2

Simple Models of Monopsony and Oligopsony

This chapter introduces some simple models of monopsony and oligopsony which form the foundation for the analysis in the rest of the book. The first three sections present some partial equilibrium models: the textbook static model of monopsony, a simple model of dynamic monopsony, and what is called a generalized model of monopsony where the firm has instruments other than the wage to influence the flow of recruits. The fourth section then presents a general equilibrium model of dynamic oligopsony (based on a simplified version of Burdett and Mortensen, 1998) to show how the framework is a fully coherent vision of the labor market as a whole. Although this model is highly stylized, it does capture the most important features of a labor market with frictions. Workers are faced with a distribution of wages so that there are good jobs and bad jobs. They try to get themselves into the good jobs but their progress resembles a game of "snakes and ladders." Sometimes they meet a "snake" and suffer the misfortune of losing their job and sometimes they find a "ladder" and have the good fortune to move to a better job. From the perspective of employers, the frictions in labor markets give them some discretion in setting wages. If they lower wages, they find it more difficult to recruit and retain workers but the existing workers do not all leave immediately and they continue to be able to recruit some workers so that they retain some workers even in the long run. The wages that employers set are influenced by competition from other employers but this competition is neither so cutthroat as to enable workers to extract all the surplus from the employment relationship, nor so weak as to enable employers to extract all the rents.

The chapter concludes by arguing that the fraction of recruits from non-employment is a good "back-of-the-envelope" measure of the extent to which workers are able to freely move between employers and, hence, of competition among employers for workers and the extent of market power possessed by employers in the labor market. Empirical evidence from the United Kingdom and the United States suggests that 45–55% of recruits were previously non-employed, a level which is likely to give employers considerable market power.

2.1 Static Partial Equilibrium Models of Monopsony

Given the lack of attention paid to monopsony in much of labor economics it is perhaps helpful to start with a quick review of the static textbook model of monopsony. In this model, the firm is assumed to face a labor supply curve that relates the wage paid, w, to the level of employment, N. Denote the supply of labor to the firm if it pays w by $N(w)$. Also, denote the inverse of this relationship by $w(N)$. Both $N(w)$ and $w(N)$ will be referred to as the labor supply curve to the individual firm. Total labor costs are given by $w(N)N$. Assume that the firm is a simple monopsonist who has to pay a single wage to all its workers (the incentives for wage discrimination are discussed in chapter 5). Assume the firm has a revenue function $Y(N)$. It wants to choose N to maximize profits which are given by

$$\pi = Y(N) - w(N)N \tag{2.1}$$

This leads to the first-order condition

$$Y'(N) = w(N) + w'(N)N \tag{2.2}$$

The left-hand side of (2.2) is the marginal revenue product of labor. The right-hand side is the marginal cost of labor, the increase in total labor costs when an extra worker is hired. The marginal cost of labor has two parts: the wage, w, that must be paid to the new worker hired and the increase in wages that must be paid to all existing workers. The solution is represented graphically in figure 2.1. Equilibrium is on the labor supply curve with the wage paid to workers being less than their marginal revenue product. Although the employer is making positive profit on the marginal worker, there is no incentive to increase employment because doing so would require increasing the wage (to attract the extra worker) and this higher wage must be paid not just to the new worker but also to all the existing workers. One particularly useful way of representing the choice of the firm is that marginal cost of labor is a mark-up on the wage, the mark-up being given by the elasticity of the labor supply curve facing the firm. Let us write the elasticity of the labor supply curve facing the firm as $\varepsilon_{Nw} = wN'(w)/N(w)$and let ε be the inverse of this elasticity. Then (2.2) can be written as

$$\frac{Y' - w}{w} = \frac{1}{\varepsilon_{Nw}} = \varepsilon \tag{2.3}$$

so that the proportional gap between the wage and the marginal revenue product is a function of the elasticity of the labor supply curve facing the firm. The gap between the wage and the marginal revenue product is what Pigou (1924) and Hicks (1932) referred to as the rate of exploitation and

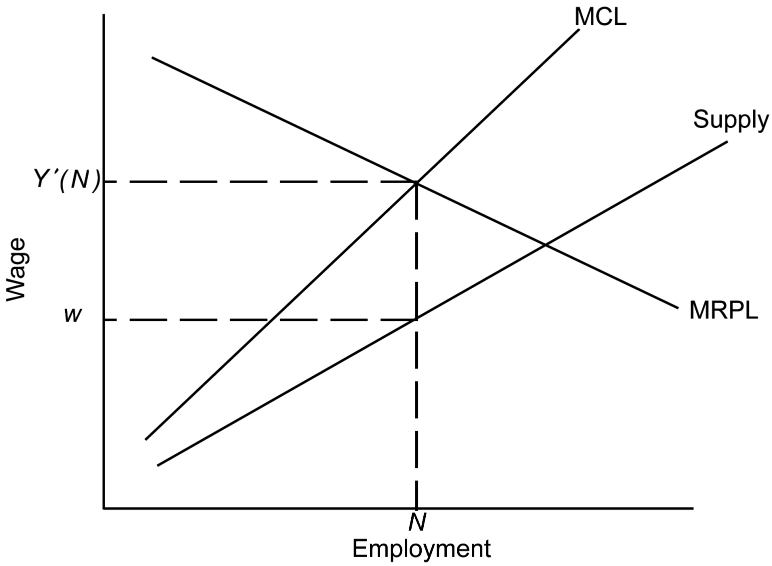

Figure 2.1 The textbook model of monopsony.

we will follow this tradition. Perfect competition corresponds to the case where $\varepsilon_{Nw} = \infty$ and $\varepsilon = 0$ in which case (2.3) says that the wage will be equal to the marginal revenue product.

Some of the comparative statics of the static monopsony model are the same as in the competitive model and some are different. For example, an increase in the marginal revenue product of labor will lead to an increase in employment and a rise in wages. The former would occur in a competitive model but the latter would not as a competitive firm would simply continue to pay the market wage. The impact of shifts in the labor supply curve to the firm are more complicated as the impact depends on how the change affects the marginal cost of labor and not just the average cost of labor. An increase in the supply of labor to the firm that keeps the elasticity the same will result in a rise in employment and a fall in wages just as in the competitive model. But, matters are more complicated if the elasticity of the labor supply curve changes as the average and marginal cost of labor can move in opposite directions; the most familiar example of this is the impact of a minimum wage. The minimum wage raises the average cost of labor but (if it is binding) reduces $w'(N)$ so its effect on the marginal cost of labor (see (2.2)) is ambiguous. In fact, one can show that a minimum wage that just binds must raise employment (a demonstration of this can be found in most labor economics textbooks).

2.2 A Simple Model of Dynamic Monopsony

One might wonder how this completely static model of the labor market corresponds to the description of the labor market in the first chapter that was based on dynamic arguments. The static and dynamic models can be linked in the following way. Assume that workers leave the firm at a rate s that depends negatively on the wage paid, and recruits arrive at the firm at a rate R that depends positively on the wage. If the firm had N_{t-1} workers last period and pays w_t this period, its labor supply will be

$$N_t = [1 - s(w_t)]N_{t-1} + R(w_t) \tag{2.4}$$

where $s(w)$ is the separation rate and $R(w)$ the recruitment rate.

