

Chapter 4

Labor Market Equilibrium

Order is not pressure which is imposed on society from without, but an equilibrium which is set up from within.

—*José Ortega y Gasset*

Workers prefer to work when the wage is high, and firms prefer to hire when the wage is low. Labor market equilibrium “balances out” the conflicting desires of workers and firms and determines the wage and employment observed in the labor market. By understanding how equilibrium is reached, we can address what is perhaps the most interesting question in labor economics: Why do wages and employment go up and down?

This chapter analyzes the properties of equilibrium in a perfectly competitive labor market. We will see that if markets are competitive and if firms and workers are free to enter and leave these markets, the equilibrium allocation of workers to firms is efficient; the sorting of workers to firms maximizes the total gains that workers and firms accumulate by trading with each other. This result is an example of Adam Smith’s justly famous **invisible hand theorem**, wherein labor market participants in search of their own self-ish goals attain an outcome that no one in the market consciously sought to achieve. The implication that competitive labor markets are efficient plays an important role in the framing of public policy. In fact, the impact of many government programs is often debated in terms of whether the particular policy leads to a more efficient allocation of resources or whether the efficiency costs are substantial.

We also will analyze the properties of labor market equilibrium under alternative market structures, such as monopsonies (where there is only one buyer of labor) and monopolies (where there is only one seller of the output). Each of these market structures generates an equilibrium with its own unique features. Monopsonists, for instance, generally hire fewer workers and pay less than competitive firms.

Finally, the chapter uses a number of policy applications—such as taxes, subsidies, and immigration—to illustrate how government policies shift the labor market to a different equilibrium, thereby altering the economic opportunities available to both firms and workers.

4-1 Equilibrium in a Single Competitive Labor Market

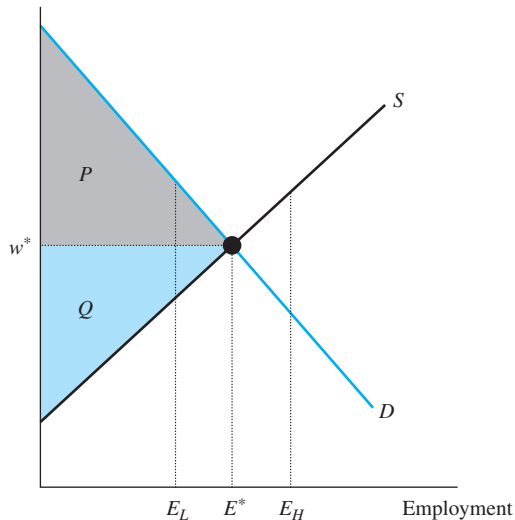
We have already briefly discussed how a competitive labor market attains equilibrium. We now provide a more detailed discussion of the properties of this equilibrium. Figure 4-1 illustrates the familiar graph showing the intersection of labor supply (S) and labor demand (D) curves in a competitive market. The supply curve gives the total number of employee-hours that agents in the economy allocate to the market at any given wage level; the demand curve gives the total number of employee-hours that firms in the market demand at that wage. Equilibrium occurs when supply equals demand, generating the competitive wage w^* and employment E^* . The wage w^* is the market-clearing wage because any other wage level would create either upward or downward pressures on the wage; there would be too many jobs chasing the few available workers or too many workers competing for the few available jobs.

Once the competitive wage level is determined in this fashion, each firm in this industry hires workers up to the point where the value of marginal product of labor equals the competitive wage. The first firm hires E_1 workers; the second firm hires E_2 workers; and so on. The total number of workers hired by all the firms in the industry must equal the market's equilibrium employment level, E^* .

FIGURE 4-1 Equilibrium in a Competitive Labor Market

The labor market is in equilibrium when supply equals demand; E^* workers are employed at a wage of w^* . In equilibrium, all persons who are looking for work at the going wage can find a job. The triangle P gives the producer surplus; the triangle Q gives the worker surplus. A competitive market maximizes the gains from trade, or the sum $P + Q$.

Dollars



Theory at Work

THE INTIFADAH AND PALESTINIAN WAGES

Throughout much of the 1980s, nearly 110,000 Palestinians who resided in the occupied West Bank and Gaza Strip commuted to Israel for their jobs. Many of these Palestinians were employed in the construction or agriculture industries.

As a result of the Intifadah—the Palestinian uprising against Israeli control of the West Bank and Gaza territories that began in 1988—there were major disruptions in the flow of these workers into Israel. Israeli authorities, for instance, stepped up spot checks of work permits and began to enforce the ban on Palestinians spending the night in Israel, while strikes and curfews in the occupied territories limited the mobility of commuting workers. Within one year, the daily absenteeism rate jumped from less than 2 percent to more than 30 percent; the average number

of work days in a month dropped from 22 to 17; and the length of time it took a commuting Palestinian to reach the work location rose from 30 minutes to three or four hours.

The Intifadah, therefore, greatly reduced the supply of Palestinian commuters in Israel. The supply and demand framework suggests that the uprising should have increased the equilibrium wages of these Palestinian workers. In fact, this is what occurred. The roughly 50 percent cut in the labor supply of Palestinian commuters increased their real wage by about 50 percent, implying that the demand elasticity for Palestinian commuters is on the order of -1 .

Source: Joshua D. Angrist, "Short-Run Demand for Palestinian Labor," *Journal of Labor Economics* 14 (July 1996): 425–453.

As Figure 4-1 shows, there is no unemployment in a competitive labor market. At the market wage w^* , the number of persons who want to work equals the number of workers firms want to hire. Persons who are not working are also not looking for work *at the going wage*. Of course, many of these persons would enter the labor market if the wage rose (and many would withdraw if the wage fell).

Needless to say, a modern industrialized economy is continually subjected to many shocks that shift both the supply and demand curves. It is unlikely, therefore, that the labor market actually ever reaches a stable equilibrium—with wages and employment remaining at a constant level for a long time. Nevertheless, the concept of labor market equilibrium remains useful because it helps us understand why wages and employment seem to go up or down in response to particular economic or political events. As the labor market reacts to a particular shock, wages and employment will tend to move toward their new equilibrium level.

Efficiency

Figure 4-1 also shows the benefits that accrue to the national economy as workers and firms trade with each other in the labor market. In a competitive market, E^* workers are employed at a wage of w^* . The total revenue accruing to the firm can be easily calculated by adding up the value of marginal product of the first worker, the second worker, and all workers up to E^* . This sum, in effect, gives the value of the total product produced by all workers in a competitive equilibrium. Because the labor demand curve gives the value of marginal product, it must be the case that the area under the labor demand curve gives the value of total product. Each worker receives a wage of w^* . Hence, the profits accruing to firms, which we call **producer surplus**, are given by the area of the triangle P .¹

¹ To simplify the discussion, assume that labor is the only factor in the production function.

Workers also gain. The supply curve gives the wage required to bribe additional workers into the labor market. In effect, the height of the supply curve at a given point measures the value of the marginal worker's time in alternative uses. The difference between what the worker receives (that is, the competitive wage w^*) and the value of the worker's time outside the labor market gives the gains accruing to workers. This quantity is called **worker surplus** and is given by the area of the triangle Q in Figure 4-1.

The total **gains from trade** accruing to the national economy are given by the sum of producer surplus and worker surplus, or the area $P + Q$. *The competitive market maximizes the total gains from trade accruing to the economy.* To see why, consider what the gains would be if firms hired more than E^* workers, say E_H . The “excess” workers have a value of marginal product that is less than their value of time elsewhere. In effect, these workers are not being efficiently used by the labor market; they are better off elsewhere. Similarly, consider what would happen if firms hired too few workers, say E_L . The “missing” workers have a value of marginal product that exceeds their value of time elsewhere, and their resources would be more efficiently used if they worked.

The allocation of persons to firms that maximizes the total gains from trade in the labor market is called an **efficient allocation**. A competitive equilibrium generates an efficient allocation of labor resources.

4-2 Competitive Equilibrium across Labor Markets

The discussion in the previous section focused on the consequences of equilibrium in a *single* competitive labor market. The economy, however, typically consists of many labor markets, even for workers who have similar skills. These labor markets might be differentiated by region (so that we can talk about the labor market in the Northeast and the labor market in California), or by industry (the labor market for production workers in the automobile industry and the labor market for production workers in the steel industry).

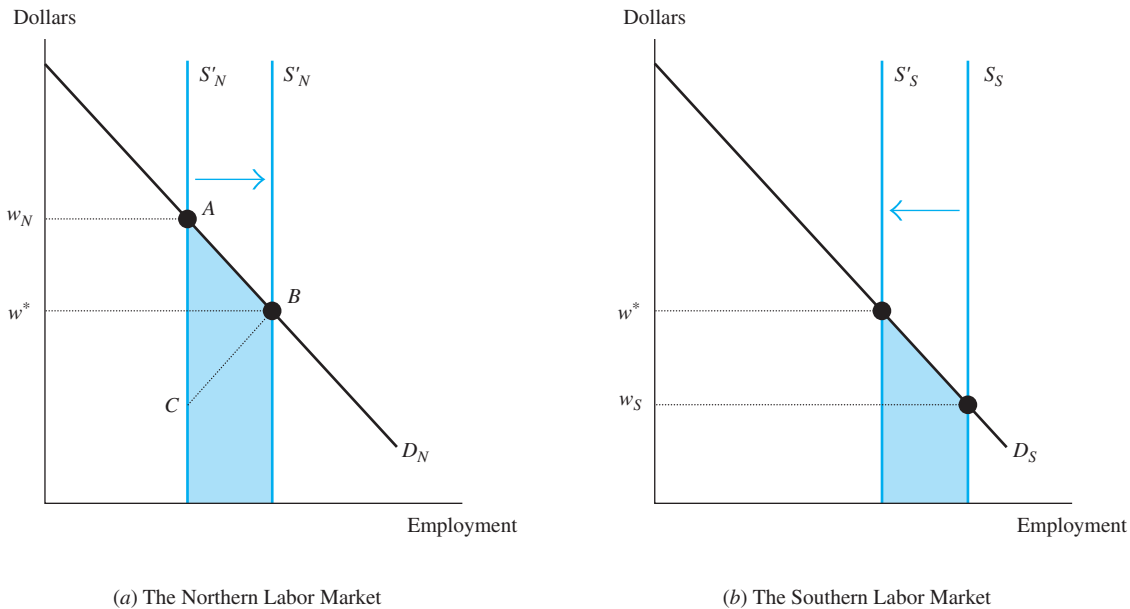
Suppose there are two regional labor markets in the economy, the North and the South. We assume that the two markets employ workers of similar skills so that persons working in the North are perfect substitutes for persons working in the South. Figure 4-2 illustrates the labor supply and labor demand curves in each of the two labor markets (S_N and D_N in the North, and S_S and D_S in the South). For simplicity, the supply curves are represented by vertical lines, implying that supply is perfectly inelastic within each region. As drawn, the equilibrium wage in the North, w_N , exceeds the equilibrium wage in the South, w_S .

Can this wage differential between the two regions persist and represent a true competitive equilibrium? No. After all, workers in the South see their northern counterparts earning more. This wage differential encourages southern workers to pack up and move north, where they can earn higher wages and presumably attain a higher level of utility. Employers in the North also see the wage differential and realize that they can do better by moving to the South. After all, workers are equally skilled in the two regions, and firms can make more money by hiring cheaper labor.

If workers can move across regions freely, the migration flow will shift the supply curves in both regions. In the South, the supply curve for labor would shift to the left (to S'_S) as southern workers leave the region, raising the southern wage. In the North, the supply curve would shift to the right (to S'_N) as the southerners arrived, depressing the northern wage. If there were free

FIGURE 4-2 Competitive Equilibrium in Two Labor Markets Linked by Migration

The wage in the northern region (w_N) exceeds the wage in the southern region (w_S). Southern workers want to move north, shifting the southern supply curve to the left and the northern supply curve to the right. In the end, wages are equated across regions (at w^*). The migration of workers reduces the gains from trade in the South by the size of the shaded trapezoid in the southern labor market, and increases the gains from trade in the North by the size of the larger shaded trapezoid in the northern labor market. Migration increases the total gains from trade in the national economy by the triangle ABC .



(a) The Northern Labor Market

(b) The Southern Labor Market

entry and exit of workers in and out of labor markets, the national economy would eventually be characterized by a single wage, w^* .

Note that wages across the two labor markets also would be equalized if firms (instead of workers) could freely enter and exit labor markets. When northern firms close their plants and move to the South, the demand curve for northern labor shifts to the left and lowers the northern wage, whereas the demand curve for southern labor shifts to the right, raising the southern wage. The incentives for firms to move across markets evaporate once the regional wage differential disappears. As long as either workers or firms are free to enter and exit labor markets, therefore, a competitive economy will be characterized by a single wage.²

Efficiency Revisited

The “single wage” property of a competitive equilibrium has important implications for economic efficiency. Recall that, in a competitive equilibrium, the wage equals the value of marginal product of labor. As firms and workers move to the region that provides the

² A study of whether a labor market can be considered competitive is given by Stephen Machin and Alan Manning, “A Test of Competitive Labor Market Theory: The Wage Structure among Care Assistants in the South of England,” *Industrial and Labor Relations Review* 57 (April 2004): 371–385.

best opportunities, they eliminate regional wage differentials. Therefore, workers of given skills have the same value of marginal product of labor in all markets.

The allocation of workers to firms that equates the value of marginal product across markets is also the sorting that leads to an efficient allocation of labor resources. To see why, suppose that a benevolent dictator takes over the economy and that this dictator has the power to dispatch workers across regions. In making allocation decisions, suppose this benevolent dictator has one overriding objective: to allocate workers to those places where they are most productive. When the dictator first takes over, he faces the initial situation illustrated in Figure 4-2, where the competitive wage in the North (w_N) exceeds the competitive wage in the South (w_S). Note that this wage gap implies that the value of marginal product of labor is greater in the North than in the South.

The dictator picks a worker in the South at random. What should he do with this worker? Because the dictator wants to place this worker where he is most productive, the worker is dispatched to the North. In fact, the dictator will keep reallocating workers to the northern region as long as the value of marginal product of labor is greater in the North than in the South. The law of diminishing returns implies that as the dictator forces more and more people to work in the North, the value of marginal product of northern workers declines and the value of marginal product of southern workers rises. The dictator will stop reallocating persons when the labor force consists only of persons whose value of marginal product exceeds the value of their time outside the labor market *and* when the value of marginal product is the same in all labor markets.

It is also easy to see how migration leads to an efficient allocation of resources by calculating the gains from trade in the labor market. Because the supply curves in Figure 4-2 are perfectly inelastic (implying that the value of time outside the labor market is zero), the total gains from trade are given by the area under the demand curve up to the equilibrium level of employment. The migration of workers out of the South reduces the total gains from trade in the South by the shaded area of the trapezoid in the southern labor market. The migration of workers into the North increases the total gains from trade in the North by the shaded area of the trapezoid in the northern labor market. A comparison of the two trapezoids reveals that the area of the northern trapezoid exceeds the area of the southern trapezoid by the size of the triangle *ABC*, implying that the total gains from trade in the national economy increase as a result of worker migration.

The surprising implication of our analysis should be clear: *Through an “invisible hand,” workers and firms that search selfishly for better opportunities accomplish a goal that no one in the economy had in mind: an efficient allocation of resources.*

Convergence of Regional Wage Levels

There is a great deal of interest in determining whether regional wage differentials in the United States (as well as in other countries) narrow over time, as implied by our analysis of labor market equilibrium. Many empirical studies suggest that there is indeed a tendency toward convergence.³

³ Robert J. Barro and Xavier Sala-i-Martin, “Convergence across States and Regions,” *Brookings Papers on Economic Activity* (1991): 107–158; and Olivier Jean Blanchard and Lawrence F. Katz, “Regional Evolutions,” *Brookings Papers on Economic Activity* 1 (1992): 1–61.

FIGURE 4-3 Wage Convergence across States

Source: Olivier Jean Blanchard and Lawrence F. Katz, "Regional Evolutions," *Brookings Papers on Economic Activity* 1 (1992): 1–61.

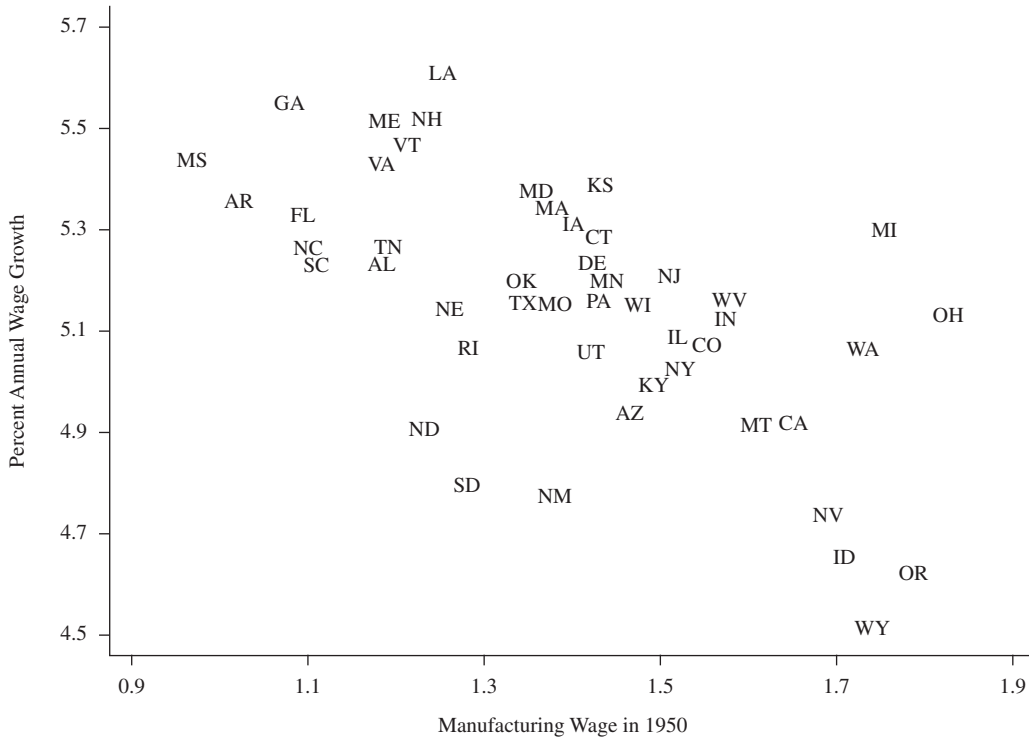


Figure 4-3 summarizes the key data underlying the study of wage convergence across states in the United States. The figure relates the annual growth rate in the state's manufacturing wage between 1950 and 1990 to the initial wage level in 1950. There is a strong negative relationship between the rate of wage growth and the initial wage level so that the states with the lowest wages in 1950 subsequently experienced the fastest wage growth. It has been estimated that about half the wage gap across states disappears in about 30 years. The evidence indicates, therefore, that wage levels do converge over time—although it may take a few decades before wages are equalized across markets.

Wage convergence also is found in countries where the workforce is less mobile, such as Japan. A study of the Japanese labor market indicates that wage differentials across prefectures (a geographic unit roughly comparable to a large U.S. county) disappear at about the same rate as interstate wage differentials in the United States: half of the regional differences vanish within a generation.⁴

⁴ Robert J. Barro and Xavier Sala-i-Martin, "Regional Growth and Migration: A Japan–United States Comparison," *Journal of the Japanese and International Economies* 6 (December 1992): 312–346. Other international studies of wage convergence within regions of a country include Christer Lundh, Lennart Schon, and Lars Svensson, "Regional Wages in Industry and Labour Market Integration in Sweden, 1861–1913," *Scandinavian Economic History Review* 53 (2005): 71–84; and Joan R. Roses and Blanca Sanchez-Alonso, "Regional Wage Convergence in Spain 1850–1930," *Explorations in Economic History* 41 (October 2004): 404–425.

Of course, the efficient allocation of workers across labor markets and the resulting wage convergence are not limited to labor markets within a country, but also might occur when we compare labor markets across countries. Many recent studies have attempted to determine if international differences in per capita income are narrowing.⁵ Much of this work is motivated by a desire to understand why the income gap between rich and poor countries seems to persist.

The empirical studies typically conclude that when one compares two countries with roughly similar endowments of human capital (for example, the educational attainment of the population), the wage gap between these countries narrows over time, with about half the gap disappearing within a generation. This result, called “conditional” convergence (because it compares countries that are already similar in terms of the human capital endowment of the workforce), does not necessarily imply that there will be convergence in income levels between rich and poor countries. The wage gap between rich and poor countries can persist for much longer periods because the very low levels of human capital in poor countries do not permit these countries to be on the same growth path as the wealthier countries.

The rate of convergence in income levels across countries plays an important role in the debate over many crucial policy issues. Consider, for instance, the long-term effects of the North American Free Trade Agreement (NAFTA). This agreement permits the unhampered transportation of goods (but not of people) across international boundaries throughout much of the North American continent (Canada, the United States, and Mexico).

