

Rat Races and Glass Ceilings: Career Paths in Organizations

Peter Bardsley
The University of Melbourne

Katerina Sherstyuk
The University of Melbourne and
The University of Hawaii.

November 2001

Abstract

In an ongoing organization, such as a large law partnership firm, employees are motivated not only by current rewards but also by the prospect of promotion, and the opportunity to influence policy and make the rules in the future. This leads to a dynamic programming problem in contract design.

We model career design in such a firm as a recursive mechanism design problem in an overlapping generations environment. Agents entering the firm may differ in their private characteristics which affect their costs of effort. We find that under recursive structure, a profit-maximizing principal offers, and promotion-motivated agents accept, “rat-race” contracts with very low wages and high effort levels. With wages driven down to zero, promotions become the main instrument to discriminate among agents in an adverse selection environment. The optimal adverse selection contract introduces a promotion barrier, or a “glass ceiling”, for the high cost agents. We thus find that the issues of inefficiently high work levels (the “rat-race”) and of unequal promotion rates (the “glass ceiling”) are intimately interconnected. We apply this framework to equal opportunity and gender discrimination in employment.

Keywords recursive contracts, mechanism design, overlapping generations, rat-race, glass ceiling.

1 Introduction

In an ongoing firm, employees are motivated not only by current rewards but also by the prospect of promotion. Promotion to a senior position brings with it the valuable opportunity to influence policy, and to make the rules in the future. In particular, it may bring the right to set the terms under which the next cohort of employees will work.

This process may be observed to some degree in many large organizations, but it is particularly clear in firms that are organized as professional partnerships; for example law firms, accountancy partnerships and management consultants. In these firms the career path leads towards ownership rights, and the ability to make management decisions unrestricted by external owners.

In such a firm, the employment contract is recursive. The contract that will be agreed today depends upon the prospects for promotion and the nature of the contracts that will be agreed tomorrow. This structure leads to a dynamic programming problem in contract design. Bardsley [3] has studied such a recursive principal-agent problem in an overlapping generations firm, assuming complete information and a homogeneous population. He found that the recursive structure typically induces an inefficient “rat-race,” with low wages and high effort levels.

In this paper we consider recursive contracts under adverse selection. Employees entering the firm have different private characteristics that affect the cost, to the employee, of exerting a given level of effort. We find that the contract imposed by the principal is characterised not only by a “rat-race,” with low wages and very high effort levels, but also by entry barriers and promotion barriers. In order to induce high effort levels from the top (low cost) type, and to discourage them from imitating the bottom (high cost) type, the principal will seek to extract all information rents from the bottom type. Since wages are already low and effort levels high, this is done by excluding the latter from employment altogether, or by reducing the probability that they will be promoted.

We use this framework to explore the implications of equal opportunity employment laws and conventions. Some employees (typically women) may have a higher outside value of time due to child care costs and household production opportunities¹, or some social groups may have suffered educational disadvantage. However the contract designer must remain blind to these characteristics. This may be due to explicit legal requirements, or it may be due to the force of social norms and conventions. We thus consider equal opportunity practice as the suppression of information, leading to an adverse selection problem. Under this adverse selection the recursive structure of the ongoing firm leads not only to a “rat-race,” but also to entry barriers and a “glass ceiling.” We thus find that the issues of high work levels (the “rat-race”) and of unequal promotion rates (the “glass ceiling”) are intimately interconnected. It should be emphasized that, although the outcome might be considered discriminatory, this “glass

¹Wood, Corcoran and Courant [19] provide evidence of the high private costs incurred by women in a large law firm.

ceiling” effect occurs without any overt discrimination. The contracts offered are completely neutral between types (in fact type is invisible to the principal when the contract is offered). Agents self select on the basis of their preferences. Curiously, we find that the suppression of information is welfare enhancing (at least relative to the very severe rat-race that would ensue under full information). However the benefit accrues entirely to the top type, typically men. The presence of the bottom type mitigates the severity of the “rat-race,” and the top type is able to earn some information rents.

The law firm is a natural laboratory in which to explore recursive contracts in an overlapping generations environment, and this is the example that we will have in mind. “Rat-races” and “glass ceilings” appear to be very prevalent in such firms. These firms are typically organized as an ongoing partnership with no outside equity. The characteristics of the large commercial law firm have been well documented in a number of studies, for example Carr and Matthewson [6], Feinberg [8], Ferrall [9], Galanter and Palay [11], Gilson and Mnookin [12], Landers Rebitzer and Taylor [14], O’Flaherty and Siow [15], and Spurr [18]. In the traditional law partnership the partners own the firm in equal shares and they face soft incentives (for example equal profit shares and permanent tenure). Associates in these firms face much harsher incentives, with an “up or out” career path. They earn a relatively low wage (compared with a partner’s profit share), and they are subject to work pressures that appear to be inefficiently high (Ferrall [9], Landers, Rebitzer and Taylor [14]; see also Akerlof [1]).

Traditionally, major law firms were an exclusively male preserve. The proportion of women associates has grown rapidly in recent years; however a glass ceiling seems to operate, and far fewer reach partnership (Spurr [18], Wood, Corcoran and Courant [19]). The issues of excessively high effort levels and of barriers to the promotion of women are very relevant in these firms (Albrecht, Björklund and Vroman [2], Blau and Kahn [5], Landers, Rebitzer and Taylor [14], Spurr [18]).

Staughton and Talmor [17] have recently considered managerial compensation as a mechanism design problem, but not in a recursive context. Sichernman and Galor [16] model career mobility with an explicit consideration of internal as well as external promotion opportunities. Several recent studies model the law firm as an overlapping generations structure, but not from a contract design perspective (see also Cremer[7]). Ferrall [9] considers a moral hazard problem, with a tournament model of promotions. Landers, Rebitzer and Taylor [14] present an interesting adverse selection model, where the selection pressure is driven by free riding between partners. We will discuss the relationship of our model with their work in more detail below.