In a steady state, total separations sN must equal recruits R so that we have

$$N(w) = \frac{R(w)}{s(w)} \tag{2.5}$$

giving us a positive long-run relationship between employment and the wage.[1] In this case the elasticity of the labor supply curve facing the firm can be written as

$$\varepsilon_{Nw} = \varepsilon_{Rw} - \varepsilon_{sw} \tag{2.6}$$

where ε_{Rw} is the elasticity of recruits with respect to the wage and ε_{sw} is the elasticity of separations with respect to the wage.

In a dynamic model, there is an important distinction between the elasticity of the short-run labor supply curve facing the employer and the long-run elasticity. The elasticity of (2.6) is the long-run elasticity of the labor supply curve facing the firm. The short-run elasticity, denoted by ε^s_{Nw}, is the elasticity of N_t with respect to w_t holding N_{t-1} fixed. Differentiating (2.4), we have

$$\begin{aligned} \varepsilon^s_{Nw} = \frac{w_t}{N_t}\frac{\partial N_t}{\partial w_t} &= -w_t s'(w_t)\frac{N_{t-1}}{N_t} + \frac{w_t R'(w_t)}{N_t} \\ &= -\varepsilon_{sw} s(w_t)\frac{N_{t-1}}{N_t} + \varepsilon_{Rw}\frac{R(w_t)}{N_t} \\ &= s(w_t)[\varepsilon_{Rw} - \varepsilon_{sw}]\frac{N_{t-1}}{N_t} + \varepsilon_{Rw}\frac{N_t - N_{t-1}}{N_t} \end{aligned} \tag{2.7}$$

In a steady state in which $N_t = N_{t-1}$, the elasticity of the short-run labor supply curve facing the firm, ε^s_{Nw}, can, using (2.6), be written as

[1] One can invert $N(w)$ to give $w(N)$, the wage the firm must pay if it wants to have N workers.

$$\varepsilon_{Nw}^{s} = s(w_t)\varepsilon_{Nw} \tag{2.8}$$

(2.8) shows that the short-run labor supply curve facing the firm is less elastic than the long-run one (as $s(w_t) < 1$), and the difference is greater the lower the separation rate.

In a dynamic monopsony model, it is not immediately clear whether the short- or long-run labor supply elasticity is most relevant for wage determination. If firms must commit to a particular wage and do not discount the future, they will be interested in maximizing steady-state profits and the formula in (2.3) will still hold where the relevant elasticity is the long-run one. But, suppose firms do discount the future at a rate D and cannot make long-term commitments on the wage. In particular, suppose the firm cannot commit itself to a particular wage for more than a single period in advance so that the promise of a particular wage this period carries no guarantee that it will be paid next period.[2] The following result (Boal and Ransom 1997) tells us about the steady-state relationship between the marginal revenue product of labor and the wage in this case.

Proposition 2.1. *In a steady state, the relationship between the marginal revenue product and the wage is given by*

$$\frac{Y'(N) - w}{w} = \varepsilon\left[1 + \frac{(1-D)(1-s)}{s}\right] = (1-D)\varepsilon^{s} + D\varepsilon \tag{2.9}$$

where ε is the inverse of long-run elasticity of the labor supply curve and ε^{s} is the inverse of the short-run elasticity.

Proof. See Appendix 2.

(2.9) says that the rate of exploitation is a weighted average of the long-run and short-run elasticities with the weight on the long-run elasticity being the discount factor. The more employers discount the future, the greater the weight given to the short-run elasticity and the larger will be the rate of exploitation (as it will be the case that $\varepsilon^{s} > \varepsilon$). The intuition is that cutting wages is more attractive when employers discount the future more heavily as the costs of this strategy (lower future labor supply) do not weigh so heavily on their minds.

One variant of (2.9) is where the length of the "period" (of wage commitment) goes to zero. If the length of period is Δ, and d and ς are the instantaneous interest and separation rates, respectively, we will have

[2] It is convenient to work in discrete time although we will consider the limit as the length of time between periods goes to zero.

$D = e^{-d\Delta}$ and $s = 1 - e^{-\varsigma\Delta}$. Then taking the limit as $\Delta \rightarrow 0$, we have

$$\frac{Y'(N) - w}{w} = \varepsilon\left[1 + \frac{d}{\varsigma}\right] \tag{2.10}$$

The difference between the right-hand sides in (2.3) and (2.10) is probably rather small for plausible values of d and ς (perhaps an annual interest rate of 5% and 20% for the labor turnover rate) although (2.10) does suggest a larger rate of exploitation than (2.3).

In the interests of simplicity, most of the theoretical analysis in this book is based on the assumption that employers do not discount the future and choose wages once-for-all. In this case, it is the long-run labor supply elasticity that is important and, as this section has demonstrated, this is likely to understate the true extent of monopsony power.

2.3 A Generalized Model of Monopsony

In the models of monopsony considered so far, there is only one way for a firm to get employment of N and that is to pay the wage $w(N)$. In reality, firms can influence their employment through other means, for example, by varying the intensity of their recruitment activity. In this section we present a simple, yet general and flexible framework for thinking about monopsony in this situation.

Define the labor cost function, which we denote by $C(w, N)$, as the cost per worker, excluding direct wage costs, of keeping employment at N when the firm pays a wage w.[3] Some examples might make the idea clearer. For example, if recruiting and training a worker costs T (independent of the number of recruits) and the separation rate is $s(w)$, a flow of sN recruits is needed to maintain employment at N so that $C(w, N) = T/s(w)$. In this case, the labor cost function is independent of N. But, if it becomes increasingly hard to recruit and train workers

[3] One can think of both perfect competition and the static model of monopsony as being particular forms (albeit, non-differentiable) of the labor cost function. The traditional static monopsony model implicitly assumes that a firm that pays a wage w incurs no recruitment costs if it wants employment less than the labor supply forthcoming at that wage, but that there is no way at all for the firm to attract more workers. So, the form of the labor cost function in this model is $C(w, N) = 0$ if $N < N(w)$ and $C(w, N) = \infty$ if $N > N(w)$. The labor cost function for the competitive labor market model is the following. If there are no recruitment/training costs then, if w^c is the competitive wage, the labor cost function for the competitive model can be thought of as $C(w, N) = 0$ if $w \geqslant w^c$, and $C(w, N) = \infty$ if $w < w^c$. This says that any amount of labor can be recruited at zero cost as long as the wage paid is at or above the competitive level, but that no labor is available at any cost if a wage below the competitive wage is offered.

then the cost of recruiting and training workers $T(R)$ will be an increasing function of R and the labor cost function will take the form $C(w, N) = T(N/s(w))/s(w)$ in which case it will depend positively on employment. The issue of whether there are diseconomies of scale in recruitment and training turns out to be of some importance.

