3

Efficiency in Oligopsonistic Labor Markets

Discussions of the partial equilibrium static model of monopsony often emphasize that the free market equilibrium is inefficient in a very particular way. Both the wage and employment are too low and full efficiency can be restored by ensuring that the wage is equal to what it would be in a perfectly competitive labor market. One way of achieving this outcome is by means of an artfully chosen minimum wage. This chapter considers whether general equilibrium models of oligopsony allow such clear-cut policy prescriptions: the conclusion is that they do not, although there is no presumption that the "free market" equilibrium is efficient.

This chapter discusses the efficiency issue using the Burdett and Mortensen (1998) model introduced in chapter 2. The simple version of the model presented in the previous chapter cannot provide an adequate analysis of efficiency because the equilibrium is fully efficient as all matches between unemployed workers and firms are consummated. Any distribution of the surplus between employer and workers is consistent with this equilibrium outcome. For example, any minimum wage up to the level of p, the marginal product of workers, results in the same outcome in terms of employment and only affects the division of the surplus between wages and profits.

But, this conclusion is the result of simplifying assumptions made for expositional reasons. Efficiency is not an interesting issue in this version of the model because very few decisions of employers and workers are free to respond to incentives. However, there are a number of ways in which one might modify the model to endogenize decisions of both firms and workers so that they can be affected by incentives.

First, the model of the previous chapter assumed that the supply of both workers and firms to the market is inelastic, that is, the number of workers and firms in the market are fixed at M_w and M_f, respectively. A natural way to introduce incentives is to assume that the supply of both firms and workers is not inelastic. For firms, the simplest way to do this is to assume that there is free entry (i.e., to go to the opposite extreme and to assume that the supply of firms to the market is perfectly elastic). For workers, one could make an analogous assumption that a fixed cost (perhaps the cost of acquiring skills necessary for employment) must be paid to enter the labor market but there are alternative ways to make the overall labor supply have some elasticity. For example, section 3.4 discusses the case where there is heterogeneity in the value of leisure (hence the reservation wage) so that not all workers will be interested in every job.

But, even once agents have decided to participate in the market there are other decisions that might be influenced by incentives. For example, firms can decide how much effort to spend in looking for recruits and workers can decide how hard to look for work. Jointly, these decisions about search intensity can be expected to determine the arrival rate of job offers, λ.

Agents may make some investments in match quality before a match is realized. For example, workers may make decisions about how much human capital to acquire before they start looking for a job. And firms may have to commit capital to jobs before they start looking for workers for those jobs. Both of these decisions will be affected by expected future returns to the agents.

All of these "margins" of decision are likely to be present in reality. But, to include all of them simultaneously in a model of the labor market is a recipe for indigestion. Consequently, this chapter presents only the simplest possible models to make the relevant points. The models examined in sections 3.1–3.5, their main conclusions, and their implications for one particular policy intervention (the minimum wage) are summarized in table 3.1.

TABLE 3.1
The Structure of the Chapter

Section	*Model*	*Efficiency of Free Market*	*Optimal Minimum Wage*
3.1	Free entry of firms (perfectly elastic supply of firms to the market)	Too many firms, employment too high	Minimum wage to ensure appropriate division of surplus
3.2	Endogenous recruitment activity of firms	Too much recruitment, employment too high	Minimum wage to ensure appropriate division of surplus
3.3	Free entry of workers (perfectly elastic supply of workers to the market)	Too few workers in labor force, employment too low	Minimum wage to ensure appropriate division of surplus
3.4	Heterogeneity in reservation wages	Employment too low	Minimum wage equal to marginal product
3.5	Heterogeneity in reservation wages + free entry of firms	Employment may be too high or too low	Minimum wage may not be desirable

The main conclusions of sections 3.1–3.5 are as follows:

- there are good reasons to believe the free market equilibrium is inefficient;
- it is hard to say anything a priori about the direction of the inefficiency;
- it is hard to make unambiguous predictions about policy from theoretical models alone.

Thus, issues of efficiency are more complicated and subtle in general equilibrium dynamic models of oligopsony than the static partial equilibrium textbook model of monopsony might suggest. The conclusion that theory provides little in the way of a guiding principle is, in many ways, an unsatisfying one as it suggests a failure to find the most general result. One might wonder whether one can find better guidance elsewhere in the literature.

Other "search" models of the labor market have discussed whether the free market equilibrium is likely to be efficient. This debate probably started with Friedman's (1968) celebrated article on the natural rate of unemployment. Commentators like Tobin (1972) argued that the free market equilibrium was likely to be inefficient and there have been papers on the subject ever since (e.g., Albrecht and Jovanovic 1986; Hosios 1990; Moen 1997; Acemoglu and Shimer 2000; among others). Many (though not all) of the papers appear to arrive at sweeping conclusions but readers must recognize that they are based on very specific models and are not robust to reasonable changes in the assumptions.[1] Perhaps there is a result to be derived about the efficiency or otherwise of the free market in a class of reasonably general models. Armed with such a result, there might also be general prescriptions about policies to get to the first-best (the Coase Theorem and the prescription to define property rights springs to mind as an example from another part of economics). But, a combination of lack of intellect and laziness on the part of this author means this book does not provide such a general result and it certainly does not exist in the current literature. My conjecture is that, whenever a model of the labor market is proposed that has strong conclusions, one will be able to provide another, observationally equivalent, model with different conclusions. If this conjecture is correct, theory alone is not going to help us very much. This conclusion is similar in spirit to that of Lucas and Prescott (1974, p. 206) who criticized Tobin's (1972) conclusion that the free market *must* be inefficient while recognizing that "the question of

[1] For example, Albrecht and Jovanovic (1986, p. 1256), conclude that "in contrast to the competitive equilibrium, the monopsonistic equilibrium is shown to be inefficient involving too much search and too little employment." This is, of course, a correct conclusion in their model but the model is not a general one and the conclusion would be relatively simple to overturn.

whether there exist important external effects in *actual* labour markets, remains, of course to be settled."