From a theoretical perspective, the recursive mechanism design problem appears to be significantly more difficult than the standard non-recursive problem. The reason is that the incentive compatibility constraints interact quite strongly with the recursive structure of the model, making many standard arguments inapplicable. To solve the problem, we introduce a regularity assumption on the

technology which is stronger than the standard single crossing property.² In our view, the analysis of recursive adverse selection contracts presents an interesting addition to the mechanism design literature.

The structure of the paper is as follows. The basic model is set out in Section 2, and its properties are analyzed in Section 3. The optimal contract for the two-type case is derived in Section 4. In Section 5 we illustrate the findings with a numerical example and apply the model to explore the issue of discrimination and glass ceilings.

2 The Model

Consider a pure adverse selection model with no moral hazard involved. There are a finite number I of types of agents in the economy. The agent's type θ_i , $1 \leq i \leq I$, affects his cost of effort, as will be defined below. There are a finite number n_i of agents of type θ_i born each period, and each agent lives for two periods. The agent's type is private information. The set of types and the number of agents of each type is common knowledge; the total number of agents born each year is $n = \sum_i n_i$.

There is a single firm, owned in each period by a single principal. If hired by the firm, the agents get involved in a production activity which is costly for them. An agent of type θ_i who produces an output $b \in R_+$ exerts an effort $e_i \in R_+$, which is a deterministic function of the output level and the agent's type: $e_i = e(b, \theta_i)$. We assume that the effort is decreasing in type: if $\theta_i < \theta_j$, then $e(b, \theta_i) > e(b, \theta_j)$ for all $b > 0$. Each agent's output is ex-post observable, but the effort is not.

An agent of type θ_i who is born at date t , is offered, with probability $\sigma_{i,t}$, an employment contract $(w_{i,t}, \pi_{i,t}, b_{i,t})$ with the firm. This contract specifies a wage $w_{i,t}$, a promotion probability $\pi_{i,t}$, and an output level $b_{i,t}$. Promotion probability $\pi_{i,t}$ is the probability, conditional upon having been hired in the first place, that the agent will be the principal in the next period. The contract $(\sigma_{i,t}, b_{i,t}, w_{i,t}, \pi_{i,t})$ thus sets out how many agents of each type are hired, what they do, what they are paid, and who is promoted. The contract will be required to satisfy incentive compatibility constraints to induce truthful revelation. Unemployed agents, and agents who reject the contract, receive a reservation utility of 0 in both periods. Agents who are not promoted leave the firm and receive 0 in the second period. The agent who is promoted to become the principal in the second period exerts no effort and enjoys the firm's current profits net of the wages paid to the employees. We assume that the agent utility is separable between periods and there is no discounting; all agents are risk neutral.

For simplicity, we consider only time-stationary contracts. Thus the subscript t can be dropped. Bardsley [3] discusses both stationary and non-stationary recursive contracts in a complete information environment. He also discusses the effect of risk aversion and of congestible supervisory effort by the principals.

²See, e.g., Fudenberg and Tirole [10], page 259, on the single crossing, or sorting, conditions.

Here we assume stationarity, risk neutrality and an inactive principal in the interest of tractability, and in order to focus more clearly on the effect of adverse selection.

It is straight forward to re-interpret this as a model of a multi-principal partnership firm under the assumption that the partners exert no effort, and they share the partnership profits equally (for details, see Bardsley [3]). Thus, although they may be of different types, the multiple principals have no difficulty in agreeing on the objectives of the firm. We will not explore this interpretation any further here, and henceforth we will retain the assumption that the number of “partners” is 1.

Given the assumption of risk-neutrality and no discounting between periods, a type θ_i agent’s ex-ante (before employment contract is offered) expected utility from employment at the firm is his wage net of the cost of effort, plus the expected payoff W from becoming the principal in the next period, times the probability of being hired in the first place:

$$U(\sigma_i, w_i, \pi_i, b_i, W; \theta_i) = \sigma_i (w_i + \pi_i W - e(b_i, \theta_i))$$

Since the principal is not directly involved in production, his objective is simply to maximize his own current payoff W , which is the firm’s total output less the wages paid to the agents. An optimal contract maximizes the payoff to the principal subject to individual rationality and incentive compatibility constraints. The design problem is thus as follows.

Problem (P) Choose $w_i, \pi_i, \sigma_i, b_i, W$ to maximize W subject to

$$W \leq \sum_i n_i \sigma_i (b_i - w_i) \tag{1a}$$

$$\sigma_i (w_i + \pi_i W - e(b_i, \theta_i)) \geq 0 \tag{1b}$$

$$\sigma_i (w_i + \pi_i W - e(b_i, \theta_i)) \geq \sigma_j (w_j + \pi_j W - e(b_j, \theta_j)) \tag{1c}$$

$$b_i \geq 0 \tag{1d}$$

$$w_i \geq 0 \tag{1e}$$

$$0 \leq \pi_i \leq 1 \tag{1f}$$

$$0 \leq \sigma_i \leq 1 \tag{1g}$$

$$\sum_i n_i \pi_i \sigma_i = 1. \tag{1h}$$

In the above, (1a) is the budget constraint, (1b) and (1c) are the individual rationality and incentive compatibility constraints, respectively, (1d-1g) are feasibility constraints, and (1h) is the stationarity constraint.³ For convenience we will sometimes write IR_i to refer to the individual rationality constraint (1b),

³Due to recursive nature of the problem, and, specifically, the presence of stationarity constraint (1h), the standard argument of “no random contracts” (see, e.g., Laffont and Tirole [13], pages 119-120) does not apply here. Hence we explicitly incorporate employment probabilities into the model.

and IC_{ij} to refer to the incentive compatibility constraint (1c). As defined above, $n = \sum n_i$ is the population size; let $\bar{n} = \sum \sigma_i n_i$ denote the size of the firm, net of the principal.