In 2000, per capita GDP in the United States was over three times as large as that in Mexico. Our analysis suggests that NAFTA should eventually reduce the huge income differential between Mexico and the United States. As U.S. firms move to Mexico to take advantage of the cheaper labor, the demand curve for Mexican labor shifts out and the wage differential between the two countries will narrow. Our discussion suggests that U.S. workers who are most substitutable with Mexican workers will experience a wage cut as a result of the increase in trade. At the same time, however, American consumers will gain from the increased availability of cheaper goods. In short, NAFTA will likely create distinct groups of winners and losers in the American economy. In fact, the available evidence suggests that manufacturing firms are now finding the Mexican labor market to be relatively expensive.⁶

Although NAFTA inevitably affects the distribution of income across the three countries, our analysis of labor market efficiency implies that the *total* income of the countries in the free-trade zone is maximized when economic opportunities are equalized across countries. In other words, the equalization of wages across the three signatories of NAFTA increases the size of the economic pie available to the entire region. *In principle*, the additional wealth can be redistributed to the population of the three countries so as to make

⁵ Robert J. Barro, “Economic Growth in a Cross-Section of Countries,” *Quarterly Journal of Economics* 105 (May 1990): 501–526; N. Gregory Mankiw, David Romer, and David N. Weil, “A Contribution to the Empirics of Economic Growth,” *Quarterly Journal of Economics* 107 (May 1991): 407–437; and Xavier Sala-i-Martin, “The World Distribution of Income: Falling Poverty and . . . Convergence, Period,” *Quarterly Journal of Economics* 121 (May 2006): 351–397.

⁶ Elisabeth Malkin, “Manufacturing Jobs Are Exiting Mexico,” *New York Times*, November 5, 2002.

everyone in the region better off. This link between free trade and economic efficiency is typically the essential point emphasized by economists when they argue in favor of more open markets.⁷

4-3 Policy Application: Payroll Taxes and Subsidies

We can easily illustrate the usefulness of the supply and demand framework by considering a government policy that shifts the labor demand curve. In the United States, some government programs are funded partly through a payroll tax assessed on employers. In 2011, firms paid a tax of 6.2 percent on the first \$106,800 of a worker's annual earnings to fund the Social Security system and an additional tax of 1.45 percent on all of a worker's annual earnings to fund Medicare.⁸ In other countries, the payroll tax on employers is even higher. In Germany, for example, the payroll tax is 17.2 percent; in Italy, it is 21.2 percent; and in France, it is 25.3 percent.⁹

What happens to wages and employment when the government assesses a payroll tax on employers? Figure 4-4 answers this question. Prior to the imposition of the tax, the labor demand curve is given by D_0 and the supply of labor to the industry is given by S . In the competitive equilibrium given by point A , E_0 workers are hired at a wage of w_0 dollars.

Each point on the demand curve gives the number of workers that employers wish to hire at a particular wage. In particular, employers are willing to hire E_0 workers if each worker costs w_0 dollars. To simplify the analysis, consider a very simple form of payroll tax. In particular, the firm will pay a tax of \$1 for every employee-hour it hires. In other words, if the wage is \$10 an hour, the total cost of hiring an hour of labor will be \$11 (\$10 goes to the worker and \$1 goes to the government). Because employers are only willing to pay a *total* of w_0 dollars to hire the E_0 workers, the imposition of the payroll tax implies that employers are now only willing to pay a wage rate of $w_0 - 1$ dollars to the workers in order to hire E_0 of them.

The payroll tax assessed on employers, therefore, leads to a downward parallel shift in the labor demand curve to D_1 , as illustrated in Figure 4-4. The new demand curve reflects the wedge that exists between the *total* amount that employers must pay to hire a worker and the amount that workers actually receive from the employer. In other words, employers take into account the *total* cost of hiring labor when they make their hiring decisions—so that the amount that they will want to pay to workers has to shift down by \$1 in order to cover the payroll tax. The payroll tax moves the labor market to a new equilibrium (point B in the figure). The number of workers hired declines to E_1 . The equilibrium wage rate—that is, the wage rate actually *received* by workers—falls to w_1 , but the *total* cost of hiring a worker rises to $w_1 + 1$.

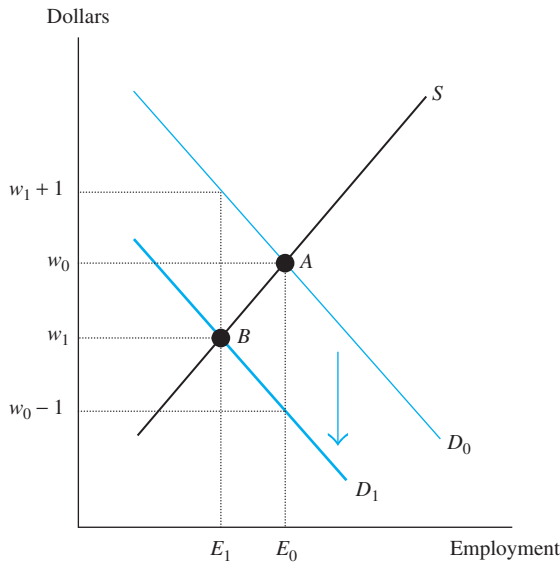
⁷ A detailed discussion of the impact of trade on labor markets is given by George Johnson and Frank Stafford, "The Labor Market Implications of International Trade," in Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics*, vol. 3B, Amsterdam: Elsevier, 1999, pp. 2215–2288.

⁸ Workers are also assessed a similar tax on their earnings, so the total tax payment is 15.3 percent on the first \$106,800 of salary, and a 2.9 percent tax on wages above \$106,800.

⁹ U.S. Bureau of the Census, *Statistical Abstract of the United States, 2012*. Washington, DC: Government Printing Office, 2011, Table 1361.

FIGURE 4-4 The Impact of a Payroll Tax Assessed on Firms

A payroll tax of \$1 assessed on employers shifts down the demand curve (from D_0 to D_1). The payroll tax cuts the wage that workers receive from w_0 to w_1 and increases the cost of hiring a worker from w_0 to $w_1 + 1$.



It is worth noting that even though the legislation clearly states that employers must pay the payroll tax, the labor market shifts part of the tax to the worker. After all, the cost of hiring a worker rises at the same time that the wage received by the workers declines. In a sense, therefore, firms and workers “share” the costs of the payroll tax.

A Tax Assessed on Workers

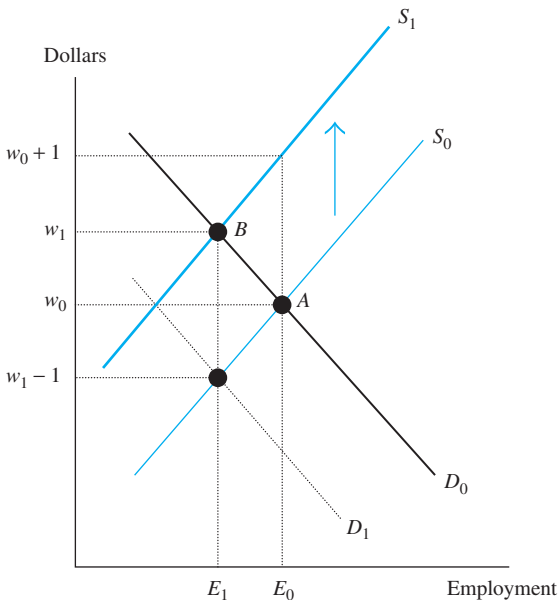
The political debate over payroll taxes often makes it appear that workers are better off when the payroll tax is assessed on the firm, rather than on the worker. In short, there seems to be an implicit assumption that most workers would rather see the payroll tax imposed on firms, whereas most firms would rather see the payroll tax imposed on workers. It turns out, however, that this assumption represents a complete misunderstanding of how a competitive labor market works. *It does not matter whether the tax is imposed on workers or firms.* The impact of the tax on wages and employment is the same regardless of how the legislation is written.

Suppose, for instance, that the \$1 tax on every hour of work had been assessed on workers rather than employers. What would the resulting labor market equilibrium look like?

The labor supply curve gives the wage that workers require to supply a particular number of hours to the labor market. In Figure 4-5, workers are willing to supply E_0 hours when the wage is w_0 dollars. The government now mandates that workers pay the government \$1 for every hour they work. Workers, however, still want to take home w_0 dollars if they supply E_0 hours. In order to supply these many hours, therefore, the workers will now want a payment of $w_0 + 1$ dollars from the employer. In effect, the payroll tax assessed on workers shifts the supply curve up by one dollar to S_1 . The payroll tax imposed on workers, therefore, creates a wedge between the amount that workers must receive from their employers if they are to offer their services in the labor market and the amount that workers get to take home.

FIGURE 4-5 The Impact of a Payroll Tax Assessed on Workers

A payroll tax assessed on workers shifts the supply curve to the left (from S_0 to S_1). The payroll tax has the same impact on the equilibrium wage and employment regardless of who it is assessed on.



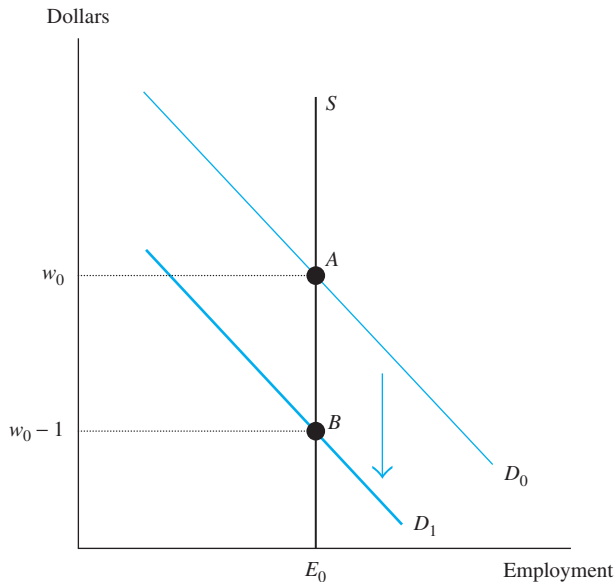
The labor market equilibrium then shifts from A to B . At the new equilibrium, workers receive a wage of w_1 dollars from the employer, and total employment falls from E_0 to E_1 . Note, however, that because the worker must pay a \$1 tax per hour worked, the actual after-tax wage of the worker falls from w_0 to $w_1 - 1$.

The payroll tax assessed on the worker, therefore, leads to the same types of changes in labor market outcomes as the payroll tax assessed on firms. Both taxes reduce the take-home pay of workers, increase the cost of an hour of labor to the firm, and reduce employment. In fact, one can show that the \$1 payroll tax will have *exactly* the same numerical effect on wages and employment regardless of who bears the legal responsibility of paying for it. To see this, note that if the \$1 payroll tax had been assessed on firms, the demand curve in Figure 4-5 would have shifted down by \$1 (see the curve D_1 in the figure). The labor market equilibrium generated by the intersection of this demand curve and the original supply curve (S_0) is the same as the labor market equilibrium that resulted when the tax was assessed on workers. If the tax were assessed on firms, the worker would receive a wage of $w_1 - 1$, and the firm's total cost of hiring a worker would be w_1 .

This result illustrates a principle that is worth remembering: The true incidence of the payroll tax (that is, who pays what) has little to do with the way the tax law is written or the way the tax is collected. In the end, the true incidence of the tax is determined by the way the competitive market operates. Even though a payroll tax assessed on the firm shifts down the demand curve, it has the same labor market impact as a revenue-equivalent payroll tax assessed on workers (which shifts up the supply curve).

FIGURE 4-6 The Impact of a Payroll Tax Assessed on Firms with Inelastic Supply

A payroll tax assessed on the firm is shifted completely to workers when the labor supply curve is perfectly inelastic. The wage is initially w_0 . The \$1 payroll tax shifts the demand curve to D_1 , and the wage falls to $w_0 - 1$.

**When Will the Payroll Tax Be Shifted Completely to Workers?**

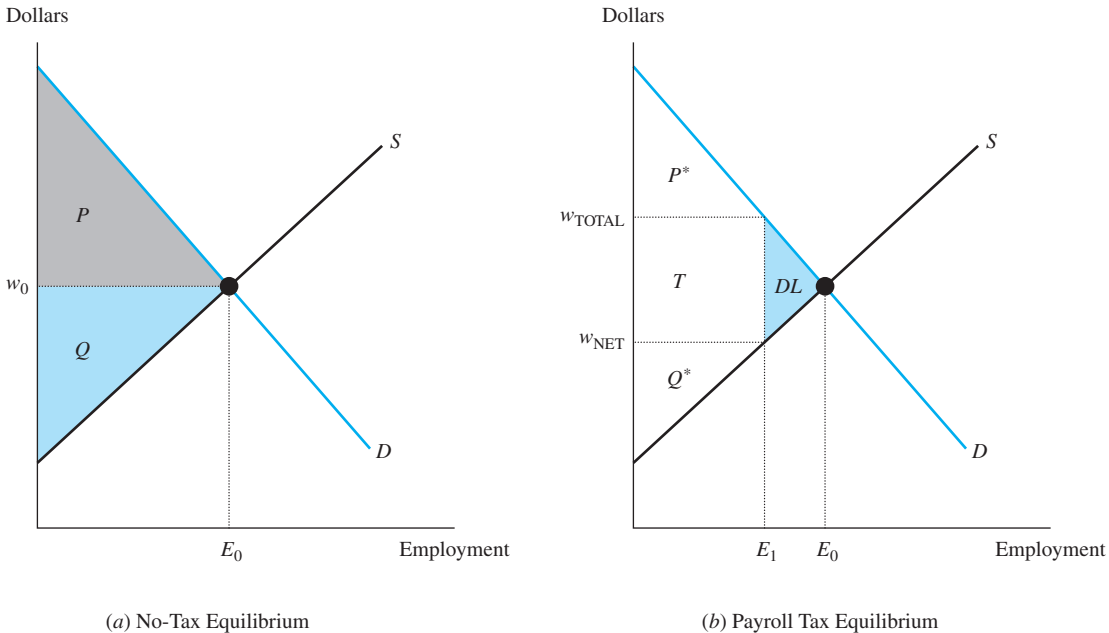
In one extreme case, the payroll tax is shifted entirely to workers. Suppose that the tax is assessed on the firm and that the supply curve of labor is perfectly inelastic, as illustrated in Figure 4-6. A total of E_0 workers are employed in this market regardless of the wage. As before, the imposition of the payroll tax shifts the demand curve down by \$1. Prior to the tax, the equilibrium wage was w_0 . After the tax, the equilibrium wage is $w_0 - 1$. The more inelastic the supply curve, therefore, the greater the fraction of the payroll taxes that workers end up paying.

As we saw in Chapter 2, labor supply curves for men are inelastic. It would not be surprising, therefore, if most of the burden of payroll taxes is indeed shifted to workers. Although there is some disagreement regarding the exact amount of this shift, some studies suggest that workers, through a lower competitive wage, pay for as much 90 percent of payroll taxes.¹⁰

¹⁰ Daniel S. Hamermesh, "New Estimates of the Incidence of the Payroll Tax," *Southern Economic Journal* 45 (February 1979): 1208–1219; Charles Beach and Frederick Balfour, "Estimated Payroll Tax Incidence and Aggregate Demand for Labour in the United Kingdom," *Econometrica* 50 (February 1983): 35–48; Jonathan Gruber, "The Incidence of Payroll Taxation: Evidence from Chile," *Journal of Labor Economics* 15 (July 1997, Part 2): S102–S135; Kevin Lang, "The Effect of the Payroll Tax on Earnings: A Test of Competing Models of Wage Determination," National Bureau of Economic Research Working Paper No. 9537, February 2003; and Patricia M. Anderson and Bruce D. Meyer, "Unemployment Insurance Tax Burdens and Benefits: Funding Family Leave and Reforming the Payroll Tax," *National Tax Journal* 59 (March 2006): 77–95.

FIGURE 4-7 Deadweight Loss of a Payroll Tax

(a) In a competitive equilibrium, E_0 workers are hired at a wage of w_0 . The triangle P gives the producer surplus and Q gives the worker surplus. The total gains from trade equal $P + Q$. (b) The payroll tax reduces employment to E_1 ; raises the cost of hiring to w_{TOTAL} ; and reduces the worker's take-home pay to w_{NET} . The triangle P^* gives the producer surplus; the triangle Q^* gives the worker surplus; and the rectangle T gives the tax revenues. The net loss to society, or deadweight loss, is given by the triangle DL .

**Deadweight Loss**

Because payroll taxes typically increase the cost of hiring a worker, these taxes reduce total employment—regardless of whether the tax is imposed on workers or firms. The after-tax equilibrium, therefore, is inefficient because the number of workers employed is not the number that maximizes the total gains from trade in the labor market.

Figure 4-7a illustrates again the total gains from trade accruing to the national economy in the absence of a payroll tax. The total gains from trade are given by the sum of producer surplus and worker surplus, or the area $P + Q$.

Figure 4-7b shows what happens to this gain when the government imposes a payroll tax. As we have seen, it does not matter if the payroll tax is imposed on firms or imposed on workers. In either case, employment declines to E_1 ; the cost of hiring a worker rises to w_{TOTAL} ; and the worker's take-home pay falls to w_{NET} . The producer surplus is now given by the smaller triangle P^* ; the worker surplus is given by the smaller triangle Q^* ; and the tax revenues accruing to the government are given by the rectangle T . The total gains from trade are given by the sum of the new producer surplus and the new worker surplus, as well as the tax revenues. After all, the government will redistribute these tax revenues in some fashion and someone will benefit from the government's expenditures. Table 4-1 summarizes the relevant information.

TABLE 4-1
Welfare
Implications of
a Payroll Tax

| | No-Tax Equilibrium | Payroll Tax Equilibrium |
|-----------------------|--------------------|-------------------------|
| Producer surplus | P | P^* |
| Worker surplus | Q | Q^* |
| Tax revenues | — | T |
| Total gain from trade | $P + Q$ | $P^* + Q^* + T$ |
| Deadweight loss | — | DL |

The comparison of Figure 4-7a and Figure 4-7b yields an important conclusion. The imposition of the payroll tax reduces the total gains from trade. There is a triangle, DL , that represents the **deadweight loss** (or *excess burden*) of the tax. Note that the deadweight loss measures the value of gains forgone because the tax forces employers to cut employment below the efficient level and has nothing to do with the cost of enforcing or collecting the payroll tax. The deadweight loss arises because the tax prevents some workers who were willing to work from being hired by employers who were willing to hire them. These forgone deals were beneficial to society because the worker's value of marginal product exceeded the worker's value of time outside the labor market.¹¹

Employment Subsidies

The labor demand curve is shifted not only by payroll taxes but also by government subsidies designed to encourage firms to hire more workers. An employment subsidy lowers the cost of hiring for firms. In the typical subsidy program, the government grants the firm a tax credit, say of \$1, for every person-hour it hires. Because this subsidy reduces the cost of hiring a person-hour by \$1, it shifts the demand curve up by that amount, as illustrated in Figure 4-8. The new demand curve (D_1) gives the price that firms are willing to pay to hire a particular number of workers after they take account of the employment subsidy. Labor market equilibrium shifts from point A to point B . At the new equilibrium, there is more employment (from E_0 to E_1). In addition, the subsidy increases the wage that workers actually receive (from w_0 to w_1), and *reduces* the wage that firms actually have to pay out of their own pocket (from w_0 to $w_1 - 1$).

The labor market impact of these subsidies can be sizable and will obviously depend on the elasticities of the labor supply and the labor demand curve. For instance, if the labor supply elasticity is 0.3 and the labor demand elasticity is -0.5 , it has been estimated that a subsidy that reduces the cost of hiring by 10 percent would increase the wage by 4 percent and increase employment by 2 percent.¹²

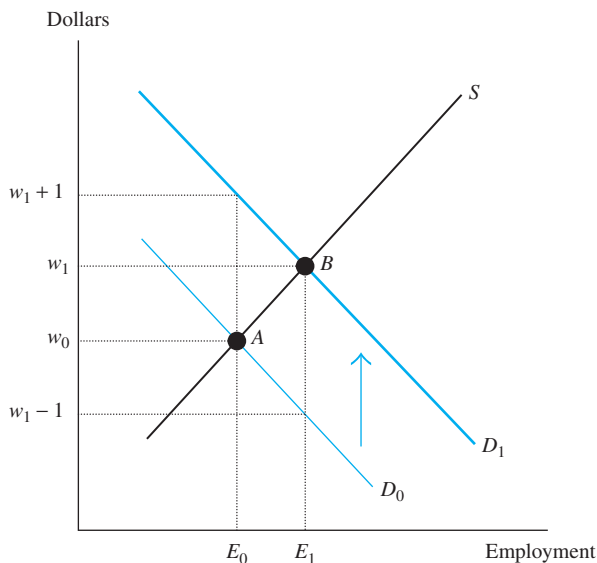
The largest employment subsidy program in U.S. history, the New Jobs Tax Credit (NJTC), began soon after the recession of 1973–1975 and was in effect from mid-1977 through 1978. The NJTC gave firms a tax credit of 50 percent on the first \$4,200 paid to a worker, as long as the firm's total wage bill rose by more than 2 percent over the

¹¹ A detailed discussion of the deadweight loss arising from various types of government regulations is given by James R. Hines Jr., "Three Sides of Harberger Triangles," *Journal of Economic Perspectives* 13 (Spring 1999): 167–188.

¹² Lawrence F. Katz, "Wage Subsidies for the Disadvantaged," in Richard B. Freeman and Peter Gottschalk, editors, *Generating Jobs*, New York: Russell Sage Press, 1998, pp. 21–53.

FIGURE 4-8 The Impact of an Employment Subsidy

An employment subsidy of \$1 per worker hired shifts up the demand curve, increasing employment. The wage that workers receive rises from w_0 to w_1 . The wage that firms actually pay falls from w_0 to $w_1 - 1$.



previous year. The firm could claim no more than \$100,000 as a tax credit for any given year. Because only the first \$4,200 of earnings were eligible for a credit, this program was designed to encourage the employment of low-wage workers.

A survey of the evidence concludes that the NJTC increased employment in the subsample of firms that were aware of the program, generating about 400,000 permanent new jobs.¹³ The total cost of the tax credit to the U.S. Treasury was roughly \$4.5 billion, so each new job cost taxpayers an average of \$11,250.

