selling occurs when firms require employees to pay a fee when they are hired. If firms are obtaining payments from new workers, their labor demand is higher for a given wage; thus the wage and employment rise as the economy moves up the no-shirking curve. If firms are able to require bonds or sell jobs, they will do so, and unemployment will be eliminated from the model.

Bonding, job selling, and the like may be limited by an absence of perfect capital markets (so that it is difficult for workers to post large bonds, or to pay large fees when they are hired). They may also be limited by workers' fears that the firm may falsely accuse them of shirking and claim the bonds, or dismiss them and keep the job fee. But, as Carmichael (1985) emphasizes, such considerations cannot eliminate these schemes entirely: if workers strictly prefer employment to unemployment, firms can raise their profits by, for example, charging marginally more for jobs. In such situations, jobs are not rationed, but go to those who are willing to pay the most for them. Thus even if these schemes are limited, they still eliminate unemployment. In short, the absence of job fees and performance bonds is a puzzle for the theory.

It is important to keep in mind that the Shapiro–Stiglitz model focuses on one particular source of efficiency wages. Neither its conclusions nor the difficulties it faces in explaining the absence of bonding and job selling are general. For example, suppose firms find high wages attractive because they improve the quality of job applicants on dimensions they cannot observe. Since the attractiveness of a job presumably depends on the overall compensation package, in this case firms have no incentive to adopt schemes such as job selling. Likewise, there is no reason to expect the implications of the Shapiro–Stiglitz model concerning the effects of a shift in labor demand to apply in this case.

As described in Section 10.8, workers' feelings of gratitude, anger, and fairness appear to be important to wage-setting. If these considerations are the reason that the labor market does not clear, again there is no reason to expect the Shapiro–Stiglitz model's implications concerning compensation schemes and the effects of shifts in labor demand to hold. In this case, theory provides little guidance. Generating predictions concerning the determinants of unemployment and the cyclical behavior of the labor market requires more detailed study of the determinants of workers' attitudes and their impact on productivity. Section 10.8 describes some preliminary attempts in this direction.

10.5 Contracting Models

The second departure from Walrasian assumptions about the labor market that we consider is the existence of long-term relationships between firms and workers. Firms do not hire workers afresh each period. Instead, many jobs involve long-term attachments and considerable firm-specific skills on the part of workers.

The possibility of long-term relationships implies that the wage does not have to adjust to clear the labor market each period. Workers are content to stay in their current jobs as long as the income streams they expect to obtain are preferable to their outside opportunities; because of their long-term relationships with their employers, their current wages may be relatively unimportant to this comparison. This section explores the consequences of this observation.

A Baseline Model

Consider a firm dealing with a group of workers. The firm's profits are

$$\pi = AF(L) - wL, \qquad F'(\bullet) > 0, \qquad F''(\bullet) < 0, \tag{10.39}$$

where L is the quantity of labor the firm employs and w is the wage. A is a factor that shifts the profit function. It could reflect technology (so that a higher value means that the firm can produce more output from a given amount of labor), or economy-wide output (so that a higher value means that the firm can obtain a higher relative price for a given amount of output).

Instead of considering multiple periods, it is easier to consider a single period and assume that A is random. Thus when workers decide whether to work for the firm, they consider the expected utility they obtain in the single period given the randomness in A, rather than the average utility they obtain over many periods as their income and hours vary in response to fluctuations in A.

The distribution of A is discrete. There are K possible values of A, indexed by i; p_i denotes the probability that $A = A_i$. Thus the firm's expected profits are

$$E[\pi] = \sum_{i=1}^{K} p_i [A_i F(L_i) - w_i L_i], \tag{10.40}$$

where L_i and w_i denote the quantity of labor and the wage if the realization of A is A_i. The firm maximizes its expected profits; thus it is risk-neutral.

Each worker is assumed to work the same amount. The representative worker's utility is

$$u = U(C) - V(L), \quad U'(\bullet) > 0, \quad U''(\bullet) < 0, \quad V'(\bullet) > 0, \quad V''(\bullet) > 0, \tag{10.41}$$

where $U(\bullet)$ gives the utility from consumption and $V(\bullet)$ the disutility from working. Since $U''(\bullet)$ is negative, workers are risk-averse.[15]

Workers' consumption, C, is assumed to equal their labor income, wL.[16] That is, workers cannot purchase insurance against employment and wage fluctuations. In a more fully developed model, this might arise because workers are heterogeneous and have private information about their labor-market prospects. Here, however, the absence of outside insurance is simply assumed.