We make the following assumptions about the population size and the technology.

Assumption 1 (no rare types) $n_i > 1$ for all i .

Assumption 2 (convexity of effort) The effort function is continuously differentiable at least twice, and

$$e(0, \theta) = 0 \tag{2a}$$

$$e_b(b, \theta) > 0 \tag{2b}$$

$$e_{bb}(b, \theta) > 0 \tag{2c}$$

$$\frac{e(b, \theta)}{b} \rightarrow \infty \quad (\text{as } b \rightarrow \infty) \tag{2d}$$

Assumption 3 (sorting conditions) If $\theta_i < \theta_j$ and $b > 0$, then

$$e_b(b, \theta_i) > e_b(b, \theta_j) \tag{3a}$$

$$e_{bb}(b, \theta_i) \geq e_{bb}(b, \theta_j) \tag{3b}$$

$$\frac{\partial}{\partial b} \left(\frac{e(b, \theta_i) - b}{e(b, \theta_i) - e(b, \theta_j)} \right) > 0. \tag{3c}$$

Assumption 4 (productive types) All types are potentially productive, where we say that a type θ is potentially productive if $e(b, \theta) < b$ for some $b > 0$.

Equation (3a) is the usual single crossing condition, which says that marginal effort is declining in type, while (3b) says that the convexity of the effort curve (which may be loosely interpreted as aversion to output uncertainty) is also declining in type. If we write $\gamma(b, \theta_i, \theta_j) = e(b, \theta_i) - e(b, \theta_j)$ for the cost advantage of type j over type i in performing task b , then our assumptions imply that this quantity is positive, strictly increasing and convex in b (we do not impose strict convexity). While the above assumptions are standard in the adverse selection literature, (3c) is an additional assumption we impose to solve the recursive model.⁴ Equation (3c) may be interpreted to say that although $\gamma(b, \theta_i, \theta_j)$ increases in b it does not grow too quickly; in fact, less quickly than $e(b, \theta_i) - b$. This assumption will be used below in proving Lemma 8 (Appendix A). For convenience we will write SC , SC' and SC'' for the sorting conditions (3a) to (3c). These sorting conditions hold for a range of plausible specifications of the effort function, including for example $e(b, \theta) = \frac{b^2}{\theta}$ and $e(b, \theta) = \frac{b}{\theta} + b^2$.

⁴Although assumption (3c) is not a necessary condition, we are unable to report on the contractual structures that may arise in the absence of this assumption. Finding a necessary and sufficient condition on the technology that would guarantee "sensible" solutions constitutes an interesting challenge for further research.

It may be helpful at this stage to compare this framework with the adverse selection model of Landers, Rebitzer and Taylor [14], which is quite different in structure. Their model is driven by free riding under equal sharing between partners, and high effort levels are imposed at the entry level to screen for good quality potential partners. They describe equilibrium contracts in which the control parameters are the wage and the effort level; promotion probabilities are held fixed, and are in fact set equal to 1. In contrast, we ignore free riding by partners, and focus instead on the dynamic inefficiencies that can arise under recursive contracting. We are interested in the way that the multiple job design instruments are combined, and we look for optimal contracts (assuming that there is a single firm with monopoly power in contract design) in which hiring probabilities, wages, effort levels and promotion probabilities are all set flexibly. It is interesting to note that, in contrast to Landers, Rebitzer and Taylor, we find that wages in fact play no role in the optimal contract, and that the compensation of agents for their effort occurs entirely through the promotion probability π (a conclusion that seems to be consistent with the stylized facts; see for example Galanter and Palay [11]). It is also interesting that both models, although so different in structure and assumptions, predict inefficiently high effort levels.

3 Basic Properties

We now turn to establishing some basic properties of the model. The approach that we will take is somewhat more roundabout than the way one would approach the standard adverse selection agency problem. The reason is that the recursive structure, which is here embedded in the stationarity constraint (1h), interacts with the incentive compatibility constraints. Consider for example the familiar proposition that all rent is extracted from the bottom agents. In the standard (non-recursive) model, the principal can simply reduce the wages of all types by a constant amount until this is so, and this will not upset any incentive compatibility constraints. In our model, it may happen that some wage is zero; in fact we will find that in the optimal contract all wages are zero. So it may be that the principal can only extract rent from an agent by increasing effort levels or by reducing the promotion probability. Either action may upset incentive compatibility. An increase in effort flows back to all agents as an increase in expected reward in the second period, but the distribution of this benefit is uneven, depending on promotion probabilities. And a change in promotion probabilities requires a complicated rebalancing of other elements of the contract, since equation (1h) imposes a sort of “balanced budget” condition in probabilities. We do in the end (in Section 4) show that all rent is extracted at the bottom, but it is not immediate, and we will need the non-standard condition (3c).

Lemma 1 *In an optimal contract $W > 0$. That is, under an optimal contract, the principal’s payoff is strictly positive.*

Proof. Let θ_1 be the lowest type, and choose b_1 such that $e(b_1, \theta_1) < b_1$. Consider the pooling contract under which all types are employed with certainty ($\sigma_i = 1$), all are paid the wage $w_i = e(b_1, \theta_1)$ and required to produce b_1 , and all are promoted with probability $\pi_i = \frac{1}{n}$. This contract is incentive compatible and individually rational, and it gives to the principal a strictly positive payoff. An optimal contract can only do better. ■

Lemma 2 *The budget constraint (1a) binds. That is, all surplus is distributed.*

Proof. Let $B = \sum_i n_i \sigma_i b_i - \sum_i n_i \sigma_i w_i - W$ be the budget surplus. Assume that $B > 0$. We will show that the principal can capture, in an incentive compatible way, a strictly positive part of this surplus.