It turns out, however, that only 27 percent of small firms were even aware of the NJTC's existence and that only 6 percent of firms actually made use of the tax credits. Because of the limited participation of firms, it is possible that only a small fraction of the employment increase can be directly attributed to the NJTC. After all, firms that had plans to expand and hire more workers had the most incentive to learn about the program and to make use of the tax credits. In other words, employment would have risen among the firms that ended up being the beneficiaries of the NJTC even if the program had not been in effect.

The Targeted Jobs Tax Credit (TJTC), which began in 1978, offers subsidies (lasting two years) to firms that hire workers from specific groups. These groups include ex-convicts, persons receiving general assistance, and Vietnam veterans. Originally, the TJTC provided

¹³ Jeffrey Perloff and Michael Wachter, "The New Jobs Tax Credit—An Evaluation of the 1977–78 Wage Subsidy Program," *American Economic Review* 69 (May 1979): 173–179; John Bishop, "Employment in Construction and Distribution Industries: The Impact of the New Jobs Tax Credit," in Sherwin Rosen, editor, *Studies in Labor Markets*, Chicago, IL: University of Chicago Press, 1981; and Mark R. Killingsworth, "Substitution and Output Effects on Labor Demand: Theory and Policy Applications," *Journal of Human Resources* 20 (Winter 1985): 142–152.

a tax credit amounting to 50 percent of first-year and 25 percent of second-year wages (up to \$6,000) for employers who hired individuals in the targeted groups. Few employers seem to have been aware of the existence of this program, and the evidence does not suggest that this particular type of targeted tax credit greatly increased the employment of targeted groups.¹⁴ One possible explanation for the failure of targeted tax credits to increase employment is that employers may attach a stigma to targeted workers and will shy away from them. The impact of this type of discrimination on the firm's demand for labor is discussed at length in Chapter 9.

Employer Sanctions as a Payroll Tax

The number of illegal immigrants residing in the United States has risen dramatically in recent years. According to the Department of Homeland Security, 11.6 million illegal immigrants lived in the United States as of January 2006.¹⁵

The 1986 Immigration Reform and Control Act (IRCA) makes it illegal for employers to “knowingly hire” illegal immigrants. Under this legislation, employers who hire these unauthorized workers are subject to potential penalties and fines. Employers who disobey the law are liable for fines that, for first-time offenders, range from \$250 to \$2,000 per illegal alien hired. Criminal penalties can be imposed on repeated violators when there is a “pattern and practice” of hiring illegal aliens. These penalties include a fine of \$3,000 per illegal alien, and up to six months in prison.

Partly because of lax enforcement, IRCA has obviously not hampered the growth of the illegal immigrant population in the United States. As a result, some states have moved to fill the void by enacting legislation that penalizes employers who hire illegal immigrants within the state. Beginning on January 1, 2008, for example, after a second offense, Arizona penalizes employers who hire illegal immigrants with the ultimate “death sentence”: by simply revoking the employer's license to operate a business within the state of Arizona.

A number of studies examine the labor market impact of the employer sanctions introduced by IRCA.¹⁶ A central idea in these studies is that employer sanctions act as a “tax” on employers. In other words, employer sanctions increase the cost of hiring a worker. After all, there is some probability that a hired worker is unauthorized to work in the

¹⁴ John H. Bishop and Mark Montgomery, “Does the Targeted Jobs Tax Credit Create Jobs at Subsidized Firms?” *Industrial Relations* 32 (Fall 1993): 289–306. An extensive discussion of how tax policy affects human capital investments is given by Pedro Carneiro and James J. Heckman, “Human Capital Policy,” in James J. Heckman and Alan B. Krueger, *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press, 2003; see also J. Michael Orszag and Dennis J. Snower, “Designing Employment Subsidies,” *Labour Economics* 10 (October 2003): 557–572.

¹⁵ Michael Hofer, Nancy Rytina, and Christopher Cambell, “Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2006,” Office of Immigration Statistics, Department of Homeland Security, August 2007, available at http://www.dhs.gov/xlibrary/assets/statistics/publications/ill_pe_2006.pdf.

¹⁶ See Deborah A. Cobb-Clark, Clinton R. Shiells, and B. Lindsay Lowell, “Immigration Reform: The Effects of Employer Sanctions and Legalization on Wages,” *Journal of Labor Economics* 13 (July 1995): 472–498; and J. Edward Taylor and Dawn Thilmany, “Labor Turnover, Farm Labor Contracts, and IRCA's Impact on the California Farm Labor Market,” *American Journal of Agricultural Economics* 75 (May 1993): 350–360. An international comparison of various employer sanction programs is given by Philip Martin and Mark Miller, “Employer Sanctions: French, German, and US Experiences,” International Migration Branch, International Labour Office, Geneva, 2000.

United States, and there is some probability that the employer will be caught and fined for this illegal hiring. In a competitive labor market, this fine is like a payroll tax that shifts the demand curve downwards as in Figure 4-4. Using this perspective, therefore, employer sanctions would *reduce* employment and *lower* wages in a competitive labor market. It would seem, therefore, that the group of workers that the employer sanctions are presumably trying to protect (namely, the “legal” or authorized workforce) is paying for the government’s attempt to penalize employers that hire unauthorized aliens.

On reflection, however, it should be evident that viewing employer sanctions as a payroll tax introduces two crucial assumptions into the discussion that may not be correct. First, the application of Figure 4-4 in the employer sanctions context assumes that employers do not know the legal status of the workers they hire with certainty—which is why the “tax” is imposed on every worker hired. Second, the labor supply curve is assumed to remain fixed so that the number of unauthorized aliens in the labor market does not respond to the imposition of employer sanctions.

The assumption that employers do not know the legal status of potential new hires does not correctly describe how many of the newer state-specific employer sanction statutes in the United States operate. Some of these laws, for example, encourage an employer to use an electronically based program to authenticate whether a new hire is, in fact, legally entitled to work in the country. The E-Verify program is an “Internet-based system operated by the Department of Homeland Security (DHS) in partnership with the Social Security Administration (SSA) that allows participating employers to electronically verify the employment eligibility of their newly hired employees.”¹⁷ In other words, the E-Verify program allows an employer to (quickly and cheaply) determine whether a job applicant is legally entitled to work in the United States simply by checking the Social Security number reported by the applicant against the federal government’s database.

Arizona’s employer sanction legislation explicitly declares that the use of the E-Verify program relieves the employer from all legal liability if they were to mistakenly hire an unauthorized alien. From the perspective of an Arizona employer, as long as the E-Verify system authenticates the employment eligibility of a job applicant, the employer may proceed with the hiring even if it turns out that the electronic authentication was mistaken and that the job applicant was indeed an unauthorized alien. As a result, there is no “tax” associated with possible detection from hiring unauthorized workers as long as the employer abides by the results of the E-Verify program. Put differently, the Arizona employer sanction program would not lower the wage of the “legal” workforce since, by definition, anyone who “passes” the E-Verify test is considered to be an authorized worker.¹⁸

Second, the assumption that the labor supply of unauthorized aliens does not respond to the employer sanctions seems to be false in the Arizona context. Newspaper reports have noted that, even at this early stage, Arizona’s program has already induced a decline in the

¹⁷ The U.S. Department of Homeland Security maintains a Web site that provides information about the E-Verify program at http://www.dhs.gov/ximgtn/programs/gc_1185221678150.shtm.

¹⁸ There is one hiring cost imposed by the Arizona employer sanction program that would indeed act as a “payroll” tax in a competitive labor market—namely, the cost that employers must incur to conduct the electronic check of Social Security numbers. The available evidence, however, suggests that this cost (on a per-worker basis) is numerically trivial.

number of unauthorized aliens choosing to settle in that state.¹⁹ The laws of supply and demand would then presumably imply that such a cut in supply increases employer competition for the remaining workers, *raising* the wage of authorized workers.

4-4 Policy Application: Payroll Taxes versus Mandated Benefits

The government can ensure that workers receive particular benefits by mandating that firms provide those benefits to their workers. In the United States, for example, the federal government mandates that employers keep the workplace safe or provide workers' compensation insurance to their workers. How do such **mandated benefits** affect labor market outcomes in terms of wages and employment?

To illustrate the basic theory, it is useful to think in terms of a specific mandated benefit; for example, the provision of spinach pie to workers during the lunch hour. Although this example might sound a bit far-fetched, it is quite useful for understanding how the labor market consequences of government mandates differ from those of payroll taxes—regardless of whether the mandate requires firms to provide spinach pie or health insurance.

Figure 4-9a illustrates how the government mandate affects labor market equilibrium.²⁰ The initial equilibrium is at point P , with wage w_0 and employment E_0 . Suppose that the mandated provision of spinach pie costs C dollars per worker. The mandated provision of this benefit results in a parallel downward shift of the demand curve to D_1 , where the vertical difference between the two demand curves is C dollars. After all, the firm is willing to hire E_0 workers only if the per-worker total cost of employment is w_0 . The mandated provision of spinach pie implies that the firm is now willing to pay each of the E_0 workers a wage of $w_0 - C$.

Consider initially the case where workers despise spinach pie—regardless of what the government says about its nutritional values. The government may mandate the firm to provide the benefit; the firms in the industry may indeed serve up a slice of spinach pie at lunchtime; but no one can force the workers to eat it. The workers simply take their slice and quickly dispose of it in the trash can. As a result, workers attach no value whatsoever to this particular benefit. The new labor market equilibrium would then be at point Q , where firms spend a total of $w_1 + C$ dollars to hire a worker (w_1 for the wage and C for the pie), and employment falls to E_1 . Note that the equilibrium resulting from a government mandate where workers attach no value to the mandated benefit is what we would have observed if the government had instead enacted a payroll tax of C dollars.

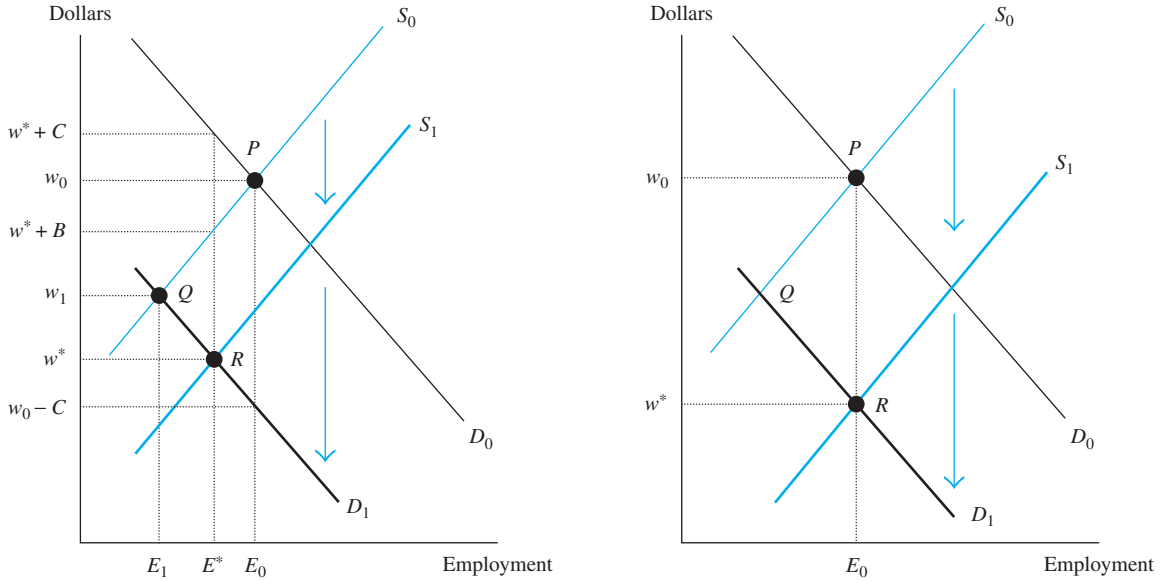
However, it is possible that the typical worker appreciates the nutritional content of the spinach pie, finds it quite tasty, and values the mandated benefit. In particular, suppose that each worker in the industry values the provision of the spinach pie at B dollars, where $B < C$. In other words, workers are willing to pay somewhat less for the spinach pie than what it costs firms to provide it. The fact that the spinach pie makes workers better off implies that the mandated benefit affects not only the demand curve, but also the supply curve. The initial

¹⁹ "Arizona Seeing Signs of Flight by Immigrants," *New York Times*, February 12, 2008.

²⁰ See Lawrence H. Summers, "Some Simple Economics of Mandated Benefits," *American Economic Review* 79 (May 1989): 177–183, for a more detailed discussion of the labor market consequences of mandated benefits.

FIGURE 4-9 The Impact of a Mandated Benefit

(a) It costs firms C dollars to provide a mandated benefit, shifting the demand curve from D_0 to D_1 . Workers value the benefit only by B dollars, so the supply curve shifts down by less. Employment at the new equilibrium (point R) is higher than would have been the case if the firm had been assessed a payroll tax of C dollars (point Q), but lower than in a no-tax equilibrium (point P). (b) When the cost of providing the mandate equals the worker's valuation, the resulting equilibrium replicates the competitive no-tax equilibrium in terms of employment, total cost of hiring workers, and total compensation received by workers.



(a) Cost of Mandate Exceeds Worker's Valuation

(b) Cost of Mandate Equals Worker's Valuation

supply curve S_0 in Figure 4-9a indicates that E_0 workers are willing to work as long as each receives a *total* compensation of w_0 dollars. Because workers value the spinach pie at B dollars, the E_0 workers are now willing to work as long as firms pay them a wage of $w_0 - B$. In effect, the mandated benefit leads to a parallel downward shift of the supply curve by B dollars, leading to the new supply curve S_1 .

Because it is costly to provide the spinach pie and because workers value this pie, the new labor market equilibrium is given by the intersection of the new supply and demand curves (point R), so that E^* workers are employed at the new equilibrium. One important result of the analysis is that although employment falls from E_0 to E^* , it falls by less than it would have fallen if the government had instead imposed a payroll tax of C dollars on firms; in that case, employment would have dropped from E_0 to E_1 .

The new equilibrium wage is w^* . But this wage does not represent the value of the employment package from the perspective of either workers or firms. It costs the firm $w^* + C$ dollars to hire a worker; and the worker values the compensation package at $w^* + B$ dollars. In contrast to the initial competitive equilibrium, workers receive less compensation and firms face higher costs. However, in contrast to the payroll tax equilibrium, both firms and workers are better off—workers earn higher compensation and firms face lower costs.

There is one special case that is of interest. Suppose that the mandated provision of a spinach pie costs C dollars to the firm *and* that workers value this pie at C dollars. In other words, workers value the mandated benefit at the same rate that it costs to provide this benefit (so that $B = C$). Figure 4-9*b* illustrates this situation. The supply curve and the demand curve both shift down by *exactly* the same amount (that is, C dollars). At the new equilibrium (point R), employment is still E_0 . Similarly, workers value their compensation package at $w^* + C$, and the firm's cost is $w^* + C$. This quantity equals the competitive wage w_0 .

The analysis of mandated benefits, therefore, reveals an important property of competitive labor market equilibrium. As long as the mandated benefit provides some value to workers, the mandated benefit is preferable to a payroll tax because it leads to a smaller cut in employment. Put differently, the government mandate reduces the deadweight loss arising from the reduced employment caused by the payroll tax. In fact, if the cost of providing the mandated benefit is exactly equal to the value that workers attach to this benefit, the mandated benefit does not create any such deadweight loss, as firms end up hiring exactly the same number of workers that would have been hired in a competitive equilibrium at exactly the same cost.

Health Insurance as a Mandated Benefit

In the United States, nearly two-thirds of persons below the age of 65 are covered by employer-provided health insurance, and nearly 16 percent do not have any health insurance coverage at all. There is a heated debate about whether employers should be required to provide health insurance to all their workers. Our discussion of payroll taxes and mandated benefits clearly suggests that mandated increases in health insurance premiums could have significant effects in the labor market, including changes in the market wage and in the number of workers employed.

A recent study estimates the magnitude of the labor market effects associated with health-related increases in hiring costs.²¹ Beginning around 2000, partly because of a substantial increase in malpractice payments, the premiums for physician malpractice insurance soared, which, in turn, greatly increased the cost of employer-provided health insurance. Since 2000, for example, the cost of employer-provided health insurance has risen by nearly 60 percent, even though the type and scope of the coverage were unchanged. These increases vary greatly across states, suggesting that one can use the state variation in malpractice payments as an instrument in a model that attempts to identify how increases in the cost of employer-provided health insurance premiums affect wages and employment.

It has been estimated that a 10 percent increase in health insurance premiums reduces the probability of employment by 1.2 percentage points, reduces the number of hours worked by 2.4 percent, and lowers the wage of workers with employer-provided health insurance by just over 2 percent. In short, the implementation of any new health insurance mandate can easily have significant repercussions on the labor market.

As an example, consider President Clinton's Health Care Reform proposal (prepared in 1993 by a task force headed by his wife, Hillary Clinton). The Clinton proposal would have required employers to pay for a large fraction of the health insurance premium of

²¹ Katherine Baicker and Amitabh Chandra, "The Labor Market Effects of Rising Health Insurance Premiums," *Journal of Labor Economics* 3 (July 2006): 609–634.

their workers. In particular, firms would have had to pay 80 percent of the costs of health insurance premiums for their workforce, with the total employer contributions being capped at 7.9 percent of the firm's payroll. Firms that employed fewer than 50 workers would have had their contributions capped at lower levels, sometimes as low as 3.5 percent of payroll.

Had it been enacted, the Clinton proposal would have been a new payroll tax on employers who did not currently provide health insurance to their workers or who provided "substandard" programs. As such, the program would have had sizable disemployment effects. In addition, because part of the tax is shifted to workers, wages would have fallen.

Our discussion suggests that the impact of payroll taxes on both employment and wages depends on the elasticities of both labor supply and labor demand. A back-of-the-envelope calculation suggests that if the labor supply curve has an elasticity of 0.2 and the labor demand curve has an elasticity of -1 , the Clinton plan would have reduced employment by 517,000 jobs and the annual earnings of the workers who were currently uninsured would have fallen by at least \$1,000.²²

The Clinton Health Care Reform proposal would likely have had many other impacts on the labor market. For example, small firms would clearly have hesitated before expanding their workforce to more than 50 workers, while firms currently employing just over 50 workers would likely have contracted (or subdivided) as they searched for ways of minimizing their financial burden.

4-5 Policy Application: The Labor Market Impact of Immigration

We now consider how government policies that restrict or favor large-scale immigration shift the supply curve and alter labor market outcomes. Because of major policy changes, the United States witnessed a major resurgence in immigration after 1965. In the 1950s, for example, only about 250,000 immigrants entered the country annually. Since 2000, over 1 million legal and illegal immigrants are entering the country annually. These sizable supply shifts reignited the debate over immigration policy in the United States.²³

There also has been a resurgence of large-scale immigration in many other developed countries. According to the United Nations, 3.1 percent of the world's population (or approximately 214 million people) now reside in a country where they were not born.²⁴ By 2010, the fraction of foreigners in the country's population was 13.1 percent in Germany, 10.7 percent in France, 13.5 percent in the United States, 21.3 percent in Canada, and 23.2 percent in Switzerland. Perhaps the key issue in the immigration debate

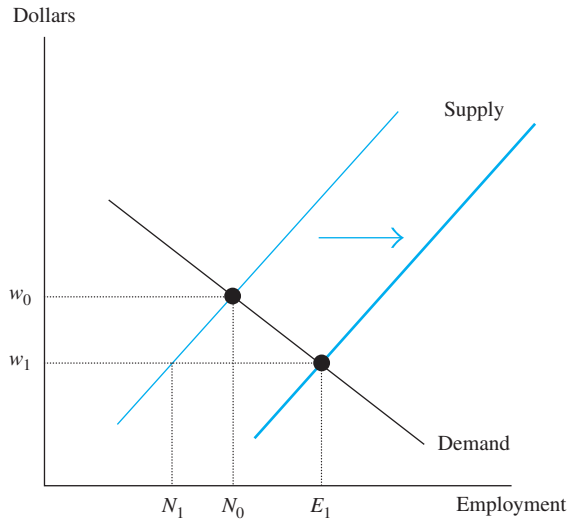
²² Alan B. Krueger, "Observations on Employment-Based Government Mandates, with Particular Reference to Health Insurance," in Lewis Solmon and Alec Levenson, editors, *Labor Markets, Employment Policy and Job Creation*, Boulder, CO: Westview Press, 1994.

²³ A good summary of the socioeconomic characteristics of immigrants in the U.S. labor market is given by Abraham T. Mosisa, "The Role of Foreign-Born Workers in the U.S. Economy," *Monthly Labor Review* 125 (May 2002): 3–14.

²⁴ United Nations, Department of Economic and Social Affairs. Trends in International Migrant Stock: The 2008 Revision, <http://esa.un.org/migration/p2k0data.asp>

FIGURE 4-10 The Short-Run Impact of Immigration When Immigrants and Natives Are Perfect Substitutes

Because immigrants and natives are perfect substitutes, the two groups are competing in the same labor market. Immigration shifts out the supply curve. As a result, the wage falls from w_0 to w_1 , and total employment increases from N_0 to E_1 . Note that, at the lower wage, there is a decline in the number of natives who work, from N_0 to N_1 .



in most receiving countries concerns the impact of immigrants on the labor market opportunities of native-born workers.²⁵

The simplest model of immigration assumes that immigrants and natives are perfect substitutes in production. In other words, immigrants and natives have the same types of skills and are competing for the same types of jobs. The impact of immigration on this labor market in the short run—with capital held fixed—is illustrated in Figure 4-10. As immigrants enter the labor market, the supply curve shifts out, increasing total employment from N_0 to E_1 and reducing wages (from w_0 to w_1). Note that fewer native-born workers are willing to work at this lower wage, so the employment of native workers actually falls, from N_0 to N_1 . In a sense, immigrants “take jobs away” from natives by reducing the native wage and convincing some native workers that it is no longer worthwhile to work.