Equation (10.41) implies that the representative worker's expected utility is

$$E[u] = \sum_{i=1}^{K} p_i [U(C_i) - V(L_i)]. \tag{10.42}$$

There is some reservation level of expected utility, u_0, that workers must attain to be willing to work for the firm. There is no labor mobility once workers agree to a contract. Thus the only constraint on the contract involves the average level of utility it offers, not the level in any individual state.

Implicit Contracts

One simple type of contract just specifies a wage and then lets the firm choose employment once A is determined; many actual contracts at least appear to take this form. Under such a *wage contract,* unemployment and real wage rigidity arise immediately. A fall in labor demand, for example, causes the firm to reduce employment at the fixed real wage while labor supply does not shift, and thus creates unemployment (or, if all workers work the same amount, underemployment). And the cost of labor does not respond because, by assumption, the real wage is fixed.

But this is not a satisfactory explanation of unemployment and real wage rigidity. The difficulty is that this type of a contract is inefficient (Leontief, 1946). Since the wage is fixed and the firm chooses employment taking the wage as given, the marginal product of labor is independent of A. But since employment varies with A, the marginal disutility of working depends on

[15] Because the firm's owners can diversify away firm-specific risk by holding a broad portfolio, the assumption that the firm is risk-neutral is reasonable for firm-specific shocks. For aggregate shocks, however, the assumption that the firm is less risk-averse than the workers is harder to justify. Since the main goal of the theory is to explain the effects of aggregate shocks, this is a weak point of the model. One possibility is that the owners are wealthier than the workers and that risk aversion is declining in wealth.

[16] If there are $\overline{L}$ workers, the representative worker's hours and consumption are in fact $L/\overline{L}$ and $wL/\overline{L}$, and so utility takes the form $\tilde{U}(C/\overline{L}) - \tilde{V}(L/\overline{L})$. To eliminate $\overline{L}$, define $U(C) = \tilde{U}(C/\overline{L})$ and $V(L) = \tilde{V}(L/\overline{L})$.

A. Thus the marginal product of labor is generally not equal to the marginal disutility of work, and so it is possible to make both parties to the contract better off. And if labor supply is not very elastic, the inefficiency is large. When labor demand is low, for example, the marginal disutility of work is low, and so the firm and the workers could both be made better off if the workers worked slightly more.

To see how it is possible to improve on a wage contract, suppose the firm offers the workers a contract specifying the wage and hours for each possible realization of A. Since actual contracts do not explicitly specify employment and the wage as functions of the state, such contracts are known as *implicit contracts*.[17]

Recall that the firm must offer the workers at least some minimum level of expected utility, u_0, but is otherwise unconstrained. In addition, since L_i and w_i determine C_i, we can think of the firm's choice variables as L and C in each state rather than as L and w. The Lagrangian for the firm's problem is therefore

$$\mathcal{L} = \sum_{i=1}^{K} p_i[A_i F(L_i) - C_i] + \lambda\left(\left\{\sum_{i=1}^{K} p_i[U(C_i) - V(L_i)]\right\} - u_0\right). \tag{10.43}$$

The first-order condition for C_i is

$$-p_i + \lambda p_i U'(C_i) = 0, \tag{10.44}$$

or

$$U'(C_i) = \frac{1}{\lambda}. \tag{10.45}$$

Equation (10.45) implies that the marginal utility of consumption is constant across states, and thus that consumption is constant across states. Thus the risk-neutral firm fully insures the risk-averse workers.

The first-order condition for L_i is

$$p_i A_i F'(L_i) = \lambda p_i V'(L_i). \tag{10.46}$$

Equation (10.45) implies $\lambda = 1/U'(C)$, where C is the constant level of consumption. Substituting this fact into (10.46) and dividing both sides by p_i yields

$$A_i F'(L_i) = \frac{V'(L_i)}{U'(C)}. \tag{10.47}$$

[17] The theory of implicit contracts is due to Azariadis (1975), Baily (1974), and Gordon (1974).

Implications

Under efficient contracts, workers' real incomes are constant. In that sense, the model implies strong real wage rigidity. Indeed, because L is higher when A is higher, the model implies that the wage per hour is countercyclical. Unfortunately, however, this result does not help to account for the puzzle that shifts in labor demand appear to result in large changes in employment. The problem is that with long-term contracts, the wage is no longer playing an allocative role (Barro, 1977; Hall, 1980). That is, firms do not choose employment taking the wage as given. Rather, the level of employment as a function of the state is specified in the contract. And, from (10.47), this level is the level that equates the marginal product of labor with the marginal disutility of additional hours of work.