For convenience we will make the temporary assumption that all types are employed, $\sigma_i > 0$ for all i , giving any agents of type θ_j for whom $\sigma_j = 0$ an equivalent null contract: $\sigma_j = 1$, $b_j = 0$, $\pi_j = 0$, $w_j = 0$, under which they are employed with certainty. This assumption will allow us to treat employed and unemployed types together, rather than by separate arguments.

We will construct a new contract, where we increase W by $\delta W > 0$ and we increase w_i by $\delta w_i \geq 0$, all other variables being kept fixed.

So choose δB such that $0 < \delta B \leq B$, and choose δW such that $0 < \delta W < \frac{\sigma_i w_i + \frac{\delta B}{n}}{\sigma_i \pi_i}$ for all i such that $\sigma_i \pi_i \neq 0$. This can be done, since there are only a finite number of types. We set $\delta w_i = \frac{\delta B}{\sigma_i n} - \pi_i \delta W$.

First we note that $\sigma_i (w_i + \delta w_i) = \sigma_i w_i + \frac{\delta B}{n} - \sigma_i \pi_i \delta W \geq 0$ by the choice of δW , so the new contract pays a non-negative wage to all types. Next we note that the change in the ex-ante return to type i is $\sigma_i \delta w_i + \sigma_i \pi_i \delta W = \frac{\delta B}{n}$, which is strictly positive and the same for all types. So the incentive compatibility and individual rationality constraints are not violated. Finally we note that the sum of the extra payments to the principal and all agents is $\delta W + \sum n_i \sigma_i \delta w_i = \delta B \leq B$, so the budget constraint is not violated. Thus the original contract could not have been optimal. ■

Remark 1 *For an unemployed type ($\sigma_i = 0$), the variables w_i , π_i and b_i can be chosen arbitrarily. We will now impose the convention that if $\sigma_i = 0$ then $w_i = \pi_i = b_i = 0$ (unemployed types are offered a null contract).*

Lemma 3 *Let i_0 be the lowest type such that $\sigma_{i_0} > 0$ (this is the lowest type that is offered an employment contract). If $i > i_0$ then $\sigma_i > 0$, and the individual rationality constraint IR_i is slack. That is, employment is sequential in type and the IR constraints do not bind, except perhaps at the bottom.*

Proof. The argument is standard. Applying the incentive constraint IC_{i_0} and the assumption that the effort is decreasing in type, we obtain:

$$\begin{aligned} \sigma_i (w_i + \pi_i W - e(b_i, \theta_i)) &\geq \sigma_{i_0} (w_{i_0} + \pi_{i_0} W - e(b_{i_0}, \theta_i)) \\ &= \sigma_{i_0} (w_{i_0} + \pi_{i_0} W - e(b_{i_0}, \theta_{i_0})) + \sigma_{i_0} (e(b_{i_0}, \theta_{i_0}) - e(b_{i_0}, \theta_i)) \\ &> 0. \end{aligned}$$

■

It is convenient to rewrite the constraints, eliminating the variables π_i and replacing them with a new variable $\rho_i = w_i + \pi_i W$, which is the agent's interim expected payoff conditional on being employed. Making this substitution we obtain an equivalent problem P' .

Problem (P') Choose $\rho_i, w_i, b_i, \sigma_i, W$ to maximize W subject to

$$W = \sum_i n_i \sigma_i (\rho_i - w_i) \quad (4a)$$

$$\sum_i n_i \sigma_i b_i \geq \sum_i n_i \sigma_i \rho_i \quad (4b)$$

$$\sigma_i \rho_i - \sigma_i e(b_i, \theta_i) \geq 0 \quad (4c)$$

$$\sigma_i \rho_i - \sigma_i e(b_i, \theta_i) \geq \sigma_j \rho_j - \sigma_j e(b_j, \theta_j) \quad (4d)$$

$$w_i \geq 0 \quad (4e)$$

$$0 \leq \rho_i - w_i \leq W \quad (4f)$$

$$0 \leq \sigma_i \leq 1 \quad (4g)$$

$$b_i \geq 0. \quad (4h)$$

By Lemma 2, the modified budget constraint equation (4b) could now be imposed as an equality. We choose not to do so at this stage, in order to sign some Kuhn Tucker multipliers below.

Theorem 1 *If $\sigma_i > 0$ then $\pi_i = 1$ or $w_i = 0$. If $\sigma_i = 1$ then $\pi_i < 1$ and $w_i = 0$. That is, all employed agents are paid zero wage except, possibly, those who are promoted with certainty.*

Proof. Assume that $\sigma_i > 0$, and that $w_i > 0$ and $\pi_i < 1$ (that is to say, $\rho_i - w_i < W$). Consider reducing w_i a little without violating either of these constraints, keeping b_i, ρ_i, σ_i and $w_j, j \neq i$, fixed (this is possible since $\pi_i < 1$). Then w_i does not occur in the three inequality constraints (4b, 4c, 4d), and by assumption the constraints (4e, 4f) are not binding. But by (4a) the objective W will increase. Thus either $w_i = 0$ or $\pi_i = 1$.

Now assume that $\sigma_i = 1$. Then

$$\begin{aligned} W &= \sum_i n_i \sigma_i (\rho_i - w_i) \\ &\geq n_i \sigma_i (\rho_i - w_i) \\ &= n_i \pi_i W. \end{aligned}$$

Thus $n_i \pi_i = 1$; but by Assumption 1, $n_i > 1$. So $\pi_i < 1$. Hence, from above, $w_i = 0$. ■

We interpret this Theorem to say that, provided it is feasible to do so, wages will always be driven down to the reservation level of zero. The intuition is that the current principal prefers to extract any surplus now, and to defer any

payment to the agents until a future period. The agents do not mind, since they are risk neutral and do not discount the future. The next principal is no worse off, since the liability may be moved forward yet again. Bardsley [3] explores this mechanism in some detail, finding that it is always feasible to drive the wage to zero provided that the expected future earnings of the firm are sufficient to compensate agents for their current effort⁵. We will see below that, in the model that we study here, wages will always be driven to zero.

