Wage differentials, discrimination and efficiency

Shouyong Shi*

Department of Economics, University of Toronto, 150 St. George Street, Toronto, Ont., Canada, M5S 3G7

Received 31 August 2005; accepted 20 January 2006
Available online 20 March 2006

Abstract

I analyze a large labor market where homogeneous firms post wages to direct the search of workers who differ in productivity. I show that the model has a unique equilibrium. The wage differential depends positively on the workers’ productivity differential only when the latter is large. When the productivity differential is small, high-productivity workers get a lower wage than low-productivity workers. This reverse wage differential remains even when the productivity differential shrinks to zero. However, the equilibrium is socially efficient. High-productivity workers always get the employment priority and higher expected wages than low-productivity workers. Although discrimination in terms of expected wages does not exist, conventional measures are likely to incorrectly find discrimination in the model.

JEL classification: J3; J6; J7

Keywords: Search; Wage differential; Discrimination

1. Introduction

Standard economic theories view wage differentials solely as a compensation for workers’ human capital or productivity. These theories have encountered great difficulties in explaining the large wage differentials in the US data. For example, Juhn et al. (1993)
have found that all observable characteristics of workers’ productivity, such as education, experience and age, can explain only one-third of the differential between the 19th and 10th percentile of the wage distribution between 1963 and 1989. On the other hand, wage differentials seem to depend statistically on seemingly irrelevant background characteristics, such as workers’ race, gender, and height. Standard theories attribute this anomalous dependence to discrimination.1

Given these difficulties, it is useful to explore alternative theories of the labor market. In a seminal paper, Mortensen (1982) provided one such alternative that emphasized search frictions and showed that prices (or wages) may depart significantly from Walrasian prices in order to achieve efficiency. Following this line of research, I use a stylized model to show that the wage differential in an efficient equilibrium has no systematic relationship with the productivity differential between workers. Nor does it represent discrimination, although it would be construed as such according to the usual practice in labor economics.

The labor market has two types of workers, one with higher productivity than the other. This observable difference between workers can be very small. All firms are identical and each needs one worker to produce. The recruiting game is one of directed search. First, firms simultaneously announce two wages, one for each type of workers, and a ranking scheme that determines which type of workers will be selected first for the employment. After observing the announcements of all firms, workers decide which firm to apply to and they cannot coordinate their applications. After receiving the applicants, each firm selects one worker according to the ranking scheme and pays the posted wage, employed workers produce and the game ends. Search is directed because each firm explicitly takes into account the effect of its announcements on the matching probability.

Because workers cannot coordinate their applications, a natural restriction on the equilibrium is that identical workers must use the same strategy. There is a unique equilibrium that satisfies this restriction. In the equilibrium, all firms use the same strategy, every worker applies to all firms with the same probability, and all firms attract both types of workers with positive probability. Also, every firm gives high-productivity applicants the priority of employment. So, high-productivity workers are more likely to be employed than low-productivity workers. They also obtain a higher expected wage, which is the product of the employment probability and the actual wage. That is, expected wages reward workers’ productivity properly.

However, actual wages do not depend on productivity in a systematic way. When the productivity differential is large, high-productivity workers get a higher wage than low-productivity workers. When the productivity differential is small, high-productivity workers get a lower wage. This reverse wage differential is a negative compensation for the high ranking that high-productivity workers get. For a high-productivity worker, the combination of the ranking and the reverse wage differential is an optimal tradeoff between the wage and the employment probability. By attracting high-productivity workers, the combination is also optimal for a firm.

1See Altonji and Blank (1999) for a survey of the facts and the literature on discrimination in the labor market. A popular spinoff is the theory of statistical discrimination. It argues that when firms are uncertain about workers’ fundamental characteristics, discrimination can be an equilibrium outcome, either because the characteristics on which discrimination is based are correlated with workers’ fundamental characteristics, or because discrimination leads to self-fulfilling separation of worker types.
When the productivity differential decreases, the reverse wage differential increases because the wage for high-productivity workers must be cut by more in order to maintain the favorable ranking for them. When the productivity differential shrinks to zero, the reverse wage differential does not disappear—Instead, it becomes most pronounced. The usual practice in labor economics would interpret this wage differential in the limit case as discrimination. This would be misleading because the correct measure of a worker’s compensation is the expected wage, not the actual wage. There is no discrimination in terms of expected wages. High-productivity workers do get higher expected wages than low-productivity workers and this differential in expected wages approaches zero when the productivity differential shrinks to zero.

I show that the equilibrium is socially efficient. That is, if a social planner tries to maximize expected aggregate output under the constraint of the same frictions as in the equilibrium, then the planner will choose the same allocation between workers and firms as in the equilibrium. Efficiency arises from the feature that a worker’s expected wage is equal to the worker’s social marginal value. Moreover, a firm’s expected profit is equal to the firm’s social marginal value, and so the equilibrium continues to be efficient when there is competitive entry of firms.

I extend the above results to a market with many types of workers and use numerical examples to illustrate the wage distribution. The numerical examples show that even a minuscule difference in productivity can lead to large differences in wages and employment probabilities.

One may wonder why a long analysis is necessary for establishing the results here. Is the reverse wage differential an obvious consequence of compensating differentials? It is not. To see this, suppose that there are two types of workers, T and S, where a type T worker’s productivity exceeds a type S worker’s by $\delta > 0$. Let the wage for a worker be $w$, the employment probability be $p$, and the expected wage be $E$. The argument of compensating differentials is nothing more than a description of a worker’s indifference curve, $pw = E$. There are infinitely many combinations of $p$ and $w$ that lie on this curve. The compensating differential argument has no prediction for which one of these combinations will arise in equilibrium. The predictive power is particularly weak in the limit case $\delta \to 0$. In this case, all possible rankings of the two types of workers are plausible and consistent with the compensating differential argument. If one knows that type T workers are ranked the first, then it is easy to claim that the resulting reverse wage differential is a compensating differential. But how does one know that firms will always rank type T workers the first in equilibrium? Why does not the compensating wage differential go the other way by ranking type S workers higher and giving them a lower wage than type T workers? Or, for that matter, why should any compensating wage differential exist at all? By establishing the uniqueness of the equilibrium, I show that the reverse wage differential is the only equilibrium outcome when $\delta$ is small, not just one of many possibilities. More importantly, the outcome is efficient while other possibilities are not. Thus, regarding the reverse wage differential as an obvious consequence of compensating differentials would miss the main points of the analysis.

This paper belongs to the search literature in general (see Mortensen, 2002). In particular, the paper borrows directed search from Peters (1991), Moen (1997), Acemoglu and Shimer (1999), Burdett et al. (2001), Julien et al. (2000), and Shi (2001, 2002). These models have either homogeneous agents on both sides of the market, or heterogeneous agents on both sides who are complementary with each other in production. The model
here lies somewhere in between—it has heterogeneous workers and identical firms. On discrimination, Black (1995) and Bowlus and Eckstein (2002) use models of undirected search and assume that some firms have prejudice against a subset of workers. It is not clear whether the resulting wage differential in these papers can sustain when search is directed or when prejudice is not assumed. Lang et al. (2002) show that discrimination arises with directed search, but they restrict each firm to post only one wage for all workers who apply to the firm. Once this artificial restriction is eliminated, I show that the outcome characterized by Lang et al. is neither an equilibrium nor efficient.

I will organize the paper as follows. In Section 2, I will describe the simple model with two types of workers and propose a candidate equilibrium. Section 3 will show that the candidate is the unique equilibrium. In Section 4, I will examine the properties of the equilibrium and show that the equilibrium is socially efficient. Section 5 will extend the model to incorporate many types of workers. I will then conclude in Section 6.

2. The model

2.1. Workers and firms

Consider a labor market with a large number of workers, \( N \). There are two types of workers, type \( T \) (for “tall”) and type \( S \) (for “short”). Type \( T \) workers are a fraction \( \gamma \) of all workers and type \( S \) a fraction \((1 - \gamma)\), where \( \gamma \in (0, 1) \). Sometimes I will use the notation \( \gamma_T = \gamma \) and \( \gamma_S = 1 - \gamma \). A worker’s type is observed by the firm immediately upon applying to a job. A type \( S \) worker produces \( y \) units of output and a type \( T \) worker produces \((1 + \delta)\)y, where \( \delta > 0 \). In a large part of the analysis, I will restrict \( \delta \) to be sufficiently small. The purpose is to examine whether a small productivity difference can generate a large wage differential.

There are also a large number of firms, \( M \), all of which are identical. For the moment, this number is fixed. I will introduce competitive entry of firms in Section 4.3. Each firm wants to hire only one worker. Denote the tightness of the market as \( \theta = N/M \). For simplicity, assume that the economy lasts for one period.

The recruiting game is as follows. First, all firms make their announcements simultaneously. Each firm \( i \) announces two wages, \( w_{iT} \) for type \( T \) workers and \( w_{iS} \) for types \( S \) workers, together with a ranking of the two types of applicants. Let \( R_i \in \{1, 0, \Phi\} \) denote this ranking or priority rule, where \( \Phi = [0, 1] \). The firm selects type \( T \) workers first if it sets \( R_i = 1 \) and type \( S \) workers first if it sets \( R_i = 0 \); If \( R_i = \Phi \), the firm is indifferent between the two types of workers.³ The firm commits to the posted wages. All workers observe the announcements by all firms and then decide which firm to apply to. This decision may be mixed strategies over the jobs. Let \( z_{ij} \) denote the probability with which a type \( j \) worker applies to firm \( i \). After receiving the applicants, a firm selects one according

²Shimer (1997) constructs a model similar to mine. Although our results overlap to some extent, his focus is to contrast the effects of different mechanics of wage determination on the division of the match surplus. In particular, he does not emphasize the reverse wage differential that can arise in the directed search environment.

³I include the ranking scheme in a firm’s announcement to ease exposition. However, firms do not have to post the ranking literally. All that is needed is that workers expect firms to use the ranking scheme to select workers after workers apply. This expectation will be fulfilled because the ranking is optimal for firms ex post, as well as ex ante, which I will establish later.
to the ranking and pays the corresponding wage. The worker produces immediately, obtains the wage and the game ends.