Now consider the optimal choice of the wage and employment. If we assume the firm has a revenue function $Y(N)$, steady-state profits can be written as

$$\pi = Y(N) - [w + C(w, N)]N \tag{2.11}$$

A more sophisticated analysis would recognize that recruitment takes time and there is a need to pay attention to the date at which costs are incurred and revenues accrue but the decision problem at the end of the day can normally be written as something that looks like (2.11) (for an explicit justification of this claim, see Manning 2001a).

A difference from the basic monopsony model is that the firm has a choice of the wage it can pay if it wants to maintain employment at N so that both wages and employment are choice variables that can vary independently of each other. Given N, it is optimal for the firm to choose w to minimize direct and indirect labor costs so let us define the function $\omega(N)$ as

$$\omega(N) = \min_{w} w + C(w, N) \tag{2.12}$$

Profits can then be written as

$$\pi = Y(N) - \omega(N)N \tag{2.13}$$

A comparison of (2.1) and (2.13) should make apparent the relationship between the model presented here and the basic monopsony model: it is that the labor supply curve $w(N)$ needs to be replaced by the labor supply curve $\omega(N)$. As $\omega(N)$ is the relevant labor supply curve, let us call it the effective labor supply curve. We can represent the decision problem for the employer as in figure 2.1 with $w(N)$ replaced by $\omega(N)$, the effective supply of labor to the firm. Unsurprisingly, it is going to be of some interest whether $\omega(N)$ is increasing in N which would give us the equivalent of an upward-sloping labor supply curve.

By application of the envelope theorem to (2.12), we have

$$\omega'(N) = C_N(w(N), N) \tag{2.14}$$

where $w(N)$ is the wage chosen if employment is N. Hence, the effective labor supply curve facing the firm is upward-sloping if the labor cost function is increasing in employment, that is, if there are diseconomies of scale in recruitment and training. If the level of employment has no impact on recruitment and training costs, then the effective labor supply curve facing the firm will be infinitely elastic and will resemble the labor supply curve in a perfectly competitive market. Given this discussion, it should be apparent that the form of the labor cost function $C(w, N)$ is of some importance.[4] The labor market is ‘‘monopsonistic’’ if $C_N > 0$ so that the non-wage costs are increasing in employment and ‘‘competitive’’ if $C_N = 0$.

There is a reasonable argument that the labor cost function $C(w, N)$ should be used in all the analysis that follows, and that analysis suggests we should focus on the effective labor supply function $\omega(N)$ rather than the labor supply function $w(N)$. However, this is hard to do as we rarely have the requisite data on non-wage labor costs like training and recruitment costs. However, in chapter 10 we present some evidence that $C_N > 0$ so that there are diseconomies of scale in recruitment activity.

All of the models considered so far in this chapter have been partial equilibrium models in which the influence of factors external to the individual firm have been buried in its labor supply curve, $N(w)$. One could introduce general equilibrium considerations by explicitly allowing the actions of other firms to affect the supply of labor to the firm (or the labor cost function). But, such an approach would inevitably be ad hoc and it is best to construct an explicit general equilibrium model of an oligopsonistic labor market to check that the model as a whole is internally consistent. This is the subject of the next section.

2.4 A General Equilibrium Model of Oligopsony

Any general equilibrium model of oligopsony must model interactions between employers in an internally consistent manner. There are a number of ways in which this has been done in the literature: for example, Bhaskar and To (1999) use a Hotelling-style location model. Here we outline another such model developed by Burdett and Mortensen (1998) which can be thought of as a general equilibrium version of the dynamic monopsony model described in section 2.2. The assumptions made about the labor market are the following.[5]

(A1) Workers: There are M_w workers all of whom are equally productive and attach equal value, b, to leisure.

(A2) Employers: There are M_f employers, each of which is assumed to

[4] More detailed analysis of the comparative statics of the generalized model of monopsony can be found in Manning (2001a).

[5] These assumptions have been chosen to be the simplest possible whilst retaining the essential features of an oligopsonistic labor market. Many of these assumptions are relaxed at various points in the book or in other papers in the literature.

be infinitesimally small in relation to the market as a whole.[6] All employers have constant returns to scale, the productivity of each worker being p. For future use, denote the ratio of firms to workers by $M = M_f/M_w$.

(A3) Wage-setting: Employers set wages once-for-all to maximize steady-state profits (which is equivalent to assuming there is no discounting). All workers within a firm must be paid the same wage. Denote the cumulative density function of wages across employers by $F(w)$ and the associated density function by $f(w)$.

(A4) Matching Technology: Both employed and non-employed workers receive job offers at a rate λ. Job offers are drawn at random from the set of firms, that is, from the distribution $F(w)$. Employed workers leave their jobs for non-employment at an exogenous job destruction rate δ_u. All workers, both employed and non-employed, leave the labor market at a rate δ_r, to be replaced by an equal number of workers who initially enter non-employment. For future use, define $S = S_u + S_r$.

These assumptions are simpler than those used in Burdett and Mortensen (1998) but capture the essence of their model. Now, consider the equilibrium in the basic model.

The Behavior of Workers

The behavior of workers in this labor market is very simple. An employed worker will move to another job whenever a wage offer above the current wage is received. A non-employed worker will accept a job whenever the wage offer received is above some reservation wage, r. As job offers arrive at the same rate whether employed or non-employed, the decision to accept a current job offer has no consequences for future job opportunities. So, the job will be taken if it makes a worker better off now than they would be if non-employed, that is, if the wage exceeds the value of leisure. Hence, the reservation wage, the lowest wage for which workers will be prepared to work will simply be equal to b, the value of leisure. Later, in chapter 9, we analyze the determinants of the reservation wage in a more complicated setup where on- and off-the-job searches differ in their effectiveness.

The Employer's Decision

The employer's decision in this model is to choose the wage to maximize profits $\pi = (p - w)N(w; F)$ where $N(w; F)$ is the steady-state level of employment in a firm that pays a wage w when the distribution of wages in the market as a whole is F. So, prior to considering the profit

[6] Note that this assumption implies that employers do not have to be "large" to possess market power.

maximization decision, we need to consider employment determination.

Employment Determination

An employer who pays a wage w will recruit workers from among the non-employed (as long as w is larger than the reservation wage b) and from workers in other firms that pay less than w. The employer will lose workers who exit to non-employment or leave the labor force or who quit to other firms that pay higher wages. In general terms, if $s(w; F)$ is the separation rate and $R(w; F)$ is the recruitment rate, we must have in a steady state that

$$s(w; F)N(w; F) = R(w; F) \tag{2.15}$$

so that $N(w; F)$ is the level of employment at which the flow of recruits equals the flow of separations. In deriving $N(w; F)$, a very useful result is the following.

Proposition 2.2. *If $\infty > \lambda/\delta > 0$, the equilibrium must be a distribution of wages without any spikes.*

Proof. See Appendix 2.

