Credentials and Wage Discrimination

Robert Wilson*
Stanford Business School, Stanford, CA, USA

Abstract
The subject is wage discrimination among workers with differing privately known reservation wages and costs of obtaining “credentials” that can be observed by a monopsonist firm. Numerical examples are studied in which workers are equally productive, and in the population their reservation wages and credentialing costs have a Normal distribution with negative correlation. According to these examples, wage discrimination is unprofitable.

I. Introduction
An early issue in theories of collective bargaining was whether a firm with monopsony power, in a labor market without a union having countervailing monopoly power, would have an incentive to discriminate among workers according to their reservation wages, paying lower wages to those with lower reservation wages, even though they were equally productive; e.g., Robinson (1933, Chapter 26, §6). That is, is screening of workers according to their reservation wages a profitable strategy for a monopsonist firm?

If workers’ reservation wages are privately known then wage discrimination depends on observable behaviors that are correlated with reservation wages. Here we focus on “credentials” that are observable or verifiable by the firm. In related contexts, Spence (1973) emphasizes the role of wages conditioned on credentials, such as general “schooling”, that are costly for workers to obtain but need not increase productivity at work. In Spence’s model, more credentials signal higher productivity when productivity is positively related to a worker’s innate ability and the cost of credentials is

*Research support from the National Science Foundation, grant SES8605666, and the Office of Naval Research is gratefully acknowledged. I am indebted to John Roberts and Peter Reiss for comments and suggestions but they bear no responsibility for errors that remain.

1Wage discrimination could be based on an observable attribute such as age, gender, or race, but we exclude this possibility here. This case is the one analyzed by Robinson (1933).
negatively related; that is, productivity and credentialing costs are negatively correlated.

Here, however, we assume that workers are equally productive, and our focus is on the possibility that more credentials can signal a higher reservation wage. As in Spence's model, correlation between reservation wages and credentialing costs could arise from their mutual dependence on an unobservable or unverifiable third factor, again interpreted as innate ability, affecting both the worker's job opportunities or mobility, and the cost of credentials. One interpretation, of course, is that both are affected by a worker's valuation of home production or leisure; in particular, negative correlation is implied if "schooling" is mostly leisure.

Nevertheless, the interpretation emphasized here is the acquisition of firm-specific skills, which are the principal credentials observed in practice. Suppose that a worker's innate ability is measured by his or her ease in mastering firm-specific skills at any firm. That is, ability facilitates qualification for any job and fosters job mobility. Therefore a worker's reservation wage, as represented by alternative employment opportunities, is positively associated with ability. According to the hypothesized nature of ability, moreover, the worker's cost of mastering firm-specific skills at any one firm is negatively associated with ability. The predicted statistical relationship is thus that workers' reservation wages and credentialing costs are negatively correlated.

A simple example illustrates. Suppose that a firm offers a job that pays a wage \( w \) but requires acquisition of \( t > 0 \) firm-specific skills, each of which costs a worker \( c \). The net value to the worker of such a job is \( w - tc \). If a population \( F \) of such firms, each offering a different job profile \((w, t)\), defines a worker's alternative opportunities for employment, then for a worker whose cost is \( c \) the imputed reservation wage is

\[
\nu(c) = \max_{(w, t) \in F} (w - tc).
\]

According to the envelope theorem, this reservation wage is a decreasing convex function of the worker's cost \( c \) provided \( F \) is invariant.\(^2\) Thus, if opportunities of this sort define workers' reservation wages, then one expects them to be negatively correlated with workers' costs of firm-specific skills at any other firm.

The literature on signaling in labor markets that Spence initiated addresses a significantly different issue than we consider here. This literature mostly considers firms without monopsonist power and workers with equal (or irrelevant) reservation wages but differing productivities. In this context, wage differentiation based on credentials arises from workers' competitive incentives to demonstrate their productivities relative to their peers: if more productive workers incur lower costs of obtaining each credential, then they have incentives to signal their greater productivities by offering more credentials. Thus, in this context, wages conditioned on unproductive credentials reflect genuine differences in workers' innate productivities.