The plan of the chapter is as follows. Section 3.1 considers the welfare properties of the equilibrium when there is free entry of firms, section 3.2 when employers can choose their level of recruitment activity. Section 3.3 then considers the case where there is free entry of workers and section 3.4 assumes some heterogeneity in the reservation wages of workers. Section 3.5 puts together the model of the sections 3.1 and 3.4, illustrating how the whole is rather different from the sum of the parts. Section 3.6 shows how easy it is to generate multiple equilibria in models of oligopsony, a feature that is potentially useful in explaining a range of phenomena from agglomeration to ghettoes.

3.1 Free Entry of Firms

In this section the basic Burdett–Mortensen model of section 2.4 is extended to the case where there is a perfectly elastic supply of firms to the market so that the number of firms is M_f is endogenous. To produce an equilibrium with a finite number of firms one needs to assume that there is a fixed cost of entry which we will denote by C_f. Firms enter until profits are equal to C_f. Using results derived in the previous chapter, equilibrium profits are given by (2.22). The number of firms affects profits as it affects M $(= M_f/M_m)$. But it also plausibly affects λ, the arrival rate of job offers for workers. Suppose that the number of matches between workers and firms is given by $m(M_w, M_f)$. It is conventional and convenient (for a recent survey, see Petrongolo and Pissarides 2001) to assume that the matching function has constant returns to scale so that the arrival rate of job offers for workers can be written as

$$\lambda = \frac{m(M_w, M_f)}{M_w} = m(1, M) \equiv \lambda(M) \tag{3.1}$$

where $\lambda(M)$ is an increasing concave function of its argument so that a fall in the number of firms reduces the arrival rate of job offers for workers. M will be given by the level at which profits are equal to C_f. Taking account of (3.1) and (2.22), one can write this free entry condition as

$$\frac{\delta\lambda(M)(p - b)}{M[\delta + \lambda(M)]^2} = C_f \tag{3.2}$$

By differentiating the left-hand side of (3.2) one can readily verify that profits are a strictly decreasing function of M if $\lambda(M)$ is an increasing concave function of M so that (3.2) defines a unique equilibrium.

Now, let us consider the efficient level of M. If the non-employment rate is u, then a fraction $(1-u)$ of workers are in employment producing p, while a fraction u are not in work which has a value b. In addition, the total fixed costs paid in the economy are $M_f C_f$. Hence, the total social surplus can be written as

$$\Omega(M) = M_w[(1-u)p + ub] - M_f C_f = M_w\left[p - \frac{\delta(p-b)}{[\delta + \lambda(M)]} - MC_f\right] \tag{3.3}$$

where the second equality follows from the non-employment rate of (2.17).[2]

The following proposition summarizes the efficiency of the free market equilibrium.

Proposition 3.1

1. *The free market has too many firms if $\lambda(M)$ is a strictly concave function of M.*

2. *The first-best can be attained by setting a minimum wage, w_m, which satisfies*

$$\frac{p - w_m}{p - b} = \varepsilon_{\lambda M} \equiv \frac{M\lambda'(M)}{\lambda(M)} \tag{3.4}$$

where M is the efficient number of firms relative to workers.

Proof. See Appendix 3.

The intuition as to why the free market equilibrium has too many firms is that some of the employment of new entrants comes not just from workers who would otherwise be unemployed but also from those employed in other firms. While this source of employment is a private gain, it has no social purpose. The second part of the proposition shows that a well-chosen minimum wage can attain the first-best, although the case for a minimum wage is that the minimum wage causes exit of firms from the market and reduces employment but these are "good" things. This argument for minimum wages is slightly curious as proponents of minimum wages do not often argue for it on the grounds that it destroys

[2] Note that this specification of the welfare function assumes risk neutrality on the part of workers. If workers are risk averse, then the wage dispersion that is characteristic of the free market equilibrium has a welfare cost and there would be a case for policies to reduce this dispersion.

jobs. One should note that policies other than the minimum wage could be used to attain the first-best. For example, one could pay unemployment benefits to ensure that the reservation wage of workers is w_{m} or one could use a profits tax.

One interpretation of (3.4) is that the share of the total surplus going to the employer in the lowest-wage firm should be equal to the elasticity of the matching function with respect to the number of firms. This is outwardly very similar to the efficiency rule derived by Hosios (1990) in the context of a Diamond–Pissarides matching model in which wages are determined by an ex post sharing rule. There is one important difference, namely that (3.4) refers only to the sharing of the surplus in the lowest-wage firm. The employer's share of the surplus in all the other firms will be lower than that given in (3.4) as they all pay wages higher than w_{m}. This is important, because knowledge of the Hosios rule might have led one to conclude that, because workers do get a share of the surplus in the basic Burdett–Mortensen model, it is not obvious a priori whether wages are too high or too low. As the above discussion has made clear, wages are unambiguously too low.

3.2 Endogenous Recruitment Activity

In this section, firms are allowed to influence the flow of recruits by expenditure on recruitment activity. For the moment, assume that the number of firms, M_{f}, is exogenously given. Denote by z the intensity of the recruitment activity of the firm. Define z so that, other things being equal, the arrival rate of workers to the firm is proportional to z. Assume that the cost of z is given by a function $c(z)$. One could interpret this cost function narrowly as the cost of advertisement but it is probably better to think of it more widely as the cost of recruitment and training new employees as the administrative costs of handling applications and the induction of new workers are typically much larger than the direct costs of job advertisements.

In equilibrium there will be some function $z(w)$ which relates recruitment activity to the wage paid. Denote by Z the average level of recruitment activity in the economy as a whole, that is, assume Z is given by

$$Z = \int z(w)dF(w) \tag{3.5}$$

Make the simplifying assumption that the rate at which job offers arrive to workers depends on Z as well as M so can be written as $\lambda(Z, M)$. Assume that an individual employer's share of these matches is given by (z/Z). The equilibrium can be described by the pair of functions

$\{z(w), F(w)\}$ which give the distribution of wages across firms and the recruitment intensity associated with each wage.

The following proposition summarizes the nature of equilibrium and its efficiency.