Notice that a worker’s strategy can depend on the worker’s type but not on the worker’s identity. The implicit assumption here is that all workers of the same type must use the same strategy. This symmetry requirement reflects the realistic feature that workers cannot coordinate their applications.\(^4\) To preserve the coordination failure, I must also assume that each firm’s offer and ranking can only depend on workers’ types but not on workers’ identities. However, identical firms can use different strategies. This allowance is necessary for examining the possibility of separation where two groups of firms each attract a distinct type of workers.

A worker maximizes the expected wage. Let \(p_{ij}\) be the probability with which a type \(j\) worker who applies to a firm \(i\) gets the job. The expected wage of the application is \(p_{ij}w_{ij}\). Denote \(E_{ij}\) as the maximum expected wage that the worker can get by applying elsewhere. That is,

\[
E_{ij} = \max_{j \neq i}(p_{rj}w_{rj}).
\]

Then, a type \(j\) worker’s optimal strategy is \(x_{ij} = 1\) if \(p_{ij}w_{ij} > E_{ij}\), \(x_{ij} = 0\) if \(p_{ij}w_{ij} < E_{ij}\), and \(x_{ij} \in (0, 1)\) if \(p_{ij}w_{ij} = E_{ij}\). The maximum expected wage that the worker obtains in the market is \(E_j = \max_i (E_{ij})\). I refer to \(E_j\) as a type \(j\) worker’s market wage.

A single firm’s strategy affects the payoff that a worker gets elsewhere. This is because a worker’s application probabilities must add up to one across the firms. If a single firm \(i\)’s strategy increases the probability with which workers apply to the firm, then it necessarily reduces the probability with which workers apply to other firms. As a result, a worker who applies to another firm will be selected more likely than before. That is, \(p_{ij}\) increases and so does \(E_{ij}\). This effect of a single firm’s strategy on workers’ expected wage elsewhere complicates the analysis significantly because I must then analyze all firms’ decisions simultaneously. Fortunately, the effect vanishes in the limit where \(N\) and \(M\) approach infinity with their ratio, \(\theta\), being fixed (see Burdett et al., 2001). I will focus on this limit economy, where each firm \(i\) takes \(E_{ij}\) as given.

It is convenient to express a worker’s strategy with a new variable \(q_{ij} = \gamma_j N x_{ij}\). Then,

\[
q_{ij} = \begin{cases} 
\infty & \text{if } p_{ij}w_{ij} > E_{ij}, \\
0 & \text{if } p_{ij}w_{ij} < E_{ij}, \\
(0, \infty) & \text{if } p_{ij}w_{ij} = E_{ij}.
\end{cases}
\] (2.2)

The first case in (2.2) cannot happen: If \(q_{ij} = \infty\), then \(p_{ij} = 0\) and so \(p_{ij}w_{ij} = 0 < E_{ij}\), which would contradict the requirement in that case. Thus, \(p_{ij}w_{ij} \leq E_{ij}\) for all \(i\) and \(j\), with equality if \(q_{ij} > 0\). Note that this result implies \(E_j = E_{ij}\), which allows me to replace \(E_{ij}\) with \(E_j\).

Given the workers’ strategies \(q\), the expected number of type \(j\) applicants received by firm \(i\) is \(q_{ij} = q_{ij} = \gamma_j N x_{ij} = q_{ij}\). When \(q_{ij} > 0\), I say that firm \(i\) attracts type \(j\) workers. I call \(q_{ij}\) the queue length of type \(j\) workers for firm \(i\), although \(q_{ij}\) is a type \(j\) worker’s strategy. Despite this coincidence, it would be incorrect to interpret that an individual worker can influence

\(^4\)In a model with two identical agents on each side of the market, Burdett et al. (2001) have shown that there is a continuum of asymmetric equilibria, but there is only one symmetric equilibrium.
other workers’ or firms’ decisions. Because the sum of $x_{ij}$ over $i$ is one, then $q$ must satisfy
\begin{equation}
\frac{1}{M} \sum_{i=1}^{M} q_{ij} = \gamma_j \theta, \quad j = T, S.
\end{equation}

Let me compute a worker’s employment probability. The probability with which firm $i$ attracts one or more type $j$ workers is \((1 - e^{-q_{ij}})\). If firm $i$ gives type $j$ workers the selection priority, then the employment probability of a type $j$ worker at the firm is $G(q_{ij})$ where
\begin{equation}
G(q) = \frac{1 - e^{-q}}{q}.
\end{equation}

If firm $i$ gives the other type $j' \neq j$ the priority, instead, then the firm will consider a type $j$ worker only if the firm does not receive any type $j'$ applicant. This event occurs with probability $e^{-q_{ij'}}$, in which case the firm chooses each of the type $j$ applicants with probability $G(q_{ij'})$. For a general ranking $R_i$, the employment probabilities of the two types of workers at firm $i$ are
\begin{equation}
p_{iT} = [R_i + (1 - R_i)e^{-q_{IS}}]G(q_{IT}), \quad p_{IS} = [1 - R_i + R_i e^{-q_{IT}}]G(q_{IS}).
\end{equation}

Because the function $G$ is continuous and decreasing, having slightly more workers apply to a firm will result in each applicant being chosen with a slightly lower probability.

Now consider a firm $i$. With the choices \((w_{iT}, w_{IS}, R_i)\), the firm’s expected profit is
\begin{equation}
\pi_i = (1 - e^{-q_{IT}})[R_i + (1 - R_i)e^{-q_{IS}}][1 + \delta(y - w_{IT})] + (1 - e^{-q_{IS}})[1 - R_i + R_i e^{-q_{IT}}](y - w_{IS}).
\end{equation}

Taking \((E_T, E_S)\) as given, the firm chooses \((w_{iT}, w_{IS}, R_i)\) to maximize $\pi_i$, subject to (2.2) for $j = T, S$. The constraint (2.2) reflects the fact that the firm takes into account the effect of its choices on workers’ decisions.

Finally, the firm’s ranking rule must be optimal ex post, as well as ex ante. That is,
\begin{equation}
R_i = \begin{cases} 
1 & \text{if } w_{IT} - w_{IS} < \delta y, \\
0 & \text{if } w_{IT} - w_{IS} > \delta y, \\
\Phi & \text{if } w_{IT} - w_{IS} = \delta y.
\end{cases}
\end{equation}

An equilibrium consists of firms’ strategies \((w_{iT}, w_{IS}, R_i)_{i=1}^{M}\), workers’ strategies \((q_{iT}, q_{IS})_{i=1}^{M}\), and the numbers \((E_T, E_S)\) such that the following requirements are met: (i) Given the firms’ strategies and the numbers \((E_T, E_S)\), each worker’s strategy is given by (2.2); (ii) Given the numbers \((E_T, E_S)\) and anticipating workers’ responses, each firm’s strategy is optimal and the ranking is ex post optimal; and (iii) the numbers \((E_T, E_S)\) obey (2.1).

2.2. A candidate equilibrium

Let me construct a candidate equilibrium and show later that it is the only equilibrium. The candidate has the following features:

- \(E1\) All firms have the same strategy, i.e., \((w_{iT}, w_{IS}, R_i) = (w_T, w_S, R)\) for all $i$;
- \(E2\) A worker applies to all firms with the same probability;
- \(E3\) All firms rank type $T$ workers above type $S$ workers, i.e., $R = 1$. 

The feature E1 also implies that \((q_{iT}, q_{iS}) = (q_T, q_S)\) for all \(i\). The feature E2 implies that the two types of workers are completely mixed across the firms. That is, all firms attract both types of workers in the sense that \(q_{iT} > 0\) and \(q_{iS} > 0\) for all \(i\).

Although the candidate has complete symmetry across firms, I will keep the subscript \(i\) for a while in order to establish the symmetry. Setting \(R_i = 1\), a firm’s maximization problem in the candidate equilibrium simplifies to:

\[
(P) \quad \max_{(w_{iT}, w_{iS})} \pi_i = (1 - e^{-q_{iT}})[(1 + \delta)y - w_{iT}] + e^{-q_{iT}}(1 - e^{-q_{iS}})(y - w_{iS}),
\]

subject to \(G(q_{iT})w_{iT} = E_T\) and \(e^{-q_{iT}} G(q_{iS})w_{iS} = E_S\). These equality constraints come from the conjectured features that \(q_{iT} > 0\) and \(q_{iS} > 0\). Using the constraints to substitute \((w_{iT}, w_{iS})\), I can write the firm’s expected profit as follows:

\[
\pi_i = (1 - e^{-q_{iT}})(1 + \delta)y + e^{-q_{iT}}(1 - e^{-q_{iS}})y - (q_{iT}E_T + q_{iS}E_S).
\]

This result expresses a firm’s expected profit as the difference between expected output and expected wage cost. Although a firm hires only one worker, the expected wage cost of type \(j\) workers is \(q_jE_j\), as if the firm hires a number \(q_j\) of such workers at a wage rate \(E_j\).

The first-order conditions for \((q_{iT}, q_{iS})\) yield

\[
y e^{-q_{iT} + q_{iS}} = E_S, \tag{2.9}
\]

\[
\delta y e^{-q_{iT}} + E_S = E_T. \tag{2.10}
\]

The solutions to these equations are indeed independent of the firm’s index \(i\). Thus, the features E1 and E2 hold if E3 holds and if all firms attract both types of workers. In this case, \(a_{iT} = a_{iS} = 1/M\), which implies the following queue lengths:

\[
q_T = \gamma \theta, \quad q_S = (1 - \gamma) \theta. \tag{2.11}
\]

Finally, the first-order conditions and the constraints in problem (P) become

\[
E_S = y e^{-\theta}, \tag{2.12}
\]

\[
E_T = E_s + \delta y e^{-\theta} = y e^{-\theta}[1 + \delta e^{(1-\gamma)\theta}], \tag{2.13}
\]

\[
w_T = y \frac{\gamma \theta [\delta + e^{- (1-\gamma)\theta}]}{e^{\theta} - 1}, \quad w_S = y \frac{(1 - \gamma) \theta}{e^{(1-\gamma)\theta} - 1}. \tag{2.14}
\]

The wages satisfy \(w_T - w_S < \delta y\) and so the ranking \(R = 1\) is ex post optimal indeed.