This discussion implies that the cost to the firm of varying the amount of labor it uses is likely to change greatly with its level of employment. Suppose the firm wants to increase employment marginally in state i. To do this, it must raise workers' compensation to make them no worse off than before. Since the expected utility cost to workers of the change is $p_i V'(L_i)$, C must rise by $p_i V'(L_i)/U'(C)$. Thus the marginal cost to the firm of increasing employment in a given state is proportional to $V'(L_i)$. If labor supply is relatively inelastic, $V'(L_i)$ is sharply increasing in L_i, and so the cost of labor to the firm is much higher when employment is high than when it is low. Thus, for example, embedding this model of contracts in a model of price determination like that of Section 6.6 would not alter the result that relatively inelastic labor supply creates a strong incentive for firms to cut prices and increase employment in recessions, and to raise prices and reduce employment in booms.

In addition to failing to predict relatively acyclical labor costs, the model fails to predict unemployment: as emphasized above, the implicit contract equates the marginal product of labor and the marginal disutility of work. The model does, however, suggest a possible explanation for *apparent* unemployment. In the efficient contract, workers are not free to choose their labor supply given the wage. Instead, the wage and employment are simultaneously specified to yield optimal risk-sharing and allocative efficiency. When employment is low, the marginal disutility of work is low and the hourly wage, C/L_i, is high. Thus workers wish that they could work more at the wage the firm is paying. As a result, even though employment and the wage are chosen optimally, workers appear to be constrained in their labor supply.

Insiders and Outsiders

One possible way of improving contracting models' ability to explain key features of labor markets is to relax the assumption that the firm is dealing with a fixed pool of workers. In reality, there are two groups of potential workers. The first group—the insiders—are workers who have some connection with the firm at the time of the bargaining, and whose interests are therefore taken into account in the contract. The second group—the outsiders—are workers who have no initial connection with the firm but who may be hired after the contract is set.

Oswald (1993) and Gottfries (1992), building on earlier work by Lindbeck and Snower (1986), Blanchard and Summers (1986), and Gregory (1986), argue that relationships between firms and insiders and outsiders have two features that are critical to how contracting affects the labor market. First, because of normal employment growth and turnover, most of the time the insiders are fully employed and the only hiring decision concerns how many outsiders to hire. This immediately implies that, just as in a conventional labor demand problem, but in sharp contrast to what happens in the basic implicit-contract model, employment is chosen to equate the marginal product of labor with the wage. To see this, note that if this condition fails, it is possible to increase the firm's profits with no change in the insiders' expected utility by changing the number of outsiders hired. Thus it cannot make sense for the insiders and the firm (who are the only ones involved in the original bargaining) to agree to such an arrangement.

The second feature of labor markets that Oswald and Gottfries emphasize is that the wages paid to the two types of workers cannot be set independently: in practice, the higher the wage that the firm pays to its existing employees, the more it must pay to its new hires. This implies that the insurance role of wages affects employment. Suppose, for example, that the insiders and the firm agree to keep the real wage fixed and so provide complete insurance to the insiders.[18] Then when the firm is hit by shocks, employment varies to keep the marginal product of labor equal to the constant real wage.

Because the wage is now playing both an insurance and an allocative role, in general the optimal contract does not make it independent of the state. Under natural assumptions, however, this actually strengthens the results: the optimal contract typically specifies a lower wage when the realization of A is higher, and so further magnifies employment fluctuations. Intuitively, by lowering the wage in states where employment is high, the insiders and the firm reduce the amount of insurance the firm is providing but also lower the average amount spent hiring outsiders. The optimal contract involves a balancing of these two objectives, and thus a somewhat countercyclical wage.[19] Thus this model implies that the real wage is countercyclical and that it represents the true cost of labor to the firm.

[18] Recall that since the marginal hiring decisions involve outsiders, the amount the insiders work is independent of the state. Thus, in contrast to what happens in the basic implicit-contract model, here a constant wage makes the insiders' consumption constant.

[19] See Problem 10.8.

The crucial feature of the model is its assumption that the outsiders' and insiders' wages are linked. Without this link, the firm can hire outsiders at the prevailing economy-wide wage. With inelastic labor supply, that wage is low in recessions and high in booms, and so the marginal cost of labor to the firm is highly procyclical.