Here we examine the case that a single firm has monopsony power in the labor market, except as reflected in workers' reservation wages.\(^3\) Moreover, workers are equally productive at this firm. Our main technical assumption is that workers' reservation wages and credentialing costs are negatively correlated. In this context a discriminating wage schedule is purely a device to exploit the firm's monopsony power, and does not manifest differences in workers' productivities. It does, however, reflect differences in workers' reservation wages and credentialing costs.\(^4\)

The model is formulated so that workers' wages and credentialing costs are additive, and for each worker the cost of each credential is constant. Similarly, the firm assigns the same value to each worker's production, possibly depending on the firm-specific skills provided. This assumes essentially that the firm has a constant-returns technology employing homogenous workers in undifferentiated jobs, and its product market is perfectly competitive. The range of credentials is assumed to be a continuum, although this is immaterial. Similarly, the population of workers is modeled as a nonatomic probability measure.

After presenting a simple example to introduce the subject, we derive the conditions that characterize a discriminating wage schedule that is optimal for the firm. These conditions are specialized to the case of a Normal distribution of worker's attributes, and then some numerical examples are presented. Based on these numerical examples, it appears that wage discrimination can be unprofitable for the firm compared to using a single wage without any credential requirements. These examples do not, of course, exclude the possibility that other distributions might enable a firm to benefit from wage discrimination.

II. A Simple Example

To motivate the subsequent analysis, consider the following simple example. A monopsonist firm buys labor in a market in which all workers

\(^2\) Similarly, considering a heterogeneous population of workers whose costs are distributed independently of their selected or feasible job profiles, the covariance between workers' reservation wages and skill-acquisition costs is \(-10\), in terms of the mean of \( t \) in the population of firms and the variance of \( c \) in the population of workers.

\(^3\) Extensions to several firms, competing directly via discriminating wages schedules offered in the same labor market, are straightforward using the methods in Oren, Smith and Wilson (1983).

\(^4\) I share with others the experience, when teaching Spence's theory of signaling in labor markets, that students often assume initially that the subject addressed is the one considered here.
are equally productive, but some have a higher reservation wage, and the firm knows that those who do can cheaply acquire verifiable credentials. For simplicity, assume that credentials do not increase a worker's productivity. Suppose that the reservation wage and the unit cost of credentials are \( v_i \) and \( c_i \) for the two types \( i = 1, 2 \) of workers. Assuming that \( v_1 < v_2 \) and \( c_1 < c_2 \), there exists a wage \( w < v_2 \) and an amount \( t > 0 \) of credentials such that each type is indifferent between its reservation wage without credentials, and the wage \( w \) and the credentials \( t \). Consequently, the firm can offer the two options \((v, 0)\) and \((w, t)\) (or perturb these slightly to induce strict preferences for the workers), and expect that the first type will select the first option and the second type will select the second option. For the firm, this is a less costly wage policy than offering the single wage \( v_2 \) that attracts both types if in the population of workers the proportion of the first type with the lower reservation wage is sufficiently large. If we normalize units so that \( v_2 = 1 \) and \( c_2 = 1 \), then the proportion of the first type must exceed \( 1/c \) for wage discrimination to be profitable for the firm.

From this example, one sees that the optimality of wage discrimination via credentials depends on the distribution of workers' attributes in the population. Our aim in the sequel is to generalize this analysis and to apply it to the case that workers' attributes are distributed according to a Normal distribution. Fixing the form of the distribution in this way, we can then examine the range of parameters for which wage discrimination is profitable.

### III. Formulation

The firm is assumed to assign a reservation price \( p^0 + c^0 t > 0 \) to the production from each such worker who provides \( t \) credentials. Thus, each worker hired at the wage \( w \) yields a profit of \( p^0 + c^0 t - w \). The firm's demand for workers is assumed initially to be unlimited; later we mention how to include constraints on the firm's capacity to employ workers.

