The 2010 Nobel Memorial Prize in Search Theory*

James Albrecht
Department of Economics, Georgetown University, Washington DC 20057, USA
albrecht@georgetown.edu

February 2, 2011

Abstract
This paper surveys search theory with an emphasis on the contributions of the 2010 Nobel Memorial Prize winners, Peter Diamond, Dale Mortensen and Christopher Pissarides.

1 Introduction

The 2010 Nobel Memorial Prize in Economic Sciences was awarded to Peter Diamond, Dale Mortensen and Christopher Pissarides “for their analysis of markets with search frictions.” What are “markets with search frictions”? How are prices set in these markets? How are quantities allocated? How have Diamond, Mortensen and Pissarides contributed to our understanding of these issues?

Search frictions are particularly important in the labor market, and I will focus my survey on this market. Unemployed workers look for jobs at the same time that firms look for workers to fill their vacancies. It takes time and effort for a worker to find a suitable employer; similarly, a firm incurs costs while trying to hire a suitable employee. One reason these search costs arise is because the labor services being traded are not standardized. Some workers are more qualified than others, some firms’ jobs require more skill than others, and often there is an important idiosyncratic component to the value of a prospective match between worker and firm. Another reason is that there are coordination frictions. Sometimes a worker applies for a job that would have been a good match had the firm not hired another worker.

---

*Thanks to Fabien Postel-Vinay and Susan Vroman for useful comments. Any errors are, of course, mine alone.
in the meantime. Similarly, sometimes a firm finds a good worker for the vacancy it wants to fill only to find that he or she has been hired elsewhere. Search frictions are important in other markets as well – the housing market is another obvious example.

Search theory has its origins in the consensus that was reached in the late 1960’s, as expressed, for example, in the “Phelps volume” (Phelps et al. 1970), that macroeconomics was in need of a solid microfoundation. The theory uses microeconomic tools to analyze the determinants of aggregate unemployment, focusing in particular on the steady-state or “natural” rate of unemployment. As has been the case with many other theoretical developments in economics, search theory was first analyzed as an individual optimization problem. It was only later that the equilibrium nature of the problem was recognized.

The individual optimization problem in search theory is essentially the optimal stopping problem of statistics. In its simplest form, an unemployed worker samples wage offers from a known distribution, $F(w)$, at Poisson rate $\alpha$. There is a cost to rejecting an offer in hand – the worker will have to spend time and/or effort to find another offer – but there is also a benefit since rejecting one’s current offer leaves open the possibility of drawing a higher wage later in the search process. In a stationary environment, this sequential search problem has a simple solution, namely, to continue sampling until a wage, $w$, is drawn such that $w \geq R$, where $R$ is the “reservation wage.” The rate at which the worker accepts a job is then simply $\alpha (1 - F(R))$, so expected unemployment duration is increasing in $R$, and much of the early search literature focused on the question of how labor market policy affects the reservation wage.

In the individual search problem, both the rate at which offers are received and the wage offer distribution are treated as exogenous, but in equilibrium models, $\alpha$ and/or $F(\cdot)$ are treated as endogenous. The equilibrium search literature has two important branches. In the first, the wage offer distribution is modeled as the equilibrium outcome of a wage-posting game played by firms. That is, each firm posts a wage – or, equivalently, makes a take-it-or-leave-it offer to job applicants – taking into account the distribution of wage offers posted by all other firms in the market and worker reservation wage strategies. This branch of the literature has focused on endogenizing the wage offer distribution and has mostly ignored the endogeneity of the offer arrival rate. The contribution of this literature has been to help us understand equilibrium wage dispersion, that is, why equally productive workers need not all be paid the same wage.

In the second branch of the equilibrium search literature, the offer arrival
rate is determined endogenously. Assume that the rate at which unemployed workers and vacant jobs meet is determined by a constant returns to scale matching function whose arguments are $u$ and $v$, the measures of unemployed workers and vacancies, respectively. Constant returns to scale implies that the rate at which the unemployed receive job offers depends only on $\theta = v/u$, that is, on “labor market tightness.” Labor market tightness is determined by a free-entry condition, namely, that firms post vacancies so long as the expected payoff from doing so is positive. This expected payoff depends on the wage to be paid once the vacancy is filled. Because of the search frictions required to match job seekers with vacancies, when a worker and firm get together, there is a surplus to be split. The standard approach is to assume that this surplus is split according to a Nash bargaining rule with an exogenous share going to the worker. In the homogeneous worker/homogeneous firm version of the model, all workers are paid the same wage. This branch of the literature, known as the Diamond-Mortensen-Pissarides (DMP) model, thus focuses on endogenizing $\alpha$ while mostly ignoring the possibility of equilibrium wage dispersion. The contribution of this literature has been to help us understand the equilibrium or “natural” rate of unemployment. It is also obvious to ask whether this natural rate is efficient, and this question has in turn generated a substantial literature.

My focus in this survey is on the development of search theory, especially on the contributions of the laureates, but there is also a substantial corresponding body of empirical and policy-oriented work. Eckstein and Van den Berg (2007) and Postel-Vinay and Robin (2006) survey empirical work in job search, focusing in particular on work that comes out of the wage-posting literature, and two surveys by Mortensen and Pissarides (1999a,b) include a discussion of labor market policy analysis through the lens of the DMP model. The rest of this survey is organized as follows. The next section discusses wage-posting models and equilibrium wage dispersion. Section 3 is devoted to the DMP model. In Section 4, I discuss efficiency in markets with search frictions. Finally, in Section 5, I conclude.