The result that the equilibrium of this model must have wage dispersion even though all agents (both workers and firms) are assumed identical is the most striking feature of Burdett and Mortensen (1998). The intuition for it is not that easy to understand but the result comes from the fact that if there is a wage paid by a non-negligible fraction of employers, then paying an infinitesimally higher wage means that the employer starts to recruit workers from these employers leading to a discontinuous jump in the number of workers but only an infinitesimal fall in profits per worker. Hence, profits must rise and the initial situation could not have been in equilibrium.

The proposition implies that the equilibrium outcome must be a wage distribution with a continuous cumulative density function, $F(w)$. As all firms are identical but, in equilibrium, choose different wages which yield the same level of profit, there is an indeterminacy in equilibrium in the sense that which firms choose which wages is not defined and one might think there is a potential problem in ensuring that the right distribution of wages results from the uncoordinated choices of firms. This is a common problem in much of economic theory where the equilibrium involves mixed strategies. But, it is not a real problem here. The smallest differences in firms will result in a fully determinate equilibrium (see chapter 8).

As it is reasonable to believe that firm heterogeneity exists, the model presented here should be thought of as the limiting equilibrium as firm heterogeneity disappears.

From the analytical point of view this proposition is extremely convenient as it means that we can restrict attention to wage distributions $F(w)$ that are continuous. From a more practical point of view, the result has both advantages and disadvantages. The advantage is that the model can explain the existence of equilibrium wage dispersion, the well-documented fact that equally productive workers receive different wages according to who they work for (see, e.g., Lester 1946, 1952; Slichter 1950; Reynolds 1951; Dunlop 1957; Krueger and Summers 1988; amongst others). The disadvantage is that we do observe concentrations of workers (or "spikes") at particular points in wage distributions, often at the minimum wage or at "round" numbers.[7]

Now, consider how we can derive the supply of labor to a firm who pays a wage w. From (2.15) it is helpful to derive the separation and recruitment rate separately. The separation rate in a firm that pays w is

$$s(w; F) = \delta + \lambda[1 - F(w)] \tag{2.16}$$

as workers leave for non-employment at a rate δ, receive other job offers at a rate λ and a fraction $[1 - F(w)]$ of these offers are better than their current wage.

Deriving the flow of recruits to the firm, $R(w; F)$, is slightly more complicated. It is helpful to first derive the non-employment rate and the distribution of wages across workers.

The Non-Employment Rate

The non-employment rate, u, is simply given by

$$u = \frac{\delta}{\delta + \lambda} \tag{2.17}$$

as workers leave employment for non-employment at a rate δ and obtain jobs at a rate λ.

The Distribution of Wages Across Workers

The distribution of wages across firms is denoted by $F(w)$. This is not the same as the distribution of wages across workers as the systematic search by workers for better-paying jobs means that they will be concentrated in

[7] Modifying the model to allow for the existence of spikes (e.g., because of mobility costs) is likely to increase the monopsonistic elements in the model so moves us even further away from the competitive model than the current framework.

higher-wage firms. Denote by $G(w; F)$ the fraction of employed workers receiving a wage w or less when the wage offer distribution is F. The following proposition shows that there is a simple relationship between G and F.

Proposition 2.3. *The fraction of workers in employment receiving a wage w or less is given by*

$$G(w; F) = \frac{\delta F(w)}{\delta + \lambda[1 - F(w)]} \tag{2.18}$$

Proof. See Appendix 2.

From inspection of (2.18) one can see that $G(w; F) < F$ for $0 < F < 1$ so that workers are concentrated in the better-paying jobs, implying that such firms must have a higher level of employment. This is easy to understand: higher-wage firms have lower separation rates and higher recruitment rates so that they have more workers in a steady state. Some special cases may help the understanding of (2.18): as $\lambda \rightarrow 0$ so that opportunities to move up the job ladder once in employment are reduced, then $G(w; F) \rightarrow F(w)$ so the distribution of wages across workers converges to the distribution of wages across firms. On the other hand, as $\lambda \rightarrow \infty$ so that opportunities to move up the job ladder once in employment come at a very fast rate then $G(w; F) \rightarrow 0$ if $F(w) < 1$ and $G(w; F) \rightarrow 1$ if $F(w) = 1$ so that all workers end up in the firm that pays the highest wage.

The Flow of Recruits to a Firm

Now let us go back to deriving the level of employment in a firm that pays w. Recruits to this firm will come from non-employment and those employed in lower-wage jobs. There are $\lambda u M_w$ non-employed workers who receive job offers which are shared equally over the M_f firms so that the flow of non-employed recruits to the firm will be $\lambda u M_w / M_f = \lambda u / M$. Similarly, there are $\lambda(1 - u)G(w; F)M_w$ workers currently earning less than w who get job offers which again are spread over the M_f firms. So, the total flow of recruits to a firm that pays w is given by

$$R(w; F) = \frac{\lambda}{M}[u + (1 - u)G(w; F)] = \frac{\delta\lambda}{M[\delta + \lambda(1 - F(w))]} \tag{2.19}$$

where the second equality follows from use of (2.17) and (2.18). Combining (2.15), (2.16), and (2.19), we finally have the following expression:

$$N(w; F) = \frac{\delta\lambda}{M[\delta + \lambda(1 - F(w))]^2} \tag{2.20}$$

for the supply of labor to the firm. This captures the most important idea in the analysis of monopsonistic labor markets, namely that the labor supply to an individual firm is increasing in the wage paid so that the labor supply curve facing an individual firm is not infinitely elastic as is assumed in perfect competition. Employment is increasing in the wage because the separation rate is decreasing in the wage (a higher wage means workers are less likely to get a better job offer) and the flow of recruits is increasing in the wage (a higher wage means there are more workers in lower-wage firms).

The wages paid by other firms are also important in determining the supply of workers to a firm. In fact, in (2.20) the position of the firm in the wage offer distribution is a sufficient statistic for the supply of labor to the firm. This is not true in more general models but one should still think of the supply of labor to the firm as being determined by the wage offered relative to the alternatives of non-employment or employment in other firms.

The Employer's Decision Revisited

Given (2.20), profits can be written as

$$\pi(w; F) = \frac{\delta\lambda(p - w)}{M[\delta + \lambda(1 - F(w))]^2} \tag{2.21}$$

Every firm will choose its wage to maximize profits.

Equilibrium

We need to find an equilibrium wage distribution $F(w)$. $F(w)$ will be an equilibrium if two conditions are satisfied:

- all wages that are offered yield the same level of profits;
- no other wage yields a higher level of profits than a wage that is offered.

For the special model considered here, one can (as Burdett and Mortensen 1998 showed) derive a closed-form expression for the equilibrium wage distribution w. The easiest way to derive this equilibrium is in stages.

Proposition 2.4. *The lowest wage offered in equilibrium is the reservation wage, b.*

Proof. See Appendix 2.

The intuition for this result is simple. There is no point in an employer paying a wage lower than b as no workers will accept such a low wage offer and employment and profits will be zero. And there is no point in the lowest-wage employer paying a wage strictly above b as, from (2.20), the supply of labor to the firm is a function of its position in the wage distribution (F) and not the actual wage paid. So, cutting the wage to b will lead to the same level of employment but higher profits per worker.