Unfortunately, the insider-outsider literature has not established that outsiders' and insiders' wages are linked. Gottfries argues that a link arises from the facts that the firm must be given some freedom to discharge insiders who are incompetent or shirking and that an excessive gap between insiders' and outsiders' wages would give the firm an incentive to take advantage of this freedom. Blanchard and Summers (1986) argue that the insiders are reluctant to allow the hiring of large numbers of outsiders at a low wage because they realize that, over time, such a policy would result in the outsiders controlling the bargaining process. But tying insiders' and outsiders' wages does not appear to be the best way of dealing with these problems. If the economy-wide wage is sometimes far below insiders', tying the insiders' and outsiders' wages is very costly. It appears that the firm and the insiders would therefore be better off if they instead agreed to some limitation on the firm's ability to hire outsiders, or if they charged new hires a fee (and let the fee vary with the gap between the insiders' wage and the economy-wide wage).

It is also possible that a link between insiders' and outsiders' wages could arise from workers' notions about fairness and the potential effects of the firm violating those notions, along the lines of the loyalty-based efficiency-wage models we discussed in Section 10.2. But in this case, it is not clear that the contracting and insider-outsider considerations would be important; the efficiency-wage forces alone might be enough to greatly change the labor market.

In short, we can conclude only that *if* a link between insiders' and outsiders' wages can be established, insider-outsider considerations may have important implications.

Hysteresis and European Unemployment

One important extension of insider-outsider models involves dynamic settings. The previous discussion assumed that the insiders are always employed. But this assumption is likely to fail in some situations. Most importantly, if the insiders' bargaining power is sufficiently great, they will set the wage high enough to risk some unemployment: if the insiders are fully employed with certainty, there is a benefit but not a cost to them of raising the wage further. And variations in employment can give rise to dynamics in the number of insiders. Under many institutional arrangements, workers who become unemployed eventually lose a say in wage-setting; likewise, workers who are hired eventually gain a role in bargaining. Thus a fall in employment caused by a decline in labor demand is likely to reduce the number of insiders, and a rise in employment is likely to increase the number of insiders. This in turn affects future wage-setting and employment. When the number of insiders is smaller, they can afford to set a higher wage. Thus a one-time adverse shock to labor demand can lead to a persistent fall in employment. The extreme case where the effect is permanent is known as *hysteresis.*

The possibility of hysteresis has received considerable attention in the context of Europe. European unemployment fluctuated around very low levels in the 1950s and 1960s, rose fairly steadily to more than 10 percent from the mid-1970s to the mid-1980s, and has shown little tendency to decline since then. Thus there is no evidence of a stable natural rate that unemployment returns to after a shock. Blanchard and Summers (1986) argue that Europe in the 1970s and 1980s satisfied the conditions for insider-outsider considerations to produce hysteresis: workers had a great deal of power in wage-setting, there were large negative shocks, and the rules and institutions led to some extent to the disenfranchisement from the bargaining process of workers who lost their jobs.

Two possible sources of hysteresis other than insider-outsider considerations have also received considerable attention. One is deterioration of skills: workers who are unemployed do not acquire additional on-the-job training, and their existing human capital may decay or become obsolete. As a result, workers who lose their jobs when labor demand falls may have difficulty finding work when demand recovers, particularly if the downturn is extended. The second additional source of hysteresis operates through labor-force attachment. Workers who are unemployed for extended periods may adjust their standard of living to the lower level provided by income maintenance programs. In addition, a long period of high unemployment may reduce the social stigma of extended joblessness. Because of these effects, labor supply may be permanently lower when demand returns to normal.

Loosely speaking, views of European unemployment fall into two camps. One emphasizes not hysteresis, but shifts in the natural rate as a result of such features of European labor-market institutions as generous unemployment-insurance benefits. Since most of those features were in place well before the rise in unemployment, this view requires that institutions' effects operate with long lags. For example, because the social stigma of unemployment changes slowly, the impact of generous unemployment benefits on the natural rate may be felt only very gradually (see, for example, Lindbeck and Nyberg, 2006). The other view emphasizes hysteresis. In this view, the labor-market institutions converted what would have otherwise been short-lived increases in unemployment into very long-lasting ones through union wage-setting, skill deterioration, and loss of labor-force attachment.[20]

10.6 Search and Matching Models

The final departure of the labor market from Walrasian assumptions that we consider is the simple fact that workers and jobs are heterogeneous. In a frictionless labor market, firms are indifferent about losing their workers, since identical workers are costlessly available at the same wage; likewise, workers are indifferent about losing their jobs. These implications are obviously not accurate descriptions of actual labor markets.

When workers and jobs are highly heterogeneous, the labor market has little resemblance to a Walrasian market. Rather than meeting in centralized markets where employment and wages are determined by the intersections of supply and demand curves, workers and firms meet in a decentralized, one-on-one fashion, and engage in a costly process of trying to match up idiosyncratic preferences, skills, and needs. Since this process is not instantaneous, it results in some unemployment. In addition, it may have implications for how wages and employment respond to shocks.