Remark 2 *The proof of Theorem 1 shows that whenever $w_i > 0$ and $\pi_i < 1$ the principal can reduce w_i and increase his payoff. As he does so, π_i will rise. The process stops when either $w_i = 0$ or $\pi_i = 1$. Recall also (Lemma 2) that if the budget constraint does not bind then the principal can redesign the contract to share some or all of the surplus with the agents through an increase in W and in the wage w_i . If $\pi_i < 1$ for all types then these steps can be combined: some of the budget surplus can be redistributed, and then the whole of the transfer to the agents can be extracted by the principal using the technique of Theorem 1. This can go on until the condition $\pi_i < 1$ is violated for some type. We summarize by saying that if $\pi_i < 1$ for all types then the principal can, at the margin, appropriate all of the budget surplus. This observation will be useful below, in the proof of Lemma 8 (Appendix A).*

Lemma 4 *There exists a type i such that $\sigma_i = 1$. That is, there is at least one type of agents who are all employed.*

Proof. Assume that $\sigma_i < 1$ for all i . Thus there exists $k > 1$ such that $\tilde{\sigma}_i = k\sigma_i < 1$. If all the σ_i are replaced by $\tilde{\sigma}_i$ then none of the constraints (4b) to (4h) are violated but the return to the principal will increase by $(k - 1)W > 0$. Thus $\sigma_i = 1$ for some i . ■

Corollary 1 *The number of agents employed is strictly greater than the number of principals: $\bar{n} > 1$. Hence not every agent is promoted.*

Proof. It follows from the preceding Lemma and from Assumption 1 that $\bar{n} = \sum_i \sigma_i n_i > 1$. ■

Thus from Theorem 1 we have

$$w_i = \max(\rho_i - W, 0) \tag{5}$$

$$\pi_i = \min\left(\frac{\rho_i}{W}, 1\right) \tag{6}$$

$$(\rho_i - w_i) = \min(\rho_i, W) \tag{7}$$

We can now use these relations to eliminate the wages w_i , reformulating the problem a third and final time.

⁵Thus, for example, wages are driven to zero in a finite horizon version of the model, except in the last few periods before the terminal date. Bardsley also finds that the zero wage result holds up under moderate degrees of risk aversion, discounting and uncertainty about the future earnings of the firm.

Problem (P'') Choose ρ_i, b_i, σ_i, W to maximize W subject to

$$W = \sum_i n_i \sigma_i \min(\rho_i, W) \quad (8a)$$

$$\sum_i n_i \sigma_i b_i \geq \sum_i n_i \sigma_i \rho_i \quad (8b)$$

$$\sigma_i \rho_i - \sigma_i e(b_i, \theta_i) \geq 0 \quad (8c)$$

$$\sigma_i \rho_i - \sigma_i e(b_i, \theta_i) \geq \sigma_j \rho_j - \sigma_j e(b_j, \theta_i) \quad (8d)$$

$$0 \leq \sigma_i \leq 1 \quad (8e)$$

$$b_i \geq 0 \quad (8f)$$

$$\rho_i \geq 0 \quad (8g)$$

The principal's return W is now defined implicitly by (8a). This equation makes very explicit the recursive nature of the problem. To check that W is well defined, consider the function $\Gamma(W) = \sum_i n_i \sigma_i \min(\rho_i, W)$, holding fixed the parameters σ_i, ρ_i . It is easy to verify that Γ is concave, that $\Gamma(0) = 0$, that $\Gamma(W)$ is bounded, and that $\Gamma'(0) = \sum_i n_i \sigma_i = \bar{n} > 1$. Thus there is a unique positive solution to (8a).

4 The Two Type Case

We now specialize to the two type case. We first verify that basic properties of the standard non-recursive model still hold in our framework.

Theorem 2 *Assume that there are two types θ_1 , and θ_2 . Then*

1. *The individual rationality constraint IR_1 binds;*
2. *The individual rationality constraint IR_2 is slack unless the lower type is unemployed ($\sigma_1 = 0$);*
3. *The incentive compatibility constraint IC_{12} is slack; and*
4. *The incentive compatibility constraint IC_{21} binds.*

Curiously, the above properties, which are well-established in a non-recursive adverse selection framework, are quite difficult to obtain in our case; the proof is given in Appendix A. It is at this stage that we need the non-standard sorting assumption (3c) (see proof of Lemma 8 in Appendix A).

We now compute the optimal contract. We start with the complete information case as a benchmark, and then move to the hidden types case.

4.1 Complete Information Contract

In the complete information environment the contract is not constrained by incentive compatibility, so the IC constraints (8d) can be dropped. It is easy to

check that the individual rationality constraints (8c) now bind, for otherwise the agents could be required to work harder. Using these constraints to eliminate the variables ρ_i , the design problem is to maximize W subject to the constraints

$$n_1\sigma_1 \min(e(b_1, \theta_1), W) + n_2\sigma_2 \min(e(b_2, \theta_2), W) = W \quad (9a)$$

$$n_1\sigma_1 (b_1 - e(b_1, \theta_1)) + n_2\sigma_2 (b_2 - e(b_2, \theta_2)) \geq 0 \quad (9b)$$

$$0 \leq \sigma_i \leq 1 \quad (9c)$$

$$b_i \geq 0. \quad (9d)$$

Theorem 3 *Under complete information, the optimal contract is characterized by the following properties⁶*

1. *Full employment:* $\sigma_1 = \sigma_2 = 1$;
2. *Zero wages:* $w_1 = w_2 = 0$;
3. *Dissipation of all surplus (the rat-race condition):* $n_1e(b_1, \theta_1) + n_2e(b_2, \theta_2) = n_1b_1 + n_2b_2$;
4. *Equality of marginal effort:* $e_b(b_1, \theta_1) = e_b(b_2, \theta_2)$;
5. *High marginal effort:* $e_b(b_i, \theta_i) > 1$, for $i = 1, 2$;
6. *Output is monotonic in type:* $b_2 > b_1$.