Proposition 3.2

1. *In equilibrium, all firms recruit at the same intensity, z, which is given by*

$$\frac{\delta\lambda(z, M)(p-b)}{M[\delta+\lambda(z, M)]^2} = zc'(z) \tag{3.6}$$

2. *The free market has too much recruitment activity, if $\lambda(z, M)$ is a strictly concave function of z.*

3. *The first-best can be attained by setting a minimum wage, w_{m}, which satisfies*

$$\frac{p-w_{\mathrm{m}}}{p-b} = \varepsilon_{\lambda z} \equiv \frac{z\lambda_z(z, M)}{\lambda(z, M)} \tag{3.7}$$

Proof. See Appendix 3.

The first part of this proposition says that (whatever their chosen wage) all firms spend the same amount on recruitment activity in equilibrium. Paying a higher wage encourages recruitment expenditure as a higher fraction of workers contacted will be interested in the job and the expected job duration of a recruit is longer. But, the profit to be made from each recruit per period is less. So, there are off-setting effects of the wage on the incentives to recruit and the proposition simply says that the equilibrium wage distribution for the case analyzed here is such that these different effects cancel out and the incentives to recruit are independent of the wage.[3]

As in the free entry case of the previous section, the intuition for the excess recruitment activity of the second part of the proposition is that the employment of an extra firm comes not just from workers who would otherwise be unemployed but also at the expense of other firms. And, as in the case of free entry, there are a number of policies that might be used to correct this inefficiency. The final part of the proposition says that a minimum wage that ensures a particular division of the surplus in the lowest-wage firm can attain full efficiency. As long as $\lambda(Z, M)$ is a strictly

[3] This intuition also suggests that the independence result will fail if there are decreasing returns to labor or if there is firm heterogeneity. See Mortensen (1998) for an analysis of this case.

concave function of z, it is optimal to have $w_m > b$. But, as in the free entry case, a binding minimum wage will reduce employment as it reduces the recruitment activities of firms. (3.7) has an obvious similarity to the Hosios rule derived for the free entry case in (3.4) except that it is now the elasticity of the arrival rate of job offers with respect to the recruitment intensity that is important. This might make us wonder what happens if we combine free entry and endogenous recruitment activity. It is fairly simple to show that a simple minimum wage now can only attain the first-best if $\varepsilon_{\lambda M} = \varepsilon_{\lambda z}$ (i.e., if the matching function can be written as $\lambda(zM)$). In this case, the two Hosios conditions (3.4) and (3.7) are identical.

So far, we have only considered decisions of firms that respond to incentives. Even then, there is reason to believe that wages are "too low" although the other side of this coin is that employment is "too high." But, as discussed in the introduction, some decisions of workers are also likely to respond to incentives. The following sections introduce this topic.

3.3 Elasticity in Labor Supply: Free Entry of Workers

The basic Burdett–Mortensen model of section 2.4 assumed that the supply of workers to the labor market is inelastic. This section introduces some elasticity into labor supply in a very simple way: by assuming that, to participate in the labor market, individuals must pay an up-front cost of C_w. This is a crude way of introducing some elasticity in labor supply (an alternative is discussed in the next section), but it does have the virtue of being a natural analogy to the way some elasticity in the supply of firms to the market was introduced in section 3.1. If pushed, one could interpret the fixed cost as the cost of acquiring the human capital necessary to get employment or as an investment in the skills necessary for job search.

Let us distinguish between the value of being unemployed V^u, the value of non-participation V^n, and the value of being employed at wage w, $V(w)$. The value functions are given by

$$\delta_r V^u = b + \lambda \int_{w_m} [V(x) - V^u] dF(x) \tag{3.8}$$

$$\delta_r V^n = b + C_w \tag{3.9}$$

$$\delta_r V(w) = w - \delta_u [V(w) - V^u] + \lambda \int_{w} [V(x) - V(w)] dF(x) \tag{3.10}$$

where w_m is the lowest wage. (3.9) captures the fact that those who choose non-participation also save the fixed cost C_w. Free entry of workers means that, in equilibrium, we must have $V^u = V^n$. The following proposition summarizes the important results.

Proposition 3.3

1. *The free market equilibrium has too few workers in the market.*

2. *The first-best can be attained by setting a minimum wage such that:*

$$\frac{p - w_m}{p - b} = \varepsilon_{\lambda M} \tag{3.11}$$

Proof. See Appendix 3.

(3.11) should, by now, be a familiar formula. The share of the surplus in the lowest-wage firm should, for efficiency, be equal to the elasticity of the arrival rate of job offers with respect to M. Because the share of workers in the surplus in the lowest-wage firm is zero in the free market equilibrium, wages are in some sense "too low" and too few individuals choose to participate in the labor market.

The effect of a minimum wage on this labor market differs from that in the labor market with free entry of firms. A binding minimum wage causes more individuals to participate in the labor market and total employment rises so that the employment/population ratio rises. But, the unemployment rate also rises as the increase in the number of agents in the labor market causes some crowding-out of job opportunities.

This section has made the supply of workers to the market perfectly elastic. Without further modification, one could not combine this model with free entry of firms as the scale of activity in the economy would then be indeterminate (for versions of this model in which there is an arbitrary degree of elasticity in the supply of both firms and workers to the market, see Manning 2001b). The next section takes a different approach to introducing some elasticity into the supply of labor to the market.

3.4 Elasticity in Labor Supply: Heterogeneity in Reservation Wages

In the models discussed so far, all workers have been assumed to have the same value of leisure, b, which, given our other assumptions, is also the reservation wage. This section modifies the model and assumes that there is some heterogeneity in b. Denote the cumulative density function of b by $H(b)$ and the associated density function by $h(b)$. It is helpful (although not essential) to assume that $H(b)$ is log-concave. For convenience, assume that $H(p) = 1$ so that all workers have $b \le p$.[4] Furthermore, assume that b is not observed by employers so that wage offers cannot be conditional on it. As a result, not all wages that are offered in equilibrium will be attractive to all workers. If a worker has value of leisure b (which will also be their reservation wage), then only a fraction $[1 - F(b)]$ of jobs will be desired.

