Search-Theoretic Models of the Labor Market: A Survey

RICHARD ROGERSON, ROBERT SHIMER, and RANDALL WRIGHT*

We survey the literature on search-theoretic models of the labor market. We show how this approach addresses many issues, including the following: Why do workers sometimes choose to remain unemployed? What determines the lengths of employment and unemployment spells? How can there simultaneously exist unemployed workers and unfilled vacancies? What determines aggregate unemployment and vacancies? How can homogeneous workers earn different wages? What are the trade-offs firms face from different wages? How do wages and turnover interact? What determines efficient turnover? We discuss various modeling choices concerning wage determination and the meeting process, including recent models of directed search.

1. Introduction

The economic fortunes of most individuals are largely determined by their labor market experiences—that is, by paths for their wages, their employers, and their intervening spells of unemployment or non-employment. Hence, economists are naturally interested in documenting the empirical behavior of wages, employment, and unemployment, and also in building models to help us understand the forces that shape these outcomes and using the models to assess the consequences of changes in policies or institutions.

While the usual paradigm of supply and demand in a frictionless labor market is useful for discussing some issues, many important questions are not easily addressed with this approach. Why do unemployed workers sometimes choose to remain unemployed, say by turning down job offers? What determines the lengths of employment and unemployment spells? How can we simultaneously have unemployed workers and unfilled vacancies? What factors determine the aggregate unemployment and vacancy rates? How can apparently homogeneous workers in similar jobs end up earning different wages? What are the trade-offs faced by firms in paying different wages? How do wages and turnover interact? What determines the efficient amount of turnover?

From its inception, search theory has provided a rigorous yet tractable framework that can be used to address these and related questions. Central to the approach
is the notion that trading frictions are important. It takes time and other resources for a worker to land a job, especially a good job at a good wage, and for a firm to fill a vacancy. There is simply no such thing as a centralized market where buyers and sellers of labor meet and trade at a single price, as assumed in classical equilibrium theory. Of course, economic models do not have to be realistic to be useful, and the supply-and-demand paradigm is obviously useful for studying many issues in labor economics. But it is equally clear that the simple supply-and-demand approach is ill suited for discussing questions such as those raised in the previous paragraph.

We argue that even the earliest search models, with their focus on a single worker, enhance our ability to organize observations on employment histories. We then examine more recent research that embeds the decision-theoretic model into an equilibrium framework. Although there are several important modeling decisions in equilibrium search theory, we argue that two questions are paramount. First, how do agents meet? In particular, is search random, so that unemployed workers are equally likely to locate any job opening, or directed, so that for example firms can attract more applicants by offering higher wages? Second, exactly how are wages determined? Do matched workers and firms bargain, or are wages posted unilaterally before they meet? We consider various alternative sets of assumptions and indicate how the choice affects predictions. We also emphasize how a common set of methods and ideas are used in all of the different approaches.

Before proceeding, we mention that search theory constitutes a very large field. In addition to labor, it has been used in monetary theory, industrial organization, finance, the economics of the marriage market, and other areas, all of which we must neglect lest this survey becomes unmanageable.1 Search has been used in much fairly technical theoretical research, and has also been a workhorse for empirical economics, but we can neither delve into pure theory nor pay attention to all of the econometric issues and empirical findings here.2 Also, while we strive to be rigorous, we emphasize issues rather models or methods per se. Hence the presentation revolves around the ways in which the framework helps us think about substantive questions like those mentioned above.

The logical structure of the paper is as follows. We begin in section 2 with the problem of a single agent looking for a job, not only because this is the way the literature started, but because it is a building block for the equilibrium analysis to follow. Even this rudimentary model is consistent with two facts that do not come out of frictionless models: it takes time to find an acceptable job; and what one ends up with is at least partially a matter of luck, which means similar agents may end up with different wages. In section 3, we describe some generalizations of this model designed to help understand turnover and labor market transitions. This section also introduces tools and techniques needed for equilibrium search theory.

We also spend some time here discussing the logical transition between decision theory and equilibrium theory. We first argue that one can reinterpret the single-agent

1 Examples in monetary economics include Nobuhiro Kiyotaki and Randall Wright (1993), Shouyong Shi (1995), and Alberto Trejos and Wright (1995); examples in the marriage literature include Dale T. Mortensen (1988), Kenneth Burdett and Melynn G. Coles (1997, 1999), and Robert Shimer and Lones Smith (2000); examples in IO include Steven C. Salop (1977), Boyan Jovanovic (1982), and Jovanovic and Glenn M. MacDonald (1994); examples in finance include Darrell Duffie, Nicole Garleanu, and Lasse Pedersen (2002) and Pierre-Olivier Weill (2004).

2 Examples of theoretical work studying the question of whether frictionless competitive equilibrium is the limit of search equilibrium as the frictions get small include Ariel Rubinstein and Asher Wolinsky (1985, 1990), Douglas Gale (1987), and Mortensen and Wright (2002). Theresa J. Devine and Nicholas M. Kiefer (1991), Kenneth I. Wolpin (1995), and Zvi Eckstein and Gerard J. van den Berg (forthcoming) survey empirical work.
model as an equilibrium of a simple economy. But in such a model several key variables, including the arrival rate and distribution of wage offers, are essentially fixed exogenously. For some issues, one may want to know how these are determined in equilibrium, and in particular how they are affected by labor market conditions or labor market policy. There is no single way to proceed, but any approach requires us to confront the two questions mentioned above: how do workers and firms meet; and how do they determine wages?

In section 4, we present a class of equilibrium models built on two main ingredients: the matching function, which determines how workers and firms get together, and the bargaining solution, which determines wages once they do. The matching function helps us in analyzing transitions from unemployment to employment as a function of, in general, the behavior of all the workers and firms in the market. Bargaining is one of the more popular approaches to wage determination in the literature, and since it is a key ingredient in many models we take some time to explain how the generalized Nash solution works and how to interpret it in terms of strategic bargaining theory.

In section 5, we consider an environment where wages are posted ex ante, rather than bargained after agents meet, and in addition where search is directed—i.e., workers do not encounter firms completely at random but try to locate those posting attractive terms of trade. Models with the combination of wage posting and directed search, called competitive search models, behave quite differently from those in section 4, although they can be used to analyze similar issues. In section 6, we consider models where wages are posted but search is once again purely random. This class of models has been widely used in research on wage dispersion and the distribution of individual employment/unemployment spells.

In section 7, we discuss efficiency. This is important because, to the extent that one wants to analyze policy, one would like to know whether the models rationalize a role for intervention in a decentralized economy. It is fine to say, for example, that a change in some variable reduces unemployment, but it is obviously relevant to know whether this improves welfare. We derive conditions under which equilibrium with bargaining is efficient. We also show that the combination of wage posting and directed search generates efficiency under quite general assumptions, providing a version of the first welfare theorem for economies with frictions. Finally, we finish in section 8 with a few concluding remarks.3

2. Basic Job Search

We begin here with the familiar discrete-time formulation of the basic job search model, and then derive its continuous-time analogue, which is used for the remainder of the paper. We then use the framework to discuss some issues relating to unemployment duration and wages.

2.1 Discrete Time

Consider an individual searching for a job in discrete time, taking market conditions as given.4 He seeks to maximize $\sum_{t=0}^{\infty} \beta^t x_t$.

---

3 Earlier surveys of search theory as applied to labor markets include Steven A. Lippman and John J. McCall (1976a), Mortensen (1986), and Mortensen and Christopher A. Pissarides (1990a, 1990b). While there is naturally some overlap, there are important differences in our approach. We build equilibrium models up from the decision problem, focusing on the role of random versus directed search and the role of bargaining versus wage posting. In particular, previous surveys do not examine directed search models, or explicitly discuss the efficiency properties of different models, as we do.

4 While it is often said that the economics of search began with George Stigler (1961), his formulation was static. The sequential job search model in this section was developed first by McCall (1970), Mortensen (1970), and Reuben Gronau (1971), although others had analyzed related problems, including Herbert A. Simon (1955), who discussed the housing market, and Samuel Karlin (1962), who discussed asset markets. The presentation in this and the next section is based on some lecture notes that contain many more details, examples and exercises, and can be found at http://www.ssc.upenn.edu/~rwright/courses/courses.html.
where $\beta \in (0, 1)$ is the discount factor, $x$, is income at $t$, and $\mathbb{E}$ denotes the expectation. Income is $x = w$ if employed at wage $w$ and $x = b$ if unemployed. Although we refer to $w$ as the wage, more generally it could capture some measure of the desirability of the job, depending on benefits, location, prestige, etc., and although we refer to $b > 0$ as unemployment insurance (UI), it can also include the value of leisure or home production.\(^5\)

We begin with the case where an unemployed individual samples one independently and identically distributed (i.i.d.) offer each period from a known distribution $F(w)$. If an offer is rejected, the agent remains unemployed that period. Assume previously rejected offers cannot be recalled, although this is actually not restrictive because the problem is stationary, so an offer that is not acceptable today will not be acceptable tomorrow. For now we assume that if a job is accepted the worker keeps it forever. Hence, we have the Bellman equations

\begin{align*}
(1) \quad W(w) &= w + \beta W(w) \\
(2) \quad U &= b + \beta \int_0^w \max\{U, W(w)\} dF(w),
\end{align*}

where $W(w)$ is the payoff from accepting a wage $w$ ($W$ stands for working) and $U$ is the payoff from rejecting a wage offer, earning $b$, and sampling again next period ($U$ stands for unemployed).

Since $W(w) = w/(1 - \beta)$ is strictly increasing, there is a unique $w_R$, called the reservation wage, such that $W(w_R) = U$, with the property that the worker should reject $w < w_R$ and accept $w \geq w_R$ (we adopt the convention that he accepts when indifferent). Substituting $U = w_R/(1 - \beta)$ and $W(w) = w/(1 - \beta)$ into (2), we have

\begin{align*}
(3) \quad w_R &= T(w_R) \\
&= (1 - \beta)b + \beta \int_0^\infty \max\{w, w_R\} dF(w).
\end{align*}

The function $T$ is easily shown to be a contraction, so there is a unique solution to $w_R = T(w_R)$. This implies that if one fixes $w_0$ and recursively defines $w_{N+1} = T(w_N)$, the sequence converges to $w_R$ as $N \to \infty$. If the initial wage is $w_0 = b$, the worker's reservation wage in the final period of a finite horizon problem, $w_N$ has the interpretation of being the reservation wage when $N$ periods of search remain, after which the worker receives either $b$ or the accepted wage $w$ forever.