Each worker in the population is described by a pair \((v, c)\) of attributes having the following interpretation. Compared to rejecting employment at the firm, the worker's net benefit is \( w - [c^0 + c] t - v \) at the wage \( w \) if the worker acquires \( t \) credentials. If \( w - [c^0 + c] t - v < 0 \) for all combinations \((w, t)\) offered by the firm, then the worker prefers to reject employment. Otherwise, the worker acquires the amount of credentials \( t \geq 0 \) that maximizes this net benefit, and obtains the wage \( w = w(t) \) offered by the firm to workers presenting \( t \) credentials. A special case is a constant (or linear) wage function for which credentials are ignored, no worker acquires credentials, and all workers whose reservation wages exceed the offered wage accept employment.

The population is described by a distribution of workers' attributes. Assuming that attributes are distributed continuously, at least in expectation, this distribution is conveniently summarized from the firm's viewpoint by a density function \( D \) having the following interpretation: \( D(\pi, c) \) is the density of workers having credentialing cost \( c^0 + c \) for whom \( v \leq p^0 - \pi \). Equivalently, if we make the transformation \( u = p^0 - v \), then it is the density of those with cost \( c^0 + c \) and \( u \geq \pi \). Thus, absent any credentials, in terms of the firm's profit margin \( \pi = p^0 - w \) over the wage \( w \) that it pays to these workers, the density function \( D(\pi, c) \) is also the firm's demand function for jobs from this segment of the labor market. Workers' costs are naturally restricted to the range \( c^0 + c \geq 0 \). Assume that \( D \) is twice differentiable, decreasing in \( \pi \), and increasing in \( c \). The last of these implies that higher credentialing costs are associated statistically with lower reservation wages.

Represent the firm's strategy as the choice of a function \( p(t) = p^0 + c^0 t - w(t) \) indicating its intended profit margin on workers offered the wage \( w(t) \) for credentials in the amount \( t \). Since the wage schedule must be concave and nondecreasing to obtain discrimination among workers, we expect that \( p \) will be a nonincreasing and convex function.\(^5\) Similarly, represent workers' strategies in terms of a function \( c^0 + c(t) \) indicating the unit credentialing cost of a worker acquiring \( t \) credentials, or more generally the greatest unit cost among those acquiring at least \( t \) credentials. We expect that the variable portion \( c(\cdot) \) will be a nonincreasing function: workers with lower credentialing costs will offer more credentials. In terms of this notation, the firm's (expected) profit is

\[
\Pi[p, c] = \int_{c(0)}^{\infty} D(p(0), c) \, dc + \int_{c(T)}^{\infty} p(t) D(p(t) + c(t), c(t)) \, dc(t)
\]

\[
+ p(T) \int_{c(T)}^{\infty} D(p(T) + cT, c) \, dc,
\]

assuming these integrals are well defined; here \( T \) is the maximum amount of credentials for which the wage schedule is specified by the firm. The three terms in this formula represent the profits from those workers accepting employment who acquire no credentials, who acquire credentials \( t \in (0, T) \), and who acquire the maximum credentials \( T \).

Given any differentiable, increasing, strictly concave wage schedule that induces a differentiable function \( c(t) \) on some interval within \((0, T)\), optimal behavior by workers entails the requirement that

\[
c(t) = w'(t) - c^0 = -p'(t),
\]

\(^5\) Concavity of the wage schedule is due here to the linear structure of workers' net benefit functions, hence it is not a property that carries over to more general formulations.
and therefore
\[ c'(t) = w''(t) = -p''(t); \]
that is, a worker acquires credentials up to the point that the resulting wage increment falls to his marginal cost. Subject to this restriction, the firm's objective is to choose its wage schedule or the equivalent schedule of marginal profits so as to maximize the profit \( \Pi[p,c] \) shown above.

Assume temporarily that the schedule \( p \) is differentiable, decreasing, and strictly convex; and that the schedule \( c \) is differentiable. We focus the analysis first on the optimization of the second term above in the formula for the profit:

\[
\int_0^t p(t)D(p(t) - p'(t)t, -p'(t))d\mu(t)dt,
\]