This section presents a model of firm and worker heterogeneity and the matching process. Because modeling heterogeneity requires abandoning many of our usual tools, even a basic model is relatively complicated. As a result, the model here only introduces some of the issues involved. This class of models is known collectively as the *Mortensen–Pissarides model* (for example, Pissarides, 1985; Mortensen, 1986; Mortensen and Pissarides, 1994; Pissarides, 2000).

Basic Assumptions

The model is set in continuous time. The economy consists of workers and jobs. There is a continuum of workers of mass 1. Each worker can be in one of two states: employed or unemployed. A worker who is employed produces an exogenous, constant amount y per unit time and receives an endogenous and potentially time-varying wage $w(t)$ per unit time. A worker who is unemployed receives an exogenous, constant income of $b \geq 0$ per unit time (or, equivalently, receives utility from leisure that he or she values as much as income of b).

Workers are risk neutral. Thus a worker's utility per unit time is $w(t)$ if employed and b if unemployed. Workers' discount rate is $r > 0$.

[20] For more on these issues, see Siebert (1997); Ljungqvist and Sargent (1998, 2006); Ball (1999a); Blanchard and Wolfers (2000); Prescott (2004); Rogerson (2008); and Alesina, Glaeser, and Sacerdote (2005).

A job can be either filled or vacant. If it is filled, there is output of y per unit time and labor costs of $w(t)$ per unit time. If it is vacant, there is neither output nor labor costs. Any job, either filled or vacant, involves a constant, exogenous cost $c > 0$ per unit time of being maintained. Thus profits per unit time are $y - w(t) - c$ if a job is filled and $-c$ per unit time if it is vacant. y is assumed to exceed $b + c$, so that a filled job produces positive value. Vacant jobs can be created freely (but must incur the flow maintenance cost once they are created). Thus the number of jobs is endogenous.

In the absence of search frictions, the equilibrium of the model is trivial. There is a mass 1 of jobs, all of which are filled. If there were fewer jobs, some workers would be unemployed, and so creating a job would be profitable. If there were more jobs, the unfilled jobs would be producing negative profits with no offsetting benefit, and so there would be exit. Workers earn their marginal product, $y - c$. If they earned more, profits would be negative; if they earned less, creating new jobs and bidding up the wage would be profitable. Thus all workers are employed and earn their marginal products. Shifts in labor demand—changes in y—lead to immediate changes in the wage and leave employment unchanged.

The central feature of the model, however, is that there are search frictions. That is, unemployed workers and vacant jobs cannot find each other costlessly. Instead, the stocks of unemployed workers and vacancies yield a flow of meetings between workers and firms. Let $E(t)$ and $U(t)$ denote the numbers of employed and unemployed workers at time t, and let $F(t)$ and $V(t)$ denote the numbers of filled and unfilled jobs. Then the number of meetings per unit time is

$$M(t) = M(U(t), V(t)), \qquad M_U > 0, M_V > 0. \tag{10.48}$$

This *matching function* proxies for the complicated process of employer recruitment, worker search, and mutual evaluation.

In addition to the flow of new matches, there is turnover in existing jobs. Paralleling the Shapiro–Stiglitz model, jobs end at an exogenous rate λ per unit time. Thus if we assume that all meetings lead to hires, the dynamics of the number of employed workers are given by

$$\dot{E}(t) = M(U(t), V(t)) - \lambda E(t). \tag{10.49}$$

Because of the search frictions, the economy is not perfectly competitive. When an unemployed worker and a firm with a vacancy meet, the worker's alternative to accepting the position is to continue searching, which will lead, after a period of unemployment of random duration, to meeting another firm with a vacancy. Likewise, the firm's alternative to hiring the worker is to resume searching. Thus, collectively, the worker and the firm are strictly better off if the worker fills the position than if he or she does not. Equivalently, the worker's reservation wage is strictly less than his or her marginal revenue product.

One immediate implication is that either workers earn strictly more than their reservation wages or firms pay strictly less than the marginal revenue product of labor, or both. In standard versions of the model, as we will see, both inequalities are strict. Thus even though every agent is atomistic, standard competitive results fail.