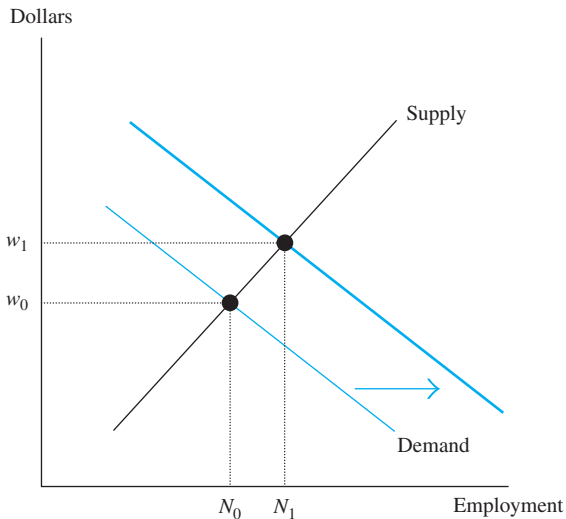
The short-run impact of immigration when native workers and immigrants are perfect substitutes, therefore, is unambiguous. As long as the demand curve is downward sloping and capital is fixed, an increase in immigration will move the economy down the demand curve, reducing the wage and employment of native-born workers.

Of course, the assumption that native workers and immigrants are perfect substitutes is questionable. It may be that immigrant and native workers are not competing for the same types of jobs. For instance, immigrants may be particularly adept at some types of labor-intensive agricultural production. This frees up the more skilled native workforce to

²⁵ An excellent description of the academic debate over how to measure the labor market impact of immigration and how this discussion has influenced the U.S. policy debate is given by Roger Lowenstein, “The Immigration Equation,” *New York Times Magazine*, July 9, 2006.

FIGURE 4-11 The Short-Run Impact of Immigration When Immigrants and Natives Are Complements

If immigrants and natives are complements, they are not competing in the same labor market. The labor market in this figure denotes the supply and demand for native workers. Immigration makes natives more productive, shifting out the demand curve even though capital is fixed. This leads to a higher native wage and to an increase in native employment.



perform tasks that make better use of their human capital. The presence of immigrants increases native productivity because natives can now specialize in tasks that are better suited to their skills. Immigrants and natives thus *complement* each other in the labor market.

If the two groups are complements in production, an increase in the number of immigrants raises the marginal product of natives, shifting up the demand curve for native-born workers. As Figure 4-11 shows, this increase in native productivity raises the native wage from w_0 to w_1 . Moreover, some natives who previously did not find it profitable to work now see the higher wage rate as an additional incentive to enter the labor market, and native employment also rises from N_0 to N_1 .

Short-Run versus Long-Run Effects

Suppose that immigrants and natives are perfect substitutes. In the short run, immigrants lower the wage but raise the returns to capital. After all, employers can now hire workers at a lower wage. Over time, the increased profitability of firms will inevitably attract capital flows into the marketplace, as old firms expand and new firms open up shop to take advantage of the lower wage. This increase in the capital stock, therefore, will shift the demand curve for labor to the right and will tend to attenuate the negative impacts of the initial labor supply shock.

The crucial question is: By how much will the demand curve shift to the right in the long run? If the demand curve were to shift just a little, the competing native workers would still receive lower wages. If, on the other hand, the demand curve were to shift to the right dramatically, the negative wage effects might disappear.

The extent of the rightward shift in the labor demand curve depends on the technology underlying the production function. To illustrate, suppose that the aggregate production function in the receiving country can be described by the well-known Cobb-Douglas production function:

$$q = AK^\alpha L^{1-\alpha} \quad (4-1)$$

where A is a constant and α is a parameter that lies between 0 and 1. Note that this Cobb-Douglas production function has the property that the aggregate economy in the receiving country has constant returns to scale: if we double labor and double capital, we double output. There is strong evidence suggesting that the aggregate U.S. economy can be reasonably described by the type of production technology specified in equation (4-1).²⁶

The theory of factor demand in a competitive labor market implies that the price of capital (which equals the rate of return to capital) is given by the value of marginal product of capital and that the wage is given by the value of marginal product of labor. For simplicity, suppose that the price of the output is set arbitrarily equal to \$1. Using elementary calculus, it is then easy to show that the value of marginal product equations for capital and labor are given by

$$r = \$1 \times \alpha AK^{\alpha-1} L^{1-\alpha} \quad (4-2)$$

$$w = \$1 \times (1 - \alpha)AK^\alpha L^{-\alpha} \quad (4-3)$$

A little algebraic manipulation shows that we can rewrite these two equations as

$$r = \alpha A \left(\frac{K}{L} \right)^{\alpha-1} \quad (4-4)$$

$$w = (1 - \alpha)A \left(\frac{K}{L} \right)^\alpha \quad (4-5)$$

The short-run effect of immigration is simply to increase the number of workers in the economy. Examination of equations (4-4) and (4-5) will show that this increase in the number of workers will raise the rate of return to capital r and will lower the wage w .

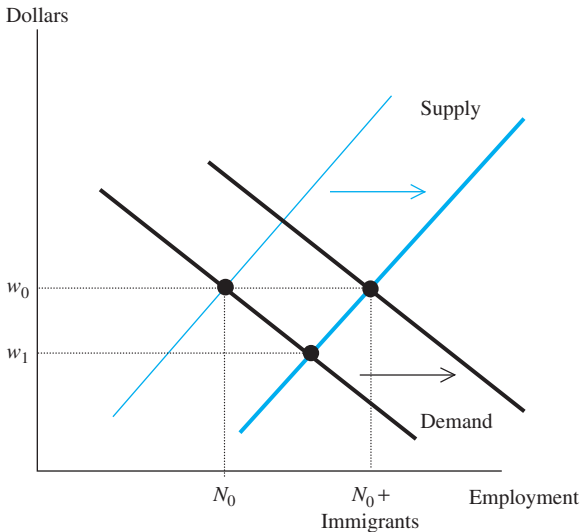
Over time, the higher rate of return to capital will induce an increase in the size of the capital stock K . Suppose that, in the long run, after all the capital adjustments that could have taken place have taken place, the rate of return to capital falls back to its “normal” level. This argument implies that the rate of return to capital is fixed in the long run at a value of r . But equation (4-4) clearly illustrates that the only way that the rate of return to capital can be fixed in the long run is if the capital-labor ratio (K/L) also is fixed in the long run. In other words, if immigration increases the number of workers by, say, 20 percent, then the capital stock also must increase by 20 percent in the long run.

This theoretical insight has very interesting (and important) implications for the labor market impact of immigration in the long run. Consider equation (4-5). If the capital-labor ratio is constant in the long run, equation (4-5) clearly shows that *the wage also must be*

²⁶ Daniel S. Hamermesh, *Labor Demand*, Princeton, NJ: Princeton University Press, 1993.

FIGURE 4-12 The Long-Run Impact of Immigration When Immigrants and Natives Are Perfect Substitutes

Because immigrants and natives are perfect substitutes, the two groups are competing in the same labor market. Immigration initially shifts out the supply curve. As a result, the wage falls from w_0 to w_1 . Over time, capital expands as firms take advantage of the cheaper workforce, shifting out the labor demand curve. If the aggregate production function has constant returns to scale, it must be the case that, after all capital adjustments have taken place, the wage is back at its initial level of w_1 . In addition, the long-run level of native employment is exactly what it was prior to the immigrant influx.



constant in the long run. In other words, immigration lowers the wage initially; over time, the capital stock increases as employers take advantage of the cheaper workforce; but, in the end, the capital stock completely adjusts to bring the economy back to where it began, with the same rate of return to capital and the same wage rate!

It is worth emphasizing that this theoretical prediction does *not* hinge on the assumption that the aggregate production function is Cobb-Douglas. The conclusion that immigration will have no long-run labor market impacts in the receiving country will hold whenever the aggregate production function has constant returns to scale.

The long-run effects are illustrated in Figure 4-12. The labor market is initially in equilibrium at a wage of w_0 and N_0 natives are employed at that wage. In the short run, the supply curve shifts to the right and the wage falls to w_1 . In the long run, the demand curve also shifts to the right—and it must shift by a sufficient amount to bring the labor market back to its pre-immigration equilibrium. In the end, the wage is again equal to w_0 . Note that, at this wage, the same number of native workers is employed as was employed prior to the immigrant influx.

We do not know how long it takes for the long run to arrive. It is unlikely that the capital stock adjusts instantaneously. We showed in Chapter 3, for example, that costs of adjustments create frictions in the speed with which employers adjust to various shocks. But the long run may not take as long as Keynes implied in his famous quip: “In the long run, we are all dead.” The key lesson from the theory is that immigration will have an adverse wage impact on competing native workers over some time period, and this impact will weaken as the economy adjusts to the immigrant influx.

Spatial Correlations

The discussion suggests a simple way to determine empirically if immigrants and natives are complements or substitutes in production. If they are substitutes, the earnings of native workers should be lower if they reside in labor markets where immigrants are in abundant supply. If they are complements, native earnings should be higher in those labor markets where immigrants tend to cluster.

Much of the empirical research that attempts to determine how immigration alters the economic opportunities of native workers is based on this implication of the theoretical analysis. These studies typically compare native earnings in cities where immigrants are a substantial fraction of the labor force (for example, Los Angeles or New York) with native earnings in cities where immigrants are a relatively small fraction (such as Pittsburgh or Nashville). The cross-city correlations estimated between wages and immigration are called *spatial correlations*. Of course, native wages would vary among labor markets even if immigration did not exist. The validity of the analysis, therefore, hinges crucially on the extent to which all the other factors that generate dispersion in native wages across cities can be controlled for when estimating a spatial correlation. These factors include geographic differences in the skills of natives, regional wage differentials, and variations in the level of economic activity. In terms of the fixed effects methodology introduced in Chapter 2, these empirical studies often include fixed effects for each city. As a result, the wage impact of immigration is being estimated by “differencing” the data within each city and observing how a city’s wage responds to changes in the number of immigrants settling in that city.

There has been a remarkable consensus in the many studies that estimate these spatial correlations.²⁷ The spatial correlation is probably slightly negative, so the native wage is somewhat lower in those labor markets where immigrants tend to reside. But the magnitude of this correlation is often very small. The evidence thus suggests that immigrants seem not to have much of an impact on the labor market opportunities of native workers.

It is often argued that African Americans are the one group whose economic progress is most likely to be hampered by the entry of immigrants into the United States.²⁸ The available evidence from the across-city studies, however, does not seem to support this claim.

²⁷ Jean B. Grossman, “The Substitutability of Natives and Immigrants in Production,” *Review of Economics and Statistics* 54 (November 1982): 596–603; Joseph G. Altonji and David Card, “The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives,” in John M. Abowd and Richard B. Freeman, editors, *Immigration, Trade, and the Labor Market*, Chicago: University of Chicago Press, 1991, pp. 201–234; and Robert J. LaLonde and Robert H. Topel, “Labor Market Adjustments to Increased Immigration,” in John M. Abowd and Richard B. Freeman, editors, *Immigration, Trade, and the Labor Market*, Chicago: University of Chicago Press, 1991, pp. 267–299. The evidence is surveyed in George J. Borjas, “The Economics of Immigration,” *Journal of Economic Literature* 32 (December 1994): 1667–1717; Rachel M. Friedberg and Jennifer Hunt, “The Impact of Immigration on Host Country Wages, Employment and Growth,” *Journal of Economic Perspectives* 9 (Spring 1995): 23–44; and David Card, “Is the New Immigration Really So Bad?” *Economic Journal* 115 (November 2005): F300–F323. Recent research also uses the cross-city comparison methodology to examine the impact of immigration on consumer prices; see Patricia Cortes, “The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data,” *Journal of Political Economy* 116 (June 2008): 381–422.

²⁸ See Daniel S. Hamermesh and Frank Bean, editors, *Help or Hindrance? The Economic Implications of Immigration for African-Americans*, New York: Russell Sage Press, 1998, for a collection of studies that analyze the impact of immigration on the economic well-being of black natives.

On the contrary, some studies report that African Americans residing in cities with relatively large numbers of immigrants actually have slightly higher wages than those residing in other labor markets.

The Mariel Boatlift

On April 20, 1980, Fidel Castro declared that Cuban nationals wishing to move to the United States could leave freely from the port of Mariel. By September 1980, about 125,000 Cubans, mostly low-skill workers, had chosen to undertake the journey. The demographic impact of the *Marielitos* on Miami's population and labor force was sizable. Almost overnight, Miami's labor force had unexpectedly grown by 7 percent. An influential study, however, indicates that the trend of wages and employment opportunities for Miami's population, including its African-American population, was barely affected by the Mariel flow.²⁹ The economic trends in Miami between 1980 and 1985, in terms of wage levels and unemployment rates, were similar to those experienced by such cities as Atlanta, Houston, and Los Angeles, cities that did not experience the Mariel flow.

Table 4-2 summarizes the evidence. In 1979, prior to the Mariel flow, the black unemployment rate in Miami was 8.3 percent. This unemployment rate rose to 9.6 percent by 1981, after the Mariel flow. Before we conclude that the Marielitos were responsible for this 1.3 percentage point increase in black unemployment in Miami, however, we have to determine what was happening in comparable cities, cities that did not experience the Mariel flow. It turns out that black unemployment was rising even faster in the control group, from 10.3 to 12.6 percent (or an increase of 2.3 points)—probably because macroeconomic conditions were worsening during that period. If anything, therefore, it seems that the Mariel flow actually slowed down the rise in black unemployment, so that the difference-in-differences calculation (or 1.3–2.3) suggests that the Mariel flow was responsible for a 1.0 percentage point *decline* in the black unemployment rate.³⁰

TABLE 4-2 Immigration and the Miami Labor Market

Sources: The Mariel flow data are drawn from David Card, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review* 43 (January 1990), p. 251. The data for the Mariel flow that did not happen are drawn from Joshua D. Angrist and Alan B. Krueger, "Empirical Strategies in Labor Economics," in Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics*, vol. 3A, Amsterdam: Elsevier, 1999, Table 7. The comparison cities are Atlanta, Houston, Los Angeles, and Tampa–St. Petersburg.

| | The Mariel Flow | | The Mariel Flow That Did Not Happen | |
|--------------------------------|-----------------|-------|-------------------------------------|-------|
| | Before | After | Before | After |
| Unemployment rate of blacks in | | | | |
| Miami | 8.3 | 9.6 | 10.1 | 13.7 |
| Comparison cities | 10.3 | 12.6 | 11.5 | 8.8 |
| Difference-in-differences | –1.0 | | +6.3 | |

²⁹ David Card, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review* 43 (January 1990): 245–257.

³⁰ It is important to point out, however, that the margin of error around this calculation is quite large, so one cannot confidently conclude that the difference-in-differences estimate is statistically different from zero.

The conclusion that even large and unexpected immigrant flows do not seem to adversely affect local labor market conditions seems to be confirmed by the experience of other countries. For instance, 900,000 persons of European origin returned to France within one year after the independence of Algeria in 1962, increasing France's labor force by about 2 percent. Nevertheless, there is no evidence that this increase in labor supply had a sizable impact on the affected labor markets.³¹ Similarly, when Portugal lost the African colonies of Mozambique and Angola in the mid-1970s, nearly 600,000 persons returned to Portugal, increasing Portugal's population by almost 7 percent. The *retornados* did not seem to have a large impact on the Portuguese economy.³²

Natural Experiments: Proceed with Caution

The Mariel study provides an excellent example of the difference-in-differences methodology: measuring the impact of immigration by comparing what happened in the labor market of interest (that is, the treated group) with what happened in labor markets that were not penetrated by immigrants (the control group). Recent research, however, has raised some questions about the interpretation of the evidence generated by these natural experiments—at least in the context of immigration.

In 1994, economic and political conditions in Cuba were ripe for the onset of a new boatlift of refugees into the Miami area, and thousands of Cubans began the hazardous journey. To prevent the refugees from reaching the Florida shore, the Clinton administration ordered the Navy to redirect all the refugees toward the American military base in Guantanamo. As a result, few of the potential migrants were able to migrate to Miami.

One can replicate the methodological design of the Mariel study by comparing Miami's labor market conditions—relative to those of control cities—before and after “the Mariel boatlift that didn't happen.”³³ It turns out that this nonevent *had* a substantial adverse impact on the unemployment rate of Miami's black workers. The black unemployment rate in Miami rose from 10.1 to 13.7 percent between 1993 and 1995 (see again Table 4-2), as compared to a decline from 11.5 to 8.8 percent in a set of comparison cities. The difference-in-differences methodology [or $3.6 - (-2.7)$] would then indicate that the unemployment rate of African Americans in Miami *rose* by 6.3 percentage points.³⁴

³¹ Jennifer Hunt, “The Impact of the 1962 Repatriates from Algeria on the French Labor Market,” *Industrial and Labor Relations Review* 45 (April 1992): 556–572.

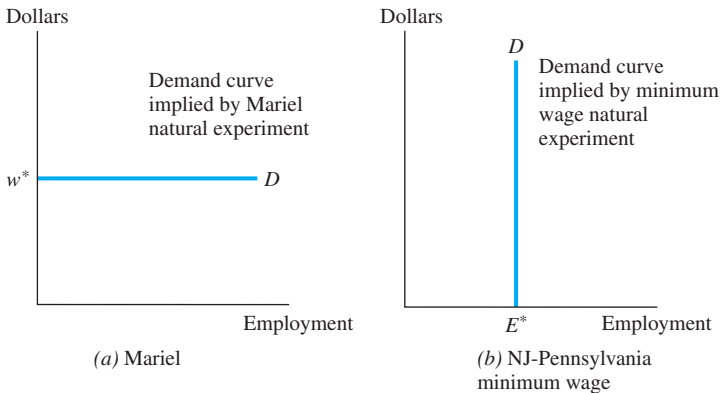
³² William J. Carrington and Pedro de Lima, “The Impact of 1970s Repatriates from Africa on the Portuguese Labor Market,” *Industrial and Labor Relations Review* 49 (January 1996): 330–347. For other international evidence, see Jörn-Steffen Pischke and Johannes Velling, “Employment Effects of Immigration to Germany: An Analysis Based on Local Labor Markets,” *Review of Economics and Statistics* 79 (November 1997): 594–604; Rachel M. Friedberg, “The Impact of Mass Migration on the Israeli Labor Market,” *Quarterly Journal of Economics* 116 (November 2001): 1373–1408; Joshua D. Angrist and Adriana D. Kugler, “Protective or Counter-Productive? European Labor Market Institutions and the Effect of Immigrants on EU Natives,” *Economic Journal* 113 (June 2003): F302–F331; and Brath Erling, Bernt Bratsberg, and Oddbjorn Raaum, “Local Unemployment and the Earnings Assimilation of Immigrants in Norway,” *Review of Economics and Statistics* 88 (May 2006): 243–263.

³³ Joshua D. Angrist and Alan B. Krueger, “Empirical Strategies in Labor Economics,” in Orley C. Ashenfelter and David Card, editors, *Handbook of Labor Economics*, vol. 3A, Amsterdam: Elsevier, 1999, pp. 1277–1366.

³⁴ Moreover, it turns out that the margin of error around this quantity is sufficiently small that the estimate is statistically significantly different from zero.

FIGURE 4-13 The Short-Run Labor Demand Curve Implied by Different Natural Experiments

(a) The analysis of data resulting from the Mariel natural experiment implies that increased immigration does not affect the wage, so that the short-run labor demand curve is perfectly elastic. (b) The analysis of data resulting from the NJ-Pennsylvania minimum wage natural experiment implies that an increase in the minimum wage does not affect employment, so that the short-run labor demand curve is perfectly inelastic.



If one interprets this finding in the traditional way, it would seem to suggest that a phantom immigrant flow greatly harmed the economic opportunities of black workers. This evidence obviously raises some questions about whether one should interpret the evidence for the Mariel boatlift that *did* happen as indicating that immigration had little impact on Miami's labor market.

In addition to these interpretation difficulties, the natural experiment approach sometimes leads to contradictory evidence—contradictions that cannot be easily resolved. For example, suppose we take the results from the original Mariel study at face value, so that we infer that immigration had little impact on the wage of native workers—even in the short run. Figure 4-13a illustrates the short-run labor demand curve implied by the Mariel study. It is a perfectly elastic curve, indicating that wages are constant regardless of the level of labor supply.

In Chapter 3, we discussed an equally famous natural experiment study that attempted to measure the impact of the minimum wage on employment in the fast-food industry.³⁵ This empirical exercise compared employment in New Jersey and Pennsylvania prior to and after the imposition of a state minimum wage in New Jersey. Since the minimum wage increased only in New Jersey, one would have expected that fast-food employment in New Jersey would have declined relative to fast-food employment in Pennsylvania. In fact, the data resulting from this natural experiment showed that no such employment decline occurred in New Jersey as a result of the increase in the minimum wage—relative to the Pennsylvania control group.

³⁵ David Card and Alan B. Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review* 84 (September 1994): 772–793.

Suppose again we take the results from the New Jersey–Pennsylvania minimum wage natural experiment at face value. We can then infer that minimum wages have little impact on employment. Figure 4-13*b* illustrates the short-run labor demand curve implied by this natural experiment. It is a perfectly inelastic curve, indicating that employment is essentially constant regardless of the level of the wage.