which poses a problem in the calculus of variations. The Euler condition provides a necessary condition that characterizes an optimal schedule; and in addition, in combination with the other terms in the profit function, transversality conditions determine the optimal endpoints. Because \( p''(t) \) occurs linearly in the integrand, the Euler condition takes a simple form that reduces to the following statement. Define the two elasticities
\[
\eta_\pi = \pi D_\pi/D \quad \text{and} \quad \eta_c = cD_c/D,
\]
evaluated at \((\pi, c) = (p(t) - p'(t)t, -p'(t))\). Then the Euler condition specifies at \( t \) that either \( p''(t) = 0 \), meaning that \( p \) is locally linear, or else:
\[
\|\eta_\pi + \eta_c\| = -1.
\]
Note that this condition resembles the usual optimal pricing condition for a monopoly, which requires that the demand elasticity is \(-1\); here, however, the market segments are not isolated and the monopsonist firm takes account also of the workers' self-selection of their credentials.

The Euler condition can be written out in full as follows. Define the function
\[
M(\pi, c) = \frac{\partial}{\partial \pi} \pi D(\pi, c) + \frac{\partial}{\partial c} cD(\pi, c).
\]
Then we obtain two conditions for each \( t \): either the optimal schedule is locally linear, or
\[
M(\pi, c) = 0,
\]
and
\[
tM_\pi(\pi, c) + M_c(\pi, c) = 0.
\]
Given \( t \), let the solution to these two equations be \((\pi(t), c(t))\). The first equation merely restates the Euler condition described above in terms of elasticities; the second equation is equivalent to the property that if \( \pi(t) = p(t) - p'(t)t \) and \( c(t) = -p'(t) \) as required, then \( \pi'(t) = c'(t)t \). Having obtained a solution to these two equations for each value of \( t \), the optimal schedule is \( p(t) = \pi(t) - c(t)t \).

There is, however, the important proviso mentioned earlier: in any interval where this schedule is not convex or these equations have no solution, the optimal schedule is linear (i.e., \( p''(t) = 0 \)) and in this case the optimal schedule is obtained by taking the linear extension of the schedule from regions where it is defined and convex. In addition, it may be in any case that the optimal schedule is everywhere constant or linear, corresponding to a wage schedule that either does not reward credentials or does not induce workers to acquire credentials. This case can be checked only by comparing the firm's profit from a discriminating schedule with the profit it obtains from a fixed wage.

It is important to note that the two equations derived above entirely determine the optimal discriminating schedule in the interval \( (0, T) \). Consequently, the transversality conditions play a role in determining the maximum credential quality \( T^0 \). One can show that the transversality conditions require that \( T^0 \) is the value of \( T \) that satisfies the equation
\[
\int_0^{T^0} \frac{d}{dT}[p(T)D(p(T) + cT, c)] dc = 0.
\]

Another Example
In order to emphasize the important role of the proviso mentioned previously, consider an example in which \( D(\pi, c) = \max\{0, c - \pi\} \) for \( \pi, c \in [0, 1] \), corresponding to the case that in the population \( \pi \) and \( c \) are uniformly distributed on the triangle where \( \pi \leq c \). In this case \( M(\pi, c) = 3(c - \pi) \) and the second equation above has no solution for \( t < 1 \); indeed the optimal schedule is any linear schedule of the form \( p(t) = 1/3 - kt \) for \( 0 \leq k \leq 1/3 \). If \( k > 0 \) then one might describe this schedule as an attempt at discrimination, but in fact all workers who accept employment offer no credentials.

Similarly, in the case that \( \pi \) and \( c \) are uniformly distributed on the triangle where \( \pi \geq c \), an optimal schedule is the constant schedule \( p(t) = 2/3 \). In both of these cases, wage discrimination by the firm based on credentials is unprofitable.

In the following section we use a Normal distribution of workers' reservation wages and credential costs in order to study wage discrimination in a wider class of cases.
IV. A Normal Distribution of Attributes