The optimal search strategy is completely characterized by (3), but we also present some alternative representations that are often seen in the literature. First, subtracting $\beta w_N$ from both sides of (3) and simplifying gives the standard reservation wage equation

\begin{align*}
(4) \quad w_R &= b + \frac{\beta}{1 - \beta} \int_{w_R}^\infty (w - w_R) dF(w).
\end{align*}

Using integration by parts, we can also write this as

\begin{align*}
(5) \quad w_R &= b + \frac{\beta}{1 - \beta} \int_{w_R}^\infty [1 - F(w)] dw,\nonumber
\end{align*}

which, as we shall see, is handy in some of the applications below.\(^6\)

\(^5\)Our worker is interested in maximizing expected discounted income. This is the same as maximizing expected utility if he is risk neutral, but also if he is risk averse and consumption markets are complete, since then he can maximize utility by first maximizing income and then smoothing consumption. The case of a risk averse agent facing incomplete markets is more difficult. Early analyses include John P. Danforth (1979) and John R. Hall, Lippman, and McCall (1979); more recent studies include Victor Valdivia (1996), James Costain (1997), Daron Acemoglu and Shimer (1999b, 2000), Martin Browning, Thomas F. Crossley, and Eric Smith (2003), and Rasmus Lento and Torben Tranæs (forthcoming).

\(^6\)Note that in the above analysis, as in most of what we do here, it is assumed that the worker knows $F$. If he has to learn about $F$ while searching, the problem gets harder and a reservation strategy may not even be optimal. For example suppose we know either: (a) $w = w_i$ with prob $1$; or (b) $w_i = w_i$ with prob $\pi$ and $w_i = w_i$ with prob $1 - \pi$. If $w_i > w_i > w_i$ and $\pi$ is small, it can be optimal to accept $w_i$ but not $w_i$ since an offer of $w_i$ signals that there is a good chance of getting $w_i$. Michael Rothschild (1974) gives conditions that guarantee a reservation strategy is optimal. See Burdett and Tara Vishwanath (1988a) for additional discussion and references.
2.2 Continuous Time

We now derive continuous-time versions of the above results. First, generalize the discrete-time model to allow the length of a period to be \( \Delta \). Let \( \beta = \frac{1}{1+r\Delta} \) and assume that the worker gets a wage offer with probability \( \alpha \Delta \) in each period.\(^7\) Then the payoffs to working and unemployment satisfy the following versions of (1) and (2)

\[
\begin{align*}
W(w) &= \Delta w + \frac{1}{1+r\Delta} W(w) \\
U &= \Delta b + \frac{\alpha \Delta}{1+r\Delta} \\
&\times \int_0^\infty \max(U, W(w)) dF(w) + \frac{1-\alpha \Delta}{1+r\Delta} U.
\end{align*}
\]

Algebra implies

\[
\begin{align*}
(6) \quad W(w) &= \Delta w + \frac{1}{1+r\Delta} W(w) \\
(7) \quad U &= \Delta b + \frac{\alpha \Delta}{1+r\Delta} \\
&\times \int_0^\infty \max(U, W(w)) dF(w) + \frac{1-\alpha \Delta}{1+r\Delta} U.
\end{align*}
\]

When \( \Delta \to 0 \), we obtain the continuous time Bellman equations

\[
\begin{align*}
(8) \quad rW(w) &= (1+r\Delta) w \\
(9) \quad rU &= (1+r\Delta) b \\
&+ \alpha \int_0^\infty \max(0, W(w) - U) dF(w).
\end{align*}
\]

Intuitively, while \( U \) is the value of being unemployed, \( rU \) is the flow (per period) value. This equals the sum of the instantaneous payoff \( b \), plus the expected value of any changes in the value of the worker's state, which in this case is the probability that he gets an offer \( \alpha \), times the expected increase in value associated with the offer, noting that the offer can be rejected.

The reservation wage \( w_R \) satisfies \( W(w_R) = U \), so equation (10) implies \( W(w) - U = (w - w_R)/r \). Substituting this into (11) gives the continuous time reservation wage equation

\[
\begin{align*}
(12) \quad w_R &= b + \frac{\alpha}{r} \int_{w_R}^\infty (w - w_R) dF(w).
\end{align*}
\]

Again one can integrate by parts to get

\[
\begin{align*}
(13) \quad w_R &= b + \frac{\alpha}{r} \int_{w_R}^\infty [1-F(w)] \, dw.
\end{align*}
\]

Although most of the models that we discuss assume fixed search intensity, it can be endogenized. Suppose a worker can affect the arrival rate of offers \( \alpha \), at cost \( g(\alpha) \), where \( g'>0 \) and \( g''>0 \). Unemployed workers choose \( \alpha \) to maximize \( rU = w_R \), where

\[
\begin{align*}
(14) \quad w_R &= b - g(\alpha) + \frac{\alpha}{r} \int_{w_R}^\infty (w - w_R) dF(w).
\end{align*}
\]

The first order condition for an interior solution is

\[
\begin{align*}
(15) \quad \int_{w_R}^\infty (w - w_R) dF(w) &= rg'(\alpha).
\end{align*}
\]

Worker behavior is characterized by a pair \((w_R, \alpha)\) solving (14) and (15). It easy to show that an increase in \( b \), e.g., raises \( w_R \) and reduces \( \alpha \).

2.3 Discussion

Traditional frictionless models assume that a worker can costlessly and immediately choose to work for as many hours as he wants at the market wage. By relaxing these extreme assumptions, search models allow us to think about unemployment and wages in a different light. Consider unemployment duration. The probability that the worker has not found a job after a spell of length \( t \) is \( e^{-Ht} \), where \( H = \alpha [1 - F(w_R)] \) is called the hazard rate and equals the product of the contact rate \( \alpha \) and the probability of accepting...
1 - F(w_b).\footnote{Because this simple model is stationary, H does not change with the duration of unemployment. Several generalizations would overturn this, the simplest being to assume a finite horizon. More interestingly, Burdett (1979) and Mortensen (1977) allow UI to vary over time. Lippman and McCall (1976b) and Lippman and John W. Mamer (1989) allow wage offers to vary over time. Salop (1973) studies systematic search where a worker first looks for opportunities that are best according to some prior and then, if unsuccessful, proceeds to other locations, typically lowering his reservation wage over time. The models discussed earlier in which the worker learns about the wage distribution also predict w_R and hence H vary over time. Bruce D. Meyer (1990) and Wolpin (1987) are examples of a large body of empirical work on hazard rates.} The average duration of an unemployment spell is therefore
\begin{equation}
D = \int_0^\infty tH e^{-\alpha t} dt = \frac{1}{H}.
\end{equation}
Also, the observed distribution of wages paid is G(w) = F(w | w \geq w_R).

Consider the impact of an increase in b, say more generous UI, assuming for simplicity here that search intensity and hence \alpha are fixed. From (12), the immediate effect is to increase w_R, which has two secondary effects: the distribution of observed wages G(w) is higher in the sense of first order stochastic dominance, since more low wage offers are rejected; and the hazard rate H is lower, which increases average unemployment duration. Hence, even this elementary model makes predictions about variables that would be difficult to generate using a theory without frictions.\footnote{We can also study changes in F or \alpha. As pointed out by Burdett (1981), for some experiments it is useful to assume F is log-concave. Suppose we increase every w in F either by a constant or proportionally. Perhaps surprisingly, \mathbb{E}[w | w \geq w_R] may actually decrease, but it can be guaranteed to increase under log-concavity; see Mortensen (1986) or Wright and Janine Loberg (1987). Suppose we increase the arrival rate. One might expect that this must raise the hazard, but since it increases \alpha, the net effect is ambiguous. One can show \partial H/\partial \alpha > 0 under log-concavity; see Christopher J. Flinn and James J. Heckman (1983), Burdett and Jan Ondrich (1985), or van den Berg (1994).}

3. Worker Turnover

Although the model in the previous section is interesting, there are important issues that it cannot address. For instance, we assumed above that when a worker accepts a job he keeps it forever. Yet according to Bruce Fallick and Charles A. Fleischman (2004), in the United States from 1994 to 2004, 6.6 percent of employment relationships ended in a given month (of these, forty percent of workers switched employers while the rest either became unemployed or left the labor force). We now generalize the framework to capture such transitions.

3.1 Transitions to Unemployment

The simplest way of generating transitions from employment into unemployment is to assume that jobs end for some exogenous reason; sometimes in the literature this is interpreted by saying that workers face layoff risk. A tractable formulation is to assume that this occurs according to a Poisson process with parameter \lambda, which for now is an exogenous constant.\footnote{An interesting extension of the basic model is to let \lambda vary across jobs, which implies the reservation strategy generally depends on the pair (w,\lambda); see Burdett and Mortensen (1980) or Wright (1987).}

Introducing exogenous separations does not affect the Bellman equation for U, which is still given by (11), but now we have to generalize (10) to
\begin{equation}
rW(w) = w + \lambda [U - W(w)].
\end{equation}
The reservation wage still satisfies W(w_R) = U, and the methods leading to (13) yield
\begin{equation}
w_R = b + \frac{\alpha}{r + \lambda} \int_{w_R}^\infty [1 - F(w)] dw.
\end{equation}
Notice that \lambda affects w_R only by changing the effective discount rate to r + \lambda. However, a worker now goes through repeated spells of employment and unemployment: when unemployed, he gets a job at rate H = \alpha [1 - F(w_R)], and while an employed he loses the job at rate \lambda.

A simple way to endogenize transitions to unemployment is to allow w to change at a given job. Suppose that this happens according to a Poisson process with parameter \lambda, and that in the event of a wage change a new
search. Suppose new offers arrive at rate \( \lambda \) whenever \( w > w_1 \). This implies \( W(w) \) is increasing, and there is a reservation wage \( w_R \) such that unemployed workers accept if \( w \geq w_R \) and employed workers quit if their wage falls to \( w < w_R \). Hence separations are decreasing in \( w \). In the simplest case where \( F(w' | w) = F(w) \) (independence), we have

\[
W(w) = w + \lambda \int_0^\infty \max \{ W(w') - W(w) \} dF(w' | w).
\]

A natural assumption is that \( F(w' | w) \) first order stochastically dominates \( F(w' | w_1) \) whenever \( w_2 > w_1 \). This implies \( W(w) \) is increasing, and there is a reservation wage \( w_R \) such that unemployed workers accept if \( w \geq w_R \) and employed workers quit if their wage falls to \( w < w_R \). Hence separations are decreasing in \( w \). In the simplest case where \( F(w' | w) = F(w) \) (independence), we have

\[
W(w) = w + \lambda \int_0^\infty \max \{ W(w') - W(w) \} dF(w).
\]

Notice that \( \lambda > \alpha \) implies \( w_R < b \); in this case, workers accept a job paying less than unemployment income and wait for the wage to change, rather than searching while unemployed. In any case the usual comparative static results, such as \( \partial w_R / \partial b > 0 \), are similar to what we found earlier.

3.2 Job-to-Job Transitions

To explain how workers change employers without an intervening spell of unemployment, we need to consider on-the-job search.\textsuperscript{11} Suppose new offers arrive at rate \( \alpha_o \) while unemployed and \( \alpha_i \) while employed. Each offer is an i.i.d. draw from \( F \). Assume that employed workers also lose their job exogenously at rate \( \lambda \). The Bellman equations are

\[
U = b + \alpha_o \int_{w_R}^\infty \left[ W(w') - U \right] dF(w')
\]

(22) \( rW(w) = w + \alpha_i \int_0^\infty \max \{ W(w') \} \]

\[ - W(w) - [U - W(w)] dF(w'). \]

The second term in (22) represents the event that an employed worker gets an offer above his current wage.