---

1The preceding paragraphs make it sound as if these two branches of the equilibrium search literature are incompatible. On the contrary, the ideas of endogenizing the wage offer distribution and endogenizing the offer arrival rate can be combined in a straightforward way. See, e.g., Mortensen (2000).
2 Wage-Posting Models

The “Diamond paradox,” presented in Diamond (1971), is the seminal idea in the wage-posting literature. The paradox (which I will present in a labor market context – Diamond used a product market setting) is as follows. Consider a market in which unemployed workers search for vacant jobs. Suppose workers are homogeneous in the sense of being equally productive (each worker, when employed, produces output $y$), having the same time cost of search, and having the same flow value of leisure, $b$, and suppose firms are also homogeneous. Assume large numbers of workers and firms so that each agent is individually negligible. What wages will firms choose to offer in this setting? Consider a candidate distribution of wage offers, $F(w)$. Since workers are all the same, they all have the same reservation wage, say $R$, when drawing from $F(w)$. Each firm then wants to deviate from $F(\cdot)$ by offering $R$. Any offer below $R$ would be rejected, and firms don’t need to offer wages above $R$ to get workers to accept. But if every firm offers $R$, then the common worker reservation wage falls. Workers are willing to accept wages slightly below $R$ because of the cost of search. Every firm then wants to deviate to the new, lower reservation wage, etc. This “process” continues until all firms offer $b$, the “Diamond monopsony wage.” More precisely, the only Nash equilibrium in the wage-posting game is the symmetric one in which all firms post $b$. The situation is even “worse” if there is a monetary cost of search. In that case, unless the first search step is free – an assumption that is made in many equilibrium search models – no equilibrium exists.

In what sense is the Diamond (1971) result a paradox? Consider the market described in the preceding paragraph, but suppose there were no search frictions in the sense that workers could draw new wage offers instantaneously. Then the market would be competitive – firms would bid workers’ wages up to $y$. The “paradox” is that a cost of search that is arbitrarily close to zero moves the equilibrium wage from the competitive to the monopsony level, that is, from $y$ to $b$. More generally, a “small” search friction can have a “large” impact on the market equilibrium. The Diamond result is also important because it made search theorists rethink their “partial-partial” approach (the phrase comes from the influential 1973 survey article by Rothschild) to unemployment duration. In the individual sequential search problem, workers are assumed to sample wages from a distribution, $F(w)$, but the Diamond paradox suggests that in an equilibrium model, this distribution may well be degenerate, rendering the individual search problem meaningless.
Diamond (1971) sparked a substantial literature that attempted to find conditions consistent with equilibrium wage (or price) dispersion. There are several ways around the paradox. One possibility is to make a different assumption about the search process. For example, in a “noisy search” model, the number of offers that a worker receives in a period is treated as a random variable. This means that a firm offering a wage to a job applicant cannot be sure that the applicant might not have a better offer in hand from another employer. Noisy search can, under some conditions, give an equilibrium with wage dispersion. Butters (1977) is a particularly nice example. Nonsequential search (Burdett and Judd 1983) can likewise lead to wage dispersion. Alternatively, one can retain the sequential search assumption and allow for some heterogeneity across workers other than in productivity. One simple approach (Salop and Stiglitz 1977) is to assume that some workers have zero (monetary) search cost, while others have a positive search cost. If the fraction of workers with the zero search cost is small, the monopsony outcome obtains; if this fraction is large, the competitive outcome is realized; for intermediate values, there is wage dispersion in equilibrium. If the distribution of search costs has no mass point at zero, however, the Diamond paradox is more difficult to resolve. Axell (1977) is the seminal paper along this line. Finally, one can assume that workers are heterogeneous with respect to b – some workers find unemployment less onerous than others do – as in Albrecht and Axell (1984). It is worth noting that most of the papers mentioned in this paragraph resolve the Diamond paradox in a product-market setting. Except for the fact that the employment relationship is typically long term while the interaction between buyer and seller in a product market is usually a one-shot affair, the issues raised by search frictions in a labor market are essentially the same as those that come up in the product market.

Burdett and Mortensen (1998) offers a more fundamental resolution of the Diamond paradox in the sense of generating equilibrium wage dispersion in a model of sequential search in which workers are *ex ante* identical.\footnote{The publication date for the Burdett-Mortensen paper is deceptive. Versions of the paper were circulated in the late 1980's.} The key to their model is on-the-job search. The reservation wage of an employed job seeker is simply his or her current wage, so the distribution of reservation wages across job seekers is a mixture of the reservation wage of the unemployed with the distribution of reservation wages across employed searchers. This reservation wage heterogeneity in turn supports equilibrium wage dispersion. The idea underlying equilibrium wage dispersion is similar.
to the one used in the noisy search and nonsequential search literatures. When a job applicant contacts a firm, that firm doesn’t know whether it is in competition with another firm for that worker’s services and, if so, what wage the other firm is posting.

To understand how Burdett and Mortensen (1998) works, consider the following slightly simplified version of their model. Suppose workers, whether unemployed or employed, receive job offers at exogenous Poisson rate \( \alpha \). An unemployed worker then accepts any offer above the flow value of leisure, \( b \), while an employed worker accepts any offer above his or her current wage.\(^3\) The Poisson rate \( \alpha \) thus governs the rate at which workers move from unemployment to employment and the rate at which they “move up the job ladder.” In steady state, there needs to be a corresponding flow back into unemployment, so suppose that jobs break up at exogenous Poisson rate \( \lambda \).\(^4\)

We seek an equilibrium distribution of wage offers, \( F(w) \), that is, a distribution consistent with profit maximization and with the steady-state conditions implied by \( \alpha \) and \( \lambda \). The only equilibrium is one with continuous wage dispersion. To see this, suppose there were a mass point in \( F(w) \) at \( w \). Then a single firm could profit by increasing its wage offer to an arbitrarily small amount above \( w \) — the firm would gain access to the mass of workers employed at \( w \) at negligible extra cost per worker. In addition, the lowest wage offered must be \( b \). The firm posting the lowest wage can only hire the unemployed, and since unemployed workers accept any \( w \geq b \), this firm profits by lowering its wage to the flow value of leisure. A similar argument shows that the support of \( F(w) \) must be connected; that is, there cannot be “holes” in the support of \( F(w) \). The “no mass points” and “no holes” results follow more or less directly from Burdett and Judd (1983).