The following proposition analyzes this case.

Proposition 3.4

1. *The free market equilibrium is inefficient with employment too low.*

2. *A minimum wage equal to p can restore full efficiency.*

Proof. See Appendix 3.

In this model, employment is, in general, too low. It is efficient to consummate all matches with $p \ge b$ yet matches are only consummated when $w \ge b$. Because $w < p$, some efficient matches are not consummated. If the number of firms (and their recruitment activities) are fixed, then attaining the first-best is simple: set a minimum wage equal to p and the inefficiency disappears. As in the other models discussed in this chapter, there is an efficiency case for a minimum wage but, now, a binding minimum wage is associated with increases in employment. This model is the closest oligopsony model to the static monopsony model as both employment and wages are too low in the free market equilibrium and full efficiency can be restored by ensuring wages are equal to marginal products.

However, this discussion has assumed that the supply of firms is completely inelastic: the next section considers what happens when we introduce an elastic supply of firms as in the model of section 3.1.

3.5 Heterogeneity in Reservation Wages and Free Entry of Firms

In this section we combine the model of the previous section with the model of section 3.1 in which there is a completely elastic supply of firms to the market.

One might have thought that because a minimum wage can improve efficiency in the two constituent models (the free entry model of section 3.1

[4] If $b > p$ there is no point in a worker being in the labor market as they will never be able to find employment at a wage acceptable to them.

and the heterogeneous reservation model of section 3.4), it must be of potential benefit in the current model. But, the following proposition shows that this is not necessarily the case.

Proposition 3.5

1. *Social surplus may be increasing or decreasing in the number of firms.*

2. *A just-binding minimum wage may increase or reduce efficiency.*

Proof. See Appendix 3.

This result is an example of the general principle of the second-best, that moving towards the first-best in one dimension may worsen welfare. In this case, a binding minimum wage, because it reduces the number of firms, tends to reduce the social surplus if there are too few firms in equilibrium, and may reduce the social surplus. The intuition for the ambiguity about the optimal number of firms is that, on top of the congestion effect which tends to lead to too many firms in the free market equilibrium, the entry of a new firm now improves the wage offer distribution which results in more workers being in employment thus increasing efficiency.

What policies can attain efficiency in this case? It should be readily apparent that a simple minimum wage can no longer lead to the first-best outcome. To consummate all efficient matches, the minimum wage would need to be equal to p but this results in no firms entering the market. So, a minimum wage of p would need to be combined with a subsidy to the entry of firms.

It should be apparent that we have quickly arrived at a point where theory provides little guidance about the nature of inefficiency in the free market equilibrium and the types of policies that might help. One could proceed further to consider more complicated models combining all the decisions we have endogenized here and even adding new ones. But, the payoff from this strategy is likely to be small as the qualitative conclusions we have drawn are unlikely to be altered.

3.6 Multiple Equilibria in Models of Oligopsony: An Application to Ghettoes

All the oligopsony models presented so far in this chapter have a single equilibrium. But, it is important to realize that it is relatively simple to produce oligopsony models with multiple equilibria that may further complicate welfare analysis. This is because of the way in which supply and demand factors interact in markets with frictions: it is possible that, in a sense to be made clearer below, "supply can create its own demand" (and vice versa).

In a frictionless competitive market, there is a sense in which it is good to be unique (because of diminishing marginal productivity). If one acquires some specialized skill that requires some specialized capital with which to work, there is no problem in meeting the person with that capital or in inducing someone to make that specific capital investment. However, in a market with frictions, uniqueness is not necessarily an advantage. If employers have to make some ex ante investment in creating jobs, they may not choose to make that investment if the chance of finding a suitable person to fill those jobs is very low. In that case, an increase in the supply of workers of a particular type may encourage employers to create jobs tailored for that type of worker and hence create its own demand. The problem is that there is no mechanism to ensure that an individual act of investment by either worker or employer will be matched by the equivalent investment on the other side of the market that is necessary for the investment to have its full effect. This problem is a potent source of multiple equilibria.

There are numerous examples of this type of model in the economics literature, probably beginning with Diamond (1982) but including Acemoglu (1998) and Machin and Manning (1997), among others. They can be used to explain a number of potentially important facts: why some countries are industrialized and others are not, and the phenomenon of agglomeration.

Here, we illustrate these ideas by presenting a simple model of how ghettoes may arise. The stylized picture of the ghetto is that of an area where both wages and employment are lower than in neighboring areas. This does seem to have something to do with the disadvantage experienced by certain ethnic groups as, in the United States, inclusion of family background variables plus good measures of educational attainment (e.g., the AFQT in the NLSY) seems to be able to eliminate much, if not all, of the observed black–white wage differential (see Neal and Johnson 1996; Altonji and Blank, 1999).[5] This suggests paying attention to pre-market factors more than labor market outcomes in looking for the origins of the black–white wage differential. Some of these pre-market factors (e.g., family background) may be the product of more explicit racial discrimination that undoubtedly existed in the past, while others may be the result of the poorer quality of education typically received by blacks. But, it is also possible that some of it represents decisions not to acquire human capital that are rational given the economic situation faced.

[5] Although there is a debate about whether the AFQT test scores themselves contain a racial bias.

As a simple example of this type of mechanism, consider the Burdett–Mortensen general equilibrium model of an oligopsonistic labor market discussed in section 2.4. But modify it, so that workers, before they enter the labor market, are assumed to have a choice about the level of human capital they acquire. This will determine their productivity p. We assume that acquiring productivity of p requires a cost of $c(p)$. We also assume that the distribution of wages facing a working of productivity p is $F(w;p)$ as given by (2.24). This implies that firms offer a different wage to workers of each quality level.

Given this, p will be chosen to maximize $V^u - c(p)$. In a frictionless market, workers with quality p would earn p with probability 1 for their entire life of expected length $(1/\delta_r)$. Hence, workers would invest to the point where $\delta_r c'(p) = 1$.

In a labor market with frictions we can prove the following proposition.