It is easy to see that \( W \) is increasing, implying that unemployed workers use a reservation wage satisfying \( W(w_R) = U \), and employed workers switch jobs whenever \( w > w_R \). Evaluating (22) at \( w = w_R \) and combining it with (21), we get

\[
w_R = b + (\alpha_o - \alpha_i) \int_{w_R}^\infty [W(w') - W(w_R)] dF(w').
\]

Observe that \( w_R > b \) if and only if \( \alpha_o > \alpha_i \). Thus, if a worker gets offers more frequently when employed than when unemployed, he is willing to accept wages below \( b \).

To eliminate \( W \) from (23), use integration by parts and insert \( W'(w) = \frac{[r + \lambda + \alpha_i [1 - F(w)]^{-1}, \] which we get by differentiating (22), to yield

\[
w_R = b + (\alpha_o - \alpha_i) \int_{w_R}^\infty \frac{1 - F(w)}{r + \lambda + \alpha_i [1 - F(w)]} dw.
\]

If \( \alpha_i = 0 \), this reduces to the earlier reservation wage equation (13). Many results, like \( \partial w_R / \partial b > 0 \), are similar to what we found above, but we also have some new predictions. For instance, when \( w_0 \) is higher, workers are less likely to accept a low \( w \), so they are less likely to experience job-to-job transitions. Thus, an increase in UI reduces turnover.\textsuperscript{12}

\textsuperscript{11} The on-the-job search model was introduced by Burdett (1978). The presentation here follows Mortensen and George R. Neumann (1984).

\textsuperscript{12} As in the basic model, we can endogenize search intensity here. Let \( g_i(\alpha) \) be the cost of achieving \( \alpha \) for an unemployed worker and \( g_i(\alpha) \) the cost for an employed worker. If \( g_i(\alpha) \leq g_e(\alpha) \) for all \( \alpha \), unemployed workers search harder than employed workers. In any case, search intensity decreases with \( w \) for employed workers.
3.3 Discussion

The model of the last subsection is a natural framework in which to analyze individual transitions between employment and unemployment, and between employers. It makes predictions about the relationships between wages, tenure, and separation rates. For example, workers typically move up the wage distribution during an employment spell, so the time since a worker was last unemployed is positively correlated with his wage. Also, workers who earn higher wages are less likely to get better opportunities, generating a negative correlation between wages and separation rates. And the fact that a worker has held a job for a long time typically means that he is unlikely to obtain a better one, generating a negative relationship between job tenure and separation rates. All of these features are consistent with the empirical evidence on turnover and wage dynamics summarized in, e.g., Henry Farber (1999).13

With a slight reinterpretation, the framework can also be used to discuss aggregate variables. Suppose there are many workers, each solving a problem like the one discussed above, with the various stochastic events (like offer arrivals) i.i.d. across workers. Each unemployed worker becomes employed at rate $H = \alpha_0[1 - F(w_R)]$ and each employed worker loses his job at rate $\lambda$, so the aggregate unemployment rate $u$ evolves according to

$$u = \dot{u} = \lambda(1 - u) - \alpha_0[1 - F(w_R)]u.$$ 

Over time, this converges to the steady state

$$u^* = \frac{\lambda}{\lambda + \alpha_0[1 - F(w_R)]}.$$

One can also calculate the cross-sectional distribution of observed wages for employed workers, denoted by $G(w)$, given any offer distribution $F(w)$. For all $w \geq w_R$, the flow of workers into employment at a wage no greater than $w$ is $u\alpha_0[F(w) - F(w_R)]$, equal to the number of unemployed workers times the rate at which they find a job paying between $w_R$ and $w$. The flow of workers out of this state is $(1 - u)G(w)[\lambda + \alpha_0[1 - F(w R)]]$, equal to the number of workers employed at $w$ or less, times the rate at which they leave either for exogenous reasons or because they get an offer above $w$. In steady state, these flows are equal. Using (25) and rearranging, we have

$$G(w) = \frac{\lambda[F(w) - F(w_R)]}{[1 - F(w_R)][\lambda + \alpha_0[1 - F(w)]]}.$$ 

We can now compute the steady state job-to-job transition rate,

$$\alpha_0\int_{w^*}^{w_R} [1 - F(w)]dG(w).$$

This type of model has been used in a variety of applications. For example, Wright (1986) uses a version with learning to discuss macro aggregates, showing how the combination of search and learning generates considerable persistence in the unemployment rate. Under the interpretation that the learning comes from a signal-extraction problem, where workers see nominal wages and have to learn the real wage, the model also generates a Phillips curve relation between inflation and unemployment. Unlike previous signal-extraction models, such as Robert E. Lucas (1972), unemployment is persistent because search provides a natural propagation mechanism.

Jovanovic (1987) considers a variation that allows workers who are not satisfied with...
their current wage to either search for a better one, or to rest and return to work when $w$ increases. This model can generate procyclical quits and productivity, along with countercyclical unemployment, as in the data. Wright and Loberg (1987) use another variation to analyze the impact of changes in taxes on unemployment and wages of workers at different points in the skill distribution. Ljungqvist and Sargent (1998) add human capital accumulation during employment and deaccumulation during unemployment and study the effects of policy. Kambourov and Manovskii (2005) use a version of the model to study life-cycle earnings profiles.

Although these applications all seem interesting, there is an old critique of the framework—which is that it is partial equilibrium because the distribution $F$ is exogenous (Rothschild 1973). From a logical point of view, this critique is not compelling, as there are various ways to embed the model into an equilibrium context without changing the results. For instance, one can imagine workers looking for wages as fishermen looking for lakes. Then $F$ is simply the distribution of fish across lakes, which is something we can logically take as fixed with respect to most policy interventions.

Another approach that involves only a small change in interpretation is to invoke the island metaphor often used in search theory. Imagine workers searching across islands, on each of which there are many firms with a constant returns to scale technology using only labor. The productivity of a randomly selected island is distributed according to $F$. If the labor market on each island is competitive, a worker on an island with productivity $w$ is paid $w$. This trivially makes the equilibrium wage distribution the same as the productivity distribution, $F$. In fact, this is the Lucas and Edward C. Prescott (1974) equilibrium search model, aside from some minor details—e.g., they allow for decreasing returns to scale, which complicates the algebra but does not change the idea.14

In summary, we think it is silly to criticize the framework as being partial equilibrium per se, since it is trivial to recast it as an equilibrium model without changing the essence. The more pertinent question is, do we miss anything of substance with these stories about lakes and islands? The models that we present below introduce firms and equilibrium considerations using more economics and less geography.

4. Random Matching and Bargaining

This section introduces a popular line of research, emanating from Pissarides (1985, 2000), used to study the determinants of arrival rates, match formation, match dissolution, and wages. We present a sequence of models that emphasize different margins, including entry, the decision to consummate matches, and the decision to terminate matches. Before we begin, there are two issues that need to be addressed: how do workers and firms meet, and how are wages determined. These models assume meetings are determined through a matching function and wages through bargaining.

4.1 Matching

Suppose that at some point in time there are $v$ vacancies posted by firms looking for workers and $u$ unemployed workers looking for jobs. Building on ideas in Peter A. Diamond (1981, 1982a, 1982b), Mortensen (1982a, 1982b), Pissarides (1984, 1985), and elsewhere, assume the flow of contacts between firms and workers is given by a matching technology $m = m(u, v)$. Assuming all workers are the same and all firms are the same, the arrival rates for unemployed workers and employers with vacancies are then given by

Early empirical work on matching goes back to Pissarides (1986) and Olivier J. Blanchard and Diamond (1989); see Barbara Petrongolo and Pissarides (2001) for a recent survey.

An interesting alternative is to assume the number of matches depends on both the flows of newly unmatched workers and firms and the stocks of existing unemployment and vacancies, as in Coles and Smith (1996; 1998). See Ricardo Lagoa (2000) for another approach.

\begin{equation}
\alpha_w = \frac{m(u,v)}{u} \quad \text{and} \quad \alpha_v = \frac{m(u,v)}{v}.
\end{equation}

It is standard to assume the function \( m \) is continuous, nonnegative, increasing in both arguments and concave, with \( m(u,0) = m(0,v) = 0 \) for all \((u,v)\). It is also convenient to assume \( m \) displays constant returns to scale, i.e., \( \lambda m(u,v) = m(\lambda u, \lambda v) \).

While alternative assumptions are interesting—e.g., Diamond emphasizes that increasing returns can generate multiple equilibria—constant returns is consistent with much empirical work.\(^{15}\) Also, constant returns generates a big gain in tractability, as it implies \( \alpha_w \) and \( \alpha_v \) depend only on the ratio \( v/u \), referred to as a measure of market tightness. Thus, \( \alpha_w \) is an increasing and \( \alpha_v \) a decreasing function of \( v/u \), and there is a continuous, decreasing, 1 to 1 relationship between \( \alpha_w \) and \( \alpha_v \).

The matching technology is meant to represent in a simple, if reduced-form, fashion the notion that it takes time for workers and firms to get together. Just as a production function maps labor and capital into output, \( m \) maps search by workers and firms into matches. There are papers that model this more deeply, some of which we discuss below, but starting with an exogenous matching function allows us to be agnostic about the actual mechanics of the process by which agents make contact. An advantage is that this is a flexible way to incorporate features that seem desirable—e.g., more search by either side of the market yields more matches—and one can regard the exact specification as an empirical issue. This might make matching a bit of a black box, but it is a common and useful approach.\(^{16}\)

\[ w \in \arg\max [W(w) - U]^\theta \times [J(y - w) - V]^{1-\theta}, \]

where \( \theta \in (0,1) \) is the worker’s bargaining power. The solution to the maximization problem satisfies

\[ \theta J'(y - w) - V W'(w) = (1 - \theta) [W(w) - U] J'(y - w), \]

which can be solved for \( w \). Since it is an important building block in this class of models, and since wage determination is one of the main themes of this essay, we want to discuss Nash bargaining carefully.

John Nash (1950) did not actually analyze the bargaining process, but took as given four simple axioms and showed that his solution is the unique outcome satisfying these axioms.\(^{17}\) The solution, while elegant and

\(^{15}\) Early empirical work on matching goes back to Pissarides (1986) and Olivier J. Blanchard and Diamond (1989); see Barbara Petrongolo and Pissarides (2001) for a recent survey.

\(^{16}\) An interesting alternative is to assume the number of matches depends on both the flows of newly unmatched workers and firms and the stocks of existing unemployment and vacancies, as in Coles and Smith (1996; 1998). See Ricardo Lagoa (2000) for another approach.