Proof. The full employment result (1) follows from Lemma 9 in Appendix B. Next, since $\sigma_i = 1$ it follows from Theorem 1 that $w_i = 0$ (and also that $\pi_i < 1$ and $\rho_i < W$), for $i = 1, 2$. This demonstrates (2). Result (3) is just a restatement of the fact that the budget constraint (9b) binds (Lemma 2).

Further, using the fact that $\rho_i < W$ (and that $\sigma_i = 1$) for all i , we see that $W = n_1e(b_1, \theta_1) + n_2e(b_2, \theta_2)$. The Lagrangean can now be written

$$\begin{aligned} \mathcal{L} &= n_1e(b_1, \theta_1) + n_2e(b_2, \theta_2) \\ &+ \alpha (n_1b_1 + n_2b_2 - n_1e(b_1, \theta_1) - n_2e(b_2, \theta_2)), \end{aligned}$$

with $\alpha \geq 0$. Thus we have the first order conditions

$$\alpha + (1 - \alpha) e_b(b_1, \theta_1) \leq 0$$

$$\alpha + (1 - \alpha) e_b(b_2, \theta_2) \leq 0.$$

If $\alpha = 0$ then we would have $b_1 = b_2 = 0$, and hence $W = 0$ in contradiction to Lemma 1. It is then also clear that $\alpha \neq 1$. Thus the first order conditions hold with equality and we can eliminate the parameter α to show $e_b(b_1, \theta_1) = e_b(b_2, \theta_2)$; this is result (4).

⁶The theorem fully applies to a model with an arbitrary number of types, $1 \leq I < \infty$. The results are stated for the two-type case for an easier comparison with the incomplete information contract (Theorem 4 below).

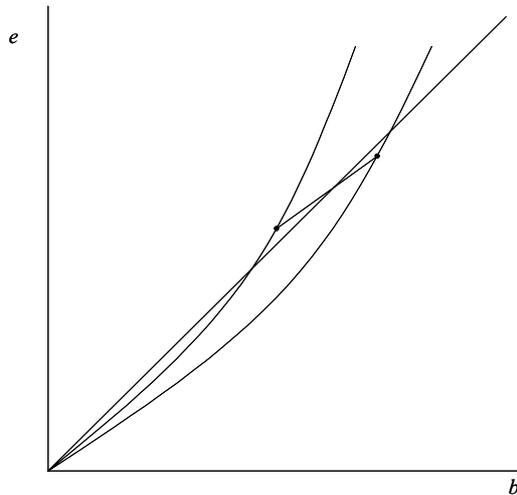


Figure 1: The full information contract. Effort curves $e(b, \theta_1)$, $e(b, \theta_2)$, and the 45° line are shown.

Let $\eta = e_b(b_1, \theta_1) = e_b(b_2, \theta_2)$ be the common slope of the effort curves at the two contract points. If $\eta \leq 1$ then, by the convexity of effort assumption (2c) the contract points would both lie strictly below the 45° line: $b_1 > e(b_1, \theta_1)$ and $b_2 > e(b_2, \theta_2)$; this would contradict result (3) above. Thus $\eta > 1$, which establishes result (5). Output monotonicity result (6) then follows from the single crossing assumption (3a). ■

The complete information contract is illustrated in Figure 1. Condition (3) of the above theorem tells us that the weighted average of the contract points $(b_1, e(b_1, \theta_1))$ and $(b_2, e(b_2, \theta_2))$ lies on the 45° line (if $n_1 = n_2$ then the weighted average is the midpoint; it always lies on the interval between the two contract points). Optimality requires that the weighted mean lie as high on the 45° line as is feasible; this is ensured by the marginal condition (4).

The figure also demonstrates that under complete information both types of agents are necessarily employed, with $b_1 > 0$, $b_2 > 0$. To see this, consider a pair of contract points $(\tilde{b}_1, e(\tilde{b}_1, \theta_1))$ and $(\tilde{b}_2, e(\tilde{b}_2, \theta_2))$ where marginal effort is equalized: $e_b(\tilde{b}_1, \theta_1) = e_b(\tilde{b}_2, \theta_2) = \eta$. If η is small then both of these points lie below the 45° line, and so does their midpoint⁷. If η is large then both of these points lie above the 45° line, and so does their midpoint. By continuity, there is an intermediate value of η for which the midpoint lies exactly on the 45° line. This is the optimal contract.

⁷For simplicity of exposition we assume that $n_1 = n_2$ so that the weighted average contract is the midpoint.

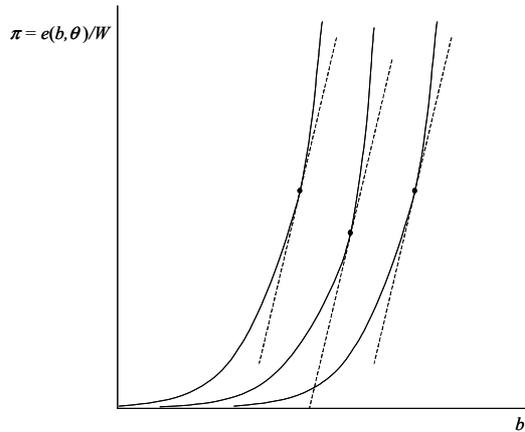


Figure 2: Non-monotonic promotion under full information. Contracts equalize marginal effort on three different effort curves.

Condition (5) of the theorem states that the marginal effort exceeds the marginal output for all types of agents; the optimal contract is thus inefficient and imposes overwork on all agent types.⁸ In fact, as the “rat-race” condition (3) clearly demonstrates, all social surplus is dissipated.