The above analysis has established the result that the candidate is unique. But the analysis has shown neither that the candidate is an equilibrium nor that an equilibrium must satisfy E1–E3. I will address these issues in the next section, whose result is summarized below:

**Theorem 2.1.** There is a unique equilibrium. The equilibrium satisfies E1–E3, where queue lengths are given by (2.11), market wages by (2.12) and (2.13), and actual wages by (2.14).

### 3. The candidate is the unique equilibrium

In this section, I will show that an equilibrium must have the features E1–E3. Let me classify the possible violations of E1–E3 into the following categories:

- **(N1)** Complete or partial separation of the two types of workers;
- **(N2)** Some or all firms do not rank type \(T\) workers strictly above type \(S\) workers.
For each of these cases, I suppose that it is an equilibrium, to the contrary, and compute the posted wages, queue lengths and market wages. Then I will construct a single firm’s deviation from this supposed equilibrium toward the one described in Theorem 2.1. I will show that this deviation is profitable, and so the possibility is not an equilibrium.

The same procedure also shows that the candidate is an equilibrium, and so it is the unique equilibrium. To see this, notice that possible deviations by a single firm from the candidate equilibrium are the limits of the above possibilities where the size of a subset of firms approaches zero. For example, consider a single firm’s deviation from the candidate equilibrium to a strategy that attracts only one type of workers. This deviation can be treated as the limit case of N1 where the size of the group of firms using such a strategy shrinks to zero. Because a firm in this group can profit by attracting both types of workers, as I propose to show, then the single firm’s deviation can be improved upon by the strategy in the candidate equilibrium. Similarly, other deviations are limit cases of N2, which are not optimal.

3.1. Separation is not an equilibrium

Consider first a special case of Case N1—complete separation. The argument can also be applied to rule out partial separation as an equilibrium, as stated at the end of this section. Suppose that the firms are divided into two groups, A and B, with group A attracting only type T workers and group B only type B workers. Such separation is the equilibrium examined by Lang et al. (2002) in a similar model under the additional restriction that each firm can post only one wage. It is important to show that separation is no longer an equilibrium once the additional restriction is eliminated. Second, the informal argument about the tradeoff between the wage and the employment probability does not automatically rule out separation as an equilibrium. With separation, each group of workers get the employment priority from the corresponding group of firms. The informal argument may even suggest that this priority can give the optimal tradeoff with wage. The fact that this is not the case indicates that the model imposes a much greater discipline on the equilibrium than the informal argument can provide. For these reasons, I will go through the detail of proving that complete separation is not an equilibrium.

Before the proof, let me explain intuitively why complete separation is not an equilibrium. Suppose the separating outcome described above is an equilibrium and consider the following deviation by a firm in group A. This firm maintains the same wage for type T workers and ranks such workers first as in the supposed equilibrium, but chooses a wage to attract type S workers as well. Type T workers will not change their strategy of applying to this firm, and so the expected profit from hiring a type T worker does not change. However, in the case where no type T worker shows up at the firm, the deviating firm can hire a type S worker and obtain additional profit. Thus, the deviation is profitable, provided that there is a feasible wage offer to attract type S workers with the lower ranking.

To verify this intuition, let me characterize the supposed equilibrium of separation. Let the posted wage be $w_A$ by a group A firm (for type T workers) and $w_B$ by a group B firm (for type S workers). Let $a$ be the fraction of firms that are in group A. Then, the expected number of applicants is $q_A = \gamma \theta / a$ for a group A firm and $q_B = (1 - \gamma) \theta / (1 - a)$ for a group B firm. Let $\hat{E}_T$ and $\hat{E}_S$ be expected wages of the two types of workers. By modifying the firm’s problem in the last section, I can formulate the maximization problem of firms in
each of the two groups. Solving these problems, I can obtain the following expressions for expected wages:

\[ G(q_A)w_A = \tilde{E}_T = (1 + \delta)ye^{-q_A}, \quad G(q_B)w_B = \tilde{E}_S = ye^{-q_B}. \] (3.1)

Also, a firm’s expected profit is as follows for the two groups, respectively:

\[ \pi_A = (1 - e^{-q_A})(1 + \delta)y - w_A = (1 + \delta)y[1 - (1 + q_A)e^{-q_A}], \]
\[ \pi_B = [1 - e^{-q_B}](y - w_B) = y[1 - (1 + q_B)e^{-q_B}]. \]

For a firm to be indifferent between being in the two groups, the supposed equilibrium must have \( \pi_A = \pi_B. \) This condition solves for the fraction \( a. \) Note that the condition implies \( q_A < q_B, \) i.e., \( a > \gamma, \) for all \( \delta > 0. \) If \( \delta = 0, \) then \( a = \gamma. \)

Now consider the deviation described earlier for a firm in group \( A. \) The firm maintains the wage \( w_A \) for type \( T \) workers and ranks these workers first, but posts \( w_{dA} \) to attract type \( S \) workers. It is clear that a type \( T \) worker will apply to the deviating firm with the same probability as in the supposed equilibrium, and so the expected number of type \( T \) workers whom the firm will receive is still \( q_A = \gamma 0/a. \) Let \( q_{dA}^d \) be the expected number of type \( S \) workers whom the deviating firm will receive. Then, a type \( S \) worker who applies to the deviating firm will be selected with probability \( e^{q_A}G(q_{dA}^d). \) Let \( w_{dA} \) and the associated queue length \( q_{dA}^d \) satisfy the following conditions:

\[ e^{-q_A}G(q_{dA}^d)w_{dA} = \tilde{E}_S, \quad ye^{-q_A+q_{dA}^d} = \tilde{E}_S. \] (3.2)

The first condition requires the deviation to give a type \( S \) applicant the same expected wage as in the market, and the second condition requires the deviation to be the best of its kind so that a type \( S \) worker’s expected output is equal to the worker’s market wage.

The deviation indeed attracts type \( S \) workers, i.e., \( q_{dA}^d > 0. \) To verify this, substituting \( \tilde{E}_S \) from (3.1) into the second equation in (3.2) yields \( q_{dA}^d = q_B - q_A, \) which is positive as shown earlier. Also, the deviation is feasible in the sense that \( w_{dA} < y. \) To verify this, combine the two equations in (3.2) to obtain \( w_{dA} = yq_{dA}^d/(e^{q_{dA}^d} - 1) < y. \) The strict inequality also implies that the profit from hiring a type \( S \) worker is positive. Finally, using the fact \( q_B > q_A \) and the formulas of \( (w_{dA}^d, w_A), \) I can show that the deviating firm’s ranking scheme is ex post optimal, i.e., \( w_A - w_{dA} < \delta y. \) Because the firm’s expected profit from attracting type \( T \) workers is the same as in the supposed equilibrium and the additional expected profit from attracting type \( S \) workers is strictly positive, the deviation is profitable. Thus, complete separation is not an equilibrium.

A similar argument shows that an equilibrium does not have partial separation in which one group of firms attract only one type of workers while other firms attract both types of workers. A firm that attracts only one type of workers can increase its expected profit by deviating to a strategy that attracts both types of workers.

3.2. All firms rank high-productivity workers the first

Because an equilibrium has no separation, all firms must attract both types of workers. The only remaining configuration that is different from the candidate equilibrium is Case N2, i.e., that some or all firms do not rank type \( T \) workers strictly above type \( S \) workers.
I will explain informally why this case is not an equilibrium. A formal proof can be found in Shi (2004).

Suppose that group A firms rank type S workers the first and group B firms rank type T workers the first. The two groups of firms face the same vector of market wages \((E_T, E_S)\) and, as established before, both need to attract the two types of workers. A firm in group A can gain from the following deviation. The firm can rank type T workers first and cut the wage offer to these workers; at the same time, the firm can raise the wage offer to type S workers to compensate for their reduced ranking. After these changes, the firm attracts more type T workers and fewer type S workers than before. However, because type T workers are more productive, the benefit from the increase in type T workers applying to the firm outweighs the loss from the reduction in type S applicants. Thus, the firm’s expected profit increases.

The same deviation is profitable when all firms rank type S workers the first, or when a fraction of firms give no priority, or when all firms give no priority. This procedure exhausts all the cases where not all firms rank type T workers strictly above type S workers.

4. Properties of the equilibrium and the social optimum

4.1. Properties of the equilibrium

The following proposition can be readily confirmed from (2.11) through (2.14).

**Proposition 4.1.** The equilibrium described in Theorem 2.1 has the following properties:

(i) A firm’s ex post profit from a type T worker is higher than from a type S worker;
(ii) A type T worker has a higher employment probability than a type S worker;
(iii) \(E_T > E_S\), and \((E_T - E_S)\) is of the same order of magnitude as \(\delta\);
(iv) \(\exists \delta_1 > \delta_0 > 0\) such that \(w_T < w_S\) for all \(\delta \in [0, \delta_0)\) and \(w_T > w_S\) for all \(\delta > \delta_1\);
(v) Define \(\Delta = (w_S/w_T) - 1\). \(\exists \delta_2 > 0\) such that, if \(0 < \delta < \min\{\delta_0, \delta_2\}\), then \((d\Delta/d\delta) < 0, (d\Delta/d\delta) > 0, (d\Delta/d\delta) > 0\).

The properties (i)–(iii) are intuitive. Property (i) is the reason why a firm selects type T workers first after matches take place. This ranking gives a high-productivity worker a higher employment probability and a higher expected wage than a low-productivity worker. The differential in expected wages is of the same order of magnitude as the productivity differential.\(^5\)

However, the actual wage is not always higher for high-productivity workers. As stated in (iv), only when the productivity differential is sufficiently large do high-productivity workers get a higher actual wage than low-productivity workers. When the productivity differential is small, workers with higher productivity get lower wages. This reverse wage differential arises from the ranking of workers. A higher ranking increases a worker’s employment probability and hence the expected wage by a discrete amount. So, when a

\(^5\)Precisely, \((E_T - E_S)/\delta\) lies in the interior of \((0, \infty)\) for all \(\delta\) including the limit \(\delta \to 0\). In contrast, \((w_T - w_S)/\delta \to -\infty\) as \(\delta \to 0\).
firm awards slightly more productive workers with a higher ranking, it can cut the wage for these workers by a discrete amount and yet still be able to attract them. By doing so, the firm can increase expected profit.