\(^{17}\) Nash actually showed that the unique outcome consistent with his axioms has \( \theta = 1/2 \). Relaxing his symmetry axiom, (R-Nash) with any \( \theta \in (0,1) \) satisfies the other axioms, and this is what is called the generalized Nash solution. See, e.g., Martin J. Osborne and Rubinstein (1990).
practical, is again a black box. However, one can provide a game-theoretic description of the bargaining process, along the lines of Rubinstein (1982), that has a unique subgame perfect equilibrium with the following property: as the time between counteroffers in the game becomes small, the equilibrium outcome converges to the prediction of the Nash solution for particular choices of the threat points and bargaining power that depend on details of the underlying game; see e.g., Osborne and Rubinstein (1990).

For instance, suppose that each agent has a given probability of proposing an offer (as opposed to responding) in each round of bargaining; everything else equal, this game generates the same outcome as the Nash solution in which the bargaining power equals that probability.

Of course this only pushes \( \theta \) back one level—where does that probability come from? One position is to say that the nature of bargaining may well differ across industries, countries, and so on, and varying \( \theta \) is one way to try and capture this. Also, at least in simple models, as we vary \( \theta \) between 0 and 1 we trace out the set of bilaterally efficient and incentive compatible employment relationships, which would seem to cover the cases of interest. Just as the matching function is not the last word on how people meet, Nash bargaining is not the last word on wage determination, but it is a useful approach.

To proceed, suppose as usual that workers and firms are risk neutral, infinitely lived, and discount future payoffs in continuous time at rate \( r \), and that matches end exogenously at rate \( \lambda \). Then we have

\[
W(w) = w + \lambda [U - W(w)] 
\]

where the last equality is derived by using (31) and (32).

From (31) and (32) we have \( W(w) - U = \frac{w - \pi_w}{r + \lambda} \) and \( f(\pi) - V = \frac{\pi - \pi_w}{r + \lambda} \), where \( \pi_w \) and \( \pi_R \) are reservation wage and profit levels for the worker and firm. Then (29) reduces to

\[
w \in \arg \max \{w - \pi_w \}^{\theta} \left[ y - w - \pi_R \right]^{1-\theta},
\]

which has solution

\[
w = w_R + \theta(y - \pi_R - w_R).
\]

This says that in terms of total lifetime expected utility, the worker receives his threat point \( U \) plus a share of the surplus, denoted \( S \) and defined by

\[
S = f(y - w) - V + W(w) - U
\]

This says that in terms of total lifetime expected utility, the worker receives his threat point \( U \) plus a share of the surplus, denoted \( S \) and defined by

\[
S = f(y - w) - V + W(w) - U
\]

where the last equality is derived by using (31) and (32).

From (31) and (32) we have \( W(w) - U = \frac{w - \pi_w}{r + \lambda} \) and \( f(\pi) - V = \frac{\pi - \pi_w}{r + \lambda} \), where \( \pi_w \) and \( \pi_R \) are reservation wage and profit levels for the worker and firm. Then (29) reduces to

\[
w \in \arg \max \{w - \pi_w \}^{\theta} \left[ y - w - \pi_R \right]^{1-\theta},
\]

which has solution

\[
w = w_R + \theta(y - \pi_R - w_R).
\]

This says that in terms of total lifetime expected utility, the worker receives his threat point \( U \) plus a share of the surplus, denoted \( S \) and defined by

\[
S = f(y - w) - V + W(w) - U
\]

where the last equality is derived by using (31) and (32).

From (31) and (32) we have \( W(w) - U = \frac{w - \pi_w}{r + \lambda} \) and \( f(\pi) - V = \frac{\pi - \pi_w}{r + \lambda} \), where \( \pi_w \) and \( \pi_R \) are reservation wage and profit levels for the worker and firm. Then (29) reduces to

\[
w \in \arg \max \{w - \pi_w \}^{\theta} \left[ y - w - \pi_R \right]^{1-\theta},
\]

which has solution

\[
w = w_R + \theta(y - \pi_R - w_R).
\]

This says that in terms of total lifetime expected utility, the worker receives his threat point \( U \) plus a share of the surplus, denoted \( S \) and defined by

\[
S = f(y - w) - V + W(w) - U
\]

where the last equality is derived by using (31) and (32).

From (31) and (32) we have \( W(w) - U = \frac{w - \pi_w}{r + \lambda} \) and \( f(\pi) - V = \frac{\pi - \pi_w}{r + \lambda} \), where \( \pi_w \) and \( \pi_R \) are reservation wage and profit levels for the worker and firm. Then (29) reduces to

\[
w \in \arg \max \{w - \pi_w \}^{\theta} \left[ y - w - \pi_R \right]^{1-\theta},
\]

which has solution

\[
w = w_R + \theta(y - \pi_R - w_R).
\]
The value of posting a vacancy is
\[ rV = -k + \alpha_J [f(\pi) - V], \]
where \( k \) is a flow cost (e.g., recruiting costs). As free entry drives \( V \) to 0, we need not keep track of \( V \), and we can rewrite (37) as
\[ (38) \quad \alpha_J f(\pi) = k. \]
The value of unemployment satisfies
\[ rU = b + \alpha_w [W(w) - U], \]
while the equations for \( W \) and \( J \) are unchanged from (31) and (32). Formally, an equilibrium includes the value functions \((f, W, U)\), the wage \( w \), and the unemployment and vacancy rates \((u, v)\), satisfying the Bellman equations, the bargaining solution, free entry, and the steady-state condition.\(^{19}\)

In terms of solving this model, one approach would be to try to find the equilibrium wage. Start with some arbitrary \( w \), solve (32) for \( f(\pi) \), and then use (38) to solve for \( \alpha_J \) and \( \alpha_w \). This determines \( W \) and \( U \). This \( w \) is an equilibrium if and only if the implied values for \( f, W, \) and \( U \) are such that the bargaining condition holds. While this works, here we bypass \( w \) by working directly with the surplus, which from (34) is
\[ (40) \quad (r + \lambda)S = y - rU. \]
Now (33) allows us to rewrite (39) as
\[ rU = b + \alpha_w \theta S, \]
and (40) gives
\[ (41) \quad (r + \lambda + \alpha_w \theta)S = y - b. \]

The next step generally in this method is to obtain expressions that characterize optimal choices for each of the decisions made outside of a match, given \( S \). Here the only such decision is whether to post a vacancy. Using (38) and the fact that bargaining implies \( f(\pi) = (1 - \theta)S \), we have
\[ (42) \quad k = \alpha_w (1 - \theta)S. \]
Equilibrium is completely characterized by (41) and (42). Indeed, we can combine them as
\[ (43) \quad r + \lambda + \alpha_w \theta = \frac{y - b}{(1 - \theta)\alpha_w}. \]
Under standard regularity conditions, a unique solution for \( \alpha_w \) exists. From this we can recover the wage,
\[ (44) \quad w = y - (r + \lambda)(1 - \theta)S. \]
Finally, the steady state unemployment rate satisfies an equation analogous to (25), accounting for the fact that all meetings result in matches:
\[ u^* = \frac{\lambda}{\lambda + \alpha_w}. \]
A number of results now follow easily. For example, an increase in \( b \) reduces the rate at which workers contact firms \( \alpha_w \), raises the rate at which firms contact workers \( \alpha_w \), reduces \( S \), and raises \( w \). The conclusion that unemployment duration and wages increase with \( UI \) is similar to what we found earlier, but here the mechanism is different. In the single-agent model, an increase in \( b \) induced the worker to raise his reservation wage, and to reduce search intensity if it is endogenous. Now an increase in \( b \) raises the bargained wage, which discourages job creation, thereby increasing unemployment duration.

4.4 Match-Specific Productivity

In the above model, it takes time for workers and firms to get together, but every contact leads to a match and \( w \) is the same in every match. This seems quite special when compared to what we did in sections 2–3, as it corresponds to workers sampling from a...
degenerate distribution. Moreover, in applications, changes in the probability that a contact leads to a match may be important. Here we extend the model so that not every contact results in a match and not every match has the same \( w \). The easiest way to proceed is to assume that when a worker and firm meet they draw match-specific productivity \( y \) from a distribution \( F \), where \( y \) is observed by both and constant for the duration of the match. From section 4.2, workers and firms agree to match if and only if \( y \geq y_R \), where \( y_R \) is characterized below.

In equilibrium, workers in a match with productivity \( y \) earn wage \( w(y) \) satisfying the Nash bargaining solution. Let \( W_y(w) \) be the value for an employed worker of a match with productivity \( y \) earning \( w \), \( J_y(y-w) \) the value for a firm with a filled job at productivity \( y \) earning profits \( y-w \), and \( S_y \) the surplus in a job with productivity \( y \). Generalizing (40) gives

\[
(45) \quad (r + \lambda)S_y = y - rU.
\]

Only the Bellman equations for an unemployed worker and the free entry condition change appreciably, becoming

\[
(46) \quad rU = b + \alpha_u \int_{y_R}^{\infty} \left[ W_y(w(y)) - U \right] dF(y)
\]

\[
= b + \alpha_u \int_{y_R}^{\infty} S_y dF(y)
\]

\[
(47) \quad k = \alpha_r \int_{y_R}^{\infty} J_y(y-w(y)) dF(y)
\]

\[
= \alpha_r \int_{y_R}^{\infty} S_y dF(y).
\]

To solve this model, combine (46) and (47) to get \( rU = b + \alpha_u \int_{y_R}^{\infty} S_y dF(y) \) and substitute into (45):

\[
(50) \quad (r + \lambda)k = \alpha_r \int_{y_R}^{\infty} (y-y_R) dF(y).
\]

We can now solve for \( y_R \) and \( \alpha_r \) from (49) and (50). The first of these equations describes an increasing relationship between \( \alpha_r \) and \( y_R \) (when it is easier for a worker to find a job, he is more willing to turn down a potential match with low productivity). The second gives a decreasing relationship between \( y_R \) and \( \alpha_r \) (when \( y_R \) is higher, matches are less profitable for firms so they post fewer vacancies). There exists a unique equilibrium under standard conditions. One can again recover the wage function: since \( w(y_R) = y_R \) and \( w'(y) = w' \) if \( y > y_R \), we have \( w(y) = y_R + \theta(y - y_R) \).

Equilibrium also determines the observed distribution of productivity across existing relationships, or equivalently, given \( w(y) \), the observed wage distribution \( G(w) \). Since the distribution of match productivity is \( F(y) \) truncated at \( y_R \), the equilibrium wage distribution \( G(w) \) is determined by \( y_R, F(y), \) and \( w(y) \).

It is easy to discuss turnover and wages. For example, an increase in \( b \) shifts (49) but not (50), resulting in an increase in \( y_R \), a reduction in \( \alpha_r \), and a reduction in \( H = \alpha_u [1 - F(y_R)] \). From a worker’s perspective, this closely resembles the single-agent problem, in the sense that he receives offers at rate \( \alpha_u \) from a given distribution, and needs to decide which to accept, except now the arrival rate and distribution are endogenous.