A Normal distribution of workers' attributes is specified by the mean and variance \( (m_u, s_u^2) \) of the marginal distribution of \( u = p^0 - v \), where in the absence of credentials \( p^0 \) is the firm's valuation of a worker and \( v \) is the worker's reservation wage; the mean and variance \( (c^0 + m_c, s_c^2) \) of the marginal distribution of workers' unit credential costs; and the correlation coefficient \( r \) for the joint Normal distribution of these two attributes. It will suffice, however, to standardize the two variances to be 1 by choosing the units in which wages and credentials are measured; thus, assume \( s_u = 1 \) and \( s_c = 1 \). Recall that the conditional distribution of \( u \) given \( c \) has mean and variance \( (m_u + r(c - m_c), 1 - r^2) \) in this case. Assume further that the correlation is positive, \( r > 0 \), which is to say that reservation wages and credential costs are negatively correlated. For the numerical examples reported below, we also assume that \( m_u > 0 \) and \( m_c > 0 \). Use \( \Phi = 1 - \Phi \) and \( \phi \) to represent the standard Normal distribution and density functions for a random variable with mean and variance \((0, 1)\), and use \( \psi(x) = \Phi(x) / \phi(x) \) to represent Mill's ratio.

In terms of standardized units, the density function \( D \) can be expressed as the product

\[
D\left( \pi, c \right) = \Phi\left( \frac{\pi - m_u - r(c - m_c)}{\sqrt{1 - r^2}} \right) \phi(c - m_c)
\]

of the conditional probability, given the variable part \( c \) of the credential cost, that \( u > \pi \), and the marginal density of \( c \). Because of this special form, we can reduce the number of parameters to two by defining

\[
\mu = \frac{m_u - m_c}{\sqrt{1 - r^2}}
\]

and by defining the normalized variables:

\[
\gamma = c - m_c,
\]

\[
\delta = \frac{\pi - rc - \mu}{\sqrt{1 - r^2}},
\]

\[
\tau = \frac{r - \tau}{\sqrt{1 - r^2}}.
\]

In terms of this normalization, the two equations that determine an optimal convex schedule take the following form:

\[
\psi(\delta) | 2\gamma + m_c | = \delta + \mu,
\]

and

\[
\psi(\delta) | 2\gamma + m_c | = \tau | 3 - \gamma | 3 + m_c | - \delta [\delta + \mu].
\]

To reduce notation further, define

\[
Q(\delta) = 2 + \left[ \frac{m_c}{2} \right]^2 - \theta(\delta) \text{ where } \theta(\delta) = \frac{\delta + \mu}{\psi(\delta)}.
\]

Then there is a solution to these equations if and only if \( Q(\delta) \geq 0 \), and the solution in terms of \( \delta \) is provided by the two formulas:

\[
\gamma = -\frac{m_c}{2} \pm \sqrt{Q(\delta)},
\]

\[
\tau = \frac{\pm 2 \psi(\delta) \sqrt{Q(\delta)}}{1 + |1 - \delta \psi(\delta)| \theta(\delta)}.
\]

where one takes the same sign + or − in both equations. Using these last two formulas, one can find the values of \( \gamma \) and \( \tau \) associated with each value of \( \delta \); converting back to the original notation then yields the schedule in terms of \( p(t) \) and \( c(t) \) as desired. Generally, for each value of \( \delta \) there are two solutions depending on the sign chosen, and both are part of the schedule.\(^6\) One wants therefore to use those values of \( \delta \) sufficiently large to produce \( r \geq 0 \) (i.e., \( r \leq r/\sqrt{1 - r^2} \)) at the one extreme using the positive sign; and those values of \( \delta \) sufficiently large to produce \( c^0 + c(t) \geq 0 \) (i.e., \( \gamma \geq -c^0 - m_c \)) at the other extreme using the negative sign. Note that the chosen sign switches at \( \tau = 0 \) (i.e., \( \tau = r \) and \( c(t) = m_c/2 \)), indicating that the range of \( \delta \) has its maximum value where \( Q(\delta) = 0 \).