Needless to say, at least one of these two demand curves must be wrong. The short-run labor demand curve cannot be both perfectly elastic and perfectly inelastic at the same time. One could perhaps argue that the data are the data—and that in a particular time and in a particular context that is what the labor demand curve looked like. Unfortunately, this approach makes the inferences from experimental evidence completely useless—as the evidence cannot then be used to predict what would happen as a result of policy shifts at other times and in other contexts.

Even more disturbing is the fact that there is an intimate connection in the type of data analysis carried out by the two specific natural experiments in question. In particular, let Δw be the change in the wage before and after the “shock” and ΔE be the corresponding change in employment. In the Mariel context, for instance, the research strategy is to essentially estimate a regression model of the following type:

$$\Delta w = \alpha \Delta E + \text{Other variables} \quad (4-6)$$

In other words, the strategy is to use data from different regions to estimate the relationship between the change in the wage over a particular time period and the corresponding immigration-induced change in supply. The key result of the Mariel study is that, essentially, there is zero correlation between the dependent and independent variables, so that the coefficient α is nearly zero. This zero correlation leads to the inference that immigration-induced changes in supply have little impact on wages.

Consider now the regression model estimated in the New Jersey–Pennsylvania minimum wage experiment:

$$\Delta E = \beta \Delta w + \text{Other variables} \quad (4-7)$$

In other words, the research strategy is to relate changes in employment to changes in the wage across regions. The key result of the minimum wage natural experiment is that there is a zero correlation between employment and (minimum-wage-induced) wage changes across regions, so that the coefficient β is essentially zero. This result is then used to infer that an increase in the minimum wage has little effect on employment.

The core empirical finding in these two natural experiments is that there is little correlation between wage changes and employment changes across different geographic areas. In one experiment (i.e., the Mariel case), this zero correlation is interpreted as indicating that immigration has no effect on wages, while in the other experiment, this same zero correlation is interpreted as indicating that minimum wages have no effect on employment. As Figure 4-13 shows, however, these two interpretations contradict each other.

In short, the evidence from “natural experiments” should be interpreted with a great deal of caution. Not only does the interpretation of the evidence depend on the importance of properly defining the “treatment” and “control” groups, but it is also important to determine whether such results are internally consistent with any underlying theoretical framework.

Do Natives Respond to Immigration?

The fact that most cross-city studies find little evidence of a sizable adverse impact of immigration on native earnings raises two important questions: Why is the evidence so different from the typical presumption in the debate over immigration policy? And why does the evidence seem to be so inconsistent with the implications of the simplest supply-demand equilibrium model? Huge shifts in supply, like those observed in the Mariel flow or those observed when nearly 10 million immigrants entered the United States during a single decade (as happened in the 1990s), *should* affect the wage level in the labor market. And it is unlikely that the “long run” arrived in Miami after only a couple of years.

An important problem with the conceptual approach that underlies the interpretation of the spatial correlations (that is, Figure 4-10 in the case of perfect substitutes and Figure 4-11 in the case of complements) is that it ignores other responses that might occur in the labor market—even abstracting from the adjustments to the aggregate capital stock. The entry of immigrants into the local labor market may well lower the wage of competing workers and increase the wage of complementary workers initially. Over time, however, natives will likely respond to immigration. After all, it is not in the best interest of native workers to sit idly by and watch immigrants change economic opportunities. All natives now have incentives to change their behavior in ways that take advantage of the altered economic landscape.

Figure 4-14 illustrates the labor markets in two different localities, Los Angeles and Pittsburgh. Initially, the native wage w_0 is the same in both cities, with equilibrium occurring at the intersection of supply curve S_0 and the demand curve in each of the cities (at points P_{LA} and P_{PT} , respectively). Los Angeles then receives an influx of immigrants. Assuming that immigrants and natives are perfect substitutes in production, the supply curve shifts in the Los Angeles market to S_1 and the wage declines to w_{LA} .

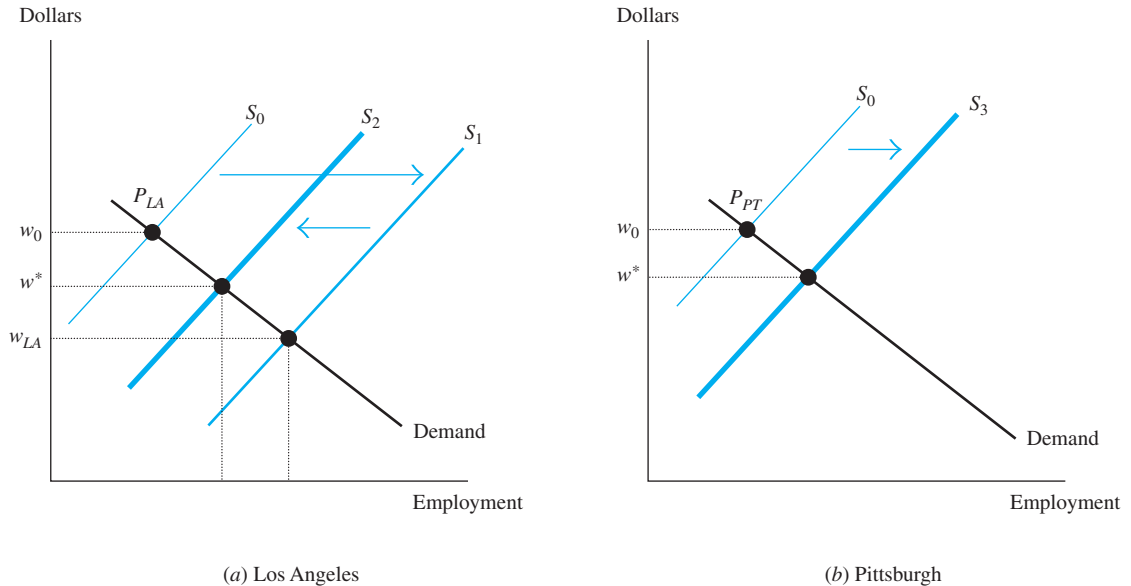
The decline in the equilibrium wage in the Los Angeles labor market is likely to induce some natives to move to Pittsburgh, a city that did not receive an immigrant flow.³⁶ As a result, the supply curve of native workers shifts in both cities. As natives move out of Los Angeles, shifting the supply curve to the left (S_2), the native wage rises slightly to w^* . As the natives move to Pittsburgh, shifting the supply curve in that market to the right (S_3), the wage of natives declines to w^* . If migration between the two cities is costless, natives will migrate until wages are again the same in the two cities. Native migration decisions, therefore, lead to an equilibrium where natives in cities with many immigrants are no worse off than natives in cities with few immigrants. This conclusion, however, disguises the fact that *all* natives, regardless of where they live, are now worse off as a result of immigration.³⁷

³⁶ For simplicity, the argument assumes that immigrants arrive in Los Angeles and remain there.

³⁷ The forces that tend to equalize economic opportunities across labor markets are reinforced by the fact that native-owned firms also will respond. For example, employers see that cities flooded by less-skilled immigrants tend to pay lower wages to less-skilled workers. Employers who demand this type of labor will want to relocate to those cities, and entrepreneurs thinking about starting up new firms will find it more profitable to open them in immigrant areas. In other words, immigration increases the returns to capitalists in the affected cities, and capital will naturally flow to the areas where the returns are the highest. The flow of jobs to the immigrant-hit areas helps cushion the adverse effect of immigration on the wage of competing workers in these localities.

FIGURE 4-14 The Native Labor Market's Response to Immigration

Initially, the two local labor markets are in equilibrium at wage w_0 . The entry of immigrants into Los Angeles shifts the supply curve from S_0 to S_1 and lowers the wage to w_{LA} . The lower wage induces some LA natives to move to Pittsburgh, shifting the supply curve back from S_1 to S_2 and shifting the supply curve in Pittsburgh to S_3 . The markets reestablish equilibrium at wage w^* . All natives earn less as a result of immigration, regardless of where they live.



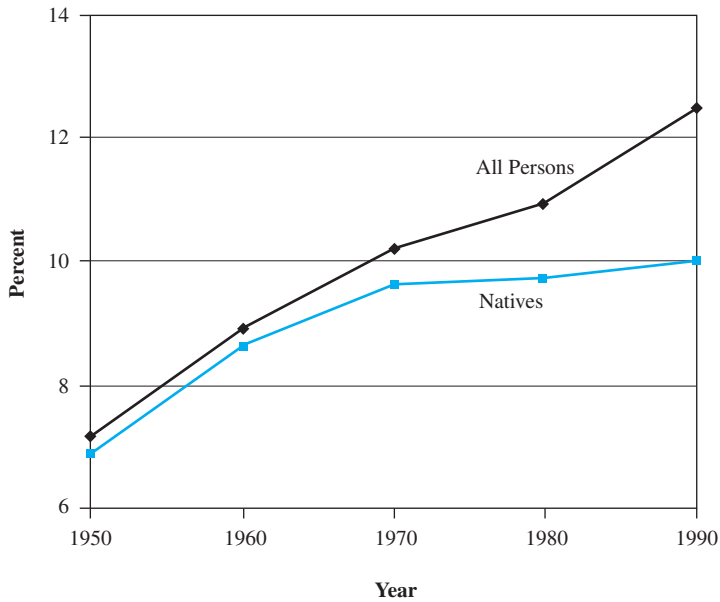
These intercity flows of labor create difficult problems if one wants to measure the labor market impact of immigration by comparing the economic opportunities of native workers in different cities. Using spatial correlations to measure the impact of immigration will not be very revealing because the flows of native-born workers effectively diffuse the impact of immigration throughout the national economy. In the end, all workers who compete with immigrants, regardless of where they live, are worse off because there are now many more such workers. Therefore, as long as natives respond to the entry of immigrants by “voting with their feet,” there is little reason to expect *any* correlation between the earnings of native workers in particular cities and the presence of immigrants. In short, the comparison of local labor markets may be hiding the “macro” impact of immigration.

The evidence on whether native migration patterns are affected by the presence of immigrants is mixed.³⁸ Figure 4-15 presents what is perhaps the most suggestive evidence of a

³⁸ Randall Filer, “The Effect of Immigrant Arrivals on Migratory Patterns of Native Workers,” in George J. Borjas and Richard B. Freeman, editors, *Immigration, Trade, and the Labor Market: Economic Consequences for the United States and Source Areas*, Chicago: University of Chicago Press, 1992, pp. 245–269. Conflicting evidence is presented in the recent studies of David Card, “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration,” *Journal of Labor Economics* (January 2001): 22–64; and George J. Borjas, “Native Internal Migration and the Labor Market Impact of Immigration,” *Journal of Human Resources* 41 (Spring 2006): 221–258.

FIGURE 4-15 Trends in California's population, 1950–1990 (Percent of U.S. Population Living in California)

Source: George J. Borjas, Richard B. Freeman, and Lawrence F. Katz, "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* 1 (1997): 27. The data refer to persons aged 18–64 who are not living in group quarters.



potential relation between immigration and native migration decisions. The resurgence of immigration in the United States began after 1968, when policy changes enacted in 1965 became effective. It seems natural, therefore, to contrast pre-1970 changes in the residential location of the native population with post-1970 changes to assess the effects of immigration on native location decisions.³⁹

Not surprisingly, the share of natives who lived in California, the major immigrant-receiving state, was rising rapidly prior to 1970. What is surprising, however, is that the share of natives living in California barely budged between 1970 and 1990. Nevertheless, California's share of the *total* population kept rising continuously until 1990, from 7 percent in 1950, to 10 percent in 1970, to 12 percent in 1990. Put differently, an extrapolation of the population growth that existed before 1970—*before the resurgence of immigration*—would have accurately predicted the state's 1990 share of the population. But whereas natives pouring into the state fueled California's population growth before 1970, immigrants alone fueled the post-1970 growth.

How should one interpret this fact? One interpretation is that around 1970, for reasons unknown, Americans simply stopped moving to California. In other words, if it were not for immigration, California's rapid population growth would have stalled in the 1970s and 1980s. An alternative—and more controversial—interpretation is that immigration into

³⁹ George J. Borjas, Richard B. Freeman, and Lawrence F. Katz, "How Much Do Immigration and Trade Affect Labor Market Outcomes?" *Brookings Papers on Economic Activity* 1 (1997): 1–67.

California essentially “displaced” the population growth that *would have occurred* in the immigrants’ absence, and this displacement effectively diffused the economic impact of immigration from California to the rest of the country.⁴⁰

Immigration and the Wage Structure

The possibility that comparisons of local labor markets do not provide valuable information about the economic impact of immigration has motivated some researchers to search for this impact by looking at the evolution of the national wage structure. A recent study analyzes the wage growth experienced by native workers belonging to groups classified in terms of educational attainment and years of work experience, and attempts to see if the wage growth experienced by these skill groups is related to the growth in the number of immigrants in the various groups.⁴¹ Put differently, the empirical exercise includes fixed effects for each skill group so that the impact of immigration on wages is being measured by “differencing” the data within each skill group.

Figure 4-16 summarizes the evidence. Each point in the scatter diagram relates the wage growth experienced by a skill group of native working men over a particular decade between 1960 and 2000 to the change in the percent of the number of workers in the group that are foreign born. There is an obvious negative correlation between the two variables. At the national level, therefore, wages grew fastest for those skill groups least affected by immigration. In fact, the data suggest that wages fall by 3 to 4 percent if immigration increases the number of workers in a skill group by 10 percent.

The national-level approach has been expanded to estimate a full-blown model that specifies the aggregate production functions linking output, capital, and the various skill groups. The structural approach typically uses the immigrant supply shock as the instrument that shifts the supply curve and that identifies the labor demand function. One benefit from this structural approach—as opposed to the simple estimation of correlations implied by the regression line in Figure 4-16—is that it allows us to estimate how the wages of a particular skill group of native workers (e.g., native college graduates) are affected by the immigration of, say, those who are high school dropouts. One can then use the own-elasticities and the cross-elasticities to simulate the impact of a particular immigrant influx on the U.S. wage structure.

⁴⁰ Recent evidence suggests that internal migration by “natives” also helped to equilibrate the labor market during the Great Depression. In the aftermath of the economic upheaval, some geographic areas began to receive a large number of in-migrants. It turns out that for every 10 new arrivals, two pre-existing residents moved out, two were unable to find a relief job, and two moved from full-time to part-time work; See Leah Platt Boustan, Price V. Fishback, and Shawn Kantor, “The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression,” *Journal of Labor Economics* 28 (October 2010): 719–746. A related study of how internal migration equilibrates wages in the UK context is given by Timothy J. Hatton and Massimiliano Tani, “Immigration and Inter-regional Mobility in the UK, 1982–2000,” *Economic Journal* 115 (November 2005): F342–F358. A recent study by Nicole Fortin, “Higher Education Policies and the College Wage Premium: Cross-State Evidence from the 1990s,” *American Economic Review* 96 (September 2006): 959–987, notes that interstate migration attenuates the measured impact of state-specific changes in the size of the high-skill population on the wage gap between college and less-educated workers in the state.

⁴¹ George J. Borjas, “The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration in the Labor Market,” *Quarterly Journal of Economics* 118 (November 2003): 1335–1374.

FIGURE 4-16 Scatter Diagram Relating Wages and Immigration for Native Skill Groups Defined by Educational Attainment and Work Experience, 1960–2000

Source: George J. Borjas, “The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration in the Labor Market,” *Quarterly Journal of Economics* 118 (November 2003): 1335–1374. Each point in the scatter represents the decadal change in the log weekly wage and the immigrant share (that is, the percent of immigrants in the workforce) for a native group of working men defined by years of education and work experience. The slope of the regression line is $-.450$, with a standard error of $.172$.

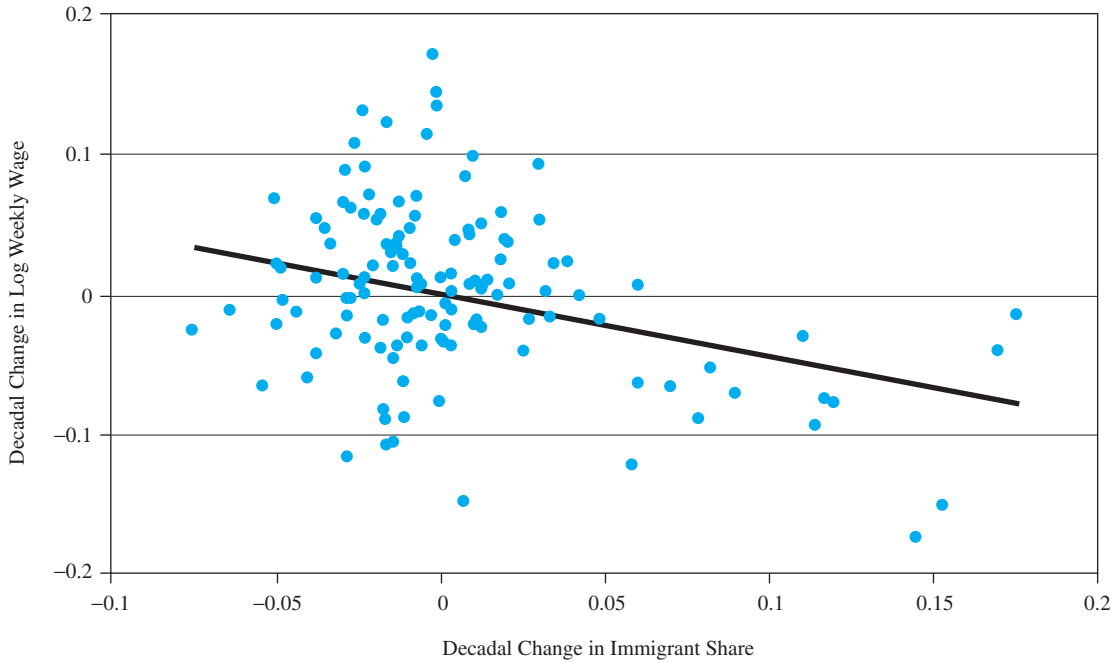


Table 4-3 summarizes the results from an influential study that uses this structural approach. Even after accounting for all the cross-effects of supply shifts on the wages of the various skill groups, the 1980–2000 immigrant influx lowered the wage of the typical worker in the United States by 3.4 percent in the short run. As implied by our theoretical analysis of long-run impacts, the predicted long-run impact must be 0.0 percent, since the typical worker in the economy is unaffected by immigration once all the capital adjustments take place. Note, however, that immigration has distributional effects even in the long run, with

TABLE 4-3 The Wage Impact of the 1980–2000 Immigrant Influx

Source: George J. Borjas and Lawrence F. Katz, “The Evolution of the Mexican-Born Workforce in the U.S. Labor Market,” in George J. Borjas, editor, *Mexican Immigration to the United States*, Chicago: University of Chicago Press, 2007.

| | Short Run | Long Run |
|-----------------------|-----------|----------|
| All native workers | -3.4% | 0.0% |
| High school dropouts | -8.2 | -4.8 |
| High school graduates | -2.2 | 1.2 |
| Some college | -2.7 | 0.7 |
| College graduates | -3.9 | -0.5 |

the average wage of high school dropouts falling by about 5 percent and the average wage of workers in the middle of the education distribution increasing slightly.⁴²

The national labor market approach also has been used to examine the link between migration and the wage structure of other countries. One particularly interesting case study examines the link between *emigration* and wages in Mexico.⁴³ Emigration (almost entirely to the United States) has quickly drained the Mexican labor market of about 10 percent of its workforce. The laws of supply and demand suggest that these labor outflows should increase wages in Mexico. As predicted by the theory, there is indeed a strong *positive* correlation between the number of emigrants in a particular skill group (again defined by education and labor market experience) and the wage growth experienced by that group. An outflow that reduces the number of workers in a skill group by 10 percent raises the wage of the workers who remained in Mexico by about 3 percent.

4-6 The Economic Benefits from Immigration

We have seen that immigrants may have an adverse impact on the job opportunities of the native workers whose skills resemble those of the immigrants. Immigrants also can make an important contribution to the receiving country. To assess the net economic impact of immigration, we must calculate the magnitude of these contributions. It turns out that there is an intimate link between the elasticity that measures the wage impact of immigration on the native workforce and the magnitude of the gains that accrue to receiving countries.

Consider the short-run supply-demand analysis presented in Figure 4-17. The supply curve of labor is given by S and the demand curve for labor is given by D . For simplicity, we assume that the labor supply curve is inelastic, so that there are N native-born workers. Competitive market equilibrium implies that the N native workers are employed at a wage of w_0 .

Recall that the labor demand curve is given by the value of marginal product schedule, so that each point on the demand curve tells us the contribution of the last worker hired. As a

⁴² Research has estimated this type of structural model by allowing for the possibility that immigrants and natives within narrowly defined skill groups are complements in production. In other words, the entry of immigrants who are, say, high school dropouts and are around 30 years old may raise the productivity of native workers who are high school dropouts and are also around 30 years old. Gianmarco Ottaviano and Giovanni Peri, "Rethinking the Effects of Immigration on Wages," NBER Working Paper No. 12497, August 2006, reports the existence of such complementarities, implying that it is possible for immigrants to raise the average wage of natives even in the long run. However, the replication study by George J. Borjas, Jeffrey Grogger, and Gordon Hanson, "Imperfect Substitution between Immigrants and Natives: A Reappraisal," NBER Working Paper No. 13887, March 2008, shows that the Ottaviano-Peri evidence is determined by their classification of currently enrolled high school students (mostly juniors and seniors) as "high school dropouts." Once these high school students are excluded from the analysis, the evidence supporting the existence of complementarities between comparably skilled immigrants and natives disappears. Recent research also explores how skill groups should be defined in the context of immigration. It turns out that if high school dropouts and high school graduates were perfect substitutes, the labor market impact of immigration on the wage structure is much smaller. See Borjas, Freeman, and Katz, "How Much Do Immigration and Trade Affect Labor Market Outcomes?"; and David Card, "Immigration and Inequality," *American Economic Review* 99 (May 2009): 1–21.