Given that the lowest wage offered is b (which is also the reservation wage), the equilibrium level of profits, π^*, can be found by using this fact in (2.21) to give

$$\pi^* = \frac{\delta\lambda(p-b)}{M[\delta+\lambda]^2} \tag{2.22}$$

The equilibrium wage distribution $F(w)$ can then be found by equating (2.21) to (2.22). After some re-arrangement this leads to the following.

Proposition 2.5. *The offered wages lie in the interval*

$$b \le w \le p - \left(\frac{\delta}{\delta+\lambda}\right)^2 (p-b) \tag{2.23}$$

and, within this interval, the equilibrium wage offer distribution is given by

$$F(w) = \frac{\delta+\lambda}{\lambda}\left[1 - \sqrt{\frac{p-w}{p-b}}\right] \tag{2.24}$$

The equilibrium wage distribution across workers, $G(w)$, is given by

$$G(w) = \frac{\delta}{\lambda}\left[\sqrt{\frac{p-b}{p-w}} - 1\right] \tag{2.25}$$

The expected wage, $E(w)$, is given by

$$E(w) = \frac{\delta}{\delta+\lambda} b + \frac{\lambda}{\delta+\lambda} p \tag{2.26}$$

Proof. See Appendix 2.

Let us now discuss some implications of these results.

2.5 Perfect Competition and Monopsony

The formulae for the equilibrium wage offer distribution and the wage distribution are not very intuitive.[8] But the formula for the expected wage is simple, saying that the expected wage is a weighted average of the marginal product of labor and the reservation wage (the value of leisure), the weight on the marginal product being an increasing function of (λ/δ), the ratio of the arrival rate of job offers to the job destruction rate.

In equilibrium, all workers get paid a wage below their marginal product (note that the upper bound for wages in (2.23) is below p). This contrasts with the perfectly competitive labor market in which workers receive a wage equal to their marginal product. One might wonder about the relationship between the equilibrium here and the perfectly competitive equilibrium. It turns out that perfect competition is a special case in which job offers arrive infinitely fast for employed workers.

Proposition 2.6. *As $(\lambda/\delta) \to \infty$, the distribution of wages across workers collapses to the perfectly competitive equilibrium in which all workers get paid their marginal product, p.*

Proof. Take the limit of (2.26) and note that $E(w) = p$ implies all workers get paid p as no workers ever get paid more than p.

This proposition corresponds well with our notion of perfect competition as a market in which there is fierce competition among employers for workers and the high arrival rate of job offers means that the threat of workers leaving if they are paid a low wage is a very real one. As the result implies that perfect competition is a special (but extreme) case of our labor market, conclusions reached using a competitive analysis are not inevitably wrong; they will be correct or nearly correct if labor market frictions are small. But it is important to correct the impression that those who believe that employers have some market power over workers are extremists—the reality is that those who believe in perfect competition are the fanatics as perfect competition is one point at the edge of the parameter space and every other point in the parameter space gives employers some monopsony power. But, although it is extreme to assume the labor market is frictionless, it may be that this is a good approximation to reality if the frictions are "low." It would be helpful to have some quick way of deciding the extent of monopsony power possessed by employers.

[8] Indeed, they should not be taken too literally as the wage distribution of (2.25) has an increasing density, a prediction that is at variance with empirical observation. There is a literature (see van den Berg and Ridder 1998; Mortensen 1998; Bontemps et al. 1999, 2000) which extends the basic model to make its predictions more consistent with the observation while preserving its qualitative features. This is discussed in more detail in section 4.8.

2.6 A Simple Measure of Monopsony Power

What limits the ability of employers to lower wages is the ability of workers to leave for another employer. So, one way to understand the result in Proposition 2.6 is that a high arrival rate of job offers makes the workers' quit threat more powerful and increases direct competition among employers for workers. The extent to which workers do freely move among employers is then likely to be a good way to measure the extent of competition in the labor market.[9] But the separation rate itself is not a good measure of labor market competition as it does not matter much to employers if workers quit freely if there is a high flow of workers recruited from non-employment to replace them. In terms of Proposition 2.6, one can see that it is (λ/δ) that is important and not just λ. A simple statistic that captures this idea is the proportion of recruits that come from other firms. The higher this proportion the more intense the competition among employers and the lower we would expect the extent of monopsony to be.

This is likely to be a good "back-of-the-envelope" measure of the extent of labor market competition in many models of the labor market but the following proposition verifies the intuition by showing that, in the simple Burdett–Mortensen model, the proportion of recruits from non-employment is a monotonic function of (λ/δ).

Proposition 2.7. *The higher the fraction of recruits from non-employment, the more monopsonistic is the labor market. The fraction of recruits from non-employment in the Burdett–Mortensen model is given by*

$$\frac{\lambda}{\delta+\lambda}\,\frac{1}{\ln\left(\dfrac{\delta+\lambda}{\delta}\right)} \tag{2.27}$$

and is monotonically decreasing in (λ/δ).

Proof. See Appendix 2.

(2.27) demonstrates that the fraction of recruits from non-employment is a function of the ratio of the job offer arrival rate to the job

[9] This implicitly assumes that threats to quit if paid low wages do, in equilibrium, turn into actual quits. Some economists are inclined to argue that threats can be important even if they are never actually carried out so may not like the statistic proposed here to measure the extent of competition in labor markets. But, Proposition 2.7 shows that, although the limiting competitive case of the Burdett–Mortensen model has no wage dispersion among workers, there is a very large amount of actual worker mobility that lies behind this.

Figure 2.2 The relationship between the weight on marginal product and the proportion of recruits from non-employment.

Notes. The weight on the marginal product in the expected wage is derived from (2.26) so the left-hand axis is $\lambda/(\lambda + \delta)$. The relationship between this variable and the fraction of recruits from non-employment is given by (2.27).

offer destruction rate so is related to the indices of monopsony power described above. The relationship between the fraction of recruits from non-employment and $\lambda/(\lambda + \delta)$ is represented in figure 2.2. Remember that, from (2.26), $\lambda/(\lambda + \delta)$ is also the weight on the marginal product in the expression for the expected wage and the labeling of figure 2.2 reflects this. As one can see, the relationship is non-linear: if the proportion of recruits from non-employment is below 25% the weight on productivity in the expected wage will be above 98%. But if the proportion of recruits from non-employment is 50%, the weight will be only 80%. This discussion suggests that a simple but crude way of getting some a priori idea of the extent of monopsony is to examine the proportion of recruits that come from non-employment.

The main labor market surveys, the Current Population Survey (CPS) for the United States and the Labour Force Survey (LFS) for the United Kingdom, can be used for this purpose. Both the CPS and LFS are rolling panels: in the CPS, individuals are in the sample for four consecutive months, followed by four months out and then another four months in. In the LFS, individuals are in the sample for five successive quarters. Neither the CPS nor the LFS allow us to directly observe the fraction of recruits that were previously in non-employment as they do not contain continuous data on employment. Our approach is to approximate the proportion by considering new recruits and recording whether their labor market status at the previous wave was employment or non-employment. As the time interval between observations on labor market status in the CPS is only a month, this is likely to be a reasonably good estimate.[10] The quarterly time interval in the LFS is perhaps more problematic.