Before proceeding further we mention three observations. One is that assuming that the range of \( t \) is a continuum is immaterial; e.g., in each of the following examples, construction of an approximate schedule based on selection of several points from the continuous schedule yields the result that the firm's profit is affected only slightly. Indeed, from numerical examples it is clear that slight administrative costs would make it useless to employ more than three or four points from the optimal schedule. Secondly, experience with numerical examples indicates that typically the transversality condition does not play an important role. Its effect is to truncate the discriminating schedule at a level of credentialing costs \( c(T^0) \) that is only slightly greater than zero. Extending the schedule further, say to zero, has a negligible effect on the firm's profit. Consequently, we truncate the schedule at \( c(t) = 0 \) in the numerical examples rather than calculate \( T^0 \) exactly. Thirdly, for the case of a Normal distribution, a constraint on the firm's capacity to employ workers has no material effect on

\(^6\) Only the positive sign arises if \( m_c < 0 \).
Table 1. Tabulation of the discriminating schedule \((m_a, s_a) = (400, 100), (m_b, s_b) = (8, 4), r^2 = 0.5, \mu = 3.657\). Population size: 0.9772. Correlation: 0.7071

<table>
<thead>
<tr>
<th>(t)</th>
<th>(p(t))</th>
<th>(\Phi(\delta))</th>
<th>(c(t))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>377.59</td>
<td>0.670</td>
<td>8.49</td>
</tr>
<tr>
<td>1.414</td>
<td>365.72</td>
<td>0.653</td>
<td>8.30</td>
</tr>
<tr>
<td>2.793</td>
<td>354.42</td>
<td>0.636</td>
<td>8.08</td>
</tr>
<tr>
<td>4.149</td>
<td>343.61</td>
<td>0.618</td>
<td>7.85</td>
</tr>
<tr>
<td>5.497</td>
<td>333.19</td>
<td>0.600</td>
<td>7.60</td>
</tr>
<tr>
<td>6.855</td>
<td>323.07</td>
<td>0.582</td>
<td>7.31</td>
</tr>
<tr>
<td>8.248</td>
<td>313.11</td>
<td>0.564</td>
<td>7.04</td>
</tr>
<tr>
<td>9.716</td>
<td>303.13</td>
<td>0.545</td>
<td>6.76</td>
</tr>
<tr>
<td>11.38</td>
<td>293.81</td>
<td>0.527</td>
<td>6.49</td>
</tr>
<tr>
<td>13.19</td>
<td>283.54</td>
<td>0.508</td>
<td>6.21</td>
</tr>
<tr>
<td>17.68</td>
<td>260.42</td>
<td>0.490</td>
<td>4.90</td>
</tr>
<tr>
<td>21.49</td>
<td>247.79</td>
<td>0.504</td>
<td>2.66</td>
</tr>
<tr>
<td>23.15</td>
<td>243.79</td>
<td>0.518</td>
<td>2.12</td>
</tr>
<tr>
<td>24.50</td>
<td>241.20</td>
<td>0.532</td>
<td>1.71</td>
</tr>
<tr>
<td>25.71</td>
<td>239.34</td>
<td>0.546</td>
<td>1.38</td>
</tr>
<tr>
<td>26.83</td>
<td>237.94</td>
<td>0.560</td>
<td>1.08</td>
</tr>
<tr>
<td>27.91</td>
<td>236.94</td>
<td>0.574</td>
<td>0.82</td>
</tr>
<tr>
<td>28.95</td>
<td>236.20</td>
<td>0.588</td>
<td>0.59</td>
</tr>
<tr>
<td>29.98</td>
<td>235.70</td>
<td>0.602</td>
<td>0.38</td>
</tr>
<tr>
<td>31.01</td>
<td>235.41</td>
<td>0.615</td>
<td>0.18</td>
</tr>
<tr>
<td>32.04</td>
<td>235.32</td>
<td>0.629</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The formulas previously derived: the constraint serves only to introduce a Lagrange multiplier \(\lambda \geq 0\) that, in effect, merely reduces \(p^0\); this shifts upward the optimal schedule \(p(t)\) by \(\lambda\) and shifts down the schedule \(w(t)\) by the same amount.

A Normal Example

In order to illustrate our main conclusion, which is based on calculation of a variety of examples, we present a single example in some detail and indicate how its features carry over to the other examples we have computed. The data for this example are shown in the heading of Table 1, which displays the resulting schedule in both tabular and graphical form. (The population size mentioned in the heading is the proportion of the population having nonnegative credentials costs; the distribution is assumed to be truncated to exclude negative costs.) The points tabulated correspond to evenly spaced values of \(\delta\); those to the left of \(t = 17.678\) are based on the positive sign, and those to the right on the negative sign. The

The examples were computed using an algorithm, programmed in the APL language, which is available from the author on request. The results are computed to an accuracy exceeding what is shown in the tables that follow.

transversality condition implies that the optimal schedule should be truncated at \(t^0 = 30\) and \(c(t^0) = 0.38\), but this affects the firm's profit only in the fourth decimal place.