⁴³ Prachi Mishra, "Emigration and Wages in Source Countries: Evidence from Mexico," *Journal of Development Economics* 82 (January 2007): 180–199. See also Abdurrahman Aydemir and George J. Borjas, "A Comparative Analysis of the Labor Market Impact of Immigration: Canada, Mexico, and the United States," *Journal of the European Economic Association* 5 (June 2007): 663–708.

Theory at Work

THE GREAT BLACK MIGRATION

The economic forces that lead to wage equalization across countries as workers leave one market and enter another apply equally well to internal migration flows within a country. Presumably, workers leave low-wage areas and move to high-wage areas, and these flows should help to equilibrate the labor market.

One particularly interesting application of the national labor market approach used to measure the wage impacts of immigration is a recent study that estimates the economic impact of the large black migration from southern to northern states in the middle of the twentieth century. Between 1940 and 1970, over 4 million blacks left the rural South and moved to the industrialized North. In 1940, more than three-quarters of blacks lived in the South. By 1970, only half of blacks lived in the South. This historic “diaspora” of southern blacks altered the course of American history and fundamentally changed race relations in the United States.

The Great Black Migration had inevitable and predictable economic effects. As a result of this labor flow, the size of the typical skill group of workers in the North (defined by education and labor market experience) increased by 5 percent between 1940 and 1960. This 5 percent increase in labor supply reduced the annual earnings of northern black men (relative to those of whites) by about 3 percent. The sizable influx of black workers to the North, in effect, improved the economic well-being of the migrants at a cost: a delay in the rate of earnings convergence between black and white workers in the North.

Sources: Leah Platt Boustan, “Competition in the Promised Land: Blacks, Migration, and Northern Labor Markets, 1940–1970,” *Journal of Economic History* 69 (September 2009): 755–782. A good account of this historic migration is given by Nicholas Lemann, *The Promised Land: The Great Black Migration and How It Changed America*, New York: Vintage Books, 1991.

result, the area under the demand curve gives the total product of all workers hired. Hence, the area in the trapezoid $ABNO$ measures the value of national income prior to immigration.

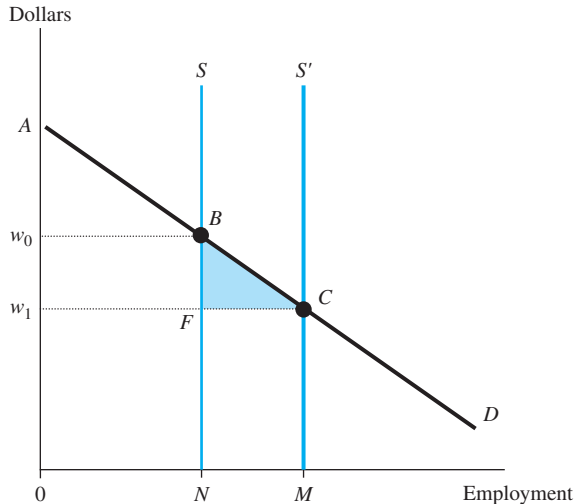
What happens to national income when immigrants enter the country? If we assume that immigrants and natives are perfect substitutes in production, the supply curve shifts to S' and the market wage falls to w_1 . National income is now given by the area in the trapezoid $ACMO$. The figure shows that the total wage bill paid to immigrants is given by the area in the rectangle $FCMN$, so that the increase in national income accruing to natives is given by the area in the triangle BCF . This triangle is the **immigration surplus** and measures the increase in national income that occurs as a result of immigration and that accrues to natives.

Why does an immigration surplus arise? Because the market wage equals the productivity of the *last* immigrant hired. As a result, immigrants increase national income by more than what it costs to employ them. Put differently, all the immigrants hired except for the last one contribute more to the economy than they get paid.

The analysis in Figure 4-17 implies that if the demand curve is perfectly elastic (so that immigrants had no impact on the native wage rate), immigrants would be paid their entire value of marginal product and natives would gain nothing from immigration. Therefore, the immigration surplus exists *only* if native wage rates fall when immigrants enter the country. Therefore, immigration redistributes income from labor to capital. In terms of Figure 4-17, native workers lose the area in the rectangle w_0BFw_1 , and this quantity plus the immigration surplus accrue to employers. Although native workers get a lower wage, these losses are more than offset by the increase in income accruing to native-owned firms.

FIGURE 4-17 The Immigration Surplus

Prior to immigration, there are N native workers in the economy and national income is given by the trapezoid $ABN0$. Immigration increases the labor supply to M workers and national income is given by the trapezoid $ACM0$. Immigrants are paid a total of $FCMN$ dollars as salary. The immigration surplus gives the increase in national income that accrues to natives and is given by the area in the triangle BCF .



Calculating the Immigration Surplus

Recall that the formula for the area of the triangle is one-half times the base times the height. Figure 4-17 then implies that the dollar value of the immigration surplus is given by

$$\text{Immigration surplus} = \frac{1}{2} \times (w_0 - w_1) \times (M - N) \quad (4-8)$$

This formula can be rewritten so as to obtain the immigration surplus as a fraction of national income. After rearranging the terms in the equation, we get⁴⁴

$$\begin{aligned} \frac{\text{Immigration surplus}}{\text{National income}} &= \frac{1}{2} \times (\% \text{ change in native wage rate}) \\ &\quad \times (\% \text{ change in employment}) \\ &\quad \times (\text{labor's share of national income}) \quad (4-9) \end{aligned}$$

where labor's share of national income is the fraction of national income that accrues to workers.

Immigrants have increased labor supply by about 10 percent in the United States. Our discussion in the previous section indicated that a 10 percent immigrant-induced increase in supply lowers the wage by about 3 to 4 percent. Finally, it is well known that labor's

⁴⁴ In particular, we can rewrite the immigration surplus as

$$\frac{\text{Immigration surplus}}{\text{National income}} = \frac{1}{2} \times \frac{w_0 - w_1}{w_1} \times \frac{M - N}{M} \times \frac{w_1 M}{\text{National income}}$$

share of national income is on the order of 0.7. This implies that immigration increases the real income of natives by only about 0.13 percent (or $0.5 \times 0.035 \times 0.10 \times 0.7$). The gross domestic product (GDP) of the United States is around \$14 trillion, so the economic gains from immigration are relatively small, about \$18 billion per year.⁴⁵

It is worth reemphasizing that this estimate of the immigration surplus is a short-run estimate. In the long run, neither the rate of return to capital nor the wage is affected by immigration. As a result, the long-run immigration surplus must be equal to zero. Immigrants increase GDP in the long run, but the entire increase in national income is paid to immigrants for their services. Ironically, in a constant-returns-to-scale economy, the economic benefits from immigration can only arise when workers in the receiving country are hurt by immigration. Equally important, *the larger the adverse wage effects, the greater the economic benefits.*

4-7 Policy Application: Hurricanes and the Labor Market

Hurricanes can be very destructive, in terms of both casualties and property damage.⁴⁶ Hurricanes develop over warm water, where the ocean's temperature exceeds 80 degrees Fahrenheit. As a result, hurricane season runs from June through November. Due to the high temperatures required to fuel the storm, most hurricanes that strike the United States first touch land in the states that surround the Gulf of Mexico or the Southeastern states, particularly Florida. In fact, all 67 counties of the state of Florida experienced some type of hurricane damage between 1988 and 2005. The hurricane threat during those years was remarkable because five of the six most damaging Atlantic hurricanes of all time hit Florida in this period.

On average, the state of Florida is typically hit by one to two hurricanes each year. Table 4-4 lists the 19 hurricanes that hit Florida between 1988 and 2005 and reports some of the key characteristics of the various hurricanes. There is clearly a lot of variation in the extent of damage unleashed by the storms. Hurricanes are categorized according to the Saffir-Simpson Scale based on their wind speed and are given a number ranging from 1 to 5. Category 1 hurricanes have wind speeds ranging from 74 to 95 miles per hour at the time of landfall. Category 2 hurricanes have wind speeds from 96 to 110 miles per hour; category 3 hurricanes have wind speeds between 111 and 130 miles per hour; and category 4 hurricanes have wind speeds between 131 and 155 miles per hour. Andrew, a category 5 hurricane, had wind speeds above 180 miles per hour when it first hit land.

Although we can predict with confidence that the hurricane season will generate some hurricanes and that Florida will likely be hit by some of these hurricanes during the course

⁴⁵ George J. Borjas, "The Economic Benefits from Immigration," *Journal of Economic Perspectives* 9 (Spring 1995): 3–22; and George E. Johnson, "Estimation of the Impact of Immigration on the Distribution of Income among Minorities and Others," in Daniel S. Hamermesh and Frank D. Bean, editors, *Help or Hindrance? The Economic Implications of Immigration for African-Americans*, New York: Russell Sage Press, 1998, pp. 17–50.

⁴⁶ The discussion in this section is based on Ariel R. Belasen and Solomon W. Polachek, "How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida," *Journal of Human Resources* 44 (Winter 2009): 251–276.

TABLE 4-4 Hurricanes Hitting Florida between 1988 and 2005

Source: Ariel R. Belasen and Solomon W. Polachek, “How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida,” *Journal of Human Resources* 44 (Winter 2009), Table 1.

| Hurricane | Category | Monetary Damage to Florida (millions) | Number of Deaths in Florida | Windspeed at Landfall (mph) | Rainfall (inches) |
|-----------------|----------|---------------------------------------|-----------------------------|-----------------------------|-------------------|
| Florence (1988) | 1 | \$0.6 | 0 | 75 | 5–10 |
| Andrew (1992) | 5 | \$43,000 | 44 | 175 | 5–7 |
| Allison (1995) | 1 | \$1.2 | 0 | 75 | 4–6 |
| Erin (1995) | 1 | \$0.5 | 6 | 87 | 5–12 |
| Opal (1995) | 3 | \$4,400 | 1 | 115 | 5–10 |
| Danny (1997) | 1 | \$100 (total to U.S.) | 0 | 80 | 2–7 |
| Earl (1998) | 1 | \$64.5 | 2 | 92 | 6–16 |
| Georges (1998) | 2 | \$392 | 0 | 103 | 8–25 |
| Irene (1999) | 1 | \$1,100 | 8 | 75 | 10–20 |
| Gordon (2000) | 1 | \$11.9 | 1 | 75 | 3–5 |
| Charley (2004) | 4 | \$15,100 | 29 | 150 | 5–8 |
| Frances (2004) | 2 | \$8,900 | 37 | 105 | 10–20 |
| Ivan (2004) | 2 | \$8,100 | 19 | 130 | 7–15 |
| Jeanne (2004) | 3 | \$6,900 (total to U.S.) | 3 | 121 | 8–13 |
| Dennis (2005) | 3 | \$2,200 | 14 | 120 | 10–15 |
| Katrina (2005) | 1 | \$115,000 (total to U.S.) | 14 | 81 | 5–15 |
| Ophelia (2005) | 1 | \$70 (total to U.S.) | 1 | 80 | 3–5 |
| Rita (2005) | 1 | \$10,000 (total to U.S.) | 2 | 80 | 2–4 |
| Wilma (2005) | 3 | \$12,200 | 35 | 120 | 7–12 |

of a decade, the exact timing and path of the hurricanes cannot be forecast. As a result, each of these hurricanes generates exogenous economic shocks to the Florida counties that are directly hit. The randomness of the path and intensity of the hurricane, therefore, provide a “natural experiment” that can be used to analyze how the economic shocks set off by such deadly storms alter labor market conditions. Because so many hurricanes have hit Florida in the past two decades, we can use the available data to estimate difference-in-differences models that examine the economic impact on the affected Florida counties relative to the economic events unfolding in the unaffected counties.

We can use the basic tools of supply and demand to easily describe what one would expect to happen when a hurricane hits a specific Florida county randomly. When a hurricane strikes the county, some people will flee—causing at least a temporary decline in the number of workers available. Of course, the duration of this cut in supply will depend on how deadly the hurricane is expected to be and how extensive the damage, in fact, was. The hurricane-induced shift in the supply curve to the left suggests that wages would rise and employment would fall in the counties directly affected by the hurricane. Many of these “refugees” would be expected to move to neighboring counties at least in the short run. This implies that the supply of labor would increase in these neighboring counties, and that the wage may actually fall (and employment increase) in these neighboring counties.

TABLE 4-5 Changes in Employment and Wages in Florida Counties hit by Hurricanes (relative to the change observed in the average Florida county)

Source: Ariel R. Belasen and Solomon W. Polachek, “How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida,” *Journal of Human Resources* 44 (Winter 2009), Table 4.

| | Percent Change in Employment | Percent Change in Earnings |
|--|------------------------------|----------------------------|
| 1. Effect of category 1–3 hurricane on county directly hit | –1.5 | +1.3 |
| 2. Effect of category 4–5 hurricane on county directly hit | –4.5 | +4.4 |
| 3. Effect of category 1–3 hurricane on neighboring county | +0.2 | –4.5 |
| 4. Effect of category 4–5 hurricane in neighboring county | +0.8 | –3.3 |

The hurricane shock also may affect the county’s labor demand curve, but it is harder to ascertain how this curve would shift. On the one hand, some firms might leave town alongside the workers, so that there would be a cutback in labor demand. On the other hand, if the hurricane destroyed a lot of the infrastructure, physical capital, and property, the reconstruction would likely shift the labor demand outwards, as firms expanded to speed up the rebuilding process.

In short, the effect of hurricanes on the labor market will depend on the relative strengths of the shift in labor demand and labor supply. Table 4-5 summarizes the key results from a careful study of the economic consequences of the 19 hurricanes that hit Florida between 1988 and 2005. The evidence seems consistent with a simple story that labor supply induced by the hurricane led to corresponding employment and wage shifts both in the county directly hit by the hurricane, as well as in surrounding counties. The wage rises in those counties that are hit by the hurricane, with the rise being stronger in counties that are hit by stronger hurricanes—suggesting that the exodus of workers is larger when the hurricane is more destructive. In fact, the wage rises by about 4 percent when a county is hit by a category 4 or 5 hurricane (relative to the wage change observed in the average Florida county at the same time). At the same time, the wage falls by a numerically similar amount in the neighboring counties—as the “surplus” labor moving to those counties increases the number of workers available.

It is worth noting that this approach to the study of data generated by natural experiments differs markedly from our earlier discussion of the impact of the Mariel supply shock or the New Jersey minimum wage increase. In each of these earlier cases, there is but *one* natural experiment to be analyzed.⁴⁷ This would be akin to injecting a particular (randomly chosen) person in the population with an experimental medicine and then comparing this person’s reaction to that of the typical noninjected person. Clearly, such a comparison may be largely driven by idiosyncratic factors—for example, the randomly chosen person just happens to be allergic to some of the chemicals in the medicine, or he had the beginnings of a cold when the injection took place. By analyzing the mean outcome of a large number of

⁴⁷ See Stephen G. Donald and Kevin Lang, “Inference with Differences-in-Differences and Other Panel Data,” *Review of Economics and Statistics* 89 (May 2007): 221–233, for a statistical discussion of this issue.

natural experiments, these idiosyncratic factors get “washed out.” As a result, the study of the average consequence of a large number of natural experiments may yield more credible estimates of the labor market consequences of particular shocks.

4-8 The Cobweb Model

Our analysis of labor market equilibrium assumes that markets adjust instantaneously to shifts in either supply or demand curves, so that wages and employment change swiftly from the old equilibrium levels to the new. Many labor markets, however, do not adjust so quickly to shifts in the underlying supply and demand curves. There is some evidence, in fact, that markets for highly skilled workers, such as engineers and other specialized professionals, exhibit systematic periods of booms and busts that dispute the notion that labor markets attain competitive equilibrium quickly and cheaply.

Consider, for example, the market for new engineering graduates. It has long been recognized that the market for newly minted engineers fluctuates regularly between periods of excess demand for labor and periods of excess supply. As a result, there is a cyclical trend in the entry wage of engineering graduates over time. In a series of studies, Richard Freeman proposed a model that showed how these trends in the entry wage could be generated.⁴⁸ Two key assumptions underlie the model: (1) It takes time to produce a new engineer and (2) persons decide whether or not to become engineers by looking at conditions in the engineering labor market *at the time they enter school*.

Figure 4-18 presents the supply and demand curves for new engineers. Initially, this entry-level labor market is in equilibrium where the supply curve S intersects the demand curve D , so that there are E_0 new engineering graduates and the entry wage is w_0 . Suppose there is a sudden increase in the demand for newly trained engineers (perhaps as a result of the race to get a man on the moon in the 1960s, or because the United States realizes that it might need a sophisticated system of missile defense in the post-9/11 environment). The demand curve for engineers shifts to D' , and engineering firms would like to hire E^* new engineers at a wage of w^* .

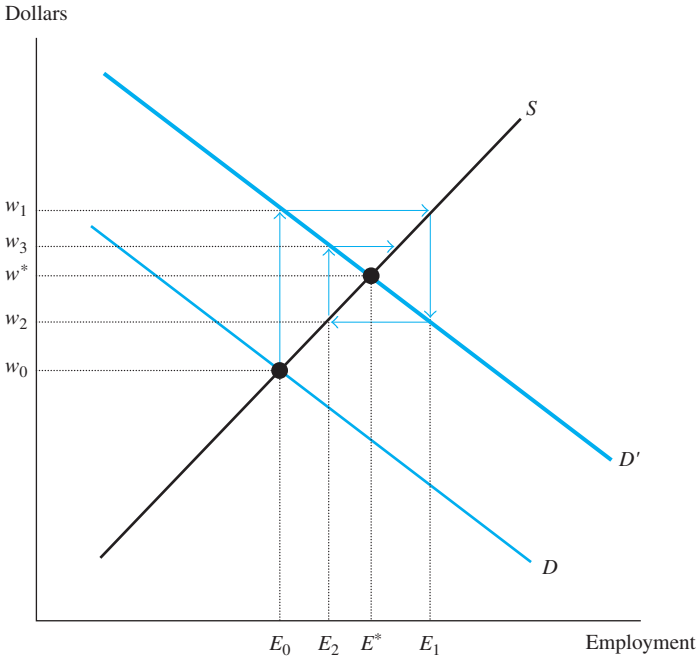
Firms will find it extremely difficult to hire this desired number of new engineers. New engineers do not come out of thin air simply because firms want to hire them. It takes time to train new engineers. Because engineering schools are only producing E_0 engineers annually, the *short-run* supply curve is perfectly inelastic at E_0 workers. The combination of this inelastic supply curve (that is, a vertical line going through E_0 workers) and the demand shift increases the entry wage of engineers to w_1 .

While all this is happening in the engineering labor market, a new generation of high school and college students is deciding whether to enter the engineering profession. These students see a relatively high wage in the engineering market and, hence, have a large incentive to become engineers. In fact, at the current wage of w_1 , a total of E_1 persons will want to enroll in engineering schools.

⁴⁸ Richard B. Freeman, “A Cobweb Model of the Supply and Starting Salary of New Engineers,” *Industrial and Labor Relations Review* 29 (January 1976): 236–246; Richard B. Freeman, “Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973,” *American Economic Review* 65 (March 1975): 27–39; and Richard B. Freeman, *The Overeducated American*, New York: Academic Press, 1976.

FIGURE 4-18 The Cobweb Model in the Market for New Engineers

The initial equilibrium wage in the engineering market is w_0 . The demand for engineers shifts to D' , and the wage will eventually increase to w^* . Because new engineers are not produced instantaneously and because students might misforecast future opportunities in the market, a cobweb is created as the labor market adjusts to the increase in demand.



After a few years, therefore, E_1 new engineers enter the marketplace. At the time in which this cohort of engineers enters the market, the short-run supply of new engineers is again perfectly inelastic at E_1 workers. Hence, the current market situation is summarized by this inelastic supply curve and the demand curve D' (assuming that demand conditions have not changed any further). Equilibrium occurs at a wage of w_2 , which is substantially below the wage that the new engineers thought they were going to get. In effect, high school and college graduates presumed that they would get a wage of w_1 dollars; therefore, there was an oversupply of engineers.

But this is not the end of the story. Still another generation of high school and college students is trying to decide whether to become engineers. At the current low wage of w_2 , the engineering profession does not look very attractive, and, hence, few persons will decide to attend engineering school. The supply curve in Figure 4-18 implies that at a wage of w_2 only E_2 persons become engineers. When these students graduate and enter the labor market, the entry wage rises to w_3 because there was an undersupply of engineers. This high wage induces the next generation of students to oversupply the marketplace, and so on.

The analysis illustrates the **cobweb** that is created around the equilibrium point as the engineering labor market adjusts to the initial demand shock. The entry wage exhibits a

systematic pattern of booms and busts as the market slowly drifts toward its long-run equilibrium wage w^* and employment E^* .⁴⁹

The Underlying Assumptions of the Cobweb Model

The cobweb model makes two key assumptions. The first is reasonable: it does take time to produce new engineers, so the supply of engineers can be thought of as being perfectly inelastic in the short run. The second is more questionable. In essence, the model assumes that students are very myopic when they are considering whether to become engineers. Students choose an engineering career based entirely on the wage they currently observe in the engineering market and do not attempt to “look into the future” when comparing their various alternatives. Potential engineers have very strong incentives to be well informed about the trends in the wage of newly minted engineers. If they knew these trends, they could easily deduce what would happen to them when their cohort enters the market. In fact, even if many of these students did not bother collecting all the relevant information, *someone would!* The information could then be sold to students, who would be willing to pay for valuable information regarding their future wage prospects.