Because a firm and a worker that meet are collectively better off if the firm hires the worker, they would be forgoing a mutually advantageous trade if the firm did not hire the worker. Thus the assumption that all meetings lead to hires is reasonable. But this does not uniquely determine the wage. The wage must be high enough that the worker wants to work in the job, and low enough that the firm wants to hire the worker. Because there is strictly positive surplus from the match, there is a range of wages that satisfy these requirements. Thus we need more structure to pin down the wage. The standard approach is to assume that the wage is determined by *Nash bargaining*. That is, there is some exogenous parameter, ϕ, where $0 \leq \phi \leq 1$; the wage is determined by the condition that fraction ϕ of the surplus from forming the match goes to the worker and fraction $1 - \phi$ goes to the firm. The specifics of how this assumption allows us to pin down the wage will be clearer shortly, when we see how to specify the parties' surpluses from forming a match.

The Matching Function

The properties of the matching function are crucial to the model. In principle, it need not have constant returns to scale. When it exhibits increasing returns, there are *thick-market effects*: increases in the resources devoted to search make the matching process operate more effectively, in the sense that it yields more output (matches) per unit of input (unemployment and vacancies). When the matching function has decreasing returns, there are *crowding* effects.

The prevailing view, however, is that in practice constant returns is a reasonable approximation. For a large economy, over the relevant range the thick-market and crowding effects may be relatively unimportant or may roughly balance. Empirical efforts to estimate the matching function have found no strong evidence of departures from constant returns (for example, Blanchard and Diamond, 1989).

The assumption of constant returns implies that a single number, the ratio of vacancies to unemployment, summarizes the tightness of the labor market. To see this, define $\theta(t) = V(t)/U(t)$, and note that constant returns imply

$$\begin{aligned} M(U(t),V(t)) &= U(t)M(1,V(t)/U(t)) \\ &\equiv U(t)m(\theta(t)), \end{aligned} \tag{10.50}$$

where $m(\theta) \equiv M(1,\theta)$. Then the *job-finding rate*—the probability per unit time that an unemployed worker finds a job—is

$$\begin{aligned} a(t) &= \frac{M(U(t),V(t))}{U(t)} \\ &= m(\theta(t)). \end{aligned} \tag{10.51}$$

Similarly, the *vacancy-filling rate* is

$$\begin{aligned} \alpha(t) &= \frac{M(U(t),V(t))}{V(t)} \\ &= \frac{m(\theta(t))}{\theta(t)}. \end{aligned} \tag{10.52}$$

Our assumptions that $M(U,V)$ exhibits constant returns and that it is increasing in both arguments imply that $m(\theta)$ is increasing in θ, but that the increase is less than proportional. Thus when the labor market is tighter (that is, when θ is greater), the job-finding rate is higher and the vacancy-filling rate is lower.

When researchers want to assume a functional form for the matching function, they almost universally assume that it is Cobb-Douglas. We will take that approach here. Thus,

$$m(\theta) = k\theta^{\gamma}, \qquad k > 0, \qquad 0 < \gamma < 1. \tag{10.53}$$

Equilibrium Conditions

As in the Shapiro–Stiglitz model, we use dynamic programming to describe the values of the various states. In contrast to how we analyzed that model, however, we will not impose the assumption that the economy is in steady state from the outset (although we will end up focusing on that case). Let $V_E(t)$ denote the value of being employed at time t. That is, $V_E(t)$ is the expected lifetime utility from time t forward, discounted to time t, of a worker who is employed at t. $V_U(t)$, $V_F(t)$, and $V_V(t)$ are defined similarly.

Since we are not assuming that the economy is in steady state, the "return" on being employed consists of three terms: a "dividend" of $w(t)$ per unit time; the potential "capital gain" on being employed from the fact that the economy is not in steady state, $\dot{V}_E(t)$; and a probability λ per unit time of a "capital loss" of $V_E(t) - V_U(t)$ as a result of becoming unemployed. Thus,

$$rV_E(t) = w(t) + \dot{V}_E(t) - \lambda[V_E(t) - V_U(t)]. \tag{10.54}$$

Similar reasoning implies

$$rV_U(t) = b + \dot{V}_U(t) + a(t)[V_E(t) - V_U(t)], \tag{10.55}$$

$$rV_F(t) = [y - w(t) - c] + \dot{V}_F(t) - \lambda[V_F(t) - V_V(t)], \tag{10.56}$$

$$rV_V(t) = -c + \dot{V}_V(t) + \alpha(t)[V_F(t) - V_V(t)]. \tag{10.57}$$

Four conditions complete the model. First, (10.49) and our assumptions about $M(\bullet)$ describe the evolution of the number of workers who are employed:

$$\dot{E}(t) = U(t)^{1-\gamma} V(t)^{\gamma} - \lambda E(t). \tag{10.58}$$

Second, recall our assumption of Nash bargaining. A worker's surplus from forming a match rather than continuing to work is $V_E(t) - V_U(t)$. Similarly, a firm's surplus from a match is $V_F(t) - V_V(t)$. Thus the Nash bargaining assumption implies

$$V_E(t) - V_U(t) = \frac{\phi}{1-\phi}[V_F(t) - V_V(t)]. \tag{10.59}$$

Third, since new vacancies can be created and eliminated freely,

$$V_V(t) = 0 \qquad \text{for all } t. \tag{10.60}$$

Finally, the initial level of employment, $E(0)$, is given. This completes the description of the model.