Proposition 3.6

1. *The optimal choice of p is given by*

$$\delta_r c'(p) = \left(\frac{\lambda}{\delta + \lambda}\right)^2 \tag{3.12}$$

   assuming an interior solution.

2. *The optimal p is increasing in λ tending to the competitive level as* $\lambda \longrightarrow \infty$.

Proof. See Appendix 3.

Unsurprisingly, Proposition 3.6 says that the faster the rate of job offers arrive, the greater the incentive to invest in human capital and that human capital investment will approach the competitive level as the labor market becomes frictionless. So we would expect p as a function of λ to be something like the line marked "investment decision of workers" in figure 3.1.

Now consider the investment level by firms. Let us model this investment decision as a simple entry decision. Assume, as in section 3.1, that entry requires a fixed investment of cost C_f, and that M_f firms enter the market. We will assume that the rate at which job offers present themselves to workers depends on M_f so that we have $\lambda(M_f)$. Invert this function to write $M_f(\lambda)$. In equilibrium, the number of firms is determined by the free entry condition so that

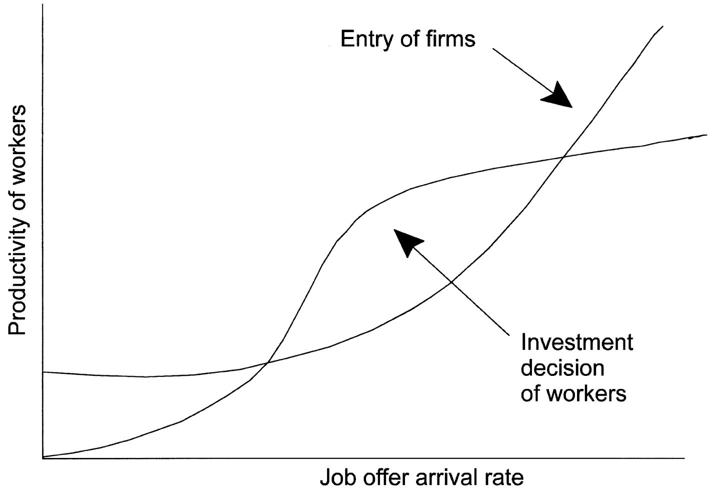

Figure 3.1 Multiple equilibria in the model of a ghetto.

$$\frac{\delta\lambda(p - b)}{M_f(\lambda)[\delta + \lambda]^2} = C_f \tag{3.13}$$

This gives λ as an increasing function of p with $\lambda \to \infty$ as $p \to \infty$. A possible outcome is given by the line marked "entry of firms" in figure 3.1. Inspection of figure 3.1 shows the possibility of multiple equilibria. As drawn, one can argue that only the high-level equilibrium is stable but one could always draw pictures with more than two stable equilibria. If this is the case then high-level equilibria will have high levels of human capital investment, high wages, high employment rates, and many firms. The differences in labor market outcomes can all be "explained" by differences in productivity but these differences in productivity are themselves the result of the poor expected labor market outcomes.

In this model, firms do not invest in the ghetto because of the poor "quality" of workers who live there. The residents do not invest in improving their skills because there are no jobs. A vicious circle is at work.

3.7 Conclusions

This chapter has considered the issue of efficiency in a number of theoretical models of oligopsonistic labor markets. It has one positive message: the free market must not be presumed efficient. But, beyond this, theory has been shown to provide little guidance about the direction of the efficiency or policies that might be expected to improve matters. Employment may be too "high," or too "low." A minimum wage may raise employment or reduce it. If a theoretical paper claims a strong conclusion about the direction of inefficiency in the free market equilibrium, then this is almost certainly because they have not considered a rich enough model in the sense that there are not enough "marginal" decisions to be influenced by incentives. This chapter has introduced three such margins: elasticity in the supply of firms to the market, elasticity in the supply of workers, and endogenous recruitment intensity, and has shown how the nature of the inefficiency induced by each of them is rather different. As discussed in the introduction, there are other margins and these may be as important. We have introduced only enough to make the point that theory alone is going to be an unreliable guide to policy-making.

The rest of the book concentrates on positive rather than normative implications of employer market power. Even though the word "monopsony" conjures up emotive images of workers being exploited (in the sense of the word used by Hicks and Pigou) by employers, one should resist such temptations. Monopsony should simply be taken to mean that the supply of labor to the firm is not perfectly elastic. Although a recurrent theme is that the perspective of monopsony encourages one to be more open-minded about the likely impact of a range of policies than a strict believer in perfect competition would be inclined to be, the book is primarily concerned with what labor market phenomena can be better understood from the perspective of monopsony.

Appendix 3

Proof of Proposition 3.1

The derivative of the social surplus with respect to M is

$$\Omega'(M) = \frac{\delta\lambda'(M)(p - b)}{[\delta + \lambda(M)]^2} - C_f \tag{3.14}$$

At the free market level of M we have, using (3.2), that

$$\Omega'(M) = C_f\left[\frac{M\lambda'(M)}{\lambda(M)} - 1\right] < 0 \tag{3.15}$$

where the final inequality follows from the fact that $\lambda(M)$ is a strictly concave function of M. (3.15) implies that there are too many firms in the free market equilibrium and that the social surplus would be maximized by having fewer firms.

A binding minimum wage of w_m becomes the lowest wage offered in equilibrium. Hence, the profits made by the lowest-wage firm (and hence the equilibrium level of profits) is given by the left-hand side of

$$\frac{\delta\lambda(M)(p - w_m)}{M[\delta + \lambda(M)]^2} = C_f \tag{3.16}$$

and the equilibrium number of firms will solve (3.16). If this is to be equal to the socially efficient level of M (the level that solves $\Omega'(M) = 0$), it must be the case that

$$\frac{p - w_m}{p - b} = \frac{M\lambda'(M)}{\lambda(M)} \tag{3.17}$$

where the right-hand side should be evaluated at the efficient level of M: this is (3.4).

Proof of Proposition 3.2

Denote by $M_w(1 - u)G(w)$ the number of workers in equilibrium who are employed at a wage w or less:[6] this is the notation used previously in chapter 2.