The wage differential responds to the market condition in an interesting way, as stated in (v). When there is a reverse wage differential as in the case of a small productivity differential, an increase in the overall ratio of workers to firms, $\theta$, increases the reverse wage differential. The explanation is as follows. When jobs become more scarce, workers value the employment probability more than the wage. Since high-productivity workers are given a higher ranking, they are willing to take a larger wage cut to maintain this higher employment probability.

Notice that the reverse wage differential increases as the productivity differential ($\delta$) decreases, as stated in (v). Thus, when the productivity differential shrinks to zero, the reverse wage differential will not shrink to zero as one might conjecture, but will get larger instead. Although unconventional, this result is quite intuitive. When the productivity advantage of one set of workers relative to other workers shrinks, a firm will maintain the employment priority for these workers only if their wage will be reduced. When the productivity advantage vanishes in the limit, the employment priority must come with a large wage cut.

More formally, the set of equilibria is discontinuous with respect to $\delta$ at $\delta = 0$ and that the limit process $\delta \to 0$ helps to select a unique equilibrium. When the economy is literally at $\delta = 0$, firms are indifferent between the two types of workers. For any ranking $R \in [0, 1]$, there are corresponding wages that are consistent with an equilibrium. That is, there is a continuum of equilibria at $\delta = 0$. However, any sequence of economies with $\delta > 0$ in the limit $\delta \to 0$ selects a unique equilibrium. This limit equilibrium has $R = 1$. In particular, the limit process does not select the equilibrium which gives all workers the same wage. As discussed in the introduction, the argument of compensating differentials is not able to make this selection.

The above results suggest that actual wages can sometimes be a bad indicator of workers’ productivity. The standard practice that attributes wage differentials to productivity differentials should be taken with caution. First, a large residual wage differential might be attributed to a statistically insignificant differential in productivity, as it is the case here when $\delta$ is small. In such a case, it is futile trying to explain the wage differential by ever expanding the list of workers’ characteristics. Second, there is nothing abnormal about a residual wage differential; rather, it is part of the equilibrium with fully rational players in a frictional labor market. Also, the residual wage differential is socially efficient, as shown later. Third, when similar workers get different wages, the ones who receive lower wage are not necessarily discriminated against. For example, when $\delta$ is small, high-productivity workers are paid a lower wage, but they get a job first and obtain higher expected wages.

4.2. The equilibrium is efficient

Since the equilibrium has unconventional features in wages, it is interesting to see whether the equilibrium is efficient. In the current context, aggregate output is an appropriate measure of social welfare because all agents are risk neutral.

Suppose that a fictional social planner tries to maximize aggregate output, subject to the same restrictions that matching frictions generate in the equilibrium. One such restriction
is that the planner must treat all workers of the same type in the same way. Thus, if the planner divides the firms into different groups to attract distinct compositions of workers, then the number of these groups should not exceed the number (two) of types of workers. Another restriction is that if the planner assigns a worker to match with a group of firms, then the worker must be matched with all firms in that group with the same probability. Both restrictions reflect the coordination failure. To describe the matching function, index the two groups of firms by $i$, where $i = A, B$. Let group $i$ firms be a fraction $a_i$ of all firms, where $a_A + a_B = 1$. Let $q_{ij}$ be the expected number of type $j$ workers attracted by each firm in group $i$, where $j = T, S$. Let $R_i$ be the ranking of the workers by a firm in group $i$, where $R_i \in \{1, 0, \Phi\}$. As in the equilibrium, the matching function facing the social planner is such that, in each group $i$, a firm receives one or more type $j$ workers with probability $(1 - e^{-q_{ij}})$. A planner’s allocation is $(a_i, q_{iT}, q_{iS}, R_i)_{i=A,B}$.

Expected output of a firm in group $i$ is

$$R_i[(1 - e^{-q_{iT}})(1 + \delta)y + e^{-q_{iT}}(1 - e^{-q_{iS}})y] + (1 - R_i)(1 - e^{-q_{iS}})y + e^{-q_{iS}}(1 - e^{-q_{iT}})(1 + \delta)y.$$ 

Because there are a large number of firms in group $i$, the above expression is also the average output per firm in group $i$. Re-arranging terms and weighting each group’s output by the group’s size, I can express expected output per firm in the economy as follows:

$$V = \sum_{i=A,B} a_i [y[1 - e^{-(q_{iT} + q_{iS})}] + \delta y(1 - e^{-q_{iT}})[R_i + (1 - R_i)e^{-q_{iS}}]]. \quad (4.1)$$

The planner chooses $(a_i, q_{iT}, q_{iS}, R_i)_{i=A,B}$ to maximize $V$, subject to $q_{iT} \geq 0$, $q_{iS} \geq 0$ and the following (resource) constraints:

$$a_A + a_B = 1, \quad a_A, a_B \in [0, 1],$$

$$a_A q_{Aj} + a_B q_{Bj} \leq \gamma_j \theta, \quad \text{for } j = T, S. \quad (4.2)$$

Here $\gamma_T = \gamma$ and $\gamma_S = 1 - \gamma$. The first constraint is self-explanatory. The second constraint is the adding-up constraint (2.3) in the current context.

The following theorem holds and a proof is supplied in Appendix A:

**Theorem 4.2.** The efficient allocation coincides with the equilibrium allocation.

The equilibrium is efficient because a worker’s expected wage takes into account the worker’s crowding-out on other workers in the matching process. To see this, let us examine the expressions for expected wages given by (2.9) and (2.10). Suppress the firm’s index $i$ in these expressions and consider first a type $S$ worker’s expected contribution. Adding a type $S$ worker contributes to a firm’s output only when the firm did not receive any other applicant, which occurs with probability $e^{-(q_{iT} + q_{iS})}$. So, a type $S$ worker’s contribution to expected output is $ye^{-q_{iT} - q_{iS}}$. Condition (2.9) equates this amount to a type $S$ worker’s expected wage. Similarly, a type $T$ worker contributes to a firm’s output by an amount $(1 + \delta y)$ if the firm did not receive any other applicant, by $\delta y$ if the firm received some type $S$ applicants and no type $T$ applicant, and by nothing if the firm received other type $T$ applicants. Since the first case occurs with probability $e^{-(q_{iT} + q_{iS})}$ and the second case occurs with probability $e^{-q_T}(1 - e^{-q_S})$, then the expected contribution of a type $T$ worker to output is $e^{-q_T}(\delta y + ye^{-q_S})$. Under (2.9), this contribution equals $(E_S + \delta ye^{-q_T})$. Condition (2.10) equates this amount to a type $T$ worker’s expected wage.
One can also compute workers’ social marginal values directly from the planner’s problem. For the planner, adding an additional worker of type \(j\) relaxes the resource constraint (4.2). The Lagrangian multiplier of this constraint in the planner’s problem, denoted \(\lambda_j\), is a type \(j\) worker’s social marginal value. One can verify that \(\lambda_j = E_j\) for both \(j = T\) and \(S\).

The efficiency property of expected wages shares the spirit of the efficient conditions described by Mortensen (1982, pp. 968–969) and Hosios (1990). Other models of directed search have shown that equilibrium wages with directed search can satisfy the Hosios condition, e.g., Moen (1997), Acemoglu and Shimer (1999) and Shi (2001). In the current context, however, having wages satisfy the Hosios condition alone is not enough for efficiency—Efficiency also requires the ranking of the different types of workers. The ranking effectively makes the wage schedule a non-linear function of productivity that resembles the outcome of an auction.\(^6\)

Let me make a few remarks on the efficiency result. First, it is expected wages, not actual wages, that serve the role of the efficient compensation scheme. Second, the wage differential (sometimes a reverse one) and the ranking are efficient. Third, separating the two types of workers in the matching process, completely or partially, is inefficient. Thus, the separating outcome described by Lang et al. (2002) is neither an equilibrium nor an efficient outcome for \(\delta > 0\), no matter how close the productivity differential is to zero.

**4.3. Equilibrium entry of firms is efficient**

So far I have assumed that the numbers of workers and firms are fixed. A natural question is whether the equilibrium will continue to be efficient when there is free entry of firms into the market. To answer this question, assume that each firm must incur a cost \(c > 0\) to enter the market, i.e., to set up a vacancy. Since the number of workers is still fixed (at a larger value \(N\)), determining the number of firms is equivalent to determining the market tightness \(\theta = N/M\). In the equilibrium, \(\theta\) is given by the free-entry condition \(\pi(\theta) = c\), where \(\pi(\theta)\) is a firm’s expected profit. In contrast, the socially efficient tightness maximizes the sum of firms’ net social values. The value created by a firm is \(V(\theta)\), which is given by (4.1). The net social value per firm is \([V(\theta) - c]\), and so the sum of firms’ net social values is \(M[V(\theta) - c] = N[V(\theta) - c]/\theta\). Maximizing this sum, the efficient tightness solves the condition: \(V(\theta) - \theta V'(\theta) = c\).

For any given \(\theta\), the equilibrium is efficient. Substituting \((q_S, q_T)\) from (2.11), \(E_S\) from (2.12) and \(E_T\) from (2.13), I can calculate \(\pi\) from (2.8) and \(V\) from (4.1) as

\[
\pi(\theta) = y[1 - (1 + \theta)e^{-\theta} + \delta[1 - (1 + \gamma \theta)e^{-\gamma \theta}]],
\]

\[
V(\theta) = y[1 - e^{-\theta} + \delta(1 - e^{-\gamma \theta})].
\]

It is easy to show that \(\pi(\theta) = V(\theta) - \theta V'(\theta)\). Thus, the following corollary holds:

**Corollary 4.3. With free entry of firms, the equilibrium continues to be socially efficient.**

The equilibrium level of entry is efficient because a firm’s expected profit in the equilibrium is equal to the firm’s social marginal value. To see this, suppose that an additional firm enters the market. It creates a value \(V\), but it also makes the market tighter.