4.5 Endogenous Separations

Mortensen and Pissarides (1994) endogenize the separation rate by incorporating on-the-job wage changes.\(^{20}\) The resulting framework captures endogenously both the

\(^{20}\) This section mimics what we did in the single-agent problem in section 3. Other extensions discussed there can also be added, including on-the-job search (Pissarides 1984, 1994), and learning (Michael J. Pries 2004, Giuseppe Moscarini 2005, and Pries and Richard Rogerson 2005).
flows into and out of unemployment. Given that these flows vary a lot across countries and over time, it allows one to begin thinking formally about factors that may account for these differences. To proceed, let \( y \) be current productivity in a match, and assume that at rate \( \lambda \) we get a new draw from \( F(y' | y) \), where \( F(y' | y) \) first order stochastically dominates \( F(y' | y_0) \) whenever \( y > y_0 \). It remains to specify the level of productivity in new matches. One can assume they start at a random \( y_0 \), but we assume all new matches begin with the same \( y_0 \).

An equilibrium is defined as the natural extension of the previous model, and we can jump directly to the equation for the surplus

\[
(51) \quad (r + \lambda) S_y = y - rU + \lambda \int_{y_0}^{y} S_y dF(y'|y).
\]

Since \( rU = b + \alpha_w \theta S_{y_0} \), this can be rewritten

\[
(52) \quad (r + \lambda) S_y = y - b - \alpha_w \theta S_{y_0} + \lambda \int_{y_0}^{y} S_y dF(y'|y).
\]

To close the model we again use free entry,

\[
(53) \quad k = \beta \alpha_w (1 - \theta) S_{y_0}.
\]

Finding equilibrium amounts to solving (53) and (52) for \( y_R \) and \( \alpha_w \).

The solution is more complicated here because we are now looking for a fixed point in a system of functional equations—(52) defines both \( y_R \) and \( S_y \) as functions of \( \alpha_w \). Nevertheless, an increase in \( \alpha_w \) reduces \( S_y \) for all \( y \) and hence raises the reservation wage \( y_R \). Thus, (52) describes an increasing relationship between \( \alpha_w \) and \( y_R \). At the same time, (53) indicates that when \( \alpha_w \) is higher \( S_{y_0} \) must be higher, so from (52) \( y_R \) must be lower, and this defines a decreasing relationship between \( \alpha_w \) and \( y_R \). The intersection of these curves gives steady-state equilibrium, which exists uniquely under standard conditions. See Mortensen and Pissarides (1994, 1999b) for details of the argument.

### 4.6 Discussion

There are many applications and extensions of this framework. David Andolfatto (1996), Monika Merz (1995, 1999), Harold L. Cole and Rogerson (1999), Wouter J. den Haan, Gary Ramey, and Joel Watson (2000), Costain and Michael Reiter (2003), Shimer (2005), and Robert E. Hall (2005) all study versions of the model quantitatively. There is a literature that uses versions of the model to study the behavior of worker and job flows across countries as well as over the business cycle, including Stephen P. Millard and Mortensen (1997), Alain Delacroix (2003), Blanchard and Pedro Portugal (2001), and Pries and Rogerson (2005).

There is also a literature that introduces heterogeneous workers and firms, including Acemoglu (1999, 2001), James Albrecht and Susan Vroman (2002), Mortensen and Pissarides (1999c), Shimer (1999), and Shimer and Smith (2000). Ricardo J. Caballero and Mohamad L. Hammour (1994, 1996) and Gadi Barlevy (2002) study whether recessions are cleansing or sullying in terms of the distribution of match quality (do they lead to more good jobs or bad jobs). Moscarini (2001) also studies the nature of match quality over the business cycle. We do not have space to do justice to all the work, but this illustrates that it is an active and productive area.

### 5. Directed Search and Posting

We now move to models where some agents can post wage offers, and other agents direct their search to the most attractive alternatives. Following Espen R. Moen (1997) and Shimer (1996), the combination of posting and directed search is referred to as competitive search. Note that it is the combination of these features that is important; in section 6 we consider wage
posting with random search, which is quite different.

The literature has proposed several equivalent approaches. One posits a group of agents called market makers who set up submarkets, with the property that any match consummated in a submarket must be at the posted wage. Within each submarket there is a constant returns matching function \( m(u,v) \), so that the arrival rates \( \alpha_w \) and \( \alpha_e \) are determined by \( q = u/v \), which is called the queue length (the inverse of market tightness).

Each unemployed worker and each firm with a vacancy take as given \( w \) and \( q \) in every submarket, and go to the one offering the highest expected utility. In equilibrium, \( q \) in each submarket is consistent with agents’ expectations and no market maker can post a different wage and attract both workers and employers.\(^{21}\)

Another approach supposes that employers themselves post wages, and unemployed workers direct their search to the most attractive firms. A high posted \( w \) attracts more applicants, which reduces workers’ contact rate \( \alpha_w \) and raises the employer’s contact rate \( \alpha_e \). In equilibrium, workers are indifferent about where to apply, at least among posted wages that attract some workers. Firms choose wages to maximize expected profit. Still another approach assumes that workers post wages, and firms direct their search to them. As we said, these approaches are equivalent, in the sense that they give rise to identical equilibrium conditions. For brevity we consider only the case where firms post wages.

5.1 A One-Shot Model

We first discuss the basic mechanism in a static setting. At the beginning of the period, there are large numbers \( u \) and \( v \) of unemployed workers and vacancies, and \( q^* = u/v \) is the queue length. Each firm with a vacancy must pay cost \( k \), and we can either assume free entry (making \( v \) endogenous) or fix the number of vacancies. Any match within the period produces output \( y \), which is divided between the worker and firm according to the posted wage. At the end of the period, unmatched workers get \( b \), while unmatched vacancies get 0. Then the model ends.

Consider a worker facing a menu of different wages. Let \( U \) denote the highest value that he can get by applying for a job at some firm. Then a worker is willing to apply to a particular job offering a wage \( w \geq b \) only if he believes the queue length \( q \) at that job—i.e., the number of workers who apply—is sufficiently small. In fact, he is willing to apply only if \( \alpha_w(q) \geq U \), where \( \alpha_w(q) \) is sufficiently large, in the sense that

\[
U \leq \alpha_w(q)v + [1 - \alpha_e(q)]b. \tag{54}
\]

If this inequality is strict, all workers would want to apply to this firm, reducing the right hand side. Therefore, in equilibrium, if any workers apply to a particular job, \( q \) adjusts to satisfy (54) with equality.

To an employer, (54) describes how a change in his wage \( w \) affects his queue length \( q \). Therefore he chooses \( w \) to maximize

\[
V = \max_{w,q} -k - \alpha_e(q)(y - w), \tag{55}
\]

taking (54) as a constraint. Note that each employer assumes he cannot affect \( U \), although this is an endogenous variable to be determined in equilibrium. Eliminating \( w \) using (54) at equality, and also using \( \alpha_e(q) = q \alpha_w(q) \), we get

\[
V = \max_q -k \tag{56}
+ \alpha_e(q)(y - b) - q(U - b).
\]

The necessary and sufficient first order condition is

\[
\alpha_e'(q)(y - b) = U - b. \tag{57}
\]

In particular, (57) implies all employers choose the same \( q \), which in equilibrium must equal the economywide \( q^* \). Hence (57)
pins down the equilibrium value of $U$. Then (54) at equality determines the market wage

$$w^* = b + \varepsilon(q^*)(y - b),$$

where $\varepsilon(q^*) \equiv \frac{d\alpha}{dq^*}$ is the elasticity of $\alpha(q^*)$, which is in $(0,1)$ by our assumptions on the matching function $m$. Comparing (58) with the results in section 4.2, notice that this wage rule operates as if the worker and firm bargained over the gains from trade, with the workers’ share $\theta$ given by the elasticity $\varepsilon(q^*)$; this has important implications for the efficiency of competitive search, as we discuss below.

Substituting (58) into (55) pins down

$$V = -k + [\alpha_s(q^*) - q^*\alpha_s'(q^*)](y - b).$$

If the number of vacancies $v$ is fixed this gives profit; or we can use free entry $V = 0$ to endogenize $v$ and hence $q^*$. In either case, the model is simple to use. Indeed, this model looks a lot like a one-shot version of the random search and bargaining model. There is a key difference, however: here the surplus share is endogenously determined.

5.2 Directed Matching

Before generalizing to a dynamic setting, we digress to discuss some related literature. James D. Montgomery (1991) studies a nascent version of the above model (see also Michael Peters 1984, 1991). He starts with two unemployed workers and two firms. First firms post wages, then each worker applies to one of them (possibly randomly). A firm receiving at least one application hires at the posted $w$; if more than one applies the firm selects at random.\(^{22}\)

Suppose both firms offer the same $w > 0$. Then there are three equilibria in the application subgame: worker 1 applies to firm 1 and worker 2 to firm 2; worker 1 applies to firm 2 and worker 2 to firm 1; and both workers use identical mixed strategies, applying to each firm with probability 1/2. One can argue that the coordination implied by the first two equilibria is implausible, at least in large labor markets, and so the mixed-strategy equilibrium is the natural outcome. This introduces a coordination friction, as more than one worker may apply for the same job.

Generalizing this reasoning, suppose there are $u$ unemployed workers and $v$ vacancies, for any $u$ and $v$. If each worker applies to each firm with equal probability, any firm gets a worker with probability $1 - (1 - \frac{1}{v})^u$. Taking the limit of this expression as $u$ and $v$ go to infinity with $q = u/v$ fixed, in a large market a fraction $\alpha_u(q) = 1 - e^{-q}$ of firms get a worker. This is a standard result in statistics. Suppose there are $u$ balls independently placed with equal probability into each of $v$ urns. Then for large $u$ and $v$, the number of balls per urn is a Poisson random variable with mean $u/v$, so a fraction $e^{-u/v}$ of the urns do not get any balls. For this reason, this process is often called an urn–ball matching function.

Because the urn–ball matching process provides an explicit microeconomic story of both meetings (a ball is put in a particular urn) and matches (a ball is chosen from that urn), it is suitable for environments with heterogeneous workers (not all balls are the same). For instance, suppose there are $u_1$ type 1 workers and $u_2$ inferior type 2 workers, with $u = u_1 + u_2$. Firms hire type 1 workers over type 2 workers when both apply; hence they hire a type 1 worker with probability $1 - e^{-u/v}$ and a type 2 worker with probability $e^{-u/v}(1 - e^{-u/v})$. They hire some worker with probability $1 - e^{-u/v}$.\(^{23}\)

There are several generalizations of this matching process. Burdett, Shi, and Wright

\(^{22}\) By assumption here firms post wages rather than more general mechanisms. Coles and Jan Eeckhout (2003) relax this (e.g., they allow $w$ to be contingent on the number of applicants who turn up) and show it does not affect the main conclusions.