Both the CPS and the LFS are address-based surveys so that individuals who move address between surveys are dropped from the sample. This is not a problem if the fraction of recruits from non-employment is the same for those who change address and those who do not, but one would like to be reassured on this. Fortunately, the LFS does contain information that allows us to infer labor market transition rates for those who move address. We would expect that, for every individual who leaves a sample address (a "mover-out" in the jargon), there is someone who moves into a sample address (a "mover-in"). As all employed workers are asked about their job tenure in their current job, and those who report being in their current address less than 3 months are asked about their labor market status 3 months ago,[11] one can use this information to compute the fraction of recruits from non-employment for the movers-in. In practice, this makes very little difference (47% of new recruits who are movers-in having come from non-employment as compared to 44% for the residential stayers).

Table 2.1 presents some statistics on the fraction of recruits from non-employment using the data sets described above. In the CPS, the fraction of new recruits who were not employed a month ago is 55%. This proportion includes the 5% who reported being on temporary lay-off last month. Quite how those on temporary lay-offs should be treated is

[10] In fact, in a labor market in steady-state one can show that the observed proportion of hires from non-employment in a period of unit length is given by $\xi[1-\exp(-(\delta/\xi)(\xi+(1-\xi)\theta))]/[1-\exp(-(\delta/\xi))][(\xi+(1-\xi)\theta)]$ where ξ is the true proportion, θ is the ratio of the rate at which the non-employed find jobs to the rate at which the employed change jobs, and δ is the rate of entry into non-employment. There is a bias to the extent that θ differs from 1. As one would expect that $\theta > 1$, one can show that one is likely to understate the proportion coming from non-employment.

[11] There is some reason to believe that the responses to such retrospective questions understate labor market transition rates. For the LFS we can get some information on this as, each spring, individuals are asked about their labor market status one year ago. For those in the final wave, this answer can be compared to the one they gave a year ago in the first wave. Overall the accuracy is high: over 95% of individuals for whom we have panel information and retrospective information gave answers to the retrospective questions that were consistent with the panel information. But this overall figure hides an important difference in the consistency of response: for those whose labor market states were the same in the two years (assuming the panel information is correct) about 98% gave consistent answers. But, for those who had made a labor market transition, the proportion of consistent answers fell to about 78% (the nature of the transition does not matter). However, we are using information on quarterly transitions here so we might expect this problem to be less serious.

TABLE 2.1
The Proportion of Recruits from Non-employment

Country	*Data Set*	*Sample*	*Period of Observation of Labor Market Status*	*Fraction of Recruits from Non-employment*
US	CPS	1998–99, age 18–60	Monthly	0.551
US	CPS	1998–99 (ignoring temp lay-offs), age 18–60	Monthly	0.525
UK	LFS	1992–99, age 18–60	Quarterly	0.443
UK	BHPS	1990–98, age 18–60	Continuous	0.465

Notes.
1. The fraction of recruits from non-employment is computed by taking all those in new jobs and computing the fraction for whom the economic activity before starting the job was non-employment.

not clear and it is hard to work out whether those now in employment have returned to the job from which they were originally laid off. So, the second row simply excludes them: this slightly lowers the proportion previously non-employed to 52%. The third row reports UK estimates from the LFS: the fraction of recruits from non-employment is lower than in the United States, approximately 45%. One might be concerned that this is the result of the fact that observations on employment are only at quarterly intervals. However, the fourth row of table 2.1 reports estimates from the British Household Panel Survey (BHPS) which has a continuous record of employment. The estimate of the proportion of recruits from non-employment is reassuringly similar to that derived from the LFS. So, it does seem that a lower proportion of recruits come from non-employment in the United Kingdom as compared to the United States: this is perhaps the result of the lower flows from employment into non-employment in the United Kingdom.

The aggregate figures in table 2.1 hide a lot of variation across individuals and over time. This is investigated in table 2.2 where probit models for whether an individual has been recruited from non-employment are reported. We first estimate equations for men and women jointly and then separately as there are important differences.

For the United States, table 2.2 suggests that the fraction of recruits from non-employment is higher for women, for young and very old workers, for the less-qualified, for those in full-time education, and for blacks (although the difference is only significant for men). Table 2.2 also shows

TABLE 2.2
Disaggregated Analysis of the Proportion of Recruits from Non-employment

	(1) US: CPS All	*(2) US: CPS Women*	*(3) US: CPS Men*	*(4) UK: LFS All*	*(5) UK: LFS Women*	*(6) UK: LFS Men*
Female	0.099			0.037		
	(0.002)			(0.003)		
Experience	0.036	−0.007	0.080	−0.057	−0.091	−0.029
1–5 years	(0.004)	(0.009)	(0.010)	(0.005)	(0.007)	
Experience	−0.015	−0.024	−0.023	−0.062	−0.039	−0.087
6–10 years	(0.004)	(0.010)	(0.011)	(0.006)	(0.008)	
Experience	−0.003	0.025	−0.032	−0.045	0.012	−0.111
11–20 years	(0.004)	(0.008)	(0.010)	(0.005)	(0.007)	
Experience	0.024	0.010	0.046	−0.039	−0.038	−0.040
31–40 years	(0.004)	(0.010)	(0.012)	(0.006)	(0.009)	(0.009)
Experience	0.066	0.062	0.049	0.098	0.127	0.077
41+ years	(0.011)	(0.027)	(0.027)	(0.009)	(0.014)	(0.012)
High school drop-	0.103	0.109	0.108	0.068	0.074	0.065
out (US), no qualifications (UK)	(0.004)	(0.009)	(0.009)	(0.005)	(0.007)	
Some college (US),	−0.046	−0.064	−0.035	−0.051	−0.026	−0.060
A levels (UK)	(0.003)	(0.008)	(0.008)	(0.005)	(0.007)	(0.006)
College degree	−0.115	−0.102	−0.108	−0.059	−0.059	−0.069
	(0.004)	(0.008)	(0.009)	(0.005)	(0.005)	(0.007)
Student	0.158	0.065	0.235	0.244	0.227	0.270
	(0.005)	(0.012)	(0.011)	(0.006)	(0.008)	(0.009)
Black	0.033	0.020	0.103	0.106	0.101	0.112
	(0.018)	(0.039)	(0.038)	(0.014)	(0.019)	(0.021)
Hispanic (US),	−0.056	−0.031	−0.048	0.104	0.104	0.103
Asian (UK)	(0.018)	(0.038)	(0.038)	(0.011)	(0.016)	(0.015)
Male	−0.130	−0.132	−0.105	−0.093	−0.015	−0.233
employment/ population ratio	(0.082)	(0.112)	(0.118)	(0.153)	(0.215)	(0.219)
Observations	172464	88797	83667	96086	48644	47442
Mean of dependent variable	0.545	0.587	0.501	0.443	0.463	0.422
Pseudo-R^2	0.032	0.021	0.041	0.040	0.035	0.057

Notes.