The next step is to develop and use the appropriate steady-state conditions. Associated with \( F(w) \), the distribution of wages offered, is \( G(w) \), the distribution of wages paid. Normalize the measure of firms to one, and let the measures of employed and unemployed workers be \( m - u \) and \( u \),

\(^3\) In Burdett and Mortensen (1998), the rates at which the unemployed and the employed receive wage offers differ. For example, it may be easier to get a job offer while unemployed. If these rates differ, then the reservation wage of the unemployed does not equal \( b \). My simplifying assumption that the arrival rate of job offers is the same for the unemployed and the employed makes it easier to present the basic idea of their model.

\(^4\) The notation is perhaps a bit awkward. Burdett and Mortensen (1998) use \( \lambda \) for the offer arrival rate; in the DMP model, \( \lambda \) typically denotes the job destruction rate. For notational consistency, I use \( \lambda \) for the job destruction rate throughout this survey.
respectively. Then steady state requires
\[ \alpha F(w)u = (\lambda + \alpha(1 - F(w))) G(w)(m - u). \tag{1} \]

The left-hand side gives the flow of workers from unemployment into jobs paying \( w \) or less; the right-hand side gives the corresponding outflow, both back into unemployment at rate \( \lambda \) and further up the wage ladder at rate \( \alpha(1 - F(w)) \). In addition, equating the flows of workers into and out of unemployment, that is, \( \alpha u = \lambda (m - u) \), gives \( \frac{u}{m - u} = \lambda/\alpha \). Combining these two steady-state conditions implies
\[ G(w) = \frac{\lambda F(w)}{\lambda + \alpha(1 - F(w))}. \tag{2} \]

Now, let \( \ell(w) \) be the steady-state employment of a firm posting \( w \). The measure of workers who are paid wages in the interval \( (w - \epsilon, w] \) is \( (G(w) - G(w - \epsilon))(m - u) \), while the measure of firms offering wages in \( (w - \epsilon, w] \) is \( F(w) - F(w - \epsilon) \). Therefore \( \ell(w) \) can be expressed as
\[ \ell(w) = \lim_{\epsilon \to 0} \left( \frac{G(w) - G(w - \epsilon)}{F(w) - F(w - \epsilon)} \right) (m - u). \]

Carrying out the requisite algebra,
\[ \ell(w) = \left( \frac{\alpha \lambda}{(\lambda + \alpha(1 - F(w)))^2} \right) m. \]

Employment is, of course, increasing in \( w \). A firm that posts a high wage attracts new hires relatively quickly and loses its employees relatively slowly.

The final step in the derivation of \( F(w) \) is to write the steady-state profit of a firm posting \( w \) as
\[ \pi(w) = (y - w) \ell(w), \tag{3} \]
where \( y \) is the flow value of output per worker. This expression illustrates the “volume-margin” tradeoff that makes it possible for firms to be indifferent across a range of wage offers. A firm that posts a low (high) wage employs relatively few (many) workers but receives a large (small) profit per worker. Since \( b \) is necessarily one of the wages posted in equilibrium, the

---

5 The measure of workers in the market, \( m \), is exogenous in this model, but \( u \) is determined endogenously.

6 The assumption that firms maximize steady-state profit requires that the discount rate go to zero, as is explicitly assumed in Burdett and Mortensen (1998). Coles (2001) shows that the model’s results also hold when the discount rate is positive.
requirement that all wages in the support of $F(\cdot)$ give equal steady-state profit gives

$$\pi(w) = \pi(b) = (y - b) \left( \frac{\alpha \lambda}{(\alpha + \lambda)^2} \right) m$$

for all $w$ in the support of $F(\cdot)$.

Substitution and rearrangement then gives

$$F(w) = \frac{\alpha + \lambda}{\alpha} \left( 1 - \left( \frac{y - w}{y - b} \right)^{1/2} \right)$$

for $b \leq w \leq y - (y - b) \left( \frac{\lambda}{\alpha + \lambda} \right)^2$. (4)

Burdett and Mortensen (1998) thus solves the Diamond paradox by adding the realistic assumption of on-the-job search. That is, it makes sense in equilibrium to think of homogeneous workers drawing wage offers from a nondegenerate distribution. The model also captures the fact that wages increase via job-to-job mobility. Wages increase with labor market experience as workers climb the wage ladder. Wage/experience profiles are concave, and mobility tends to decreases with experience. Finally, the model predicts that larger firms pay higher wages, as in reality.

On the other hand, from equations (2) and (4) it is clear that the equilibrium density of wages paid implied by the model is upward sloping, contrary to reality. This empirical problem with the basic model has been addressed by adding continuous firm heterogeneity, that is, a continuous distribution, $\Gamma(y)$, to the model. Bontemps, Robin and Van den Berg (2000) show that any wage distribution (subject to some implementability conditions) can be supported by an appropriately chosen $\Gamma(y)$. In this sense, the Burdett and Mortensen model, extended to allow for continuous firm heterogeneity in productivity, can fit the observed wage distribution perfectly. There remains, however, the question of whether the productivity distribution required to fit the wage data is realistic.\(^7\)