Steady-State Equilibrium

Characterizing the full dynamic path of the economy starting from arbitrary initial conditions is complicated by the potentially time-varying paths of the V's. We will therefore focus mainly on the steady state of the model. The assumption that the economy is in steady state implies that all the $\dot{V}(t)$'s and $\dot{E}(t)$ are zero and that $a(t)$ and $\alpha(t)$ are constant.

We solve the model by focusing on two variables, employment (E) and the value of a vacancy (V_V). We will first find the value of V_V implied by a given level of employment, and then impose the free-entry condition that V_V must be zero.

We begin by considering the determination of the wage and the value of a vacancy given a and α. Subtracting (10.55) from (10.54) (with the $\dot{V}(t)$ terms set to zero) and rearranging yields

$$V_E - V_U = \frac{w-b}{a+\lambda+r}. \tag{10.61}$$

Similarly, (10.56) and (10.57) imply

$$V_F - V_V = \frac{y-w}{\alpha+\lambda+r}. \tag{10.62}$$

Our Nash-bargaining assumption (equation [10.59]) implies that $V_E - V_U$ equals $\phi/(1-\phi)$ times $V_F - V_V$. Thus (10.61) and (10.62) imply

$$\frac{w-b}{a+\lambda+r} = \frac{\phi}{1-\phi}\,\frac{y-w}{\alpha+\lambda+r}. \tag{10.63}$$

Solving this condition for w yields

$$w = b + \frac{(a+\lambda+r)\phi}{\phi a + (1-\phi)\alpha + \lambda + r}(y-b). \tag{10.64}$$

To interpret (10.64), first consider the case when a and α are equal. Then the wage is $b+\phi(y-b)$: fraction ϕ of the difference between output and the value of leisure goes to the worker, and fraction $1-\phi$ goes to the firm. When a exceeds α, workers can find new jobs more rapidly than firms can find new employees, and so more of the output goes to the worker. When α exceeds a, the reverse occurs.

Recall that we want to focus on the value of a vacancy. Equation (10.57) states that rV_V equals $-c+\alpha(V_F - V_V)$. Expression (10.62) for $V_F - V_V$ therefore gives us

$$rV_V = -c + \alpha\frac{y-w}{\alpha+\lambda+r}. \tag{10.65}$$

Substituting expression (10.64) for w into this equation and performing straightforward algebra yields

$$rV_V = -c + \frac{(1-\phi)\alpha}{\phi a + (1-\phi)\alpha + \lambda + r}(y-b). \tag{10.66}$$

In this expression, a and α are endogenous. Thus the next step is to express them in terms of E. In steady state, $\dot{E}(t)$ is zero, and so the number of new matches per unit time must equal the number of jobs that end per unit time, λE (equation [10.49]). Thus the job-finding rate, $a = M(U,V)/U$, is given by

$$a = \frac{\lambda E}{1-E}, \tag{10.67}$$

where we use the fact that the mass of workers is 1, so that $E+U=1$.

The vacancy-filling rate, α, is $M(U,V)/V$ (equation [10.52]). We again know that in steady state, $M(U,V)$ equals λE. To express α in terms of E, we therefore need to find the V that implies $M(U,V) = \lambda E$ for a given E. Using the fact that $M(U,V) = kU^{1-\gamma}V^{\gamma}$, we can derive

$$V = k^{-1/\gamma}(\lambda E)^{1/\gamma}(1-E)^{-(1-\gamma)/\gamma}, \tag{10.68}$$

$$\alpha = k^{1/\gamma}(\lambda E)^{(\gamma-1)/\gamma}(1-E)^{(1-\gamma)/\gamma}. \tag{10.69}$$

For our purposes, the key features of (10.67) and (10.69) are that they imply that a is increasing in E and that α is decreasing. Thus (10.66) implies that rV_V is a decreasing function of E. As E approaches 1, a approaches infinity and α approaches zero; hence rV_V approaches $-c$. Similarly, as E approaches zero, a approaches zero and α approaches infinity. Thus in this case rV_V approaches $y-(b+c)$, which we have assumed to be positive. This information is summarized in Figure 10.6.