\(^{23}\) Therefore an increase in the number of undesirable type 2 workers does not affect the matching rate for type 1 workers, but an increase in the number of desirable workers adversely affects undesirable workers.
(2001) let some vacancies hire multiple workers. Albrecht, Pieter Gautier, and Vroman (2003), and Albrecht, Gautier, Serene Tan, and Vroman (2004) let workers make multiple applications. If workers can simultaneously apply for every job, this effectively flips the urn–ball problem around, so a worker is employed if at least one firm offers him a job; see Benoit Julien, John Kennes, and Ian King (2000). The intermediate case in which workers can apply for a subset of jobs delivers additional possibilities. In general, these directed search models provide a way to get inside the black box of the matching process.24

5.3 A Dynamic Model

To get something like the basic Pissarides model with directed search, start with an unemployed worker. Suppose he anticipates an unemployment–vacancy ratio $q$ and a wage $w$. Then

$$rU = b + \alpha_u(q)(W(w) - U) \tag{60}$$

$$rW(w) = w + \lambda(U - W(w)). \tag{61}$$

It is convenient to combine these into

$$rU = b + \frac{\alpha_u(q)(w - rU)}{r + \lambda}. \tag{62}$$

Similarly, for firms

$$rV = -k + \alpha_v(q)[J(y - w) - V] \tag{63}$$

$$rf(y - w) = y - w + \lambda[V - J(y - w)]. \tag{64}$$

Free entry yields

$$k = \frac{\alpha_v(q)(y - w)}{r + \lambda}. \tag{65}$$

Now suppose firms choose $w$ and $q$ to maximize $rV$, or equivalently by free entry, to maximize $\alpha_v(q)(y - w)$. They take (62) as given. Eliminating $w$ using this constraint and again using $\alpha_v(q) = qa_v(q)$, this reduces to

$$\max_q \alpha_v(q)\frac{y - rU}{r + \lambda} - q(rU - b). \tag{66}$$

The necessary and sufficient first order condition,

$$\alpha_v'(q)\frac{y - rU}{r + \lambda} = rU - b, \tag{67}$$

has a unique solution, so all firms choose the same $q$. Eliminating $U$ and $w$ from (62), (65), and (67) we get an implicit expression for $q$,

$$\frac{r + \lambda + \alpha_v'(q)}{\alpha_v(q) - qa_v'(q)} = \frac{y - b}{k}. \tag{68}$$

This pins down the equilibrium $q$, or equivalently the arrival rates $\alpha_u$ and $\alpha_v$. Under standard conditions the solution is unique. We can again perform standard exercises, such as changing $b$, with similar results: here this raises the $u/v$ ratio, which reduces $\alpha_u$, raises $\alpha_v$, and increases $w$. But although the conclusions are similar to those reached in the previous section, the mechanism is quite different. An increase in $b$ in this model makes workers more willing to accept an increase in the risk of unemployment in return for an increase in $w$. Firms respond by offering workers what they want—fewer jobs at higher wages.

5.4 Discussion

Competitive search equilibrium theory provides arguably a more explicit explanation of the matching process and of wage determination than the bargaining models in section 4. Nash bargaining says $w$ divides the surplus into exogenous shares; here workers face a trade-off between a higher wage and a lower probability of getting a job, while firms face a trade-off between profit and the probability of hiring. Competition among wage setters, whether these be firms, workers, or
market makers, leads to a point along the indifference curves of agents trading off wages and arrival rates.\textsuperscript{25}

However, a potential disadvantage of these models—indeed, of any model that assumes posting—is that it is a strong assumption to say that agents commit to the posted terms of trade. If the markets is really decentralized, what prevents them from trying to bargain for a different $w$ after they meet? Again, this comes up in any posting model. In the context of the wage posting model in the next section, Coles (2001) shows explicitly how to prevent firms from reneging on their posted wage using reputational considerations, but there is more work to be done on the issue.

In terms of other extensions, Coles and Eeckhout (2000), Shi (2001, 2002), and Shimer (forthcoming) add heterogeneity, which introduces wage dispersion among heterogeneous workers, and among identical workers at different firms. Acemoglu and Shimer (1999b) allow workers to be risk-averse, which means it is no longer possible to solve the model explicitly. They show that an increase in risk aversion reduces wages, while an increase in $b$ raises it. Building on this, Acemoglu and Shimer (2000) show that UI can enhance productivity. Much more research is currently being done on directed search models. Again, while we cannot do justice to all of the work in the area, we want to say that it appears to have great potential.

6. Random Matching and Posting

We now combine random matching, as in section 4, with posting, as in section 5.\textsuperscript{26} This class of models has been used extensively in the literature on the pure theory of wage dispersion, which tries to understand how workers with identical productivity can be paid different wages. Note that the models in the previous two sections generate wage dispersion only if workers are either ex ante or ex post heterogeneous. A pure theory of wage dispersion is of interest for a couple of reasons. First, the early literature suggested that search is relevant only if the distribution from which you are sampling is nondegenerate, so theorists were naturally led to study models of endogenous dispersion. Second, many people see dispersion as a fact of life, and for them the issue is empirical rather than theoretical.

As Mortensen (2003, p. 1) reports, “Although hundreds if not thousands of empirical studies that estimate so-called human capital wage equations verify that worker characteristics that one could view as indicators of labor productivity are positively related to wages earned, the theory is woefully incomplete in its explanatory power. Observable worker characteristics that are supposed to account for productivity differences typically explain no more that 30 percent of the variation in compensation.” Eckstein and van den Berg (forthcoming) argue that “equilibrium search models provide a framework to empirically analyze the sources of wage dispersion: (a) workers heterogeneity (observed and unobserved); (b) firm productivity heterogeneity (observed and unobserved); (c) market frictions. The equilibrium framework can . . . empirically measure the quantitative importance of each source.”\textsuperscript{27}

Diamond (1971) is an early attempt to construct a model of dispersion, and although it did not work, it is useful to understand why. Consider an economy where homogeneous workers each face a

\textsuperscript{25} In section 7, we show that competitive search equilibrium achieves the optimal trade-off between these factors.

\textsuperscript{26} The remaining combination—directed search and bargaining—is not so interesting, at least if firms are homogeneous, since there is nothing for workers to direct their search toward. Lawrence Uren (2004) studies directed search and bargaining with heterogeneous firms.

\textsuperscript{27} As they report, van den Berg and Geert Ridder (1998) estimate that up to 25 percent of wage variability is attributable to frictions, in the sense that this is what would emerge from a posting model that ignored heterogeneity, while Fabien Postel-Vinay and Jean-Marc Robin (2002) estimate up to 50 percent.
standard search problem. We do not give all the particulars, since the model is a special case of what is described in detail below, but the key is that the offer distribution $F$ is generated by wage-posting firms, each of which has a constant returns technology with labor as the only input, with productivity $y$. A firm hires any worker it contacts who is willing to accept its posted $w$. Consider an individual firm. Letting $F$ be the distribution of wages posted by other firms, he wants to maximize expected profit taking $F$ as given. An equilibrium is defined as a distribution such that every wage posted with positive probability earns the same profit, and no other wage earns greater profit.

Diamond provides a rather striking result: there is a unique equilibrium, and in it all employers set the same wage, $w = b$. The proof is simple. Given any $F$, since workers are homogeneous they all choose the same reservation wage $w_R$. Clearly, no firm posts $w < w_R$ as this would mean it cannot hire, and no firm posts $w > w_R$ as it can hire every worker it contacts at $w = w_R$. To see why it turns out that $w = b$, consider an individual firm when all firms are posting $w > b$. If it deviates and offers a wage slightly less than $w$, it still hires every worker it meets. Since this is true for any $w > b$, we must have $w = b$ in equilibrium. The model not only fails to rationalize wage dispersion, it fails to explain why workers are searching in the first place!

6.1 Heterogeneous Leisure

Why might one expect to find pure wage dispersion? One answer is that search frictions produce a natural trade-off for a firm: while posting higher wages lowers your profit per worker, it could allow you to hire workers faster, and so, in the long run, you get more of them. In Diamond’s model, this trade-off is nonexistent, since when you increase your wage above $w_R$ there is no increase in your hiring rate. Albrecht and Bo Axell (1984) allow for heterogeneity in workers—not in productivity but in their value of leisure—which leads to heterogeneity in reservation wages and this makes the above trade-off operational.

Consider two types of workers, some with $b = b_1$ and others with $b = b_2 > b_1$. For any wage distribution $F$ there are two reservation wages, $w_1$ and $w_2 > w_1$. If $W_i(w)$ is the value of a type $i$ worker who is employed at wage $w$ and $U_i$ is the value of an unemployed type $i$ worker, these wages satisfy $W_i(w_1) = U_1$ and $W_i(w_2) = U_2$. Generalizing our logic from the Diamond model, no firm posts a wage other than $w_1$ or $w_2$. It is possible that these two wages could yield equal profit, however, since low-wage firms can hire only workers with $b = b_1$ while high-wage firms can hire everyone they contact. In the Albrecht–Axell model, equal profits at two different wages can occur in equilibrium for a large set of parameters.

To see how this works, normalize the measure of firms to 1, and let the measure of workers be $L = L_1 + L_2$ where $L_i$ is the measure with $b_i$. Let $\sigma$ be the endogenous fraction of firms posting $w_2$. Any candidate equilibrium wage distribution is completely summarized by $w_1$, $w_2$, and $\sigma$. Observe first that the reservation wage of type 2 workers is $w_2 = b_2$. It is clear that their reservation wage cannot be any lower, since otherwise they would prefer to remain unemployed. On the other hand, if their reservation wage exceeds $b_2$, an argument like the one we used for the Diamond model ensures that a firm could reduce $w_2$ and still attract these workers.

To determine $w_1$, note that type 1 workers accept both $w = w_1$ and $w = w_2$, and so given the arrival rate $\alpha_n$ their value functions satisfy

$$rU_1 = b_1 + \alpha_n(1-\sigma)[W_1(w_1) - U_1] + \alpha_n\sigma[W_1(w_2) - U_1]$$

$$rW_1(w_1) = w_1 + \lambda[U_1 - W_1(w_1)]$$

Albrecht and Axell do not actually have firms earning equal profit, but allow $y$ to vary across firms and look for a cutoff $y^*$ such that firms with $y < y^*$ pay $w = w_1$, and those with $y > y^*$ pay $w = w_2$; the economic implications are basically the same.
An alternative to having firms maximize expected discounted profit is to maximize the value of posting a vacancy, which is more in line with sections 4 and 5; see Mortensen (2000). We adopt the criterion in the text because it facilitates comparison with the model in the next subsection.

An alternative to having firms maximize expected discounted profit is to maximize the value of posting a vacancy, which is more in line with sections 4 and 5; see Mortensen (2000). We adopt the criterion in the text because it facilitates comparison with the model in the next subsection.