\(^7\)To be more precise about “implementability” and the realism of the productivity distribution required to fit the Burdett and Mortensen (1998) model, note the first-order condition for profit maximization (from equation 3) implies

$$y - w = \frac{\lambda + \alpha (1 - F(w))}{2\alpha f(w)}.$$ 

Using (2), this is equivalent to

$$y - w = \frac{\lambda + \alpha G(w)}{2\alpha g(w)}.$$ 

Two problems can arise. First, $y$ may not increase with $w$ (implementability) if $g'(w)$ is too large, which can be a problem in applications at the left tail of $G(\cdot)$. Second, $y - w$ may get arbitrarily large at high wages (realism) if $g(w)$ is small enough.
The Burdett and Mortensen model has been extended in several directions, in particular by Burdett and Coles (2003) and Postel-Vinay and Robin (2002). In Burdett and Mortensen (1998), a firm that wants to reduce worker turnover can do so by paying a higher wage. However, the firm could achieve the same end at a lower cost by posting a wage-tenure contract; in particular, by offering a wage-tenure contract that specifies a low wage up to some tenure $t^*$ with a promise of a high wage thereafter. This is directly analogous to the argument that the “shirking problem” in the efficiency wage model (Shapiro and Stiglitz 1984) could be solved by requiring workers to post a bond, and the basic issue is the same, namely, whether the firm can commit credibly to its promise. Assuming credible commitment is possible, that is, that firms can offer wage-tenure contracts, in equilibrium all firms post the same step function, implying a two-point equilibrium wage distribution (Stevens 2004). However, Burdett and Coles (2003) retain the continuous wage distribution of Burdett and Mortensen (1998) by generating equilibrium wage-tenure contract dispersion. The extra assumption needed to get this result is risk aversion (and, of course, that workers can neither borrow nor save). On the one hand, firms prefer the extreme backloading of the step function; on the other hand, workers prefer a constant wage through time. The optimal contract is a compromise between these two extremes, and in equilibrium there is a smooth, upward sloping wage-tenure profile. Different firms offer different contracts but only in terms of where their contracts start on the equilibrium profile. Burdett and Coles (2003) thus retains wage growth via job-to-job mobility – a worker at a firm that offered a low starting salary will switch to a firm with a higher starting salary if his or her current wage is below the new firm’s starting salary – and adds wage growth within the firm, that is, a tenure effect.

Postel-Vinay and Robin (2002) also generates tenure effects, albeit for a different reason. In Burdett and Mortensen (1998), incumbent employers don’t react when their workers receive outside offers. Some possible justifications for this assumption are (i) firms may not be able to match an outside offer for one worker without also increasing the wages it pays its other workers, (ii) firms don’t want to encourage on-the-job search, and (iii) outside offers aren’t necessarily credible. In reality, however, firms (sometimes but not always) do react to outside offers. Postel-Vinay and Robin (2002) explores the implications of having firms respond to outside offers by allowing Bertrand competition between the current employer and the prospective future employer. As in the extended version of Burdett and Mortensen (1998), suppose productivity is continuously distributed as $\Gamma(y)$ across firms. Then Postel-Vinay and Robin (2002) retains the following results from Burdett
and Mortensen (1998): (i) the distribution of wages paid is continuous, (ii) larger firms tend to pay higher wages, and (iii) wages tend to increase with labor market experience but at a decreasing rate. In addition, their model offers some new results: (i) there is within-firm wage dispersion, (ii) wages tend to increase with tenure, and (iii) workers sometimes accept wage cuts to move to more productive firms. The intuition for the last result is that a worker may be willing to move to a new firm at a lower wage if that new firm offers substantially more attractive promotion possibilities.

Summing up, Burdett and Mortensen (1998) resolves the Diamond (1971) paradox in an appealing way. This model, and the later papers that it inspired, provide a search-theoretic structural interpretation of (i) why equally productive workers may receive different wages both within and across firms, (ii) why larger firms tend to pay higher wages, and (iii) why wages tend to increase with experience and tenure. These papers have generated a substantial body of empirical work on job search.

It is worth noting, however, that the question of unemployment, which originally motivated the equilibrium search literature, “got lost” in this strand of the wage-posting literature. There is no endogenous search unemployment in Burdett and Mortensen (1998). The unemployed draw wage offers from a nondegenerate distribution, but in equilibrium, none of these offers are ever rejected. This can be fixed by adding non-productivity-related worker heterogeneity, e.g., in $b$, as in Albrecht and Axell (1984), to the basic model, and indeed Burdett and Mortensen did precisely that in their 1998 paper, but the general point remains that the literature that flows out of their paper is more focused on equilibrium wage dispersion than on unemployment. Instead, most search economists view unemployment through the lens of the Diamond-Mortensen-Pissarides model.

3 The Diamond-Mortensen-Pissarides Model

The early one-sided search models looked at unemployment duration as a question of how many offers an unemployed worker would reject before receiving an acceptable one. These models thus took a (labor) supply side view of unemployment. The Diamond-Mortensen-Pissarides model reverses this focus. In the simplest version of DMP, there is no issue about offers being rejected. Rather, unemployment duration is determined by how long it takes a worker to receive an offer, and in turn the arrival rate of offers is determined by how many vacancies firms choose to post. In this sense, DMP takes a (labor) demand side view of unemployment.
The Diamond-Mortensen-Pissarides model was designed to understand the equilibrium or “natural” rate of unemployment. Abstracting from movements into and out of the labor force, the equilibrium level of unemployment is determined by incidence, the rate at which employed workers move into unemployment, and duration, the inverse of the rate at which workers flow in the opposite direction. The earliest version of DMP, which was mostly developed in the 1980’s, concentrates on duration, treating incidence as exogenous. A later version of DMP (Mortensen and Pissarides 1994) jointly endogenizes duration and incidence.

The basic DMP model has three elements. First, unemployed workers and vacancies meet one another at Poisson rate \( m(u, v) \), where \( u \) is the measure of unemployed workers and \( v \) is the measure of vacancies in the economy. The matching function \( m(\cdot) \) is a “black box”; that is, it is not given any microfoundation. This function is assumed to exhibit constant returns to scale, so the rate at which an unemployed worker meets a firm with a vacancy is 
\[
\frac{m(u, v)}{u} = m(1, \theta) \equiv \alpha(\theta),
\]
where \( \theta = v/u \) is “labor market tightness.” Similarly, the rate at which a vacancy is contacted by an unemployed worker is 
\[
\frac{m(u, v)}{v} = \frac{m(u, v) u}{v} = \frac{\alpha(\theta)}{\theta}.
\]
Assume that \( \frac{\alpha(\theta)}{\theta} \) is increasing and that \( \alpha(\theta) \) is decreasing in \( \theta \). Second, labor market tightness is determined by free entry of vacancies. As firms post more vacancies, \( \theta \) increases, and the expected profit associated with vacancy posting falls. Labor market tightness adjusts until the value of a vacancy equals zero. Third, when a worker and a firm match, they enjoy a surplus relative to what they would get were they to remain unmatched. DMP assumes that the division of this surplus is determined by Nash bargaining, and this in turn determines the worker’s wage.