Consider the flow of recruits to a firm that spends z on recruitment and offers a wage w. The fraction of total matches of workers to firms is given by the share of this firm in total recruitment activity, that is, by (z/Z). Of these matches only those involving employed workers in a job currently paying less than w or who are unemployed result in recruitment so that the flow of recruits to the firm is given by

$$R(w, z) = \frac{z}{Z}\left[\frac{\lambda(Z, M)}{M}(1 - u)G(w) + \frac{\lambda(Z, M)}{M}u\right] \tag{3.18}$$

Steady-state employment will be given by $N(w) = R/s(w)$ where $s(w)$ is the separation rate so that profits can be written as

$$\begin{aligned}\pi(w, z) &= (p - w)\frac{(z/Z)[\lambda(Z, M)(1 - u)G(w) + \lambda(Z, M)u]}{M[\delta + \lambda(1 - F(w))]} - c(z)\\ &\equiv \frac{z}{Z}\pi(w) - c(z)\end{aligned} \tag{3.19}$$

where $\pi(w)$ is the term in (3.19) that does not involve z. (w, z) will be chosen by firms to maximize (3.19). For this profit maximization to be well defined, we obviously require that $c(\cdot)$ be convex, otherwise profits can be increased without bound. All combinations of (w, z) that are offered in equilibrium must yield the same level of profit. Maximizing

[6] These functions will also depend on $\{z(w), F(w)\}$ but this is suppressed to keep the notation simple.

(3.19) with respect to z while holding w constant leads to the conclusion that the optimal z will be a positive function of π and, by the envelope theorem, total profits must then be an increasing function of π. As all firms must make the same level of profits in equilibrium, $\pi(w)$ and, hence z, must be constant across firms. Given this fact, a comparison of (3.19) and (2.21) shows that, conditional on the chosen z, the distribution of wages must be the same as in the basic model derived in section 2.4.

Is the level of recruitment activity efficient? The social surplus can be written as

$$\begin{aligned}\Omega(z, M) &= M_{\mathrm{w}}[(1-u)p + ub] - M_{\mathrm{f}}c(z) - M_{\mathrm{f}}C_{\mathrm{f}} \\ &= M_{\mathrm{w}}\left[p - \frac{\delta(p-b)}{[\delta + \lambda(z, M)]} - MC_{\mathrm{f}} - Mc(z)\right] \end{aligned} \tag{3.20}$$

The first-order condition for the derivative of Ω with respect to z is

$$\Omega_z(z, M) = \frac{\delta\lambda_z(z, M)(p-b)}{[\delta + \lambda(z, M)]^2} - Mc'(z) \tag{3.21}$$

where a subscript denotes a derivative with respect to that variable. At the free market level of z we have, using (3.6), that

$$\Omega_z(z, M) = Mc'(z)\left(\frac{z\lambda_z(z, M)}{\lambda(z, M)} - 1\right) < 0 \tag{3.22}$$

where the final inequality follows from the fact that $\lambda(z, M)$ is assumed to be a concave function of z. (3.22) implies that there is excessive recruitment activity in the free market equilibrium and that the social surplus would be increased by reducing recruitment activity.

A binding minimum wage of w_{m} becomes the lowest wage offered in equilibrium. Hence, the level of z chosen by firms in the free-market equilibrium will be given by

$$\frac{\delta\lambda(z, M)(p - w_{\mathrm{m}})}{M[\delta + \lambda(z, M)]^2} = zc'(z) \tag{3.23}$$

If this is to be equal to the socially efficient level of z, from (3.21) it must be the case that

$$\frac{p - w_{\mathrm{m}}}{p - b} = \frac{z\lambda_z(z, M)}{\lambda(z, M)} \tag{3.24}$$

where the right-hand side should be evaluated at the efficient level of z.

Proof of Proposition 3.3

Differentiate (3.10) to yield

$$\frac{\partial V(w)}{\partial w} = \frac{1}{\delta + \lambda(1 - F(w))} \tag{3.25}$$

Now integrating the integral term in (3.8) by parts and using (3.25), we can obtain

$$\int_{w_m} [V(x) - V^u]dF(x) = [V(w_m) - V^u] + \int_{w_m} \frac{\partial V(x)}{\partial x}[1 - F(x)]dx \tag{3.26}$$

Now, using (3.10), we have that

$$[V(w_m) - V^u] = \frac{w_m - b}{\delta + \lambda} \tag{3.27}$$

Using (3.25), (3.26) and (3.27) in (3.8) then leads to the following expression for V^u:

$$\delta_r V^u = b + \frac{\lambda}{\delta + \lambda}(w_m - b) + \int_{w_m} \frac{\lambda[1 - F(x)]dx}{\delta + \lambda(1 - F(x))} \tag{3.28}$$

Integrating the final integral term in (3.28) by parts yields

$$\delta_r V^u = b + \frac{\lambda}{\delta + \lambda}(w_m - b) + \int_{w_m} (x - w_m)\frac{\delta\lambda f(x)dx}{[\delta + \lambda(1 - F(x))]^2} \tag{3.29}$$

Using (2.20), this can be written as

$$\delta_r V^u = b + \frac{\lambda}{\delta + \lambda}(w_m - b) + M\int_{w_m} (x - w_m)N(x)f(x)dx \tag{3.30}$$

Using the expression for the aggregate non-employment rate in (2.17), this can be written as

$$\begin{aligned} \delta_r V^u &= b + \frac{\lambda}{\delta + \lambda}(w_m - b) + \frac{\lambda}{\delta + \lambda}[E(w) - w_m] \\ &= b + \frac{\lambda}{\delta + \lambda}(E(w) - b) \end{aligned} \tag{3.31}$$

Now, in equilibrium, it must be the case that

$$(p - w)N(w) = (p - w_m)N(w_m) \tag{3.32}$$

for all offered wages. Taking expectations, using (2.17) and (2.20) for the lowest-wage firm and rearranging, yields

$$E(w) = p - \frac{\delta}{\delta + \lambda}(p - w_m) \tag{3.33}$$

Substituting this into (3.31) leads to

$$\delta_r V^u = b + \frac{\lambda}{\delta+\lambda}(w_m - b) + \frac{\lambda^2}{(\delta+\lambda)^2}(p - w_m) \tag{3.34}$$