The cobwebs are generated, in effect, because the students are misinformed. They do not fully take into account the history of wages in the engineering labor market when choosing a career. Students who do take into account the entire history of wages are said to have **rational expectations**. If students had rational expectations, they would be much more hesitant to enter the engineering labor market when current wages are high and much more willing to enter when current wages are low. As a result, the cobweb might unravel.

The evidence provides strong support of cobwebs in many professional markets, so it seems as if students systematically misforecast future earnings opportunities.⁵⁰ It is worth noting, however, that students are not alone in misforecasting the future. There is some evidence that even professionals tend to have difficulty predicting future earnings opportunities.⁵¹ The inherent uncertainty in forecasting the future might force students to place too heavy a weight on the wages they currently observe, and thus generate cobwebs in professional labor markets.

4-9 Noncompetitive Labor Markets: Monopsony

Up to this point, we have analyzed the characteristics of labor market equilibrium in competitive markets. Each firm in the industry faces the same competitive price p when trying to sell its output, regardless of how much output it sells. Moreover, each firm in the

⁴⁹ Although our analysis indicates that wages and employment in the engineering market drift toward their equilibrium levels over time, depending on the values of the elasticities of supply and demand, the cobweb model can generate booms and busts where wages and employment diverge *away* from equilibrium.

⁵⁰ Evidence on how students forecast the wages of their future professions is provided in Julian R. Betts, “What Do Students Know about Wages? Evidence from a Study of Undergraduates,” *Journal of Human Resources* 31 (Winter 1996): 27–56; and Jeff Dominitz and Charles F. Manski, “Eliciting Student Expectations of the Returns to Schooling,” *Journal of Human Resources* 31 (Winter 1996): 1–26.

⁵¹ Jonathan Leonard, “Wage Expectations in the Labor Market: Survey Evidence on Rationality,” *Review of Economics and Statistics* 64 (February 1982): 157–161.

industry pays a constant wage w to all workers, regardless of how many workers it hires. We now begin the study of the properties of labor market equilibrium under alternative market structures.

A **monopsony** is a firm that faces an upward-sloping supply curve of labor.⁵² In contrast to a competitive firm that can hire as much labor as it wants at the going price, a monopsonist must pay higher wages in order to attract more workers. The one-company town (for example, a coal mine in a remote location) is the stereotypical example of a monopsony. The only way the firm can convince more townspeople to work is to raise the wage so as to meet the reservation wage of the nonworkers.

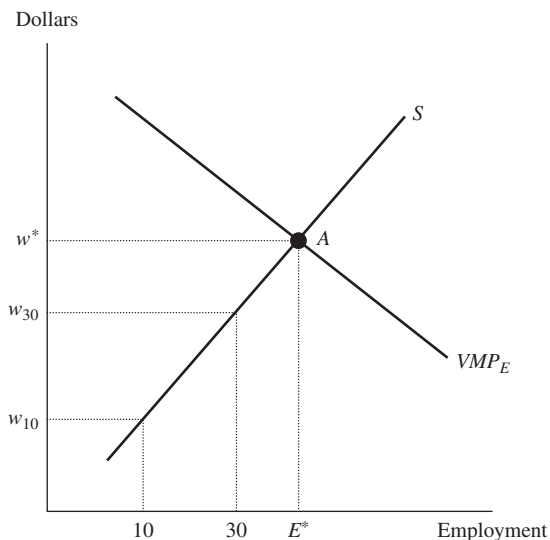
Although it is tempting to dismiss the relevance of the monopsony model because one-company towns are rare in a modern and mobile industrialized economy, it turns out that a particular firm may have an upward-sloping supply curve—the key feature of a monopsony—even when it faces a great deal of competition in the labor market. The circumstances that give rise to upward-sloping supply curves for seemingly competitive firms will be discussed in detail below.

Perfectly Discriminating Monopsonist

We consider two types of monopsonistic firms: a *perfectly discriminating* monopsony and a *nondiscriminating* monopsony. Consider first the case of a perfectly discriminating monopsony. Figure 4-19 illustrates the labor market conditions faced by this firm. As noted

FIGURE 4-19 The Hiring Decision of a Perfectly Discriminating Monopsonist

A perfectly discriminating monopsonist faces an upward-sloping supply curve and can hire different workers at different wages. The labor supply curve gives the marginal cost of hiring. Profit maximization occurs at point A . The monopsonist hires the same number of workers as a competitive market, but each worker gets paid his reservation wage.



⁵² A detailed conceptual and empirical examination of the monopsony model is given by Alan Manning, "A Generalised Model of Monopsony," *Economic Journal* 116 (January 2006): 84–100.

above, the monopsonist faces an upward-sloping labor supply curve. In addition, a *perfectly discriminating monopsonist can hire different workers at different wages*. In terms of the labor supply curve in the figure, this monopsonist need only pay a wage of w_{10} dollars to attract the 10th worker, and must pay a wage of w_{30} to attract the 30th worker. As a result, the supply curve of labor is identical to the marginal cost of hiring labor.

Because a monopsonist cannot influence prices in the output market, it can sell as much as it wants of the output at a constant price p . The revenue from hiring an extra worker equals the price times the marginal product of labor, or the value of marginal product. Hence, the labor demand curve for the monopsonist, as for a competitive firm, is given by the value of marginal product curve.

Regardless of whether firms operate in a competitive market or not, a *profit-maximizing firm should hire workers up to the point where the dollar value of the last worker hired equals the cost of hiring that last worker*. A perfectly discriminating monopsonist will then hire up to the point where the last worker's contribution to firm revenue (or VMP_E) equals the marginal cost of labor. Put differently, market equilibrium occurs at point A , where supply equals demand. The perfectly discriminating monopsonist hires E^* workers, exactly the same employment level that would have been observed if the labor market were competitive. The wage w^* , however, is *not* the competitive wage. Rather, it is the wage that the monopsonist must pay to attract the last worker hired. All other workers receive lower wages, with each worker receiving his or her reservation wage.

Nondiscriminating Monopsonist

A *nondiscriminating monopsonist must pay all workers the same wage, regardless of the worker's reservation wage*. Because the nondiscriminating monopsonist must raise the wage to all workers when he wishes to hire one more worker, the labor supply curve no longer gives the marginal cost of hiring. The numerical example in Table 4-6 illustrates this point. At a wage of \$4, no one is willing to work. At a wage of \$5, the firm attracts one worker, total labor costs equal \$5, and the marginal cost of hiring that worker is \$5. If the firm wishes to hire two workers, it must raise the wage to \$6. Total labor costs then equal \$12, and the marginal cost of hiring the second worker increases to \$7. As the firm expands, therefore, it incurs an ever-higher marginal cost.

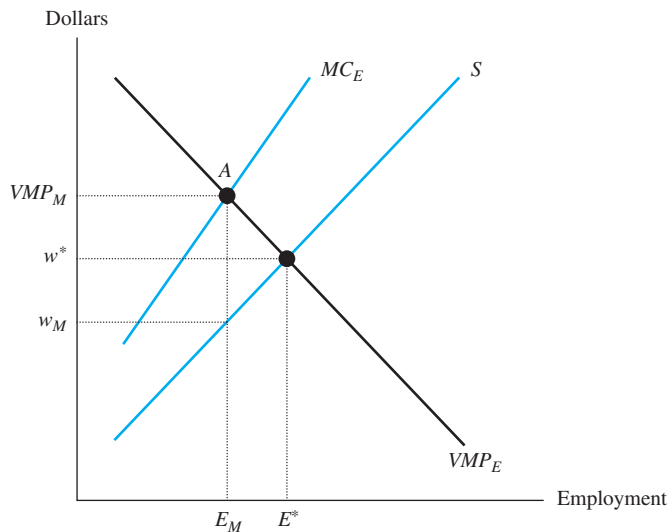
Figure 4-20 illustrates the relation between the labor supply curve and the marginal cost of labor curve for a nondiscriminating monopsonist. Because wages rise as the monopsonist tries to hire more workers, the marginal cost of labor curve (MC_E) is upward sloping, rises even faster than the wage, and lies above the supply curve. As we have seen, the marginal cost of hiring involves not only the wage paid to the additional worker but also

TABLE 4-6
Calculating
the Marginal
Cost of Hiring
for a Non-
discriminating
Monopsonist

| Wage (w) | Number of Persons Willing to Work at That Wage (E) | $w \times E$ | Marginal Cost of Labor |
|--------------|--|--------------|------------------------|
| \$4 | 0 | \$0 | — |
| 5 | 1 | 5 | \$5 |
| 6 | 2 | 12 | 7 |
| 7 | 3 | 21 | 9 |
| 8 | 4 | 32 | 11 |

FIGURE 4-20 The Hiring Decision of a Nondiscriminating Monopsonist

A nondiscriminating monopsonist pays the same wage to all workers. The marginal cost of hiring exceeds the wage, and the marginal cost curve lies above the supply curve. Profit maximization occurs at point *A*; the monopsonist hires E_M workers and pays them a wage of w_M .



the fact that all workers previously hired must now be paid a higher wage.⁵³ The profit-maximizing monopsonist hires up to the point where the marginal cost of labor equals the value of marginal product, or point *A* in the figure. If the monopsonist hires fewer than E_M workers, the value of marginal product exceeds the marginal cost of labor, and the firm should hire additional workers. In contrast, if the monopsonist hires more than E_M workers, the marginal cost of labor exceeds the contribution of workers to the firm and the monopsonist should lay off some employees. Therefore, the profit-maximizing condition for a nondiscriminating monopsonist is given by

$$MC_E = VMP_E \quad (4-10)$$

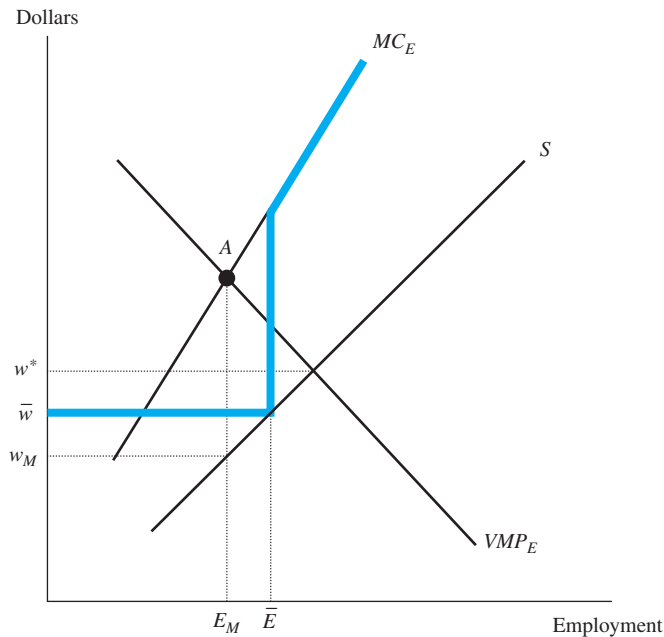
Note that the labor supply curve indicates that the monopsonist need only pay a wage of w_M to attract E_M workers to the firm.

The labor market equilibrium illustrated in Figure 4-20 has two important properties. First, a nondiscriminating monopsonist employs fewer workers than would be employed if the market were competitive. The competitive level of employment is given by the intersection of supply and demand, or E^* workers. As a result, there is underemployment

⁵³ Using calculus, it can be shown that the relationship between the wage and the marginal cost of hiring is given by $MC_E = w\left(1 + \frac{1}{\sigma}\right)$, where σ is the labor supply elasticity (that is, the percentage change in quantity supplied for a given percentage change in the wage). A competitive firm faces a perfectly elastic labor supply curve, so that the labor supply elasticity is infinite and the marginal cost of labor equals the wage. If the labor supply curve is upward sloping, the elasticity of labor supply will be positive and the marginal cost of labor exceeds the wage.

FIGURE 4-21 The Impact of the Minimum Wage on a Nondiscriminating Monopsonist

The minimum wage may increase both wages and employment when imposed on a monopsonist. A minimum wage set at \bar{w} increases employment to \bar{E} .



in a monopsony. Put differently, the allocation of resources in a nondiscriminating monopsony is not efficient.

Second, the monopsonistic wage w_M is less than the competitive wage, w^* , and also is less than the value of the worker's marginal product, VMP_M . In a monopsony, therefore, workers are paid less than their value of marginal product and are, in this sense, "exploited."

Monopsony and the Minimum Wage

The imposition of a minimum wage on a monopsonistic market can *increase* both wages and employment. In Figure 4-21, the nondiscriminating monopsonist is initially in equilibrium at point A , hiring E_M workers at a wage of w_M dollars. Suppose the government imposes a wage floor of \bar{w} . The firm can now hire up to \bar{E} workers at the minimum wage (because these workers were willing to work for a wage at or below the minimum). In other words, the marginal cost of labor is equal to the minimum wage as long as the firm hires up to \bar{E} workers. If the firm wants to hire more than \bar{E} workers, the marginal cost of hiring reverts back to its old level (because the monopsonist must pay more than the minimum wage to all workers hired). The marginal cost of labor curve, therefore, is now given by the bold line in the figure: a perfectly elastic segment up to \bar{E} workers and the upward-rising segment beyond that threshold.

A profit-maximizing monopsonist will still want to equate the marginal cost of hiring with the value of marginal product of labor. As drawn in Figure 4-21, the monopsonist hires \bar{E} workers and pays them the minimum wage. Note that the minimum wage legislation increased both the employment level of the firm (from E_M to \bar{E}) and the wage received by

workers (from w_M to \bar{w}). Moreover, there is no unemployment in the labor market. Everyone who is looking for work at a wage of \bar{w} can find it.

In fact, Figure 4-21 suggests that the government can do even better. It could set the minimum wage at the competitive level w^* (where supply equals demand). The monopsonistic firm would then employ the same number of workers that would be employed if the market were competitive, workers would be paid the competitive wage, and there would be no unemployment. A well-designed minimum wage, therefore, can completely eliminate the market power of monopsonists and prevent the exploitation of workers.

In the last chapter, we noted the evidence that—at least in the fast-food industry—minimum wage increases do not seem to result in a reduction in the number of persons employed in that industry. In contrast, some of the evidence indicated that these fast-food establishments may have *increased* their employment after the minimum wage was imposed. It has been suggested that these positive employment effects of minimum wages occurred because the fast-food industry is a monopsony in terms of employing unskilled teenage workers. Because these youths have few other alternatives, some argue that fast-food restaurants could provide the “one-company” environment that can generate a monopsony.

Could a Competitive Firm Have an Upward-Sloping Labor Supply Curve?

The one-company town is the classic example of a firm that faces an upward-sloping labor supply curve. If this type of firm wishes to expand, it has to raise the wage to attract more persons into the workforce. This situation gives “monopsony power” to the single firm in the industry: the ability to pay its workers less than the value of marginal product, allowing the firm to make excess profits.

It turns out, however, that individual firms might have some degree of monopsony power even when there are many firms in the labor market competing for the same type of labor. We have argued that one channel through which a competitive equilibrium is eventually attained is worker mobility—workers moving across firms to take advantage of better job opportunities. When firms in one market pay relatively high wages, the mobility of workers across markets reduces the wage gap and eventually equilibrates wages throughout the economy. The “law of one price,” in effect, depends crucially on the assumption that workers can costlessly move from one job to another.

It is probably the case, however, that workers incur substantial costs when they switch from one job to another. These costs are incurred as workers search for other jobs and as the workers move themselves and their families to unfamiliar economic and social environments. The presence of mobility costs implies that it does not make sense for a worker to accept every better-paying job offer that comes along. The mobility costs, after all, could well exceed the pay increase that the worker would get if he were to change jobs. As a result, mobility costs introduce a great deal of inertia into the labor market. A firm wishing to expand production and hire more workers will have to pay a wage premium that would induce workers already employed in other firms to quit those jobs, incur the mobility costs, and join the firm. In effect, mobility costs help generate an upward-sloping supply curve for a firm. A firm wishing to hire more and more workers will have to keep raising its wage to compensate workers for the costs incurred as they switch jobs.

A firm also may have an upward-sloping supply curve if the employer finds it harder to monitor its workers as employment rises. The larger the firm and the more workers it employs, the larger the possibilities for workers to “shirk” their responsibilities on the job

and go undetected. It has been suggested that a possible solution to this monitoring problem is to offer the workers a higher wage. This high wage would make workers realize that they have much to lose if they are caught shirking and are fired from their jobs. According to this argument, therefore, workers who are highly paid would have much less incentive to shirk on the job. As the firm expands its employment and finds it more difficult to monitor its workers, the firm may want to pay a higher wage to keep the workers in line. In fact, there is a great deal of evidence suggesting that larger firms pay higher wages.⁵⁴

The crucial insight to draw from this discussion is that upward-sloping supply curves for particular firms may arise even when there are many firms competing for the same workers. In short, many firms in competitive markets could have some degree of monopsony power.⁵⁵

The realization that monopsony power need not be restricted to the extreme case of a one-company town has led to a resurgence of research that attempts to estimate the labor supply elasticity to a given firm.⁵⁶ A recent study, for instance, examines how the supply of registered nurses (RNs) to a particular hospital responds to changes in the RN wage.⁵⁷

Before 1991, the U.S. Department of Veteran Affairs (VA) had a national pay scale that roughly determined RN wages in all of its facilities, regardless of whether those facilities were in high or low cost-of-living areas. This policy obviously affected the VA's ability to recruit nurses in high-wage regions, particularly during the 1980s when RN wages were rising rapidly. As an example, the starting RN hourly wage in Milwaukee in 1990 was \$11.20 in non-VA hospitals and \$11.65 in VA hospitals, so that the VA wage offer was quite competitive. In contrast, the starting RN hourly wage in San Francisco was \$16.30, but the VA starting wage lagged far behind at \$14.00.

The Nurse Pay Act of 1990 attempted to fix this problem by changing how the VA set wages in local facilities. In particular, the act tied the VA wage offer to the wages that prevailed in the local labor market. If the wage in VA hospitals were below the prevailing wage, the RN wage in the VA hospital would be raised immediately. However, if the wage in VA hospitals were above the prevailing wage, the VA wage would be held constant in nominal terms until the two wages reached parity. As a result, the law generated wage changes in VA hospitals that would presumably differentially change the supply of workers to each of these hospitals. In other words, the act would have mandated a rapid wage increase in the wage in VA hospitals in San Francisco, presumably attracting many new potential workers to those facilities, but little wage change in the VA hospitals in Milwaukee, where the supply of RNs would have remained relatively constant.

The difference-in-difference exercise reported in Table 4-7 illustrates how it is possible to use the enactment of the Nurse Pay Act of 1990 as an instrument to estimate the labor supply elasticity to VA hospitals. Between 1990 and 1992, the wage of RNs changed by

⁵⁴ Charles Brown, James Hamilton, and James Medoff, *Employers Large and Small*, Cambridge, MA: Harvard University Press, 1990.

⁵⁵ Note that the labor supply elasticity that is of interest in a study of monopsony—measuring the rate at which the firm must increase wages to attract more workers—differs conceptually from the labor supply elasticity that gives the relation between hours of work and wages for an individual worker. As a result, the empirical evidence on labor supply elasticities presented in Chapter 2 is of little use in attempting to measure the degree of monopsony power enjoyed by particular firms.

⁵⁶ See Alan Manning, *Monopsony in Motion*. Princeton, NJ: Princeton University Press, 2003, for an excellent summary of the models and estimation approaches.

⁵⁷ Douglas O. Staiger, Joanne Spetz, and Ciaran S. Pibbs, "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment," *Journal of Labor Economics* 28 (April 2010): 211–236.

TABLE 4-7 RN Wages and Employment, 1990–1992

Source: Douglas O. Staiger, Joanne Spetz, and Ciaran S. Phibbs, “Is There Monopsony in the Labor Market? Evidence from a Natural Experiment,” *Journal of Labor Economics* 28 (April 2010), p. 223.

| | VA Hospitals | Non-VA Hospitals |
|---------------------------------|--------------|------------------|
| Percent change in wage | 12.5 | 9.9 |
| Percent change in RN employment | 8.3 | 5.6 |

12.5 percent in VA hospitals and by 9.9 percent in non-VA hospitals, or a difference of 2.6 percentage points. At the same time, these wage changes led to a sizable increase in 8.3 percent in the number of RNs working at VA hospitals but only to a 5.6 percent in the number of RNs working at non-VA hospitals, or a difference of 2.7 percentage points. Recall that the labor supply elasticity is defined as the ratio of the percent change in the number of workers employed to the percent change in the wage, or $2.7 \div 2.6$, which is approximately equal to 1. In other words, a 1 percent increase in the wage that VA hospitals pay would attract 1 percent more nurses to those hospitals.

A number of recent studies use a similar methodology to estimate the labor supply elasticity to specific firms, and the findings tend to be remarkably similar.⁵⁸ For example, a study of the Norwegian teacher market documents that the labor supply elasticity of Norwegian teachers is about 1.4, while a study of schoolteachers in Missouri suggests that the elasticity is around 3.7. The crucial point about all of these estimates is that they are far below infinity, which would be the observed labor supply elasticity if the market were competitive—the firm would then face a constant wage regardless of the number of workers employed.