The model is closed by a steady-state condition. Normalize the size of the labor force to one, of which a fraction \( 1-u \) is employed and \( u \) is unemployed. Workers move from unemployment to employment at endogenous rate \( \alpha(\theta) \) and in the reverse direction at exogenous rate \( \lambda \). In steady state, \( \alpha(\theta) u = \lambda (1-u) \) or
\[
u = \frac{\lambda}{\lambda + \alpha(\theta)}.
\]
Equation (5) emphasizes that the fundamental equilibrium object in DMP is labor market tightness. Once \( \theta \) is determined, the unemployment rate follows residually. Note also that since \( \theta = v/u \), equation (5) gives an equilibrium relationship between vacancies and unemployment.
The basic model is straightforward to solve. The model is set in continuous time. Workers and firms are assumed to be homogeneous. Each firm consists of a single job or, equivalently, there are constant returns to scale in production. There is a unit measure of workers, and each worker moves back and forth between unemployment and employment. Let $U$ and $N$ be the values, that is, the expected discounted lifetime utilities, associated with these two states. On the firm side, there is an arbitrarily larger measure of potential jobs. Jobs that are in the market are either vacant or filled with corresponding values $V$ and $J$. The model is easiest to see by writing expressions for the four values. The worker values $U$ and $N$ are determined by

$$rU = z + \alpha(\theta)(N - U)$$

$$rN = w + \lambda(U - N).$$

Discounting the future at rate $r$, the flow value of unemployment equals an instantaneous return (e.g., the flow value of home production), $z$, plus the instantaneous probability, $\alpha(\theta)$, of changing state (finding a job) times the capital gain associated with doing so, $N - U$. Similarly, the flow value of employment equals the wage plus $\lambda$ times the capital loss associated with a move back into unemployment. On the firm side,

$$rV = -c + \frac{\alpha(\theta)}{\theta}(J - V)$$

$$rJ = y - w + \lambda(V - J).$$

The firm incurs flow cost $c$ while its vacancy is unfilled, but at rate $\alpha(\theta)/\theta$ it hires a worker and realizes a capital gain of $J - V$. A filled job gives flow surplus $y - w$, but at rate $\lambda$ the match ends and the firm suffers a capital loss of $V - J$. In equilibrium, free entry of vacancies gives $V = 0$, which in turn implies

$$c = \frac{\alpha(\theta)}{\theta} \left( \frac{y - w}{r + \lambda} \right).$$

This “job creation condition” gives an equilibrium relationship between labor market tightness and the wage. From a firm’s perspective, the higher is $w$, the less willing it is to post a vacancy, that is, the lower is $\theta$.

The fact that it takes time and resources for a worker and a firm to find one another means that when the pair finally forms a match, there is a surplus relative to the value of continued search. Using equation (9), the surplus realized by the firm relative to the value of continuing to post
a vacancy (with the latter equal to zero in equilibrium) is \( J = \frac{y - w}{r + \lambda} \). Similarly, using equation (7), the surplus realized by the worker relative to the value of remaining unemployed is \( N - U = \frac{w - rU}{r + \lambda} \). The total surplus to be shared is thus \( \frac{y - rU}{r + \lambda} \), and the wage determines what fraction of that surplus goes to the worker. DMP simply sets this share exogenously. That is, the model is closed by assuming that the worker gets an exogenous share \( \beta \) of the total surplus, that is, \( w - rU = \beta(y - rU) \). Equivalently,

\[
w = \beta y + (1 - \beta) r U,
\]

or, after substitution,

\[
w = \beta y + (1 - \beta) \frac{(r + \lambda) z + \alpha(\theta) w}{r + \lambda + \alpha(\theta)}.
\] (11)

This “wage curve” can be derived as the solution to a generalized Nash bargaining problem with an exogenous worker bargaining share of \( \beta \). Equation (11) gives a second equilibrium relationship between \( \theta \) and \( w \). From the worker’s perspective, the higher is \( \theta \), the easier it is to get a new job offer. A high value of \( \theta \) thus puts the worker in a relatively strong bargaining position when negotiating with a firm about how to split the surplus that would be generated by their match. That is, from the worker’s perspective, \( w \) increases with \( \theta \).

The steady-state equilibrium of this model is the \((\theta, w)\) pair that solves equations (10) and (11), and the equilibrium unemployment rate then follows from equation (5). The DMP model is easy to use for comparative steady-state analysis (how do parameter changes shift the job creation and wage curves?), and it is the tool of choice for most macro labor economists for understanding how various labor market policies affect the natural rate of unemployment. DMP is simple in the sense that the only explicit decisions are firms’ choices about whether or not to post vacancies, but this simplicity is a virtue because it makes it possible to extend the model in a variety of useful directions.

The history of the Diamond-Mortensen-Pissarides model is difficult to sort out in the sense that it is not easy to say who came up with which idea first, but it is certainly the case that the general label is justified. The matching function idea seems to have first been suggested in Phelps (1968); Pissarides (1979) and Diamond and Maskin (1979) developed the idea more or less in the way that it is used in DMP. The key idea of closing the model
with a simple free entry condition seems to originate in Pissarides (1979). The idea of applying the Nash bargaining solution to determine the division of surplus in a match between a worker and a firm seems to have first been made explicit in Diamond (1982b), although related surplus-sharing ideas are present in Mortensen (1978) and Diamond and Maskin (1979). These ideas were explicitly combined in a series of papers by Pissarides in the mid-1980's (Pissarides 1984a,b and 1985), and especially in Pissarides (1990/2000), which is the universally recognized reference in the field.