\(^6\)See Shi (2002) for a similar result in a different environment.
for existing firms and hence crowds out these firms’ output. A single firm’s entry affects the
tightness by an amount $\partial \theta / \partial M = -\theta / M$ and a change in the tightness affects each firm’s
output by an amount $V'(\theta)$. Thus, the additional firm crowds out existing firms’ output by
an amount $\theta V'$. The firm’s social marginal value is $(V - \theta V')$. This value is equal to a
firm’s expected profit in the equilibrium. Therefore, the equilibrium compensates firms, as
well as workers, efficiently.

To appreciate the corollary, let us examine a regulation that requires all firms to pay the
same wage to all workers but allows firms to rank workers according to productivity. When the number of firms is fixed, the equilibrium with the regulation is efficient because it generates the same allocation of workers to firms as the equilibrium without the regulation. However, a firm’s expected profit with the regulation is not equal to the firm’s social marginal value. Once firms are allowed to enter, the equilibrium with the regulation will be inefficient.

5. Extension to many types of workers

In this section, I extend the model to incorporate many types of workers. This allows me
to examine how wages vary when the workers’ productivity varies in a wide range.

5.1. The equilibrium and its properties

Let there be $J \geq 2$ types of workers, which are indexed by $j = 1, 2, \ldots, J$. The
productivity of a type $j$ worker is $y_j$ and the fraction of workers of type $j$ is $\gamma_j$, where
$y_1 > y_2 > \cdots > y_J$ and $\sum_{j=1}^J \gamma_j = 1$. For this subsection, let me return to the specification
where the number of firms is fixed. As before, I examine only the symmetric equilibrium,
where all workers of the same type use the same strategy. In the presence of a large number
of types of workers, it is difficult to establish the uniqueness of the equilibrium. Thus, I will
only establish the existence of an equilibrium resembling the one in the simple model, i.e.,
an equilibrium in which all firms attract all types of workers and rank the workers
according to productivity.

To examine this equilibrium, I need to examine deviations that do not attract all types of
workers or do not rank the workers according to productivity. To do so, I introduce the
following general notation. Use the notation $i > i'$ to mean that a firm ranks type $i$ workers
above type $i'$ workers in the hiring process. For any integer $K$ with $1 \leq K \leq J$, I use $K$
stand for both the number and the ordered set $\{1, 2, \ldots, K\}$, where the ordering is
$1 > 2 > \cdots > K$. As in the simple model, it is not optimal for a firm to give the same
ranking to two types of workers. Thus, I will restrict attention to strategies, including
deviations, that have strict ranking over the types of workers. A strategy is defined by three
characteristics: a strictly ordered set of types, $C$, a vector of wages, $w$, and the associated
vector of queue lengths, $q$. The set $C$ contains the types of workers whom a firm attracts

7Strictly speaking, $q$ should be treated as the workers’ strategy, rather than a firm’s. However, the use of $q$ as a
firm’s choice is convenient and it does not change the analysis. As it is clear from the simple model, a firm that
intends to attract a type of workers must offer a combination of a wage and a queue length that gives the worker
the expected market wage. Under this constraint, the firm’s choice of $w$ effectively determines $q$. 
and the ordering of these types. It is useful to express \( C \) as

\[
C = (j_k)_{k=1}^K, \quad j_k \in J, \quad K \leq J,
\]

so that the ranking in \( C \) is according to the index \( k \), with \( j_1 \) being ranked the first. The vectors \( w \) and \( q \) are sorted according to the ordering in \( C \). Given \((C, w, q)\), it is convenient to define

\[
Q_{jk} = \sum_{i=1}^{k} q_{ji}.
\]

(5.1)

\( Q_{jk} \) is the expected number of workers whom the firm attracts and whose ranking is higher than or equal to \( k \) in \( C \). Clearly, \( Q_{j0} = 0 \).

An equilibrium consists of a strategy, \((C, w, q)\), where the ordering in \( C \) is strict, and a vector of expected market wages, \( E \), such that (i) given \( E \), the strategy is optimal for each firm, and (ii) the strategy induces \( E \). As said earlier, I will focus on the particular equilibrium which resembles the equilibrium in the simple model. This equilibrium has the following features: (a) All firms use the same strategy; (b) Every firm attracts all types of workers; and (c) \( C = J \), i.e., every firm ranks the workers according to productivity.

Feature (a) implies that a worker applies to all firms with the same probability, and so the queue length of type \( j \) workers at a firm is

\[
q_j = y_j^{0} \quad \text{for all} \ j.
\]

(5.2)

To characterize this equilibrium, consider an individual firm. With probability \( e^{-\Omega_j^{-1}} \), the firm receives no applicant whose productivity is higher than \( y_j \). Thus, the probability with which the firm successfully hires a type \( j \) worker is \( e^{-\Omega_j^{-1} (1 - e^{-\gamma_j})} = e^{-\Omega_j^{-1}} - e^{-\Omega_j} \).

Similarly, a type \( j \) worker who applies to the firm will be selected with the following probability:

\[
p_j = e^{-\Omega_j^{-1}} G(q_j) = [e^{-\Omega_j^{-1}} - e^{-\Omega_j}] / q_j.
\]

(5.3)

Taking the vector \( E \) as given, the firm chooses \((w, q)\) to solve the following problem:

\[
\max \pi = \sum_{j=1}^{J} (e^{-\Omega_j^{-1}} - e^{-\Omega_j})(y_j - w_j)
\]

s.t. \( p_j w_j \geq E_j \) for all \( j \).

(5.4)

Because \( q_j > 0 \) for all \( j \), then (5.4) holds with equality; that is,

\[
w_j = E_j / p_j \quad \text{for all} \ j.
\]

(5.5)

These wages are feasible: Because \( G(q) > e^{-q} \) for all \( q > 0 \), then \( p_j > e^{-\Omega_j} \) from (5.3) and \( p_j w_j = E_j < p_j y_j \) from (5.6). Substituting wages from (5.5), the first-order conditions for \( q \) yield

\[
E_j = y_j e^{-\Omega_j} - \sum_{i=j+1}^{J} y_i (e^{-\Omega_i^{-1}} - e^{-\Omega_i}) \quad \text{for all} \ j.
\]

(5.6)

The compensation scheme (5.6) generalizes (2.12) and (2.13). The right-hand side is a type \( j \) worker’s expected contribution to output. The first term is a type \( j \) worker’s contribution to output when the firm does not receive any applicant whose productivity is higher than or
equal to $y_j$. The second term is expected output that a type $j$ worker crowds out on the workers of lower productivity. Thus, a worker’s expected wage is equal to the worker’s social value—the worker’s expected output minus the expected crowding-out on other workers’ output.

The above analysis shows that the features (a)–(c) induce unique values of $(w, q, E)$. The following theorem shows that they are equilibrium values (see Appendix B for a proof):

**Theorem 5.1.** The strategy $(J, w, q)$ and the vector $E$, given by (5.2), (5.5) and (5.6), form an equilibrium. In this equilibrium, $p_j > p_{j+1}$, $E_j > E_{j+1}$, $w_j < y_j$ and $w_j - y_j < w_{j+1} - y_{j+1}$ for all $j$. However, $w_j < w_{j+1}$ if $(y_j - y_{j+1})$ is small.

Let me outline the procedure that shows why the particular $(J, w, q)$ and $E$ form an equilibrium. There are only two possible types of deviations by a single firm from the equilibrium. One is that the firm does not rank the workers according to productivity and the other is that the firm attracts only some but not all types of workers. From such a deviation, I can construct a further deviation toward the equilibrium strategy and show that it improves the deviator’s expected profit relative to the original deviation. By repeating this argument, I can construct a finite sequence of deviations, each of which improves upon the previous one and is closer to the equilibrium strategy. The last deviation in this sequence is the equilibrium strategy. This procedure shows that the original deviation is not profitable and so the strategy described in the above Theorem is an equilibrium.

Consider a deviation of the first type, denoted as $(C, q^d, w^d)$ where $C = (j_k^d)_{k=1}^K$ for $1 \leq K \leq J$. Let $j_s$ and $j_{s+1}$ be two types of workers in $C$ whose relative ranking is opposite to their ranking of productivity. That is, $y_{j_s} < y_{j_{s+1}}$. A further deviation can be constructed as follows. The firm can switch the rankings of these two types while maintaining the same wages and queue lengths for other types in $C$. This further deviation is closer to the equilibrium strategy than the original deviation. Also, it attracts more type $j_{s+1}$ workers and fewer type $j_s$ workers. Since type $j_{s+1}$ workers are more productive than type $j_s$ workers, the gain from the further deviation outweighs the loss, and so this further deviation improves upon the original deviation.

Consider a deviation of the second type, denoted in the same way as above. Let $j^*$ be the type of workers whom the deviating firm does not attract, where $j_s < j^* < j_{s+1}$. I can construct a further deviation that attracts type $j^*$ workers and ranks type $j^*$ between $j_s$ and $j_{s+1}$. This further deviation gives the same wages and queue lengths to all the workers in the set $C$, except type $j_{s+1}$. In addition, the sum of the queue lengths of type $j^*$ and type $j_{s+1}$ workers in the further deviation is equal to the queue length of type $j_{s+1}$ workers in the original deviation. The gain from the higher productivity workers (type $j^*$) outweighs the loss from the lower productivity workers (type $j_{s+1}$), and so this further deviation improves upon the original deviation.

The equilibrium characterized above has similar properties to the equilibrium in the simple model. First, the equilibrium is efficient as expected wages internalize the matching externalities. Second, higher productivity workers have a higher employment probability and higher expected wage. Third, actual wages do not depend on productivity in a systematic way. Actual wages decrease in productivity when the productivity differential $(y_j - y_{j+1})$ is small for all $j$. The opposite pattern is possible when $(y_j - y_{j+1})$ is large for all $j$. Non-monotonic patterns can also arise. For example, if $(y_j - y_{j+1})$ is large for large $j$ and
decreases sharply as \( j \) decreases, then actual wages increase first when productivity increases from the lowest level but decrease when productivity is sufficiently high.