In the free market equilibrium we must have $V^u = V^n$ and $w_m = b$. Using (3.9) and (3.34), the free market entry condition can be written as

$$\frac{\lambda(M)^2}{(\delta+\lambda(M))^2}(p - b) = C_w \tag{3.35}$$

To derive the social surplus, assume that wages are always equal to p so all the social surplus goes to workers. Then, the value functions, (3.8) and (3.10), become

$$\delta_r V^u = b + \lambda[V(p) - V^u] \tag{3.36}$$

$$\delta_r V(p) = p + \delta_u[V^u - V(p)] \tag{3.37}$$

which implies that

$$\delta_r V^u = b + \frac{\lambda(p-b)}{\delta+\lambda} \tag{3.38}$$

Now the total social surplus is $(V^u - V^n)M_w$ as V^u is the lifetime expected utility of a worker entering the labor market. This, using the fact that $M = M_f/M_w$, yields

$$\Omega(M) = \left[\frac{\lambda(M)}{\delta+\lambda(M)}(p-b) - C_w\right]\frac{M_f}{M} \tag{3.39}$$

Differentiating (3.42), we obtain

$$\Omega'(M) = -\left[\frac{\lambda(M)}{\delta+\lambda(M)}(p-b) - C_w\right]\frac{M_f}{M^2} + \frac{\delta\lambda_M(M)}{[\delta+\lambda(M)]^2}(p-b)\frac{M_f}{M} \tag{3.40}$$

which, using (3.35) to eliminate $(p - b)$ in (3.40) yields, after some rearrangement,

$$\Omega'(M) = \frac{\delta M_f C_w}{\lambda M^2}[\varepsilon_{\lambda M} - 1] < 0 \tag{3.41}$$

so that there are too few workers entering the labor market. This proves part 1 of Proposition 3.3.

Now consider part 2. If we set $\Omega'(M) = 0$, then (3.40) can be written as

$$\frac{\lambda(M)}{\delta+\lambda(M)}(p-b) - C_w = \frac{\delta M\lambda_M(M)}{[\delta+\lambda(M)]^2}(p-b) \tag{3.42}$$

If $V^u = V^n$, then from (3.34) and substituting the expression for C_w, the equilibrium condition for efficiency can be written as

$$\frac{\lambda(M)}{\delta+\lambda(M)}(p-w_{\mathrm{m}})-\frac{\lambda(M)^2}{[\delta+\lambda(M)]^2}(p-w_{\mathrm{m}})=\frac{\delta M\lambda_M(M)}{[\delta+\lambda(M)]^2}(p-b)=C_{\mathrm{w}} \tag{3.43}$$

which, on rearrangement, leads to (3.11).

Proof of Proposition 3.4

In proving this proposition, it is helpful to first prove the following lemma on the equilibrium of the model.

Lemma 3.1

1. *If a firm pays a wage w, the supply of labor to it, N(w) will be given by*

$$N(w;F)=\frac{\delta\lambda H(w)}{M[\delta+\lambda(1-F(w))]^2} \tag{3.44}$$

2. *The lowest wage offered in the free market equilibrium, w_0, is the solution to*

$$w_0=\mathrm{argmax}(p-w)H(w) \tag{3.45}$$

3. *The equilibrium level of profits, π^*, is given by*

$$\pi^*=\frac{\delta\lambda(p-w_0)H(w_0)}{M[\delta+\lambda]^2} \tag{3.46}$$

4. *The equilibrium wage offer distribution is found by solving*[7]

$$\frac{(p-w)H(w)}{[\delta+\lambda(1-F(w))]^2}=\frac{(p-w_0)H(w_0)}{[\delta+\lambda]^2} \tag{3.47}$$

Proof. For workers with a reservation wage b, the non-employment rate will be

$$u(b)=\frac{\delta}{\delta+\lambda[1-F(b)]} \tag{3.48}$$

as only a fraction $[1-F(b)]$ of job offers are acceptable.

[7] One might be concerned that the solution $F(w)$ to (3.47) need not be a legitimate distribution function as it is possible that it decreases for some w. If this occurs, the equilibrium wage distribution is the upper envelope of the solution to (3.47). But log-concavity of $H(w)$ is sufficient to ensure the solution to (3.47) is a legitimate distribution function.

Define $M(w)$ to be the number of workers with a wage less than w. By analogy to the argument used in deriving Proposition 2.3, we must have

$$[\delta + \lambda(1 - F(w))]M(w) = \lambda M_{\mathrm{w}} \int^{w} [F(w) - F(b)]u(b)h(b)db \quad (3.49)$$

as unemployed workers with reservation wage b accept jobs paying w or less at a rate $\lambda[F(w) - F(b)]$. The supply of labor to an individual firm can then be written as

$$[\delta + \lambda(1 - F(w))]N(w; F) = \frac{\lambda}{M_{\mathrm{f}}}\left[M(w) + M_{\mathrm{w}} \int^{w} u(b)h(b)db\right] \quad (3.50)$$

Substituting (3.48) and (3.49) into (3.50) leads, after some rearrangement to (3.44) which proves part 1.

Using (3.44), profits can be written as

$$\pi(w; F) = \frac{\delta\lambda(p - w)H(w)}{M[\delta + \lambda(1 - F(w))]^2} \quad (3.51)$$

Suppose the lowest wage is above w_0 as defined by (3.45). Then profits in the lowest-wage firm can be increased by cutting wages to w_0 as $F(w)$ will still be equal to zero. Similarly if the lowest wage is below w_0 then profits can be increased by increasing the wage paid as $(p - w)H(w)$ must rise and $F(w)$ cannot fall.

In equilibrium, all offered wages must yield the same level of profit. Using (3.51) and (3.46) leads to (3.47). This completes the proof of the lemma.

The proof of Proposition 3.4 is now very simple. As long as some offered wage is below the highest value of b, some efficient matches will not be consummated. A minimum wage of p guarantees that this does not happen.