The basic Diamond-Mortensen-Pissarides model has been taken in several directions. It is, for example, straightforward to allow for capital, for stochastic match output, and for out-of-steady-state dynamics. I will briefly discuss three variations on the basic model – increasing returns to scale in the matching function, endogenous separations, and business cycle fluctuations – that have been particularly influential.

The Diamond (1982a) “coconuts model” considers the implications of increasing returns to scale in the matching function. Increasing returns are important because they can lead to multiple equilibria. Diamond (1982a) captured the idea beautifully in the following parable. Consider an island economy in which coconuts are the only good. Consuming a coconut gives a fixed level of utility, but there is a taboo against consuming one’s own coconut. Instead, the inhabitants of the island need to search for a tree with a coconut to harvest, climb the tree, pick the coconut, and then search for another islander who is also carrying around a coconut to trade. Production opportunities, that is, trees with coconuts, are found at an exogenous rate, but there are increasing returns to scale in the search for trading partners. That is, the more islanders who are looking to trade, the easier it is for any one individual to find a trading partner. The decision that has to be made in this economy is which production opportunities to undertake. That is, how high a tree is it worth climbing to harvest a coconut? This is where increasing returns in the search for trading partners matters. If everyone else only picks low-hanging fruit, then there will be relatively few islanders searching for a trading partner, and it won’t be worthwhile for any one individual to climb a high tree. On the other hand, if everyone else climbs high, then there will be many searchers, and it will be worthwhile for the individual to climb high as well. Increasing returns in matching is a classic example (perhaps the first clearly articulated one) of multiple equilibria induced by a coordination externality. These multiple equilibria are Pareto-rankable, suggesting a role for policy.

Mortensen and Pissarides (1994) extend the basic DMP model to incorporate endogenous job destruction. As in the basic model, unemployed work-
ers match with vacancies at endogenous rate \( \alpha(\theta) \). When a match forms, it is “state of the art” in the sense that match productivity starts at its highest possible value. Over the life of the match, however, shocks arrive (as usual, at an exogenous Poisson rate) to match productivity. The match ends when its productivity falls below an endogenous threshold value which is defined by the condition that the joint value of continuing the match equals the sum of the individual values of going it alone. Mortensen and Pissarides (1994) has been influential because it provides a nice theoretical counterpart to the job destruction/job creation data compiled and analyzed by Davis and Haltiwanger (e.g., Davis and Haltiwanger 1992). The Mortensen and Pissarides model is also important from a policy perspective since endogenous job destruction is an essential component of any model designed to understand the effects of firing restrictions and mandatory severance pay on unemployment and wages.

Finally, DMP has been extended to the analysis of cyclical fluctuations in unemployment. Pissarides (1985) is the seminal paper. As “macro labor,” it is natural to ask whether the model is consistent with observed time series on labor market aggregates. That is, does DMP “match the business cycle facts?” (Cole and Rogerson 1999). In an influential paper, Shimer (2005a) argues that the answer is “no.” Specifically, \( \theta \) is very much less responsive to fluctuations in \( y \) in a calibrated version of the model than it is in the data. There is, of course, considerable controversy about how much less responsive the elasticity of \( \theta \) with respect to \( y \) is in the calibrated model than in the data. The target value for this elasticity is 19.1% in Shimer (2005a), and his calibration predicts an elasticity of 1.71%, whereas the comparable figures in Pissarides (2009) are 7.56% and 3.62%. Nonetheless, even using the Pissarides (2009) numbers, the model misses its target by a factor of almost 2. Shimer (2005a) and many others argue that the inconsistency between the data and the model has primarily to do with the Nash bargaining assumption – that wages are too flexible in DMP and absorb too much of the impact of productivity fluctuations. Pissarides (2009) argues, on the other hand, that wages in new matches are in fact quite flexible and that the key to reconciling DMP with the business cycle facts has to do with the specification of how vacancy posting costs vary with \( \theta \). In general, the question of how best to reconcile DMP with the time series variation in unemployment, vacancies and wages – or, indeed, whether some other macro labor model, e.g., one based on the Lucas and Prescott (1974) paper, might be more appropriate for the analysis of cyclical unemployment – is currently a very active area of research.

Since I noted that meaningful search unemployment has been “lost” in
the strand of the wage posting literature that focuses on on-the-job search, it is only fair also to note that wage dispersion is not central to the DMP model. In the simplest version of the model, all workers are paid the same wage, and although there is a distribution of wages in Mortensen and Pissarides (1994), their version of the model is not consistent (and was not designed to be consistent) with the facts about how wages evolve over workers’ careers that Burdett and Mortensen (1998) and related models fit so well. To get meaningful equilibrium wage dispersion in DMP requires adding on-the-job search to the model. The standard approach to incorporating on-the-job search into DMP assumes in effect that when an employed job seeker moves from one firm to another, the worker quits his or her old job before bargaining with the new employer. So, the outside option for the worker in the negotiation with the new employer is $U$, that is, the same as if the job had been found by an unemployed searcher. This shuts down the possibility of between-employer competition, which is precisely the source of wage dispersion in the Burdett and Mortensen model. Restoring between-employer competition in models of wage bargaining with on-the-job search is not straightforward. Shimer (2005c) shows that, if the wage is bargained over once and for all at the time of match formation and remains fixed (non-renegotiable) for the duration of the match - this is one way of generalizing the Burdett-Mortensen wage-posting assumption to a bargaining environment - , then the wage affects the worker’s quit rate. The wage therefore does more than simply determine how match surplus is divided – it also affects the size of the surplus by affecting expected match duration. This makes match surplus a non-convex function of the wage and invalidates the Nash bargaining solution used in the DMP model. An alternative approach, immune from the Shimer problem, is taken by Cahuc, Postel-Vinay and Robin (2006). They allow wages to be renegotiated by mutual consent over the course of a job spell. The bargaining game they analyze implies that the outside option for an employed job seeker is the full surplus that would have been generated had the worker chosen the loser in the between-firm competition.