This schedule does not, however, reflect an optimal wage policy for the firm. In Table 2 we tabulate some relevant summary results based on this discriminating schedule and compare them with the best constant schedule that has \(p(t) = 31.1\), as well as the constant schedule \(p(t) = 35.6\) that yields the same supply (0.67) of workers as the discriminating schedule does. Both of these are "one-point schedules" since only one point on the schedule is chosen by workers; also shown is a two-point schedule that uses the two endpoints of the discriminating schedule. It is clear from these results that the discriminating schedule is inferior for the firm. In particular, the constant schedule \(p(t) = 35.6\) yields the same supply of workers, yet its profit of 238 exceeds the firm's profit of 229 from the discriminating schedule.

The principle conjecture that one might entertain is that a sufficiently strong correlation would suffice to make the discriminating wage schedule profitable. In Table 3 we show that this is not the case: although profits

Table 2. Summary of results \((m_a, s_a) = (400, 100), (m_b, s_b) = (8, 4), r^2 = 0.5\). Population size: 0.9772. Correlation: 0.7071

<table>
<thead>
<tr>
<th>(t) (range)</th>
<th>(p(t)) (avg.)</th>
<th>Supply</th>
<th>Firm's profit</th>
<th>Credential costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrim. Schedule:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t = 0)</td>
<td>378</td>
<td>0.38</td>
<td>145</td>
<td>15</td>
</tr>
<tr>
<td>(0 &lt; t &lt; 32)</td>
<td>294</td>
<td>0.29</td>
<td>84</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>378</td>
<td>0.67</td>
<td>211</td>
<td>35</td>
</tr>
<tr>
<td>Two-Point Schedule:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t = 0)</td>
<td>378</td>
<td>0.56</td>
<td>211</td>
<td>35</td>
</tr>
<tr>
<td>(t = 32)</td>
<td>235</td>
<td>0.10</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>355</td>
<td>0.66</td>
<td>236</td>
<td>35</td>
</tr>
<tr>
<td>One-Point Schedule:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t = 0)</td>
<td>356</td>
<td>0.67</td>
<td>238</td>
<td>35</td>
</tr>
<tr>
<td>Best One-Point Schedule:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t = 0)</td>
<td>311</td>
<td>0.81</td>
<td>252</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 3. Comparison of summary results for different correlations \((m_a, s_a) = (400, 100), (m_b, s_b) = (8, 4)\). Population size: 0.9772

<table>
<thead>
<tr>
<th>Corr, (r^2)</th>
<th>Supply</th>
<th>Single price</th>
<th>Discerning schedule</th>
<th>Cred. costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.01)</td>
<td>0.54</td>
<td>388</td>
<td>209</td>
<td>208</td>
</tr>
<tr>
<td>(0.1)</td>
<td>0.58</td>
<td>377</td>
<td>220</td>
<td>372</td>
</tr>
<tr>
<td>(0.3)</td>
<td>0.63</td>
<td>366</td>
<td>231</td>
<td>357</td>
</tr>
<tr>
<td>(0.5)</td>
<td>0.67</td>
<td>356</td>
<td>238</td>
<td>342</td>
</tr>
<tr>
<td>(0.7)</td>
<td>0.72</td>
<td>341</td>
<td>246</td>
<td>322</td>
</tr>
<tr>
<td>(0.9)</td>
<td>0.82</td>
<td>308</td>
<td>253</td>
<td>288</td>
</tr>
<tr>
<td>(0.95)</td>
<td>0.87</td>
<td>285</td>
<td>249</td>
<td>222</td>
</tr>
</tbody>
</table>
increase as the correlation increases, up to a point \( r^2 = 0.9 \), they do not increase sufficiently to surpass the profit obtainable from the single price that yields the same total supply of workers. A principle reason for this may be the increasingly large credential costs imposed on workers, which must be compensated by wages. These costs appear to account for the faster rate of increase in the profit from a single price yielding the same supply.