Output monotonicity is a consequence of equality of marginal efforts and the single crossing assumption (3a). Note, however, that output monotonicity does not necessarily imply reward (promotion probability) monotonicity. Under the optimal contract promotion probabilities are increasing in effort, hence all depends on the shape of the effort curve. Figure 2 illustrates. This has the consequence that, in this model, low ability types may be promoted more often than high ability types. Since principals do nothing, except design the contract and absorb the surplus, there is no productivity reason to promote high types (in contrast to Landers, Rebitzer and Taylor [14]). There may be good reasons to do otherwise, in order to induce high effort levels from high cost types. One can show that promotion probability is increasing in type provided that the effort function meets the following regularity assumption: $e_{\theta}(b, \theta) e_{bb}(b, \theta) \geq e_b(b, \theta) e_{b\theta}(b, \theta)$. This assumption is satisfied for many reasonable specifications

⁸We use ex-ante Pareto efficiency criterion to evaluate the contract efficiency. A contract is efficient if no generation of agents can be made better off without reducing the welfare of another generation, and no agent within a generation can be made better off without hurting someone else. Since the current principal’s utility is already accounted for as the utility of one of the agents in the previous generation, only the utilities of agents (potential employees) enter the efficiency criterion. Under the optimal stationary contract described above, all individual rationality constraints bind, and hence the ex-ante utility of every generation of agents is zero. It is easy to show that there exists a feasible contract that yields to every generation a positive utility (consider, for example, a pooling contract described in the proof of lemma 1); the inefficiency follows. See Bardsley [3] for a detailed treatment of inefficiency of the “rat-race” recursive contracts in the homogeneous agents world. The argument fully extends to the multiple type setting.

of the effort function, including for example $e(b, \theta) = \frac{b^2}{\theta}$ and $e(b, \theta) = \frac{b}{\theta} + b^2$.

4.2 Optimal Contract under Adverse Selection

The optimal contract under adverse selection is best described in terms of virtual effort functions

$$\begin{aligned}\varepsilon_1(b) &= e(b, \theta_1) + \frac{n_2}{n_1} \gamma(b) \\ \varepsilon_2(b) &= e(b, \theta_2),\end{aligned}$$

where

$$\gamma(b) \equiv e(b, \theta_1) - e(b, \theta_2)$$

is type θ_2 cost advantage over type θ_1 in producing output b . We note that, by assumptions (3a) and (3b), $\gamma(b)$ is positive for $b > 0$, increasing and convex. Hence, the virtual effort functions $\varepsilon_1(b)$, $\varepsilon_2(b)$ are positive for $b > 0$, increasing and convex as well.

Theorem 4 *The optimal contract is characterized by the following properties*

1. *Sequential employment: $\sigma_2 = 1$, but $\sigma_1 = 1$ or $\sigma_1 = 0$;*
2. *The size of the firm may be endogenously bounded;*
3. *Zero wages: $w_1 = w_2 = 0$;*
4. *Dissipation of all virtual surplus: $n_1 \varepsilon_1(b_1) + n_2 \varepsilon_2(b_2) = n_1 b_1 + n_2 b_2$;*
5. *Equality of marginal virtual effort: $\varepsilon_1'(b_1) = \varepsilon_2'(b_2) > 1$;*
6. *High marginal effort at the top: $e_b(b_2, \theta_2) = \varepsilon_2'(b_2) > 1$;*
7. *Output is monotonic in type: $b_2 > b_1$;*
8. *Positive information rents at the top, while all rents are extracted at the bottom. All information rents are in the form of increased promotion prospects.*

Proof. The sequential employment result (1) is established in Appendix C (lemmas 10 and 11). Now consider the Problem (P''). The IR_2 and IC_{12} constraints can be dropped (because they are either redundant (if $\sigma_1 = 0$) or slack (if $\sigma_1 > 0$), while the IR_1 and IC_{21} constraints bind. We can use these binding constraints to get rid of ρ_i 's (as in the proof of lemma 11, Appendix C). Further, using the fact that $\sigma_2 = 1$, and $\sigma_1 = 1$ or $\sigma_1 = 0$, and writing the model in terms of virtual effort rather than actual effort we have: Maximize W subject to

$$\begin{aligned}n_1 \sigma_1 \varepsilon_1(b_1) + n_2 \varepsilon_2(b_2) &= W \\ n_1 \sigma_1 (b_1 - \varepsilon_1(b_1)) + n_2 (b_2 - \varepsilon_2(b_2)) &\geq 0 \\ \sigma_1 &\in \{0, 1\} \\ b_i &\geq 0.\end{aligned}$$

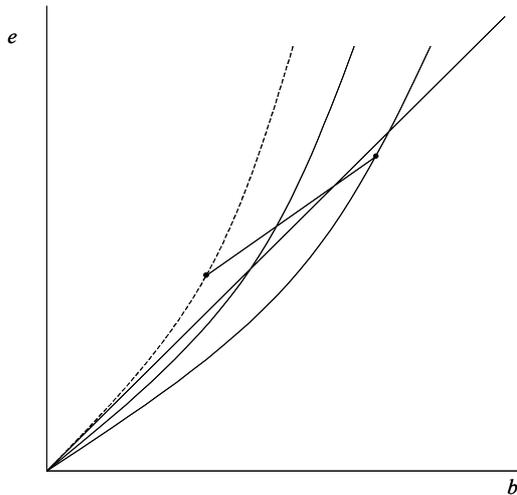


Figure 3: Possible non-existence of a full employment contract under adverse selection. The effort curves $e(b, \theta_1)$, $e(b, \theta_2)$, the virtual effort $\varepsilon(b, \theta_1)$ (dashed), and the 45° line are shown

We notice that this is almost exactly like the complete information model specified by constraints (9a)-(9d), with the result that $\sigma_2 = 1$, $\sigma_1 \in \{0, 1\}$ incorporated, and with virtual effort in place of actual effort. Now the virtual effort functions satisfy all of the Assumptions 2 and 3 made of the effort curves, with the possible exception of Assumption 3c (which is irrelevant here), and Assumption 4 (Productive Types). When we measure productivity in terms of virtual effort we cannot guarantee that type θ_1 is potentially productive, because the virtual effort curve may lie entirely above the 45° line (see Figure 3). Thus Assumption 4 may fail for type θ_1 , leading to $\sigma_1 = 0$ in an optimal contract. The example in Section 5.1 below shows that either $\sigma_1 = 1$ or $\sigma_1 = 0$ can occur. Thus information effects may bound the size of the firm, and types are employed sequentially from the top down. This demonstrates (2).