Some recent studies have also examined the long run behavior of the firm by observing the reaction of quit rates and recruitment rates changes in the firm’s wage over time. Not surprisingly, there is an intimate relationship between a firm’s monopsony power and the sensitivity of quit and recruitment rates to the firm’s wage. The study of these types of responses also suggests that the elasticity of labor supply at the firm level is in the range of 2 to 4, again far below what one would expect if the firm had no monopsony power.⁵⁹

4-10 Noncompetitive Labor Markets: Monopoly

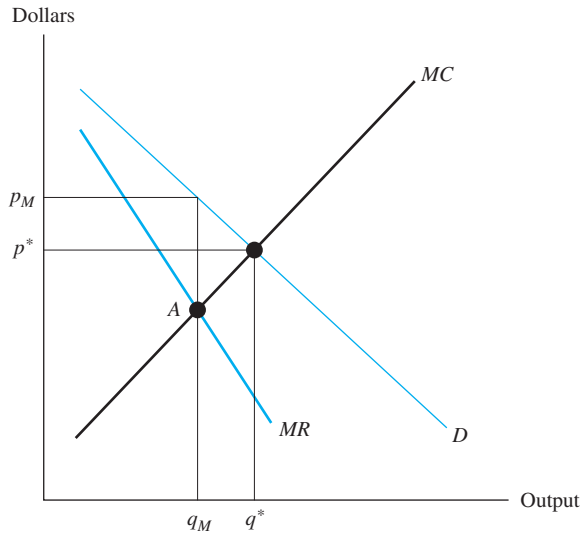
A monopsonist’s hiring decision influences the wage because the supply curve for labor is upward sloping. The more workers hired by a monopsony, the higher the wage that the firm will have to pay. We now consider hiring decisions in firms that influence the price of the output they sell. The simplest example of such a market structure is a **monopoly**, when there is only one seller in the market. As illustrated in Figure 4-22, the monopolist, unlike a competitive firm, faces a downward-sloping demand curve for his or her output.

⁵⁸ Torberg Falch, “The Elasticity of Labor Supply at the Establishment Level,” *Journal of Labor Economics* 28 (April 2010): 237–266; and Michael Ransom and David P. Sims, “Estimating the Firm’s Labor Supply Curve in a “New Monopsony” Framework: Schoolteachers in Missouri,” *Journal of Labor Economics* 28 (April 2010): 331–355.

⁵⁹ Orley C. Ashenfelter, Henry Farber, and Michael R. Ransom, “Labor Market Monopsony,” *Journal of Labor Economics* 28 (April 2010): 203–210, provides a survey of this literature and a very accessible summary of the dynamic approach.

FIGURE 4-22 The Output Decision of a Monopolist

A monopolist faces a downward-sloping demand curve for his output. The marginal revenue from selling an additional unit of output is less than the price of the product. Profit maximization occurs at point A ; a monopolist produces q_M units of output and sells them at a price of p_M dollars.



Because the price of the output falls as the monopolist expands production, the marginal revenue associated with selling an additional unit of output is not equal to the output price p . If the monopolist wants to sell an extra unit of output, he must lower the price not only for that customer but for all other customers who wish to purchase the good.⁶⁰ As a result, the marginal revenue is less than the price charged for that last unit and declines as the monopolist attempts to sell more output. Figure 4-22 shows that the marginal revenue curve (MR) for a monopolist is downward sloping and lies below the demand curve (D).⁶¹

A profit-maximizing monopolist produces up to the point where marginal revenue equals the marginal cost of production (or point A in the figure). The monopolist produces q_M units of output and charges a price of p_M dollars per unit because this is the point on the demand curve that indicates how much consumers are willing to pay to purchase q_M units. Finally, note that the monopolist produces less output than would have been produced if the industry had been competitive. In a competitive market, q^* units of output are exchanged at a price of p^* dollars. A monopolist, therefore, sells less output at a higher price.

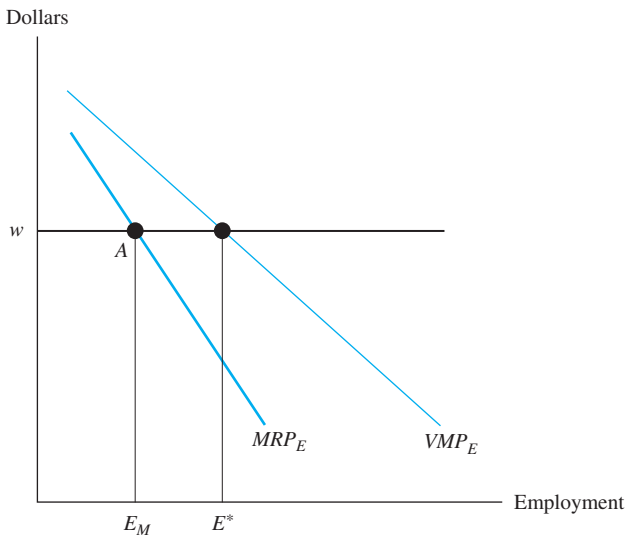
We can now derive the implications of monopoly power in the output market for the firm's labor demand curve and hiring decision. A monopolist, like any other profit-maximizing

⁶⁰ This type of monopolist is called a *nondiscriminating monopolist* because the firm charges the same price to all customers. Monopolists who can charge different prices to different customers are called *perfectly discriminating monopolists*.

⁶¹ Using calculus, it can be shown that the relationship between marginal revenue and price is given by $MR = p \left(1 + \frac{1}{\eta} \right)$ where η is the elasticity of demand for the output (that is, the percentage change in quantity demanded for a given percentage change in price). In a perfectly competitive market, the firm faces a perfectly elastic demand curve, so that the elasticity of output demand is infinite and, hence, $MR = p$. A monopolist faces a downward-sloping demand curve, so that η is negative and $MR < p$.

FIGURE 4-23 The Labor Demand Curve of a Monopolist

The marginal revenue product gives the worker's contribution to a monopolist's revenues (or the worker's marginal product times marginal revenue), and is less than the worker's value of marginal product. Profit maximization occurs at point *A*; the monopolist hires fewer workers (E_M) than would be hired in a competitive market.



firm, hires up to the point where the contribution of the last worker hired equals the cost of hiring. For a monopolist, the additional revenue from hiring an extra person equals the worker's marginal product *times* the marginal revenue received from the last unit of output sold. This variable is called the **marginal revenue product** of labor (MRP_E) and equals

$$MRP_E = MR \times MP_E \quad (4-11)$$

Note that the marginal revenue product of labor is less than the value of marginal product for a monopolist because the marginal revenue from selling the last unit of output (MR) is less than the price of the output.

Figure 4-23 illustrates the monopolist's hiring decision. Because the monopolist's actions can only influence prices in the market for the output, the monopolist can hire as much labor as it wants at the market wage w . A profit-maximizing monopolist hires E_M workers, where the wage equals the marginal revenue product of labor. If the firm hires fewer workers, an additional worker hired would generate more revenue than it would cost to hire him. Conversely, if the firm hires more than E_M workers, the last worker hired generates less revenue than it costs to employ him. The profit-maximizing condition for a monopolist is given by

$$MRP_E = w \quad (4-12)$$

Note that a monopolist ends up hiring fewer workers (E_M) than would be hired if this industry were competitive. A competitive firm hires up to the point where the wage equals the value of marginal product, or E^* workers in Figure 4-23.

There is some evidence suggesting that monopolists (such as utility companies) and other firms that can influence price (such as firms in industries where production is highly concentrated in a small number of firms, or an **oligopoly**) pay higher wage rates than the competitive

wage.⁶² Workers employed in these highly concentrated industries earn about 10 percent more than comparable workers in competitive industries. Many regulated monopolists can pass on the costs of production to consumers. As a result, these firms have little incentive to hold down costs. The monopolists may then be willing to pay high wages to attract workers with attributes that they deem desirable (such as educational pedigrees, race, or looks).

Summary

- A competitive economy where a homogeneous group of workers and firms can freely enter and exit the market has a single equilibrium wage across all labor markets.
- There is no unemployment in a competitive labor market because all workers who wish to work can find a job at the going wage.
- A competitive equilibrium leads to an efficient allocation of resources. No other allocation of workers to firms generates higher gains from trade.
- A fraction of the payroll taxes imposed on firms is passed on to workers. The more inelastic the labor supply curve, the higher the fraction of payroll taxes that is shifted to workers.
- The payroll tax creates a deadweight loss.
- A payroll tax has the same impact on wages and employment regardless of whether it is imposed on workers or on firms.
- In the short run, immigration reduces the wage of workers who have skills similar to those of immigrants and increases the wages of workers who have skills that complement those of immigrants. In the long run, these wage effects are attenuated as the capital stock adjusts to the presence of immigrants.
- The evidence does not suggest that workers living in cities penetrated by immigrants earn much less than workers in cities where few immigrants reside. This result might arise because native workers respond to immigration by migrating from the immigrant cities to the nonimmigrant cities, thereby diffusing the impact of immigration over the national economy. Immigrants do seem to have an adverse impact on native wages at the national level.
- In the short run, immigration redistributes wealth from workers to employers, but the net income of natives increases.
- Markets for professional workers are sometimes characterized by systematic booms and busts, or cobwebs.
- A nondiscriminating monopsonist hires fewer workers than would be hired in a competitive labor market and pays them a lower wage.
- The imposition of a minimum wage on a monopsony can increase both the wage and the number of workers employed.
- A particular firm may have some monopsony power, even in labor markets that may seem competitive, when workers find it costly to move across firms.
- A monopolist hires fewer workers than would be hired in a competitive product market but pays the market wage.

⁶² Ronald G. Ehrenberg, *The Regulatory Process and Labor Earnings*, New York: Academic Press, 1979; James Long and Albert Link, "The Impact of Market Structure on Wages, Fringe Benefits and Turnover," *Industrial and Labor Relations Review* 36 (January 1983): 239–250; and John S. Heywood, "Labor Quality and the Concentration-Earnings Hypothesis," *Review of Economics and Statistics* 68 (May 1986): 342–346.

Key Concepts

| | | |
|---------------------------|-----------------------------|----------------------------|
| cobweb, 186 | invisible hand theorem, 144 | monopsony, 188 |
| deadweight loss, 157 | mandated benefits, 161 | oligopoly, 196 |
| efficient allocation, 147 | marginal revenue | producer surplus, 146 |
| gains from trade, 147 | product, 196 | rational expectations, 187 |
| immigration surplus, 180 | monopoly, 194 | worker surplus, 147 |

Review Questions

1. What is the producer surplus? What is the worker surplus? Show that a competitive market equilibrium maximizes the gains from trade.
2. Discuss the implications of equilibrium for a competitive economy containing many regional markets when labor and firms are free to enter and exit the various markets. Why is the resulting allocation of labor efficient?
3. Show what happens to producer surplus, worker surplus, and the gains from trade as workers migrate from a low-wage to a high-wage region.
4. Describe the impact of a payroll tax on wages and employment in a competitive industry. Why is part of the tax shifted to workers? What is the deadweight loss of the payroll tax?
5. Why does the payroll tax have the same impact on wages and employment regardless of whether it is imposed on workers or on firms?
6. How do mandated benefits affect labor market outcomes? Why do these outcomes differ from those resulting from a payroll tax? What is the deadweight loss arising from mandated benefits?
7. Do immigrants reduce the wage of native workers? Do immigrants “take jobs away” from native workers?
8. What is the immigration surplus?
9. Describe the trends in wages and employment implied by the cobweb model for the engineering market. What would happen to the cobwebs if an economics consulting firm sold information on the history of wages and employment in the engineering market?
10. Describe the hiring decision of a perfectly discriminating monopsonist and of a non-discriminating monopsonist. In what sense do monopsonists “exploit” workers?
11. Show how the imposition of a minimum wage on a monopsony can increase both wages and employment.
12. Describe the hiring decision of a monopolist.

Problems

- 4-1. Figure 4-9 discusses the changes to a labor market equilibrium when the government mandates an employee benefit for which the cost exceeds the worker’s valuation (panel a) and for which the cost equals the worker’s valuation (panel b).
 - a. Provide a similar graph to those in Figure 4-9 when the cost of the benefit is less than the worker’s valuation and discuss how the equilibrium level of employment and wages has changed. Is there deadweight loss associated with the mandated benefit?

- b. Why is the situation in which a mandated benefit would cost less than the worker's valuation less important for public policy purposes than when the cost of the mandated benefit exceeds the worker's valuation?
- 4-2. In the United States, labor supply tends to be inelastic relative to labor demand, and according to law, payroll taxes are essentially assessed evenly between workers and firms. Given the above situation, are workers or firms more likely to bear the additional burden of an increased payroll tax in the United States? Could this burden be shifted to the firms by assessing the increase in payroll taxes on just firms rather than having firms and workers continue to be assessed payroll taxes equally?
- 4-3. Suppose the supply curve of physicists is given by $w = 10 + 5E$, while the demand curve is given by $w = 50 - 3E$. Calculate the equilibrium wage and employment level. Suppose now that the demand for physicists increases to $w = 70 - 3E$. Assume the market is subject to cobwebs. Calculate the wage and employment level in each round as the wage and employment levels adjust to the demand shock. (Recall that each round occurs on the demand curve—when the firm posts a wage and hires workers.) What are the new equilibrium wage and employment level?
- 4-4. The 1986 Immigration Reform and Control Act (IRCA) made it illegal for employers in the United States to knowingly hire illegal aliens. The legislation, however, has not reduced the flow of illegal aliens into the country. As a result, it has been proposed that the penalties against employers who break the law be substantially increased. Suppose that illegal aliens, who tend to be less-skilled workers, are complements with native workers. What will happen to the wage of native workers if the penalties for hiring illegal aliens increase?
- 4-5. a. What happens to wages and employment if the government imposes a payroll tax on a monopsonist? Compare the response in the monopsonistic market to the response that would have been observed in a competitive labor market.
- b. Suppose a firm is a perfectly discriminating monopsonist. The government imposes a minimum wage on this market. What happens to wages and employment?
- 4-6. An economy consists of two regions, the North and the South. The short-run elasticity of labor demand in each region is -0.5 . Labor supply is perfectly inelastic within both regions. The labor market is initially in an economywide equilibrium, with 600,000 people employed in the North and 400,000 in the South at a wage of \$15 per hour. Suddenly, 20,000 people immigrate from abroad and initially settle in the South. They possess the same skills as the native residents and also supply their labor inelastically.
- a. What will be the effect of this immigration on wages in each of the regions in the short run (before any migration between the North and the South occurs)?
- b. Suppose 1,000 native-born persons per year migrate from the South to the North in response to every dollar differential in the hourly wage between the two regions. What will be the ratio of wages in the two regions after the first-year native labor responds to the entry of the immigrants?
- c. What will be the effect of this immigration on wages and employment in each of the regions in the long run (after native workers respond by moving across

regions to take advantage of whatever wage differentials may exist)? Assume labor demand does not change in either region.

- 4-7. A firm faces perfectly elastic demand for its output at a price of \$6 per unit of output. The firm, however, faces an upward-sloped labor supply curve of

$$E = 20w - 120$$

where E is the number of workers hired each hour and w is the hourly wage rate. Thus, the firm faces an upward-sloped marginal cost of labor curve of

$$MC_E = 6 + 0.1E$$

Each hour of labor produces five units of output. How many workers should the firm hire each hour to maximize profits? What wage will the firm pay? What are the firm's hourly profits?

- 4-8. Polly's Pet Store has a local monopoly on the grooming of dogs. The daily inverse demand curve for pet grooming is

$$P = 20 - 0.1Q$$

where P is the price of each grooming and Q is the number of groomings given each day. This implies that Polly's marginal revenue is

$$MR = 20 - 0.2Q$$

Each worker Polly hires can groom 20 dogs each day. What is Polly's labor demand curve as a function of w , the daily wage that Polly takes as given?

- 4-9. The Key West Parrot Shop has a monopoly on the sale of parrot souvenir caps in Key West. The inverse demand curve for caps is

$$P = 30 - 0.4Q$$

where P is the price of a cap and Q is the number of caps sold per hour. Thus, the marginal revenue for the Parrot Shop is

$$MR = 30 - 0.8Q$$

The Parrot Shop is the only employer in town and faces an hourly supply of labor given by

$$w = 0.9E + 5$$

where w is the hourly wage rate and E is the number of workers hired each hour. The marginal cost associated with hiring E workers, therefore, is

$$MC_E = 1.8E + 5$$

Each worker produces two caps per hour. How many workers should the Parrot Shop hire each hour to maximize its profit? What wage will it pay? How much will it charge for each cap?

- 4-10. Ann owns a lawn-mowing company. She has 400 lawns she needs to cut each week. Her weekly revenue from these 400 lawns is \$20,000. Given an 18-inch-deck push mower, a laborer can cut each lawn in two hours. Given a 60-inch-deck riding

mower, a laborer can cut each lawn in 30 minutes. Labor is supplied inelastically at \$5.00 per hour. Each laborer works eight hours a day and five days each week.

- a. If Ann decides to have her workers use push mowers, how many push mowers will Ann rent and how many workers will she hire?
 - b. If she decides to have her workers use riding mowers, how many riding mowers will Ann rent and how many workers will she hire?
 - c. Suppose the weekly rental cost (including gas and maintenance) for each push mower is \$250 and for each riding mower is \$1,800. What equipment will Ann rent? How many workers will she employ? How much profit will she earn?
 - d. Suppose the government imposes a 20 percent payroll tax (paid by employers) on all labor and offers a 20 percent subsidy on the rental cost of capital. What equipment will Ann rent? How many workers will she employ? How much profit will she earn?
- 4-11. The immigration surplus, though seemingly small in the United States, redistributes wealth from workers to firms. Present a back-of-the-envelope calculation of the losses accruing to native workers and of the gains accruing to firms. Do these calculations help explain why some segments of society are emotional in their support of changes in immigration policy that would either increase or decrease the immigrant flow?
- 4-12. Labor demand for low-skilled workers in the United States is $w = 24 - 0.1E$ where E is the number of workers (in millions) and w is the hourly wage. There are 120 million domestic U.S. low-skilled workers who supply labor inelastically. If the United States opened its borders to immigration, 20 million low-skill immigrants would enter the United States and supply labor inelastically. What is the market-clearing wage if immigration is not allowed? What is the market-clearing wage with open borders? How much is the immigration surplus when the United States opens its borders? How much surplus is transferred from domestic workers to domestic firms?
- 4-13. Consider the policy application of hurricanes and the labor market that was presented in the text.
- a. How do labor demand and labor supply typically shift following a natural disaster?
 - b. The data on changes in employment and wages in Table 4-5 suggest that the magnitude of relative shifts in labor demand and labor supply depend on the severity of the natural disaster. According to the data, does labor demand shift more relative to labor supply in mild or in extreme natural disasters. Provide intuition for this finding.
- 4-14. Suppose the Cobb-Douglas production function given in Equation 4-1 applies to a developing country. Instead of thinking of immigration from a developing to a developed country, suppose a developed country invests large amounts of capital (foreign direct investment, or FDI) in a developing country.
- a. How does an increase in FDI affect labor productivity in the developing country? How will wages respond in the short-run?
 - b. What are the long-run implications of FDI, especially in terms of potential future immigration from the developing country?

- 4-15. A number of empirical studies suggest that labor demand is very elastic while labor supply is very inelastic. Assume too that payroll taxes are about 15 percent and legislated to be paid half by the employee and half by the employer.
- What would happen to worker wages if payroll taxes were eliminated?
 - What would happen to employment costs paid by firms if payroll taxes were eliminated?
 - What would happen to producer and worker surplus if payroll taxes were eliminated? Which measure is relatively more sensitive to payroll taxes? Why?
 - Why might workers not want payroll taxes eliminated?

Selected Readings

- Joshua D. Angrist, “Short-Run Demand for Palestinian Labor,” *Journal of Labor Economics* 14 (July 1996): 425–453.
- Orley C. Ashenfelter, Henry Farber, and Michael R. Ransom, “Labor Market Monopsony,” *Journal of Labor Economics* 28 (April 2010): 203–210.
- Ariel R. Belasen and Solomon W. Polachek, “How Disasters Affect Local Labor Markets: The Effects of Hurricanes in Florida,” *Journal of Human Resources* 44 (Winter 2009): 251–276.
- Olivier Jean Blanchard and Lawrence F. Katz, “Regional Evolutions,” *Brookings Papers on Economic Activity* 1 (1992): 1–61.
- George J. Borjas, “The Economic Benefits from Immigration,” *Journal of Economic Perspectives* 9 (Spring 1995): 3–22.
- George J. Borjas, “The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration in the Labor Market,” *Quarterly Journal of Economics* 118 (November 2003): 1335–1374.
- David Card, “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review* 43 (January 1990): 245–257.
- David Card, “Is the New Immigration Really So Bad?” *Economic Journal* 115 (November 2005): F300–F323.
- Jonathan Gruber, “The Incidence of Payroll Taxation: Evidence from Chile,” *Journal of Labor Economics* 15 (July 1997, Part 2): S102–S135.
- Prachi Mishra, “Emigration and Wages in Source Countries: Evidence from Mexico,” *Journal of Development Economics* 82 (January 2007): 180–199.
- Douglas O. Staiger, Joanne Spetz, and Ciaran S. Phibbs, “Is There Monopsony in the Labor Market? Evidence from a Natural Experiment,” *Journal of Labor Economics* 28 (April 2010): 211–236.

Web Links

The Web site of the Bureau of Citizenship and Immigration Services (BCIS) contains information on U.S. immigration policy: www.uscis.gov/graphics/index.htm.

The Department of Homeland Security reports detailed statistics on immigration to the United States: <http://www.dhs.gov/ximgtn/statistics/>.