the expected earnings gap between the two sectors. To drive the expected earnings exactly into equality would require a movement of just over 3 units of labor back to the union sector. Moving all 4 units back to the union sector would result in a situation where

$$W_U F_U = (12)(0.778) = 9.33 < 10 = W_N F_N,$$

which is not an equilibrium allocation of labor.

Given that the amount of wait unemployment is just over 3 units of labor in this case, the wage in the non-union sector will remain slightly below the original level of $10$, meaning that again the relative wage advantage will overstate the absolute effect of the union on wages. Note that it is possible, however, for the final equilibrium to occur where the wage in the non-union sector is above its original level, meaning that the relative wage advantage will understate the absolute effect of the union. The comparison of relative and absolute union wage effects can be further complicated by outward shifts in the demand for union labor that may accompany union bargaining efforts. If such shifts come at the expense of non-union labor, they lead the relative wage to again understate the absolute wage effect of the union.

Empirical evidence places the union relative wage advantage in the range of 10% to 20%, with some evidence pointing to a decrease in non-union wages. This suggests that the spillover effects (along with the demand shift effects just mentioned) tend to outweigh the threat and wait unemployment effects illustrated in this example.
In interpreting the effects of unions on wages, it is important to keep in mind that wages are only part of the total compensation package received by workers. Evidence also suggests that union workers receive higher levels of employee benefits and that employee benefits account for a greater percentage of total compensation. However, union workers also tend to work under conditions that are more structured, inflexible, and involve a faster pace and greater risk of injury. Thus, some of the difference in compensation between union and non-union workers may reflect a compensating differential for the less desirable working conditions. This issue serves as a vivid reminder that the same sets of problems that hamper the measuring of discriminatory wage differentials also applies when measuring the returns to unionization. That is, to truly measure even the relative wage advantage due to unions one must be careful to control for differences in productive characteristics and working conditions. Even then, there may be unmeasurable or unobservable differences that account for some part of the gap, leading to an overstatement of the union relative wage advantage.

Perhaps the most controversial issue surrounding unions is their effect on productivity and output. Clearly, some union activities reduce productivity and output. The labor market spillovers that result from higher union wages lead to a lower total output than before the wage change. Restrictive work rules and staffing requirements secured by unions to make the demand for labor more inelastic serve to move the firm away from its cost-minimizing combination of inputs. If a firm wishes to stay at the original cost level it must then reduce its total output. In addition, there may be lost output from strikes.

An alternative view, however, stresses that unions can have positive effects on productivity and output by giving workers an avenue for communicating directly with management. In a perfectly competitive labor market, workers who are unhappy with their jobs communicate with management by quitting. If enough workers quit, ultimately management is forced to make changes to attract workers. Quits, however, tend to erode worker loyalty and morale, and cause firms to lose their hiring and firm-specific training investments. High turnover also makes it impossible for firms to use productivity-enhancing compensation policies like pay sequencing and efficiency wages that depend on the existence of internal labor markets. By providing workers with a way to speak with a collective voice, unions can serve as an alternative and less costly adjustment mechanism. Some empirical studies suggest that unions increase productivity enough to offset the increased costs that come from higher union wages. However, such findings conflict with observations of continued employer resistance to unions and unfavorable stock market reaction to unionization efforts. These observations are consistent with a belief that unions increase costs and reduce profitability.

From a normative point of view, the question is: do unions enhance or reduce social welfare? If there is lost production due to strikes and excessive hiring due to restrictive work rules, then clearly the outcome is not Pareto efficient. Also, if the marginal productivity of the last worker employed in each industry is not equal, greater output at lower cost could be achieved by reallocating workers. If unions raise wages and thus reduce employment in some sectors, this marginal productivity condition cannot be achieved, because they are preventing some lower-wage workers (who have lower productivity due to excess labor in the low-wage sector) from moving into the high-wage sector (where they would be more productive, because there is less labor).

Others argue that unions enhance social welfare by giving workers a collective voice in establishing working conditions. Individuals may find it too costly to speak out or to leave their job if they are unhappy, and thus there are Pareto-efficient transactions that may not take place. Secondly, as in the efficient-contracts section, if employers are monopolists, unionism may involve redistribution of some of the monopolist’s profit (to the workers) rather than a reduction in employment. This is an equity issue but has no impact on overall allocative efficiency. Finally, better management-worker communications through the union’s collective voice might lead to productivity increases.