The equilibrium wage distribution depends on the pattern of productivity and the distribution of workers. To compute the wage distribution, recall that each type \( j \) worker is employed with probability \( p_j \). Since type \( j \) workers are a fraction \( \gamma_j \) of the labor force, those who are employed (at wage \( w_j \)) are the following fraction of the labor force:

\[
\gamma_j p_j = \frac{1}{\theta} (e^{-\theta r_{j-1}} - e^{-\theta r_j}) \quad \text{where} \quad \Gamma_j = \sum_{j=1}^{J} \gamma_j.
\]

The total number of employed workers of all types is \( \left( \frac{1}{\theta} e^{\theta y}/c \right)^J = y \), which is independent of the distribution of types. Therefore, the density of workers employed at wage \( w_j \) is

\[
f_j = \frac{\gamma_j p_j}{\sum_{j=1}^{J} \gamma_j p_j} = \frac{\gamma_j p_j \theta}{1 - e^{-\theta}}, \quad (5.7)
\]

Because \( p_j > p_{j+1} \), then \( f_j/\gamma_j > f_{j+1}/\gamma_{j+1} \). That is, the higher the productivity of a type of workers, the larger the fraction of these workers will be employed. If all types of workers are uniformly distributed in the labor force, then the density of wages is an increasing function of productivity. However, because wages are not necessarily an increasing function of productivity, the density of wages can be increasing, decreasing, or non-monotonic in wages. In the next subsection, I will provide some numerical examples to illustrate this distribution.

### 5.2. Numerical examples

Before providing the numerical examples, let me introduce competitive entry of firms into the market as in Section 4.3. A firm enters the market after incurring a cost \( c > 0 \). In the equilibrium, the market tightness, \( \theta \), is determined by the free-entry condition \( \pi(\theta) = c \). As in the simple model, the equilibrium continues to be efficient with entry.

I use two examples to illustrate the wage distribution. In particular, these examples show that the model can generate a hump-shaped density of the wage distribution as the one documented in the literature (e.g., Mortensen, 2002). Since the model is not calibrated to the data, these examples are only illustrative. In both examples, I will set \( J = 20 \), \( y_1 = 100 \), \( c = 0.2y_J \) and \( y_j = y_1(1 + jA)^{1-j} \) for all \( j \). I will choose different values of \( y_J \) in the two examples, which will determine \( A \). The two examples will also differ in the distribution of workers in the labor force. These differences will imply different tightness and wage distribution in the two examples.

**Example 1.** \( y_J = 98 \) and \( \gamma_j = f_j^{0.3}/\sum_{j=1}^{J} f_j^{0.3} \). These imply \( A = 5.319 \times 10^{-5} \) and \( \theta = 0.819 \).

In this example, the difference between the highest and the lowest productivity is small, and the distribution of workers’ productivity in the labor force is a decreasing function. Fig. 1 depicts the equilibrium in this example. In the upper panel, the variable on the horizontal axis is an increasing function of productivity. This panel shows that the employment probability is an increasing function of productivity, as the theory predicts.

---

8Obviously, there are only a finite number of wage levels in the equilibrium distribution and so the phrase “density function” really means the frequency function.
Also, because the productivity differential is small, the wage rate is a decreasing function of productivity. Notice the magnitude of this reverse wage differential. Although the lowest productivity is only 2% lower than the highest productivity, the least productive workers’ wage is twice as much as the most productive workers’ wage! This large reverse wage differential is countered by an even larger positive differential in the employment probability. As a result, high-productivity workers are rewarded properly with a higher expected wage.

![Equilibrium and wage density in Example 1](image)

**Fig. 1.** The equilibrium and the wage density in Example 1: \( j \): worker’s type, with a lower \( j \) corresponding to higher productivity; \( w_j \): wage of type \( j \) workers, normalized by \( y_1 \); \( p_j \): employment probability of a type \( j \) worker; \( \gamma_j \): fraction of type \( j \) workers in the labor force; \( sw \): wages sorted in an ascending order; \( f \): wage density.
In Fig. 1, I depict the wage density and the distributional density of workers’ productivity. The wage density is hump-shaped. Because the wage rate is a decreasing function of productivity in this example, the shape of the wage density implies that more workers with medium productivity are employed than workers with either very high-productivity or very low-productivity. There are very few low-wage workers employed because these workers are high-productivity workers who are a small fraction of the labor force. There are also very few high-wage workers employed because these workers are low-productivity workers, who have a very low ranking and hence a very low employment probability.

The shape of the wage density sharply contrasts with the distribution of workers in the labor force, $\gamma$. In the lower panel of Fig. 1, the plot of $\gamma$ is the hypothetical density of wages when all workers are employed. In contrast to the humped shape of the actual density, this hypothetical density is increasing. For any wage level, if the actual density exceeds the hypothetical density, the workers at that wage are employed at a rate higher than the average employment rate; if the actual density falls below the hypothetical density, the workers at that wage are employed at a rate lower than the average employment rate. Thus, in this example, low-wage (high-productivity) workers are employed at a rate higher than the average rate while high-wage (low-productivity) workers are employed at a rate lower than the average rate.

In the above example, the hump-shaped wage density comes with a reverse wage differential. The following example shows that this is not a general feature.

**Example 2.**

$$y_j = 50 \quad \text{and} \quad \gamma_j = \frac{[y(J + 11 - j)]^4}{\sum_{j=1}^{J}[y(J + 11 - j)]^4}.$$  These imply $\Delta = 1.858 \times 10^{-3}$ and $\theta = 0.696$.

In contrast to Example 1, this example has a much larger difference between the highest and the lowest productivity. As a result, the wage rate is an increasing function of productivity, as depicted in the upper panel of Fig. 2. Also, the distribution of workers in the labor force has a single peak at an intermediate level of productivity, rather than being a decreasing function of productivity. Notice that the wage differential is small, in contrast to the large difference in productivity. The large difference in productivity induces a large difference in the employment probability, and hence in the expected wage.

The density of the wage distribution is still hump-shaped, as depicted in the lower panel of Fig. 2. The hypothetical wage density with all workers being employed, $\gamma$, is also hump-shaped. In comparison with this hypothetical density, the equilibrium wage density peaks at a high wage and a larger mass is distributed at higher wages. Thus, low-wage workers are employed at a rate lower than the average employment rate while high-wage workers are employed at a rate higher than the average rate. This result reflects the fact that high-wage workers in this example are high-productivity workers who are employed with a higher probability.

6. **Conclusion**

I have analyzed a stylized model that has some features one often observes in the labor market. For example, workers and firms cannot coordinate their decisions, wages play an important role to direct workers’ search, and firms rank their applicants. I show that the model has a unique equilibrium. The wage differential in the equilibrium depends positively on the productivity differential between workers only when the latter is large. When the
productivity differential is small, high-productivity workers get a lower wage than low-productivity workers. This reverse wage differential remains even when the productivity differential shrinks to zero. However, the equilibrium is socially efficient. High-productivity workers always have the employment priority and higher expected wages. Discrimination does not exist when a worker’s compensation is correctly measured by the expected wage.
Labor economists often do not measure discrimination with expected wages or lifetime earnings. Instead, the conventional measure is differentials in actual wages among similar workers. Using this measure, one will find discrimination here. In fact, the more similar are the workers, the larger is the differential in actual wages. This result should provide a strong caution against using the conventional measure of discrimination. It is for this reason that the word discrimination appears in the title of the paper. More generally, this paper shows that actual wages can be an unreliable measure of workers’ productivity when matching frictions exist in the labor market.

The reverse wage differential may contrast with empirical findings. For example, one may observe that slightly more educated workers are both paid higher wages and more likely employed. This finding should not lead one to quickly dismiss the reverse wage differential. In economics, one does not dismiss a theoretical result purely on the ground that it contradicts an empirical finding, especially when the result is the unique and efficient equilibrium. Instead, one should ask why the market does not produce the efficient allocation as the model describes? If one worker is only slightly more productive than another, giving the first worker both a higher ranking and a higher wage than the second worker is not optimal for a firm. What are the elements in reality that induce firms to reward workers in the observed pattern but that are missing in the model?

One possible element may be the time horizon. In this paper, workers and firms play the recruiting game for only one period. In reality, however, the game may repeat for many periods. If a more productive worker is paid less initially, he will be more likely to search on the job than a less productive worker. As a result, the wage of a productive worker will grow quickly over time and will soon surpass the wage of a less productive worker. Then, the average wage of more productive workers in the population of employed workers may exceed that of less productive workers. Notice that this process of on-the-job search produces a steeper wage path for more productive workers than for less productive workers, despite the absence of learning-by-doing and human capital accumulation. Unfortunately, on-the-job search also produces wage dispersion among identical workers, which complicates the analysis of directed search enormously (see Delacroix and Shi, forthcoming). Perhaps a quantitative analysis can be conducted.

Acknowledgements

I thank two referees and an editor for valuable comments. I have also received comments from Mike Peters, Rob Shimer, Peter Norman, Audra Bowlus, and Michael Baker. I gratefully acknowledge the financial support from the Bank of Canada Fellowship and from the Social Sciences and Humanities Research Council of Canada. The opinion expressed here is my own and it does not represent the view of the Bank of Canada.