Proof of Proposition 3.5

Suppose the social planner is choosing the number of firms. They obviously want to maximize employment for all those with $p > b$ so that the social surplus can be written as

$$\Omega(M) = \left[\frac{\lambda(M)}{\delta + \lambda(M)} \int (p - b)h(b)db - MC_{\mathrm{f}}\right]M_{\mathrm{w}} \quad (3.52)$$

and the first-order condition for the efficient level of M can be written as

$$\Omega'(M) = \left[\frac{\delta\lambda'(M)}{[\delta + \lambda(M)]^2} \int (p - b)h(b)db - C_{\mathrm{f}}\right]M_{\mathrm{w}} = 0 \quad (3.53)$$

Using (3.46), we have that at the free market equilibrium, C_f is equal to the level of profits given by (3.47). Hence, the derivative of the social surplus with respect to M is

$$\Omega'(M) = \frac{\delta\lambda(M)}{M[\delta + \lambda(M)]^2}\left[\varepsilon_{\lambda M}\int(p - b)h(b)db - (p - w_0)H(w_0)\right]M_w \tag{3.54}$$

The sign of this is ambiguous. One the one hand, $\varepsilon_{\lambda M} < 1$ which tends to make (3.54) negative implying there are too many firms. On the other hand,

$$(p - w_0)H(w_0) = \int^{w_0}(p - w_0)h(b)db < \int^{w_0}(p - b)h(b)db$$
$$< \int(p - b)h(b)db \tag{3.55}$$

which tends to make (3.54) positive. This proves the first part of the proposition.

If the unemployment rate of workers with reservation wage b is $u(b)$, then total surplus per worker can be written as

$$\Omega = \int(p - b)h(b)db - \int(p - b)u(b)h(b)db - MC_f \tag{3.56}$$

Now, if the lowest wage paid is w_m (that might be determined by a binding minimum wage), then using (3.48), this can be written as

$$\Omega = [p - E(b)] - \frac{\delta}{\delta + \lambda}\int^{w_m}(p - b)h(b)db - \int_{w_m}^{\hat{w}}\frac{\delta(p - b)h(b)db}{\delta + \lambda[1 - F(b)]}$$
$$- \int_{\hat{w}}(p - b)h(b)db - MC_f \tag{3.57}$$

where $\hat{w}$ is the highest wage.

Now, from (3.47) we can eliminate $F(b)$ from the third term on the right-hand side of (3.57) and write Ω as

$$\Omega = [p - E(b)] - \frac{\delta}{\delta + \lambda}\int^{w_m}(p - b)h(b)db$$
$$- \frac{\delta\sqrt{(p - w_m)H(w_m)}}{\delta + \lambda}\int_{w_m}^{\hat{w}}\frac{(p - b)h(b)db}{\sqrt{(p - b)H(b)}}$$
$$- \int_{\hat{w}}(p - b)h(b)db - MC_f \tag{3.58}$$

and the free entry condition can be written as

$$\frac{\delta\lambda(M)(p - w_m)H(w_m)}{M[\delta + \lambda(M)]^2} = C_f \tag{3.59}$$

From the free entry condition, we have that the effect of the minimum wage on the number of firms is given by

$$\left[\frac{\lambda'(M)}{\lambda(M)} - \frac{1}{M} - \frac{2\lambda'(M)}{\delta + \lambda}\right]\frac{\partial M}{\partial w_m} = -\frac{\Psi'(w_m)}{\Psi(w_m)} \tag{3.60}$$

where $\Psi(w) \equiv (p - w)H(w)$. Note that a minimum wage set at w_0 (i.e., just binds) will have no effect on the number of firms as $\Psi'(w_0) = 0$. But a minimum wage above w_0 will have $\Psi'(w_m) < 0$.

Now, differentiating (3.58), we have that

$$\begin{aligned}\frac{\partial \Omega}{\partial w_m} = &\left[\frac{\delta}{[\delta + \lambda]^2}\int^{w_m}(p - b)h(b)db\right.\\ &\left.+ \frac{\delta\sqrt{(p - w_m)H(w_m)}}{[\delta + \lambda]^2}\int_{w_m}^{\hat{w}}\frac{(p - b)h(b)db}{\sqrt{(p - b)H(b)}}\right]\lambda'(M)\frac{\partial M}{\partial w_m}\\ &- \frac{\delta\Psi'(w_m)}{2\sqrt{\Psi(w_m)}[\delta + \lambda]}\int_{w_m}^{\hat{w}}\frac{(p - b)h(b)db}{\sqrt{(p - b)H(b)}} - C_f\frac{\partial M}{\partial w_m}\end{aligned} \tag{3.61}$$

Using (3.60) to eliminate the term in $\Psi'(w_m)$, one can write this as

$$\begin{aligned}\frac{\partial \Omega}{\partial w_m} = \frac{\partial M}{\partial w_m}&\left[\frac{\delta\lambda'(M)}{[\delta + \lambda]^2}\int^{w_m}(p - b)h(b)db\right.\\ &\left.+ \frac{\delta\sqrt{\Psi(w_m)}[\varepsilon_{\lambda M} - 1]}{2M[\delta + \lambda]}\int_{w_m}^{\hat{w}}\frac{(p - b)h(b)db}{\sqrt{(p - b)H(b)}} - C_f\right]\end{aligned} \tag{3.62}$$

If the minimum wage just binds, all these terms are zero as they all involve $\partial M/\partial w_m$ which is zero at that point. But, for strictly binding minimum wages we have that $(\partial M/\partial w_m) < 0$ so that the sign of (3.62) is determined by the sign of the terms in the square brackets on the right-hand side. The first term is positive and the others negative, making the overall sign ambiguous.

Proof of Proposition 3.6

From Proposition 3.4 and (3.34) we know that, for the case where the lowest wage is equal to b, the value of being non-employed can be written as

$$\delta_r V^u = b + \left(\frac{\lambda}{\delta + \lambda}\right)^2 (p - b) \tag{3.63}$$

Maximizing $\delta_r V^u - c(p)$ then leads to (3.12).