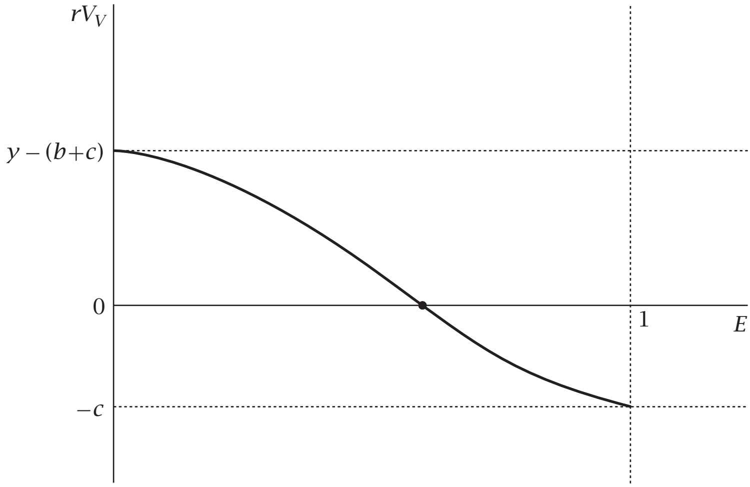

FIGURE 10.6 The determination of equilibrium employment in the search and matching model

The equilibrium level of employment is determined by the intersection of the rV_V locus with the free-entry condition, which implies $rV_V = 0$. Imposing this condition on (10.66) yields

$$-c + \frac{(1-\phi)\alpha(E)}{\phi a(E) + (1-\phi)\alpha(E) + \lambda + r}(y-b) = 0. \tag{10.70}$$

where the functions $a(E)$ and $\alpha(E)$ are given by (10.67) and (10.69). This expression implicitly defines E, and thus completes the solution of the model in the steady-state case.

Extensions

This model can be extended in many directions. Here we discuss a few of the most important.[21]

One major set of extensions are ones that introduce greater heterogeneity. Although search and matching models are motivated by the enormous variety among workers and jobs, the model we have been considering assumes that both workers and jobs are homogeneous. A simple way to introduce heterogeneity and a reason for search and matching is to suppose that when a worker and a job meet, the worker's productivity, y, is not certain but is drawn from some distribution. This assumption implies that if the realized level of productivity is too low, the meeting does not lead to a match being formed but to continued search by both sides. Moreover, the

[21] Many of these extensions are surveyed by Rogerson, Shimer, and Wright (2005).

cut-off level of productivity is endogenous, so that the fraction of meetings that lead to jobs depends on the underlying parameters of the economy and may be time-varying. Similarly, if the worker's productivity in the job is subject to shocks, the break-up rate, which is exogenous and constant in the basic model, can be endogenized.

Another extension in the same spirit is to allow workers to continue searching even when they are employed and firms to continue searching even when their positions are filled. The result is that some of workers' transitions are directly from one job to another and that firms sometimes replace a worker with another.

Another important set of extensions involves making the process of search and information flow more sophisticated. In the basic model, search is completely random. But in practice, workers have some information about jobs, and they focus their search on the jobs that look most appealing. That is, to some extent search is not random but *directed*. Likewise, firms and workers generally do not bargain over compensation from scratch each time a worker is hired; many firms have wage policies that they are to some extent committed to. That is, to some extent wages are *posted*. Since one effect of posting wages is to affect workers' search, it is natural to combine the assumption that wages are posted with the assumption that search is directed. Such models are known as *competitive search models*.

10.7 Implications

Unemployment

Search and matching models offer a straightforward explanation for average unemployment: it may be the result of continually matching workers and jobs in a complex and changing economy. Thus, much of observed unemployment may reflect what is traditionally known as *frictional* unemployment.

Labor markets are characterized by high rates of turnover. In U.S. manufacturing, for example, more than 3 percent of workers leave their jobs in a typical month. Moreover, many job changes are associated with wage increases, particularly for young workers (Topel and Ward, 1992); thus at least some of the turnover appears to be useful. In addition, there is high turnover of jobs themselves. In U.S. manufacturing, at least 10 percent of existing jobs disappear each year (Davis and Haltiwanger, 1990, 1992). These statistics suggest that a nonnegligible portion of unemployment is a largely inevitable result of the dynamics of the economy and the complexities of the labor market.

Unfortunately, it is difficult to go much beyond this general statement. Existing theoretical models and empirical evidence do not provide any clear way of discriminating between, for example, the hypothesis that search and