Appendix A. Proof of Theorem 4.2

I can assume $a_A > 0$ and $a_B > 0$ without loss of generality. To see this, suppose that an allocation has $a_A = 1$, $a_B = 0$ and $(d_{IT}, q_{IS}, R_i)_{i=A,B}$. Then, an alternative allocation $(a_i^*, q_{IT}^*, q_{IS}^*, R_i^*)_{i=A,B}$ can be constructed as follows: $a_A^* = a_B^* = \frac{1}{2}$, $q_{AT}^* = q_{BT}^* = q_{AT}$, $d_{AS}^* = d_{BS}^* = d_{AS}$, and $R_A^* = R_B^* = R_A$. This alternative allocation is equivalent to the original allocation, except that it re-labels a half of the firms in the original group $A$ as group $B$. 
Let $\lambda_j$ be the Lagrangian multiplier of (4.2) in the planner’s maximization problem. Then, the first-order conditions of $q_{iT}$ and $q_{iS}$ are as follows:

$$\lambda_T/y \geq e^{-q_{iT} + q_{iS}} + \delta e^{-q_{iT}}[R_i + (1 - R_i)e^{-q_{iS}}], \quad "=" \text{ if } q_{iT} > 0, \quad (A.1)$$

$$\lambda_S/y \geq e^{-q_{iT} + q_{iS}} - (1 - e^{-q_{iT}})(1 - R_i)e^{-q_{iS}}, \quad "=" \text{ if } q_{iS} > 0. \quad (A.2)$$

The first-order condition of $a_A$ (taking $a_B = 1 - a_A$ as a constraint) is

$$0 = \delta(1 - e^{-q_{iT}})[R_A + (1 - R_A)e^{-q_{AS}}] - \delta(1 - e^{-q_{BT}})[R_B + (1 - R_B)e^{-q_{BS}}] + [e^{-(q_{BT} + q_{BS})} - e^{-(q_{AT} + q_{AS})}] - (q_{AT} - q_{BT})\lambda_T/y - (q_{AS} - q_{BS})\lambda_S/y. \quad (A.3)$$

The following features can be easily deduced from the planner’s problem. First, from inspecting the objective function, I can infer that the choice $R_i = 1$ is efficient whenever $q_{iT} > 0$ and $q_{iS} > 0$. Second, it is efficient to utilize both types of workers, in the sense that $q_{Aj} + q_{Bj} > 0$ for both $j = T, S$. To see this, suppose $q_{AS} = q_{BS} = 0$, to the contrary. Then, (4.2) does not bind and so $\lambda_S = 0$. In this case, type $S$ workers are not assigned to match with any firm, and so the ranking $R$ is irrelevant. Set $R_i = 1$. Then (A.2) implies $0 \geq e^{-q_{iT}}$ for $i = A, B$, which is a contradiction. Similarly, the choices $q_{AT} = q_{BT} = 0$ are not efficient.

With the above features, the efficient allocation must be one of the following cases (up to re-labeling): (i) $q_{iT} > 0$ and $q_{iS} > 0$ for both $i = A$ and $B$; (ii) $q_{AT} > 0$ and $q_{AS} > 0$ but $q_{BT} = 0 < q_{BS}$; (iii) $q_{AT} > 0$ and $q_{AS} > 0$ but $q_{BT} > 0 = q_{BS}$; (iv) $q_{AT} > 0 = q_{AS}$ and $q_{BT} = 0 < q_{BS}$. In the remainder of the proof, I will first show that case (i) is a solution to the planner’s problem and it yields the same allocation as the equilibrium described in Theorem 2.1. Then I will show that cases (ii) through (iv) are not efficient.

Consider case (i) first. In this case, $R_i = 1$ for both $i = A$ and $B$. Also, the first-order conditions for $q_{iT}$ and $q_{iS}$ hold as equality for both $i = A$ and $B$. These first-order conditions and constraint (4.2) together yield $q_{AT} = q_{BT} = \gamma \theta$ and $q_{AS} = q_{BS} = (1 - \gamma) \theta$. Thus, the two groups are identical, and so the choice of $a$ is irrelevant. This allocation is the equilibrium allocation described in Theorem 2.1.

In case (ii), $R_A = 1$. Also, (A.2) holds as equality for both $i = A$ and $B$, which yields $q_{AT} + q_{AS} = q_{BT} + q_{BS}$. (A.1) holds with equality for $i = A$. Combining (A.1) and (A.2) for $i = A$ yields $\lambda_T - \lambda_S = \delta e^{-q_{AT}}$. Substituting these results into (A.3), I have

$$0 = \delta[1 - (1 + q_{AT})e^{-q_{AT}}].$$

Because $1 > (1 + z)e^{-z}$ for all $z > 0$, the right-hand side of the above condition is strictly positive. This is a contradiction, and so case (ii) is not a solution to the planner’s problem.

In case (iii), $R_A = 1$, (A.1) holds as equality for both $i = A$ and $B$, and (A.2) holds as equality for $i = A$. Thus, $\lambda_T = ye^{-q_{AT}}(e^{-q_{AS} + \delta})$, $\lambda_S = ye^{-(q_{AT} + q_{AS})}$ and

$$q_{BT} = q_{AT} - \ln\left(\frac{\delta + e^{-q_{AS}}}{1 + \delta}\right).$$

Substituting these conditions into (A.3) and simplifying, I have

$$0 = -\ln\left(\frac{\delta + e^{-q_{AS}}}{1 + \delta}\right) - \frac{q_{AS}}{1 + \delta e^{q_{AS}}}. $$

The right-hand side increases in $q_{AS}$ and, at $q_{AS} = 0$, its value is 0. Because $q_{AS} > 0$, then the above condition is violated. Thus, case (iii) is not a solution to the planner’s problem.
Finally, consider case (iv). In this case, (A.1) holds as equality for \( i = A \) and (A.2) as equality for \( i = B \). Thus, \( \lambda_T = (1 + \delta) e^{-q_{AT}} \) and \( \lambda_S = ye^{-q_{BS}} \). (A.3) yields
\[
0 = (1 + q_{BS}) e^{-q_{BS}} + \delta - (1 + \delta)(1 + q_{AT}) e^{-q_{AT}}. \tag{A.4}
\]
Because no type \( S \) worker applies to a group \( A \) firm, \( R_A \) can be set to 1. Also, (A.2) holds as inequality for \( i = A \). Substituting \( \lambda_S = e^{-q_{BS}} \) and \( R_A = 1 \), this inequality implies \( q_{BS} \leq q_{AT} \). Since the function \((1 + z)e^{-z}\) is decreasing for all \( z > 0 \), the result \( q_{BS} \leq q_{AT} \) and the condition (A.4) imply \( 0 \geq \delta[1 - (1 + q_{AT})e^{-q_{AT}}] \). This cannot hold for \( q_{AT} > 0 \). Thus, case (iv) is not a solution to the planner’s problem. \( \square \)

**Appendix B. Proof of Theorem 5.1**

Before proving that the strategy described in the theorem is an equilibrium, let me verify the properties of the equilibrium stated in the theorem. Using (5.3), I have
\[
p_j - p_{j+1} = e^{-q_j} \left( \frac{e^{q_j} - 1}{q_j} - \frac{1 - e^{-q_{j+1}}}{q_{j+1}} \right) > 0. \tag{B.1}
\]
The inequality follows from the facts that \( e^z - 1 > z > e^{-z} \) for all \( z > 0 \). From (5.6), I have \( E_j - E_{j+1} = (y_j - y_{j+1})e^{-q_j} > 0 \). Then, (5.5) implies
\[
w_j - w_{j+1} = (y_j - y_{j+1}) \frac{q_j}{e^{q_j} - 1} - E_{j+1} \left( \frac{1}{p_{j+1}} - \frac{1}{p_j} \right). \tag{B.2}
\]
Because \( q_j < e^{q_j} - 1 \) and \( p_j > p_{j+1} \), the above condition implies \( w_j - w_{j+1} < y_j - y_{j+1} \) for all \( j \). Thus, \( w_j - y_j < w_{j+1} - y_{j+1} \) for all \( j \). In turn, this implies that \( w_j < y_j \) for all \( j \) if and only if \( w_j < y_J \). The latter condition indeed holds, because (5.5) implies
\[
w_j = \frac{E_j}{p_j} = y_j \frac{q_j}{e^{q_j} - 1} < y_J.
\]
The final property of the equilibrium is that \( w_j < w_{j+1} \) if \( (y_j - y_{j+1}) \) is small. To show this, note from (B.1) that \( (p_j - p_{j+1}) \) is bounded strictly above zero even when \( (y_j - y_{j+1}) \) approaches zero. Thus, (B.2) implies that \( w_j < w_{j+1} \) when \( (y_j - y_{j+1}) \) is sufficiently small.

Now I prove that there is no profitable deviation from the strategy described in the theorem. Consider a deviation by a single firm to the strategy \((C, q^d, w^d)\), where \( C = (j_k)_{k=1}^K \) and \( 1 \leq K \leq J \). As stated in the text, focus on strict ranking so that \( j_1 > j_2 > \cdots > j_K \). The deviation must satisfy the following properties for all \( k \in K \):
\[
e^{-\Theta_{j_k}^d} - e^{-\Theta_{j_k}^d} \frac{q_j^d}{p_j^d} w_{j_k}^d = E_{j_k}, \tag{B.3}
\]
\[
y_{j_k} e^{-\Theta_{j_k}^d} - \sum_{i=k}^{K} y_j (e^{-\Theta_{j_i}^d} - e^{-\Theta_{j_k}^d}) = E_{j_k}, \tag{B.4}
\]
where \( \Theta_{j_k}^d \) is defined similarly to (5.1), with \( q^d \) replacing \( q \). The first condition is type \( j_k \) workers’ participation constraint, and the second condition comes from the first-order condition of \( w_{j_k}^d \). If one of these conditions were violated, then the deviation could be improved upon by a further deviation to the strategy that satisfies these conditions.
Lemmas B.1 and B.2 below show that the only two possible types of deviations from the equilibrium can be improved upon by a further deviation toward the equilibrium strategy. By the argument in the text, this shows that the deviations are not profitable.

**Lemma B.1.** A deviation that does not rank workers according to productivity can be improved upon by another deviation that attracts the same types of workers as in the original deviation but that ranks the workers according to productivity.