On balance, however, the Diamond-Mortensen-Pissarides model has been extremely successful and influential. If I have been successful in my description of the model, I have made it seem relatively straightforward and user-friendly. One of the model’s many virtues is that it can be used to analyze a wide range of issues. It is, for example, the tool of choice for understanding how the labor market fits into more general macro models, e.g., Andolfatto 1996, Merz 1995, and Gertler and Trigari 2009, but the basic ideas of the model apply much more broadly. In my own work, for example, I have used
variations on DMP to look at the effects of skill-biased technical change on unemployment and wages (Albrecht and Vroman 2002) and to analyze the effects of labor market policies in developing countries (Albrecht, Navarro and Vroman 2009). A list of applications that use DMP could go on and on. My point is simply that this is a very useful model. Of course, the fact that a model seems straightforward ex post does not mean that its development was easy. Quite the contrary – Diamond, Mortensen and Pissarides have made an important contribution by capturing the essence of complicated labor markets with search frictions in a concise and clean model.

4 Efficiency

Is the natural rate of unemployment efficient? Phelps (1972) and Prescott (1975) posed this question in the context of wage-posting models. In wage-posting models that assume an exogenous arrival rate of job offers, efficiency issues only arise at the level of a prospective match between an unemployed worker and a vacant job. The question is whether too many, too few, or just the right number of these matches form. In the basic model with homogeneous workers and homogeneous firms, so long as market work is more productive than home work, the efficient outcome is one in which all possible matches are consummated. In this sense, the Diamond (1971) monopsony wage outcome is efficient as is the Burdett and Mortensen (1998) equilibrium, whereas equilibria in which some offers are rejected in favor of further search are inefficient as in, for example, Albrecht and Axell (1984). One can make the efficiency issue in wage-posting models more interesting by assuming private information about match quality (Albrecht and Jovanovic 1986), but until recently – and then only in a competitive or directed search framework, on which more below – the issue of efficiency in wage-posting models has essentially been moribund.

The efficiency of equilibrium in the Diamond-Mortensen-Pissarides model has, in contrast, been extensively investigated over the past thirty years. The efficiency issue in DMP is, in a sense, the opposite of the one considered in the wage-posting framework. Given any level of labor market tightness, the DMP approach assumes pairwise efficiency. That is, when an unemployed worker meets a vacancy, the match is formed if and only if the joint surplus from doing so is greater than the value of continued search by the two parties. The efficiency question instead is whether the arrival rate of job offers is the correct one, that is, whether the equilibrium value of $\theta$ equals the value that a social planner would choose. To see the issue, consider a
firm that is deciding whether to post a vacancy. There are two externalities associated with this choice. On the one hand, if the firm posts the vacancy, it increases the expected time required to fill the vacancies that are already on the market. The firm’s private calculus does not take into account the “congestion externality” that it imposes on other firms. On the other hand, when the firm posts the vacancy, it makes the other side of the market, that is, the unemployed job seekers, better off. The firm also does not take this positive “thick market externality” into account. There is a tradeoff between two potentially countervailing externalities, and the social planner level of labor market tightness is the level that gets this tradeoff just right.

DMP assumes that worker-firm matches form if and only if doing so is mutually advantageous but does not specify precisely how the match surplus is divided between the two parties. That is, the precise value of $\beta$, the share of the surplus that goes to the worker, is not pinned down by the model. Of course, varying $\beta$ changes the incentive that firms have to post vacancies. Hosios (1990) gives a general condition on $\beta$ that just balances the congestion and thick market externalities. This “Hosios condition” builds on earlier analysis by Diamond (Diamond and Maskin 1979, Diamond 1981 and 1982b), Mortensen (1982), and Pissarides (1984a,b).

The Hosios condition is easiest to understand in a simplified one-period version of DMP in which all workers start out unemployed. Consider a social planner who chooses how many vacancies to post in this economy. The social planner’s problem is

$$\max_v m(1,v)y + (1 - m(1,v))z - cv.$$  

A social planner who posts $v$ vacancies can expect $m(1,v)$ matches to form; equivalently, in this simplified model, the planner can expect a fraction $m(1,v)$ of the unemployed to find a job. Each of these matched workers produces output $y$; those who remain unmatched each generate home production $z$. On the other hand, each vacancy entails a posting cost of $c$. The efficient level of vacancy creation solves

$$m_v(1,v)(y - z) = c.$$  

The corresponding free-entry level of vacancies is determined by

$$\frac{m(1,v)}{v}(y - w) = c.$$  

In the one-period setting, the worker’s outside option is just $z$, so

$$w = \beta y + (1 - \beta)z,$$
and the free-entry level of vacancies solves

\[
\frac{m(1,v)}{v} (1 - \beta)(y - z) = c.
\]

Efficiency thus requires

\[
\beta^* = 1 - \frac{m_v(1,v)}{m(1,v)/v}.
\] (12)

The Hosios condition sets the firm’s share of the surplus equal to the elasticity of the matching function with respect to \(v\). This condition generalizes in a straightforward way to the infinite-horizon, steady-state DMP economy (e.g., Pissarides 2000, Ch. 8).

The Hosios condition turns out to be quite general. In the basic DMP model, the only decision variable is how many vacancies to post, but the same condition applies with endogenous job destruction, with endogenous search intensity (on either or both sides of the market), with endogenous job rejection brought about by idiosyncratic match-specific productivity, etc. There are, however, situations in which equation (12) does not give efficiency, for example, if firms make \textit{ex ante} investments. Suppose the firm chooses \(y\) at cost \(k'(y) > 0\). With \textit{ex post} bargaining, there is a “holdup problem” since the worker gets a fraction of the surplus generated by an investment that the firm undertook by itself. In this situation, \(\beta\) cannot be set so that both \(y\) and \(\theta\) are chosen efficiently. The Hosios condition also does not apply when there are “composition externalities” in the market. Suppose, for example, that low-skill and high-skill workers are searching in the same market, that is, potentially “getting in each other’s way.” In this situation, when a firm hires, say, a high-skill worker, it imposes an externality by changing the mix of types in the pool from which other firms hire. Nor does the Hosios condition give efficiency when the assumption of constant returns to scale in the matching function and/or in production does not hold.