Unemployment rates for the two types are

\[ u_1 = \frac{\lambda}{\alpha_u + \lambda} \text{ and } u_2 = \frac{\lambda}{\alpha_u + \lambda}. \]

For firms, the expected value of contact-

\[ \Pi_1 = \frac{L_1 u_1 y - w_1}{L_1 u_1 + L_2 u_2 + r + \lambda}, \]

\[ \Pi_2 = \frac{y - b_2}{r + \lambda}, \]

where for now we are taking \( \alpha_u \) as given. Inserting \( u_1, u_2 \) and \( w_1 \), after some algebra, we see that \( \Pi_2 - \Pi_1 \) is proportional to

\[ T(\sigma) = (r + \lambda + \alpha_u \sigma) \{ (y - b_2) : \]

\[ \times [\lambda L_1 + (\alpha_u + \lambda) L_2] - (y - b_1) \lambda L_1 \}

\[ - r \alpha_u \sigma L_1 (b_2 - b_1). \]

For an equilibrium, we need

\[ \sigma = 0 \text{ and } T(0) < 0; \sigma = 1 \text{ and } T(1) > 0; \text{ or } \sigma \in (0,1) \text{ and } T(\sigma) = 0. \]

It is easy to show there exists a unique solution to (76), and \( 0 < \sigma < 1 \) if and only if \( y < y < \bar{y} \) where

\[ \bar{y} = b_2 + \frac{\lambda L_1 (b_2 - b_1)}{(\alpha_u + \lambda) L_2} \text{ and } \]

\[ \bar{y} = \frac{\lambda L_1 (b_2 - b_1)}{r + \alpha_u + \lambda (\alpha_u + \lambda) L_2}. \]

When productivity is low all firms pay \( w_1 = b_1 \), when it is high all firms pay \( w_2 = b_2 \), and when it is intermediate there is wage dispersion. When \( \sigma \in (0,1) \), we can solve \( T(\sigma) = 0 \) for \( \sigma \) and use (72) to solve for \( w_1 \) and the distribution of paid wages explicitly.\(^{30}\)

Of course, all of this is for given arrival rates, and the value of \( \sigma \) that solves (76) depends on \( \alpha_u \). Using the matching function, we can endogenize

\[ \alpha_u = \frac{m(L_1 u_1 + L_2 u_2)}{L_1 u_1 + L_2 u_2}, \]

where \( L_1 u_1 + L_2 u_2 \) is the number of unemployed workers and we assume that all firms are always freely posting a vacancy, so that \( v = 1 \), with the idea being that each one will hire as many workers as it can get. Since \( u_2 \) depends on \( \sigma \), so does \( \alpha_u \). An equilibrium is a pair \( (\alpha_u, \sigma) \) satisfying (76) and (78).\(^{31}\)

6.2 On-the-Job Search

In the previous section, firms may pay higher wages to increase the inflow of workers. There also exist models where firms may pay higher wages to reduce the outflow of workers (e.g., Burdett, Lagos, and Wright 2003). In Burdett and Mortensen (1998), both margins are at work, and firms that pay higher wages both increase the inflow and reduce the outflow of

\[ \sigma > 0 \text{ and } T(\sigma) < 0 \text{ or } \sigma < 0 \text{ and } T(\sigma) > 0. \]

30 Clearly, we get at most two wages here, but the argument can be generalized to many types of workers (Eckstein and Wolpin 1990).

31 See Damien Gaumont, Martin Schindler, and Wright (forthcoming) for details, and for several other variations on the general theme of Allsrecht–Axell models. They also discuss a problem with models based on ex ante heterogeneity: Given type 2 workers get no surplus, if there is any search cost \( \varepsilon > 0 \), they drop out of the market, leaving only type 1 workers. Then we are back to Diamond and the distribution collapses (and of course the problem applies with any number of types). Gaumont, Schindler, and Wright also discuss models that avoid this problem.
workers. We now present this model in detail, which is relatively easy here, because it is based on on-the-job search, and we can make substantial use of results derived in section 3.

The arrival rates are $\alpha_0$ and $\alpha_1$ while unemployed and employed, and every offer is a random draw from $F(w)$. For ease of presentation, we begin with the case $\alpha_0 = \alpha_1 = \alpha$, which implies $w_R = b$ by (24), and return to the general case later. Since all unemployed workers use a common reservation wage, and clearly no firm posts $w < w_R$, the unemployed accept all offers and we have $u = \lambda/(\lambda + \alpha)$ as a special case of (25).

Also, the distribution of paid wages is

$$G(w) = \frac{\lambda F(w)}{\lambda + \alpha[1 - F(w)]},$$

a special case of (26).

If a firm posts $w \geq w_R$, a worker contacts accepts if he is currently unemployed or currently employed but at a lower wage, which occurs with probability $u + (1 - u)G(w)$. The employment relationship then yields flow profit $y - w$ until the worker leaves either due to an exogenous separation or a better offer, which occurs at rate $\lambda + \alpha[1 - F(w)]$. Therefore, after simplification, the expected profit from $w$ is

$$\Pi(w) = \frac{\lambda(y - w)}{(\lambda + \alpha[1 - F(w)])(r + \lambda + \alpha[1 - F(w)])}.$$  

Again, equilibrium requires that any posted wage yields the same profit, which is at least as large as profit from any other wage. Clearly no firm posts $w < w_R = b$ or $w > y$. In fact, one can show that the support of $F$ is $[b, \bar{w}]$ for some $\bar{w} < y$, and there are neither gaps nor mass points on the support.\footnote{Suppose there were a mass point at some $\bar{w}$. Then a firm posting $\bar{w} + \varepsilon$ would be able to hire away any worker it contacts working from a $\bar{w}$ firm, increasing its revenue discretely with only an $\varepsilon$ increase in cost, which means $\bar{w}$ does not maximize profit. Suppose there were a gap in $F$, say between $\bar{w}$ and $\bar{w}'$. Then a firm posting $\bar{w}'$ could lower its wage and reduce cost without reducing its inflow or increasing its outflow of workers, and again $\bar{w}'$ could not maximize profit.}

We now construct $F$ explicitly. The key observation is that firms earn equal profits from all posted wages, including the lowest $w = b$: $\Pi(w) = \Pi(b)$ for all $w \in [b, \bar{w}]$. Since $F(b) = 0$, $\Pi(b) = \lambda\alpha y_b / (\alpha + \lambda)(r + \alpha + \lambda)$. Combining this with (80) gives an equation that can easily be solved for $F(w)$. In the simplest case where $r = 0$, the result is

$$F(w) = \frac{\lambda + \alpha}{\alpha} \left(1 - \frac{y - w}{y - b}\right).$$

We know the lower bound is $b$, and the upper bound $\bar{w}$ can easily be found by solving $F(\bar{w}) = 1$. This yields the unique distribution consistent with equal profit for all wages posted.

In words, the outcome is as follows. All unemployed workers accept the first offer they receive, and move up the wage ladder each time a better offer comes along, but also return to unemployment periodically due to exogenous layoffs. There is a nondegenerate distribution of wages posted by firms $F$ given by (81), and of wages earned by workers $G$, given by inserting $\bar{w}$ into (79). The model is consistent with many observations concerning worker turnover, and also concerning firms, including, e.g., the fact that high wage firms are bigger. See Mortensen (2003) for details.

There are many interesting extensions. First, with $\alpha_0 \neq \alpha_1$, the same methods lead to

$$F(w) = \frac{\lambda + \alpha}{\alpha_1} \left(1 - \frac{y - w}{y - w_R}\right),$$

where now $w_R$ is endogenous (with $\alpha_0 = \alpha_1$ we knew $w_R = b$). To determine $w_R$, integrate (24) to get

$$w_R = \frac{(\lambda + \alpha_1)^2 b + (\alpha_0 - \alpha_1)\alpha_1 y}{(\lambda + \alpha_1)^2 + (\alpha_0 - \alpha_1)\alpha_1}.$$  

One can check that in the limit as $\alpha_1 \to 0$, $\bar{w} = w_R$, which means there is a single wage, $w = w_R = b$. This is the Diamond result as a special case when there is no on-the-job search. Also, in the limit as $\alpha_1 \to \infty$, $\bar{w} = y$ and...
\( G(w) = 0 \) for all \( w < y \); hence all workers earn \( w = y \). Moreover, as \( \alpha_i \to \infty \), clearly \( u \to 0 \). Hence, the competitive solution also emerges as a special case when \( \alpha_i \) and \( \alpha_i \) get large.

One can let firms be heterogeneous with respect to \( y \). In equilibrium there is a distribution of wages paid by each type of firm, and all firms with productivity \( y_2 \) pay more than all firms with \( y_1 < y_2 \). Thus, higher productivity firms are more likely to hire and less likely to lose any worker. With heterogeneous firms, van den Berg (2003) shows there may be multiple equilibria. Perhaps more significantly, firm heterogeneity is important empirically, because with homogeneous firms the equilibrium wage distribution (81) has an increasing density, which is not in the data. With heterogeneity \( F \) can have a decreasing density. Another very important extension is to allow firms whose workers contact rivals to make counteroffers, as in Postel-Vinay and Robin (2002).

Margaret Stevens (2004) allows firms to post wage–tenure contracts, rather than simply a constant wage. She shows firms have an incentive to back load wages to reduce turnover. If workers can make an up-front payment for a job, an optimal contract extracts an initial fee and then pays \( w = y \). If firms are homogeneous this contract eliminates all voluntary quits. In equilibrium all firms demand the same initial fee, and this leaves unemployed workers indifferent about accepting the position. Stevens also shows that if initial payments are impossible, say because of liquidity constraints, there is an equilibrium where all firms offer a contract that pays the worker 0 for a fixed period and then pays \( w = y \). If firms are homogeneous, they extract all of the surplus, there are no job-to-job transitions, and all contracts are identical.

However Burdett and Coles (2003) show that Stevens’s results can be overturned if workers desire smooth consumption. They allow firms to commit to wage–tenure contracts, but assume workers are risk-averse and do not have access to financial markets (they must consume \( w \) each period). They prove that all equilibrium wage–tenure contracts are described by a common baseline salary scale, which is an increasing, continuous relationship between the wage and tenure. Firms offer different contracts in the sense that they start workers at different points on the scale. Thus, consumption smoothing reintroduces wage dispersion, both in the sense that workers get different wages depending on their tenure, and in the more fundamental sense that firms offer different contracts.

6.3 Discussion

The two models of wage dispersion we have presented are based on worker heterogeneity and on-the-job-search, respectively. Of course, one can integrate them. This is important empirically, because although on-the-job-search models (with heterogeneous firms) do a good job accounting for wages, they do less well accounting for individual employment histories, especially the fact that hazard rates tend to decrease with the length of unemployment spells. Models with worker heterogeneity do better at accounting for this, but less well for wages. An integrated model can potentially account for both; see Christian Bontemps, Robin, and van den Berg (1999). Many other extensions and applications are possible, and this is a productive area for both theoretical and empirical research.\(^{33}\)

7. Efficiency

We now move on to efficiency. Of course, in an economy with many agents there are many Pareto optimal allocations. Here we focus on those that maximize the sum of agents’ utility, or equivalently, that maximize

\[^{33}\text{We mention some related work. Masters (1999) studies a wage-posting model and uses it to analyze changes in the minimum wage. Delacroix and Shi (forthcoming) consider directed search in an environment similar to Burdett–Mortensen and show it also gives rise to wage dispersion, although for different reasons. Finally, some people consider as an alternative to random search models where workers are more likely to meet large firms, as in Burdett and Vishnawathi (1988b).}\]
the present discounted value of output net of the disutility of work and search costs.