**Proof.** Consider the deviation \((C^*, w^d, q^d)\) described above and suppose that the ranking in \(C\) is not according to productivity. Then, there exists a number \(s = \max\{k \in K : y_{jk} \leq y_{jk+1}\}\). That is, type \(j_s\) and type \(j_{s+1}\) are the lowest ranked pair which exhibits a relative ranking opposite to productivity. Consider a further deviation from this deviation \((C^*, q^s_{nk}, w^s_{nk})\) for all \(k \in K\). This further deviation attracts the same types of workers as those contained in \(C\) but it reverses the ranking between type \(j_s\) and type \(j_{s+1}\) workers. That is, \(n_k = j_s\) for all \(k \neq \{s, s + 1\}\), \(n_s = j_{s+1}\) and \(n_{s+1} = j_s\). Use \(q^s\) to construct \(Q^s\) similarly to (5.1). Furthermore, let this further deviation have the following properties: (i) For all \(k \neq \{s, s + 1\}\), \(w^s_{nk} = w^d_{js}\) and (B.3) is satisfied with \((q^s_{nk}, Q^s_{nk}, w^s_{nk})\) replacing \((q^d_{jk}, Q^d_{jk}, w^d_{jk})\); (ii) For \(k \in \{s, s + 1\}\), \((q^s_{nk}, Q^s_{nk}, w^s_{nk})\) solve the following problem:

\[
\begin{align*}
\max_{q^s_{nk}} & \sum_{k=s}^{s+1} (e^{-Q^s_{nk-1}} - e^{-Q^s_{nk}}) (y^s_{nk} - w^s_{nk}) \\
\text{s.t.} & \quad e^{-Q^s_{nk-1}} - e^{-Q^s_{nk}} \frac{q^s_{nk}}{w^s_{nk}} = E_{nk}, \quad k = s, s + 1, \\
& \quad q^s_{ns} + q^s_{n_{s+1}} = q^d_{js} + q^d_{j_{s+1}}. 
\end{align*}
\]  

(B.5)

For \(k \leq s - 1\), the two deviations have the same wages and both satisfy (B.3). Comparing the condition (B.3) for the two deviations and working from \(k = 1\), I obtain \(Q^s_{nk} = Q^d_{jk}\) for all \(k \leq s - 1\), which implies \(q^s_{nk} = q^d_{nk}\) for all \(k \leq s - 1\). With (B.6), I have \(Q^s_{n_{s+1}} = Q^d_{j_{s+1}}\). Using this result and again comparing the condition (B.3) for the two deviations, I have \(Q^s_{nk} = Q^d_{jk}\) and hence \(q^s_{nk} = q^d_{nk}\) for all \(k \geq s + 2\). Thus, the only difference between the two deviations lies in the employment probabilities and wages of type \(j_s\) and type \(j_{s+1}\) workers.

To show that the further deviation improves upon the original deviation, substitute (B.5) and (B.6) for \((w^s_{ns}, w^s_{n_{s+1}}, q^s_{n_{s+1}})\) and use the facts that \(n_s = j_{s+1}\), \(n_{s+1} = j_s\), \(Q^s_{n_{s-1}} = Q^d_{j_{s-1}}\) and \(Q^s_{n_{s+1}} = Q^d_{j_{s+1}}\). Then, the objective function of the above maximization problem becomes

\[
(e^{-Q^s_{j_{s+1}}} - e^{-Q^s_{n_s}}) y^s_{j_{s+1}} + q^s_{n_s} E_{j_{s+1}} + (e^{-Q^s_{n_s}} - e^{-Q^s_{j_{s+1}}}) y^s_{j_s} - (q^d_{j_s} + q^d_{j_{s+1}} - q^s_{n_s}) E_{j_s}.
\]

The first-order condition for \(q^s_{n_s}\) yields

\[
0 = (y^s_{j_{s+1}} - y^s_{j_s}) e^{-Q^s_{n_s}} + (E_{j_s} - E_{j_{s+1}}). 
\]

From (B.4), I can derive the following relationship:

\[
E_{j_s} - E_{j_{s+1}} = -(y^s_{j_{s+1}} - y^s_{j_s}) e^{-Q^s_{j_{s+1}}}. 
\]  

(B.7)

The above two conditions imply \(Q^s_{n_s} = Q^d_{j_s}\). Then, \(q^s_{n_s} = Q^s_{n_s} - Q^s_{n_{s-1}} = Q^d_{j_s} - Q^d_{j_{s-1}} = q^d_{j_s}\). The constraint (B.6) yields \(q^s_{n_{s+1}} = q^d_{j_{s+1}}\).
Let $\pi^d$ be the firm’s expected profit under the original deviation and $\pi^*$ under the further deviation. Then

$$\pi^d = \sum_{k=1}^K (e^{-Q_{jk}} - e^{-Q_{jk}^d})(y_{jk} - w_{jk}^d),$$

and $\pi^*$ is given similarly. Subtract the two

$$\pi^* - \pi^d = \sum_{k=s+1}^{s+1} [(e^{-Q_{n_k}^d} - e^{-Q_{n_k}})(y_{nk} - w_{nk}^*) - (e^{-Q_{n_k}^d} - e^{-Q_{n_k}})(y_{nk} - w_{nk}^*)].$$

Substituting $(Q^*, q^*)$ from above, wages from (B.5) and $E$ from (B.7), I have

$$\pi^* - \pi^d = (y_{j_{s+1}} - y_{j_s})e^{-Q_{j_s}}[e^{Q_{j_s}} - (2 + q_{j_s}^d - q_{j_{s+1}}^d) + e^{-q_{j_{s+1}}}].$$

Since $e^\varepsilon > 1 + \varepsilon$ and $e^{-\varepsilon} > 1 - \varepsilon$ for all $\varepsilon > 0$, the expression in $[\cdot]$ above is positive. Because $y_{j_{s+1}} > y_{j_s}$, then $\pi^* > \pi^d$. □

**Lemma B.2.** A deviation that does not attract all types of workers can be improved upon by a further deviation that attracts all types.

**Proof.** Let the original deviation be $(C, w^d, q^d)$, where $C = (j_k)_{k=1}^K$ and $j_k \in J$. Because the deviation does not attract all types, then $K < J$. With Lemma B.1, I can assume that $C$ ranks the workers by productivity, i.e., $j_1 < j_2 < \cdots < j_K$. Also, since any deviation that does not satisfy (B.3) and (B.4) can be improved upon, the deviation must satisfy these conditions. Let $j^* = \max\{j \in J \setminus K\}$. There are two cases to consider: $j^* > j_K$ and $j^* < j_K$.

If $j^* > j_K$, construct the further deviation $(C^*, w^*, q^*)$ as follows. Set $C^* = \{j_1, j_2, \ldots, j_K, j_{K+1}\}$, where $j_{K+1} = j^*$. For every $k \leq K$, set the wage $w_{jk}^* = w_{jk}^d$ and let the constraint (B.3) hold. For $k = K + 1$, let (B.3) and (B.4) hold. Then, $q_{jk}^* = q_{jk}^d$ for all $k \leq K$. This further deviation increases the firm’s expected profit, provided $w_{j_{K+1}}^* < y_{j_{K+1}}^*$. The latter condition can be verified.

Now examine the case $j^* < j_K$. Let $s < K$ be such that $j_{s-1} < j^* < j_s$. From the original deviation, I construct a further deviation $(C^*, w^*, q^*)$, where $C^* = (n_k)_{k=1}^{k+1}$ and $n_k$ is given as

$$n_k = \begin{cases} j_k & \text{if } k \leq s - 1, \\ j^* & \text{if } k = s, \\ j_{k+1} & \text{if } k \geq s + 1. \end{cases}$$

That is, the type $j^*$ is inserted between the types $j_{s-1}$ and $j_s$. Define $Q_{n_k}^* = \sum_{i=1}^k q_{n_i}^*$. Let the strategy in the further deviation satisfy: (i) For all $k \leq s - 1$, $w_{nk}^* = w_{jk}^d$ and (B.3) is satisfied with $(q_{nk}^*, Q_{nk}^*, w_{nk}^*, K + 1)$ replacing $(q_{jk}^d, Q_{jk}^d, w_{jk}^d, K)$; (ii) For all $k \in \{s, s + 1, \ldots, K + 1\}$, (B.3) and (B.4) are satisfied with $(q_{nk}^*, Q_{nk}^*, w_{nk}^*, K + 1)$ replacing $(q_{jk}^d, Q_{jk}^d, w_{jk}^d, K)$.

As in the proof of Lemma B.1, property (i) implies $Q_{n_k}^* = Q_{jk}^d$, $q_{nk}^* = q_{jk}^d$, and $w_{nk}^* = w_{jk}^d$ for all $k \leq s - 1$. For $k \geq s$, recall that $n_{k+1} = j_k$. Subtract (B.4) for $j_k$ in the original deviation and for $n_{k+1}$ in the further deviation, I have

$$0 = y_{jk}^* D_{jk} - \sum_{i=k+1}^K y_{ji}^* (D_{ji} - D_{ji}) \quad \text{for all } k \geq s,$$
where $D_{jk} = e^{-Q^*_{nk}} - e^{-Q^d_{jk}}$. Changing the index $k$ to $k+1$ and subtracting the resulted equation from the above equation, I obtain $(y_{jk} - y_{jk+1})D_{jk} = 0$. Since $y_{jk} > y_{jk+1}$, then $D_{jk} = 0$ for all $k \geq s$. That is, $Q^*_{nk+1} = Q^d_{jk}$, all $k \geq s$. In particular, $Q^*_{nK+1} = Q^d_{jk}$. Then, $Q^*_{nk+1} - Q^*_{nk} = Q^d_{jk} - Q^d_{jk}$, all $k \geq s$. Working from $k = K - 1$ to $k = s$, the above equation yields

$$q^*_{nk+1} = q^d_{jk} \quad \text{for all } k \geq s + 1.$$ 

$$q^*_{ns} + q^*_{ns+1} = Q^*_{ns+1} - Q^*_{ns} = Q^d_{js} - Q^d_{js-1} = q^d_{js}.$$ 

Since (B.3) holds in both deviations, the above equalities imply $w^*_{nk+1} = w^d_{jk}$, all $k \geq s + 1$. Thus, the only difference between the two deviations lies in type $n_s$ and type $n_{s+1}$ workers.

Let $\pi^d$ be the firm’s expected profit with the original deviation and $\pi^*$ with the further deviation. Then

$$\pi^* - \pi^d = -(e^{-Q^d_{js-1}} - e^{-Q^d_{js}})(y_{js} - w^d_{js}) + \sum_{k=s}^{s+1} (e^{-Q^*_{nk-1}} - e^{-Q^*_{nk}})(y_{nk} - w^*_{nk}).$$

Substitute (B.3) for $w^d$ and its counterpart for $w^*$, substitute (B.7), and use the above relationships between $(Q^*, q^*)$ and $(Q^d, q^d)$. Then,

$$\pi^* - \pi^d = (y_{ns} - y_{ns+1})e^{-Q^*_{ns-1}}[1 - (1 + q^*_{ns})e^{-g^*_{ns}}] > 0.$$ 

Thus, the original deviation can be improved upon by including the type $n_s = j^*$ in the set of workers to attract. This completes the proof of Lemma B.2 and hence of Theorem 5.1. 

References