In any event, it is just happenstance if \(\beta\) equals the Hosios value in the DMP model. Therefore a natural question to ask is whether the wage implied by \(\beta = \beta^*\) in DMP can be implemented by a different wage-setting mechanism. An affirmative answer was given by Moen (1997). (A simplified version of) his model retains the basic DMP framework but assumes that each firm announces and commits itself to a wage, as opposed to having wages determined \textit{ex post} by a Nash bargain. Each worker observes all wage announcements and applies to the most attractive vacancy. Firms create vacancies (and announce wages) until the value of doing so equals zero.
The outcome is one of competitive search equilibrium – competitive in the sense that because workers can observe all wage announcements, firms have no market power. Moen (1997) proves that competitive search equilibrium is efficient. That is, the competitive search equilibrium level of vacancy creation equals the level that a social planner would choose. Further, in equilibrium, all firms announce the same wage, which equals the wage that would be determined in DMP using Nash bargaining with $\beta$ at the Hosios value. In fact, competitive search also solves the holdup problem in the DMP model. Acemoglu and Shimer (1999) show that competitive search equilibrium with ex ante investment by firms is efficient. They also note that efficiency does not require every worker to observe all wage announcements – it is enough for every worker to see at least two posted wages.

The key difference between Moen (1997) and DMP is the way that search is conducted. In DMP (and in the wage-posting models discussed in Section 2), search is assumed to be random. Unemployed job seekers and vacancies are assumed simply to “bump into one another.” In particular, an attractive vacancy is no more likely to receive job applications than is an unattractive one. In Moen (1997), search is directed in the sense that each worker decides where to apply. If different vacancies announce different wages, then each worker takes into account (i) the probability that his or her application will be successful and (ii) the wage that will be received conditional on getting the job. The worker faces a “volume-margin tradeoff” (probability of getting the job versus the wage received by the successful applicant); so too does the firm (probability of hiring a worker versus the wage it has to pay if it fills the job). The intuition behind Moen’s efficiency result is that competition forces the rate at which workers are willing to trade off volume for margin to equal the corresponding rate for firms.

Competitive/directed search is currently a very active area of research. This work is closely related to earlier papers on “ex ante” pricing and “competing mechanisms,” especially those of Peters (1984, 1991). In particular, Peters’s work gives a careful discussion of the conditions under which the large-economy framework used in directed search models makes sense; that is, under what conditions can we assume that an individual firm can announce its wage without worrying about the possibility that other firms

---

8 The two terms, “competitive” and “directed” search are often used interchangeably. I think the following distinction is useful. Suppose firms posting wage $w$ attract applicants at Poisson rate $\theta$. In a competitive search model, it is assumed that this implies an offer probability of $\alpha(\theta)$ for each worker. As in DMP, $\alpha(\theta)$ is treated as a “black box.” In a directed search model, a specific search process is assumed, often urn-ball matching, and $\alpha(\theta)$ is given a microfoundation.
might adjust their posted wages in response? (A similar point is also addressed in Burdett, Shi and Wright 2001.) In this sense, Peters’s work provides a microfoundation for the directed search literature.

The directed search approach is attractive for two reasons. The first is descriptive realism – it seems natural to assume that workers have at least some idea about which prospective employers are the more attractive ones and that they direct their search accordingly. Of course, there is a question about how precise workers’ information about their search options needs to be to enforce full competition across vacancies for their applications. Second, the basic directed search model gives a useful benchmark for efficiency analysis. The equilibrium level of labor market tightness and the social planner level coincide in a directed search model with homogeneous workers and homogeneous firms in which each worker can see at least two wage announcements and in which each firm is fully committed to paying its announced wage. We can ask which departures from the basic model preserve efficiency and which lead to inefficient outcomes. For example, when the assumption that each worker applies to only one vacancy is relaxed (as is done in Albrecht, Gautier and Vroman 2006 and in Galenianos and Kircher 2009), the directed search equilibrium is generally inefficient. When we relax the assumption that workers and firms are homogeneous, the directed search equilibrium may or may not be efficient (Shi 2001, Shimer 2005b). The issue now is not only whether labor market tightness is at the optimal level but also whether the right types of jobs are created and whether the various worker types direct their search correctly. Finally, with worker heterogeneity and adverse selection, directed search equilibrium is in general inefficient, even if firms can post contracts that separate the worker types (Guerrieri, Shimer and Wright 2010).

Summing up, the equilibrium value of labor market tightness in the Diamond-Mortensen-Pissarides model is in general not equal to the corresponding social optimum. Equivalently, the natural rate of unemployment is generically inefficient in the DMP framework. Efficiency requires that the worker share parameter in the Nash bargain over match surplus equal the Hosios value, and in DMP, there is no particular reason that this equality should hold. However, if instead of assuming that wages are determined \textit{ex post} by Nash bargaining, we assume that wages are determined \textit{ex ante} through competitive/directed search, then the DMP equilibrium is efficient. The question of how robust the efficiency of competitive search equilibrium is to departures from the baseline model is an important part of the current research agenda in search theory.
5 Conclusion

The three Nobel laureates have made significant contributions to our understanding of how markets with search frictions function. The Diamond-Mortensen-Pissarides model provides a practical tool for understanding aggregate unemployment. The model gives clear insights about the link between labor market policy and aggregate labor market outcomes, and the basic DMP framework has been extended in a variety of useful directions. At the same time, the contributions of the laureates are much more general. Although I have organized my survey around the labor market, other markets are also characterized by substantial search frictions. Consumer search is an important feature of many product markets, search theory has contributed to our understanding of the housing market, and an important part of monetary theory takes a search-theoretic perspective. Search theory is an active area of research, and this research is based to a large extent on the work of Peter Diamond, Dale Mortensen and Christopher Pissarides.

References