1. The sample is all those who have just started jobs. The dependent variable is a binary variable taking the value one if the individual was recruited from non-employment.
2. The sample period is 1994–2000 for the CPS and 1992–2000 for the LFS.
3. Reported coefficients are marginal effects. Standard errors in parentheses.
4. Regressors are, as far as possible, defined in the same way for the US and UK data. Where there are unavoidable differences (in education and race), the appropriate variables are defined in the first column.
5. Student is defined as anyone who has not completed full-time education. Qualifications are coded as zero for these individuals. The omitted education category for the United States is a high school graduate and for the United Kingdom someone with O levels.
6. The CPS regressions also include month, year and state dummies. The LFS regressions also include month, year and region dummies.

very similar results for the United Kingdom. Women also have a rather different experience profile in both countries, being more likely to have been recruited from non-employment when they have between 11 and 20 years of experience: this is likely to be associated with withdrawal from and return to the labor market connected with having children. What is worth noting is that those groups that do badly in the labor market in terms of wages also do badly in terms of more frequently being recruited from non-employment. This is in line with the basic prediction of the theory where competition among employers for workers is more intense (and wages end up higher) when a lower fraction of recruits are from non-employment.

The estimated models in table 2.2 also include the prime-age male employment/population ratio to see whether there is cyclical variation in the proportion of recruits from non-employment. For the United Kingdom, the results in Burgess (1993) suggest that the proportion of recruits from non-employment falls as the labor market tightens as job-to-job moves are very pro-cyclical (although Fallick and Fleischman 2001 conclude this is not true for the United States). The results in table 2.2 provide some weak support for this conclusion although the coefficient is never significantly different from zero.[12]

In this section we have proposed that the fraction of recruits from non-employment is a crude but simple measure of the extent of competition among employers for workers. Differences in this measure across different types of workers also mirror wage differences as the theory predicts (although others might also do so).

2.7 Positive and Normative Aspects of Monopsony and Oligopsony

In an oligopsonistic equilibrium, workers are "exploited" in the sense of that term used by Hicks and Pigou: that is, they receive a wage less than their marginal product. But, the word "exploitation" has emotive power that is unfortunate in the current context. In the static model of monopsony it makes sense to use the marginal product of labor as a point of comparison for the wage. The efficient outcome is to set the wage equal to the marginal product and a minimum wage set at that level leads to a first-

[12] Although one explanation for this is that most of the variation in the employment/population ratio is absorbed by the time and regional dummies. The United Kingdom does show a remarkable fall in the proportion of recruitments from non-employment from almost 50% in 1992 to just above 35% in 1999, consistent with the rise in employment in the same period. However, there is no noticeable trend in the fraction of recruits from non-employment in the United States over a similar period.

best outcome. However, that is not necessarily true in the models of oligopsony discussed here. Even though Proposition 2.7 says that, in a frictionless market, workers would get paid their marginal product, one cannot wish away the existence of frictions and, given their existence, it is not clear that efficiency would be best served by raising wages to the marginal product. The bulk of this book is positive: about how we can achieve a better understanding of a wide variety of labor market phenomena from the distribution of wages to the provision of training to the impact of minimum wages and trade unions by recognizing that employers have non-negligible market power over their workers. There is little normative content: no judgment is made about whether these things are "good" or "bad," although, as we shall see, the approach taken here does suggest approaching many issues with a more open mind than a fanatical believer in perfect competition might be inclined to do. This emphasis on the positive aspects of the subject is not because the normative issues are unimportant but because the normative concerns are sufficiently complex that it is simply not credible to be able to draw normative conclusions from theoretical introspection or from casual empirical analysis. Justifying this conclusion is the subject of the next chapter: this can be skipped for those who are only interested in the positive implications of oligopsony.

2.8 Implications and Conclusions

The models of this chapter have been highly stylized with assumptions chosen for analytical convenience more than for realism. Nonetheless, they do convey the essence of a labor market with frictions in which employers set wages influenced in part by competition from other employers but in which this competition is not so cutthroat as to enable workers to extract all the surplus from the employment relationship, nor so feeble as to enable employers to get all the surplus.

A lot of the analysis in this chapter has been very formal. But, one should not allow this to distract attention away from the basic insights into the workings of labor markets that the monopsonistic approach provides. The rest of this book is concerned with the determinants of prices and quantities in the labor market, a traditional pre-occupation of microeconomics. The study of prices is essentially the study of the distribution of wages while the study of quantities is the study of the level and distribution of unemployment, the level of employment in firms, and of the quality of labor (as influenced by the acquisition of human capital). The implications of monopsony or oligopsony are summarized briefly for these issues.

In perfect competition we normally think of the distribution of wages as being determined by the distribution of marginal products[13] and attempt to explain wage differentials in cross-section and over time in this way. In a monopsonistic labor market, marginal productivity continues to be an important explanation of wages but other factors are also important. Perhaps the simplest way to see this is to look at the expression for the expected wage in (2.26). Marginal product, p, appears but so does:

- the value of leisure (the reservation wage);
- job offer arrival rates;
- job destruction rates.

As monopsony gives the labor economist a wider menu of possible explanations of the distribution of wages, we might hope for a richer explanation than can be provided when constrained by the straitjacket of perfect competition. In chapters 5 through 8, we show how such an approach can improve our understanding of the distribution of wages. For example, chapter 7 attempts to explain part of the gender wage gap in terms of the different labor market transition rates of men and women. There is also one final factor that is important in influencing wages: luck. The existence of wage dispersion among identical workers means that there is likely to be some part of the distribution of wages that can never be explained by economic factors: some workers will simply have been in the right place at the right time.

The differences in the determinants of quantities in the labor market might appear to be less dramatic, largely because search models are already commonly used to understand both theoretical and empirical aspects of unemployment. For example, in the model presented in this chapter, the level of non-employment is influenced by the job offer arrival rate when non-employed, the job destruction rate, and the level of wages relative to the reservation wage. This is not very different from the usual list of suspects although a strict perfect competition approach would suggest that only a comparison of the marginal product with the value of leisure is relevant.

The rest of the book aims to demonstrate how we can gain a better understanding of labor markets by this less dogmatic approach based on the perspective that employers have some market power.

Before we move on to these positive and empirical issues, the next chapter is concerned with more normative concerns, for example, is the oligopsonistic labor market efficient? and, if not, is the inefficiency of any

[13] Abstracting from compensating wage differentials (discussed in chapter 8) and the fact that marginal products may themselves be endogenous, varying with the level of employment.

particular type? and are there any policy interventions that might be expected to improve the operation of the labor market? This discussion is entirely theoretical: for those uninterested in it, one can summarize the conclusions now:

- the oligopsonistic labor market is not generally efficient;
- it is hard a priori to say anything about the direction of the inefficiency;
- it is hard to make a strong theoretical case for any particular policy intervention.