7.1 A One-Shot Model

Initially, \( u \) workers are unmatched. Firms choose how many vacancies \( v \) to post, each at cost \( k \), and then \( m(u,v) \) matches form. Let \( q = u/v \) and assume constant returns, so \( m(u,v) = v m(q,1) = v \alpha_s(q) \). Each match produces \( y \) at an opportunity cost \( b \) to each worker. Assume provisionally that wages are determined by bargaining, \( w = \theta y + (1 - \theta) b \). Then the economy ends. Firms post vacancies until the free entry condition

\[
(84) \quad \alpha_s(q)(1-\theta)(y - b) = k
\]

holds. This is a static version of the basic Pissarides model.

Now consider a planner who posts vacancies to maximize output, net of workers’ opportunity cost and the cost of posting vacancies. Equivalently, he chooses \( q = u/v \), knowing that each worker will contact a vacancy with probability \( \alpha_s(q) \), to maximize

\[
(85) \quad u \alpha_s(q)(y - b) - \theta v k = \frac{u}{q} [\alpha_s(q)(y - b) - k].
\]

The necessary and sufficient condition for a solution is

\[
(86) \quad [\alpha_s(q) - q \alpha_s'(q)](y - b) = k.
\]

Denote the planner’s solution by \( q^* \).

The following is immediate from (84) and (86): the planner’s solution and the decentralized solution coincide if and only if

\[
(87) \quad \theta^* = \varepsilon(q^*),
\]

where \( \varepsilon(q^*) = \frac{\alpha_s''(q^*)}{\alpha_s'(q^*)} \) was defined in section 5 as the elasticity of \( \alpha_s(q) \). This is the Hosios condition (Arthur J. Hosios 1990), determining the share of the surplus that must go to workers for bargaining to be efficient.\(^{34}\)

Alternatively, in terms of wages, equilibrium is efficient if \( w = w^* = \theta y + (1 - \theta) b \). If \( \theta \) is too high, for example, then \( w > w^* \) and \( \varepsilon \) is too low.

Now consider the competitive search model from section 5. From (58), in competitive search equilibrium \( w = w^* \), and equilibrium is necessarily efficient. That is, the Hosios condition holds endogenously—although agents do not bargain, the surplus is still being split, and the equilibrium split implies efficiency. To understand why, it is useful to think in terms of competition between market makers, as discussed above. With free entry, market makers effectively maximize workers’ expected utility, \( \alpha_s(q)(w - b) \), recognizing that firms must break even to participate, \( \alpha_s(q)(y - w) = k \). Using the constraint to eliminate \( w \), this is identical to the planner’s problem.

Now consider extending the model to endogenize workers’ search intensity. The total number of matches is \( m(\bar{s}u,v) \), where \( \bar{s} \) is average intensity. Assuming constant returns, a firm hires with probability \( \alpha_s(\bar{s}q) = m(\bar{s}q,1) \), while a worker with search intensity \( s \) gets hired with probability \( s \alpha_s(\bar{s}q) = s \alpha_s(\bar{s}q) / \bar{s}q \). An unemployed worker chooses \( s \) to maximize

\[
(88) \quad b - g(s) + \frac{s \alpha_s(\bar{s}q) \theta(y - b)}{\bar{s}q}.
\]

In equilibrium all workers choose the same \( s = \bar{s} \), where

\[
(89) \quad g'(\bar{s}) = \frac{\alpha_s(\bar{s}q) \theta(y - b)}{\bar{s}q},
\]

and free entry by firms implies

\[
(90) \quad k = \alpha_s(\bar{s}q)(1-\theta)(y - b).
\]

The planner solves

\[
(91) \max_{\bar{s}} b - g(\bar{s}) + \frac{s \alpha_s(\bar{s}q)(y - b) - k}{q} u,
\]

which has necessary and sufficient conditions

\[
(92) \quad g'(\bar{s}) = \alpha_e'(\bar{s}q)(y - b)
\]

\[
(93) \quad k = [\alpha_s(\bar{s}q) - \bar{s}q \alpha_e'(\bar{s}q)](y - b).
\]

\(^{34}\) Because \( \alpha_s(q) = q \alpha_s'(q) \), we have \( \frac{\alpha_s''(q)}{\alpha_s'(q)} = 1 \), hence the condition can equivalently be stated as firms’ bargaining power \( 1 - \theta \) must equal the elasticity of \( \alpha_s(q) \).
Notice something interesting: (89) and (92) coincide if the Hosios condition $\theta = \varepsilon(\tilde{s} q)$ holds, while (90) and (93) coincide under the same condition. That is, bargaining equilibrium achieves efficient search intensity and entry under the Hosios condition. This seems genuinely surprising, as there are two variables to be determined, $\tilde{s}$ and $\varepsilon$, and only one parameter, $\upsilon$.

Since we have already shown that in competitive search equilibrium we get the Hosios condition endogenously, competitive search equilibrium achieves efficient search intensity and entry. As suggested above, market makers effectively choose the terms of trade to ensure that both workers and firms behave optimally. There is nothing special about endogenous entry decisions or search intensity. We could consider various other extensions (e.g., match-specific productivity, and so on). Using the fact that $Y(e) = a_0 + a_1 e$ and the envelope theorem, we can differentiate both sides of (94) to get

$$\begin{align*}
(96) \quad r a_1 &= y - b + \frac{k}{q} - a_1 \left[ \frac{\alpha_e'(q)}{q} + \lambda \right].
\end{align*}$$

Combining (95) and (96) to eliminate $a_1$, we arrive at

$$\begin{align*}
(97) \quad \frac{r + \lambda + \alpha_e'(q)}{\alpha_e(q) - q\alpha_e'(q)} &= \frac{y - b}{k}.
\end{align*}$$

This completely characterizes the optimal policy $q = q(e)$; notice that in fact $q$ does not depend on the state $e$, only on exogenous parameters.

How does this compare with equilibrium? Recall that equilibrium in section 4.3 satisfies (43). It is easy to see that the solutions are the same, and hence bargaining equilibrium coincides with the planner’s solution, if and only if $\theta = \varepsilon(q)$. Now looking at (68) from the competitive search version of the model, we see that competitive search equilibrium is necessarily efficient. Hence the results from the static model generalize directly, and again carry over to more general models with endogenous search intensity, match-specific productivity, and so on.

### 7.3 Discussion

We have demonstrated in several contexts that competitive search is efficient, while bargaining is efficient if and only if the Hosios condition holds. Although these results are in some sense general, it would be

$$\begin{align*}
(94) \quad rY(e) &= \max_{\tilde{s}} \ v e + b(1 - e) \\
&= \frac{k}{q}(1 - e) + Y'(e) \left[ \frac{\alpha_e'(q)(1 - e)}{q} - \lambda e \right].
\end{align*}$$

One can show that $Y(e)$ is affine; i.e., $Y(e) = a_0 + a_1 e$ for some constants $a_0$ and $a_1$, where $a_0$ can be interpreted as the value of an unemployed worker and $a_1$ the surplus from a match.\(^{35}\)

The first order condition from (94) simplifies to

$$\begin{align*}
(95) \quad k &= a_1 \left[ \alpha_e(q) - q\alpha_e'(q) \right].
\end{align*}$$
misleading to suggest that we can get efficient equilibria in all search models. A general understanding of the nature of efficiency in these models is still being developed. Mortensen and Wright (2002) provide some results, but only for constant returns matching functions; without this, equilibria are unlikely to be efficient. Shimer and Smith (2001) show that in some models with heterogeneity, even with constant returns, the efficient outcome may not even be compatible with a steady state, but exhibits cycles.

The efficiency of search equilibria with ex ante investments is an interesting topic. For example, in Acemoglu (1996) and Masters (1998), agents decide on how much capital to acquire prior to searching. Generically there is no $\theta$ that makes the equilibrium outcome under bargaining efficient—the value of $\theta$ that provides the right incentive for investment in human capital by workers is different from the value of that provides the right incentive for firms (whether firms choose the number of vacancies, the types of jobs to create, or physical capital). However, Acemoglu and Shimer (1999a) show competitive search equilibrium with ex ante investments is efficient. Another case where there is no $\theta$ such that bargaining yields the efficient outcome is discussed by Smith (1999), who assumes firms have concave production functions and hire a large number of workers, but bargain with each one individually.

One can also ask about the efficiency of the wage-posting models in section 6. In general, if firms commit to pay the same $w$ no matter the circumstances, it is unlikely to yield efficient outcomes. In Albrecht and Axell (1984), e.g., a planner would want all meetings to result in matches, which does not happen. In the simplest Burdett and Mortensen (1998) model, with $\alpha_0 = \alpha_1$, efficiency does result because all meetings involving unemployed workers result in matches and other meetings are irrelevant from the planner’s perspective. In extended versions, however, there is no reason to necessarily expect efficiency (Mortensen 2000). In general, there is much more work to be done on this topic.

8. Conclusion

In contrast to standard supply-and-demand models, search theory emphasizes frictions inherent in the exchange process. Although there is no one canonical search model, and versions differ in terms of wage determination, the matching process, and other assumptions, we have tried to show that there is a common framework underlying all of the specifications. We have used the different models to discuss several issues, mainly related to worker turnover and wages. We have also used them to discuss some simple policy experiments, such as an increase in UI. Different models emphasize different margins along which such policies work, including the choice of reservation wage, search intensity, entry, and so on.

The approach, as we have seen, is consistent with many interesting observations. For a start it predicts the unemployment rate is not zero, which is perhaps a low hurdle but not one that can be met by many alternative theories. The reason the model predicts this is obvious, but also obviously correct—it takes time for workers to find jobs. It also takes time for firms to fill vacancies, and a nice property of these models is that there can be coexistence of unemployed workers and unfilled vacancies. The framework can be used to organize thinking about individual transitions between unemployment and employment, as well as job-to-job transitions.

Different search-based models can be used to help explain the relationships between tenure, wages, and turnover, including versions that incorporate on-the-job search, learning, or human capital. Several models can be used to discuss the distribution of wages, and some of the models are consistent with observations such as the fact that high wage firms tend to have
more workers, or high productivity firms tend to pay more. We also discussed the welfare properties of the models. Bargaining models achieve efficiency if and only if the bargaining weights satisfy a condition related to the matching technology; while in competitive search models, this condition emerges endogenously.

We have only scratched the surface, but hopefully we have conveyed the main ideas. The goal was never to claim that search theorists fully understand all of the issues, and indeed there are many interesting directions for future research. One involves further quantifying the models, using both micro and aggregate data. On a more conceptual front, the models we have presented do not offer a cogent theory of the firm, and do not distinguish between worker and job flows. The welfare results are incomplete. Finally, we have emphasized the importance of both wage determination and the matching function. What is the right model of wages? What are the mechanics that determine matching? Perhaps this survey will stimulate additional research on these important questions.

REFERENCES


Garnaut, Damien, Martin Schindler, and Randall Wright. Forthcoming. “Alternative Models of Wage Dispersion,” in Structural Models of Wage and


E. S. Phelps et al., eds. New York: W. W. Norton, 124–66.
Rubinstein, Ariel, and Asher Wolinsky. 1990.