V. Conclusion

One cannot conclude from the examples presented that wage discrimination aimed at screening workers according to their reservation wages is always unprofitable; for, as we saw in the first introductory example, a bimodal distribution with strong correlation and a sufficiently large proportion of workers with low reservation wages and high credential costs can make discrimination profitable. However, these last examples emphasize that wage discrimination is definitely not a foregone conclusion. The features of these examples have been repeated in the variety of others we have calculated, leading to the conjecture that wage discrimination is invariably inferior in the case of a Normal distribution. The reason why is perhaps clear: a Normal distribution is unimodal and it has a relatively thin "tail", so the proportion of workers with low reservation wages and high credential costs is insufficient to justify the higher wages that must be paid to those workers with low credential costs who present ample credentials.

In the literature on screening and signaling, an implication of those results that derive from monopoly or monopsony in the formulation is that, in addition to the usual inefficiencies, monopolists impose signaling costs on others through their attempts to discriminate via wages or prices, or on themselves if (say) customers infer quality attributes of a product from prices or other observables. The implication of the examples above is that one must be cautious in drawing this conclusion: it may be that a monopolist or monopsonist foresees that discrimination is too costly to make it worthwhile compared to the alternative of foregoing discrimination altogether. In this respect, our conclusions parallel those reached by Rothschild and Stiglitz (1976) in the case of perfectly competitive insurance markets, for which they showed that the discriminating schedule may be Pareto-dominated by a nondiscriminating one. Similarly, in the case of job market signaling by workers with differing productivities, Spence showed that a nondiscriminating "pooling" equilibrium is Pareto superior.

The matter of wage discrimination has renewed relevance lately due to the current interest in "efficiency" wages that exceed reservation wages due to incentive effects; cf. Akerlof and Yellen (1986). For example, a wage premium provides workers an incentive to avoid shirking, if the consequence of detection is termination and uncertain chances for re-employment elsewhere. The main models assume identical workers; if workers are heterogeneous, however, then discriminating wage schedules could be offered as an alternative to the efficiency-wage hypothesis. That is, it could be argued that the differences between actual wages and imputed reservation wages reflect credentialing costs and the incentive-compatibility constraints imposed by workers' self-selection. A (strictly concave) wage schedule that discriminates among workers' reservation wages on the basis of credentials typically provides a wage that exceeds the sum of the reservation wage and the cost of the requisite credentials. Nevertheless, as we have seen, the precise form of this alternative hypothesis must depend on whether a discriminating wage schedule is profitable for the firm; if it is not, then a fixed wage is optimal and the only effect is the usual one that the wage is the lowest reservation wage among those workers accepting employment.

Lastly, it is worth mentioning a context in which discriminating schedules are frequently used. Many airlines, especially in the U.S. since deregulation of the industry, offer lower fares on (partially or entirely nonrefundable) tickets purchased in advance. The model developed here applies to this context if one interprets \( t \) as the number of days before departure that a ticket is purchased, \( p(t) \) as the airline's price for such a ticket (assuming zero marginal cost), and \( u - ct \) as the traveler's reservation price depending on the duration of advance purchase. In particular, \( c \) is the traveler's expected opportunity cost per day of early commitment to the departure date, and one supposes plausibly that \( u \) and \( c \) are positively correlated in the population of potential customers. The fact that such schedules are commonly observed and yet the model does not assuredly predict that they are more profitable than fixed fares, prompts a consideration of whether the formulation is accurate. Indeed, preliminary calculations indicate that a nonlinear formulation (e.g., the traveler's reservation price is \( uv^{-\alpha} \)) may suffice. Regarding discriminating wage schedules, this emphasizes that our conclusions may be sensitive to the assumed linearity of workers' net benefit functions. Linearity has been accepted in the literature as a plausible hypothesis since Spence first adopted it, but apparently there is no empirical evidence to establish its accuracy.

References


First version submitted May 1987;
final version received October 1987.