These conclusions then justify the approach in the rest of the book which is to use the perspective of employer market power to understand a wide variety of labor market phenomena, without making any value judgment as to whether the world could be improved by an appropriate policy intervention.

Appendix 2

Proof of Proposition 2.1

At any date t the state variable for the firm will be the labor force that it had last period, N_{t-1}. Define a value function $\Pi(N_{t-1})$ to be the maximized discounted value of future profits from date t onwards. So, using dynamic programming arguments, we have

$$\Pi(N_{t-1}) = \max_{(w_t,N_t)} Y(N_t) - w_t N_t + D\Pi(N_t) \qquad (2.28)$$

This needs to be maximized subject to a dynamic labor supply curve of (2.4). Note that this dynamic labor supply curve depends only on the current wage: this implicitly assumes workers are myopic. An alternative would assume it depends on the value of the job.

Taking the first-order condition of (2.28) with respect to w_t and taking account of the dependence of N_t on w_t leads to the following first-order condition:

$$\left[Y'(N_t) - w_t + D\Pi'(N_t)\right]\frac{\partial N_t}{\partial w_t} - N_t = 0 \qquad (2.29)$$

We also have the envelope condition which allows us to derive the derivative of the value function. Differentiating (2.35) we have that

$$\Pi'(N_{t-1}) = \left[Y'(N_t) - w_t + D\Pi'(N_t)\right]\frac{\partial N_t}{\partial N_{t-1}} \qquad (2.30)$$

If the firm is in a steady state (and it is an interesting question whether there is a steady state) where wages and employment are constant, then we can solve (2.30) for $\Pi'(N)$ which leads to

$$\Pi'(N) = \frac{\dfrac{\partial N_t}{\partial N_{t-1}}[Y'(N) - w]}{1 - D\dfrac{\partial N_t}{\partial N_{t-1}}} \tag{2.31}$$

Substituting this into (2.29) and re-arranging, one can derive

$$\begin{aligned} \frac{Y'(N) - w}{w} &= \left[1 - D\frac{\partial N_t}{\partial N_{t-1}}\right]\frac{N_t}{w(\partial N_t/\partial w_t)} \\ &= \left[1 - D + D\left(1 - \frac{\partial N_t}{\partial N_{t-1}}\right)\right]\frac{N_t}{w(\partial N_t/\partial w_t)} \\ &= \ (1 - D)\varepsilon^s + D\varepsilon \end{aligned} \tag{2.32}$$

where the last equality follows from the fact that differentiation of (2.4) implies that

$$1 - \frac{\partial N_t}{\partial N_{t-1}} = s$$

and the relationship between the short- and long-run elasticities implied by (2.8).

Proof of Proposition 2.2

Suppose there is a mass of firms offering the wage w. If $w = p$, then all these firms must be making zero profits. A firm that lowers its wage can make higher profits as long as it can retain some workers in steady state, that is, as long as its separation rate is finite and its recruitment rate positive. A non-zero, finite value of (λ/δ) guarantees this.

If there is a mass of firms paying $w < p$ then consider what happens if a firm deviates by paying an infinitesimally higher wage. Profit per worker is only infinitesimally reduced but the number of workers is measurably higher (as long as $\lambda > 0$) as recruits now come from workers in all the firms who continue to pay w. Hence, profits must rise and the initial situation could not have been in equilibrium.

Proof of Proposition 2.3

The simplest way to prove Proposition 2.4 is by equating inflows and outflows from the group of workers earning w or less. The outflow rate from this group will be $[\delta + \lambda(1 - F(w))]$ as workers leave the group either to non-employment or to better-paying jobs. Recruits to this group must come from non-employment as no workers who earn more than w will ever accept a wage offer less than w. There are $M_w u$ non-employed workers who receive offers less than w at a rate $\lambda F(w)$. So the flow of recruits to jobs paying less than w will be $\lambda F(w) M_w u$. Equating inflows and outflows, we then have

$$[\delta + \lambda(1 - F(w))](1 - u)G(w; F)M_w = \lambda F(w) u M_w \qquad (2.33)$$

as total employment of those earning w or less will be $(1 - u)G(w; F)M_w$. Using (2.17) leads, after some re-arrangement, to (2.18).

Proof of Proposition 2.4

Suppose a firm pays below b. This firm will have no workers so will make zero profits which cannot be an equilibrium.

Suppose the lowest wage offered is strictly above b. The lowest-wage firm will only recruit workers from non-employment at a rate $(\lambda u/M)$ and will lose workers whenever they get another job offer, that is, at a rate $(\delta + \lambda)$. So, employment in the lowest-wage firm will be given by

$$\frac{\lambda u}{M[\delta + \lambda]} = \frac{\delta \lambda}{M[\delta + \lambda]^2} \qquad (2.34)$$

that is, independent of the wage offered. If the lowest-wage firm cuts its wage (but not below b) the recruitment and separation rate will be unchanged and hence so will employment. But, profit-per-worker will rise so profits will increase. This means the original situation could not have been in equilibrium.

Proof of Proposition 2.5

Equating (2.21) and (2.22) leads to (2.24) after some re-arrangement. The right-hand side of (2.23) is then just the value of the wage that makes $F = 1$. (2.25) comes from (2.18) and (2.24).

Now the expected wage can be written as

$$E(w) = \frac{M_f}{M_f} \frac{\int w N(w; F) f(w) dw}{\int N(w; F) f(w) dw} = p - \frac{M_f \int (p - w) N(w; F) f(w) dw}{M_w (1 - u)}$$

$$= p - \frac{M \int \pi^* f(w) dw}{(1 - u)} = p - \frac{M\pi^*}{(1 - u)} = p - \frac{\delta(p - b)}{\delta + \lambda} \qquad (2.35)$$

where the second equality follows from the fact that the two denominators in the first line are both expressions for total employment; the third equality follows from the fact that, in equilibrium, $(p - w)N$ must be the same for all firms; and the final equality follows from (2.22) and (2.17).

Proof of Proposition 2.6

The recruits to position F in the wage distribution, $R(F)$, can be written as

$$R(F) = \lambda u M_{\mathrm{w}} + \lambda M_{\mathrm{w}}(1-u)G(F) = \frac{\lambda\delta M_{\mathrm{w}}}{\delta + \lambda(1-F)} \qquad (2.36)$$

where $G(F)$ is the fraction of workers employed at position F or below and is given by (2.18). Note that the position in the wage distribution is a sufficient statistic for the number of recruits so we do not have to worry about the actual wage paid. Using the fact that F must be distributed uniformly over the unit interval, we have that the total flow of recruits in the economy is given by

$$R = \int_0^1 R(f)df = M_{\mathrm{w}}\int_0^1 \frac{\lambda\delta df}{\delta + \lambda(1-f)} = M_{\mathrm{w}}\delta \ln\left(\frac{\delta+\lambda}{\delta}\right) \qquad (2.37)$$

As the flow of recruits from non-employment is $\lambda M_{\mathrm{w}} u$, this gives (2.27) for the fraction of recruits from non-employment.