The proofs of (3) to (7) are exactly the same as the proof of Theorem 3, parts (2) to (6), using the virtual effort curves rather than the actual effort curves. Part (8) follows from parts (1) and (2) of Theorem 2, and from result (3). ■

The optimal contract is illustrated in Figure 3. Under adverse selection high type agents are hired before low type agents, and low type agents may not be hired at all, even though they are potentially productive. This is because hiring type θ_1 agents means that information rents must be paid to type θ_2 agents. Thus information effects may endogenously limit the size of the firm. It is interesting to note that there is a discrete jump in the hiring behavior.

If even one low type agent is hired then information rents must be paid to all higher agents, so if you hire one you might as well hire them all. It is clear from Figure 3 why there may be no full employment contract that meets the first order conditions. If the virtual effort curve $\varepsilon_1(b_1)$ lies above the 45° line, and if the contract points $(b_1, \varepsilon_1(b_1))$ and $(b_2, \varepsilon_2(b_2))$ are chosen to equalize marginal virtual cost, then it is possible that the weighted average of the contract points will lie always above the 45° line, which would violate the budget constraint.

It is also clear why $e_b(b_2, \theta_2) > 1$. For assume that $e_b(b_2, \theta_2) \leq 1$. If both types are employed then the contract point $(b_1, \varepsilon_1(b_1))$ must lie at a point on the virtual effort curve where $\varepsilon'_1(b_1) = \varepsilon'_2(b_2) = e_b(b_2, \theta_2) \leq 1$. But by convexity, marginal virtual effort exceeds average virtual effort, so $\frac{\varepsilon_1(b_1)}{b_1} < 1$. Thus the contract point must lie below the 45° line. This is impossible, by the same argument as in the complete information case. If only type θ_2 is employed then the contract point $(b_2, \varepsilon_2(b_2))$ must lie where the effort curve intersects the 45° line, and it is clear then that $e_b(b_2, \theta_2) = \varepsilon'_2(b_2) > 1$.

It remains true, even in the incomplete information case, that although output must be increasing in type it is possible that the promotion probability is not increasing in type. However the virtual effort curve $\varepsilon'_1(b_1)$ is steeper than the actual effort curve $e(b_1, \theta_1)$, so under adverse selection it is more likely that high types will be promoted more often. The reason is that under complete information the reward (promotion probability) need only compensate for costs, and cost may decline with type even though output increases. But in the adverse selection case the high type also earns information rents, so it is more likely that the reward increases with type.

5 Rat Races and Glass Ceilings

We have seen that in an overlapping generations firm with recursive contracts the principal will impose a “rat-race” contract. Wages are driven down to the reservation level of zero (see Theorem 1 and the following discussion). Effort will be driven up to inefficiently high levels; if there are no information constraints effort will be driven up to the point where the surplus is entirely dissipated (Theorems 3 and 4). As in most principal-agent models, all rent is extracted from the bottom. Since wages are already zero and effort is high, this can only be done by reducing the promotion probability of the bottom type (a “glass ceiling”). Thus “rat-race” and “glass ceiling” effects are intimately interconnected. In order to sustain very high effort levels at the top, the bottom type is driven to the participation margin by low promotion rates.

5.1 Example

We illustrate these features of the optimal contract with a numerical example. Let the effort functions be

$$\begin{aligned} e(b, \theta_1) &= \tau b + b^2 \\ e(b, \theta_2) &= b^2 \end{aligned}$$

where $0 < \tau < 1$. Let $\mu = \frac{n_1}{n_1+n_2}$ be the proportion of type θ_1 agents in the population.

It is straight forward to calculate that if $\mu < \frac{\tau}{2}$ then only the top type is employed ($\sigma_1 = 0, \sigma_2 = 1$), and the top type receives a pure rat-race contract that extracts all rent and dissipates all surplus: $b_2 = 1, e_b(b_2, \theta_2) = 2$ and $\frac{e(b_2, \theta_2)}{b_2} = 1$. If $\mu \geq \frac{\tau}{2}$ then there is full employment ($\sigma_1 = \sigma_2 = 1$), with output levels

$$\begin{aligned} b_1 &= \frac{\mu + \sqrt{\mu(\mu - 2\mu\tau + \tau^2)}}{2\mu} - \frac{\tau}{2\mu} \\ b_2 &= \frac{\mu + \sqrt{\mu(\mu - 2\mu\tau + \tau^2)}}{2\mu}. \end{aligned}$$

Population Mix (%)		Output		Average Effort		Marginal Effort	
Type θ_1	Type θ_2	b_1	b_2	$\frac{e(b_1, \theta_1)}{b_1}$	$\frac{e(b_2, \theta_2)}{b_2}$	$e_b(b_1, \theta_1)$	$e_b(b_2, \theta_2)$
10	90	-	1.00	-	1.00	-	2.00
20	80	-	1.00	-	1.00	-	2.00
30	70	0.00	1.00	-	1.00	0.60	2.00
40	60	0.17	0.92	0.77	0.92	0.94	1.84
50	50	0.26	0.86	0.86	0.86	1.12	1.72
60	40	0.32	0.82	0.92	0.82	1.23	1.63
70	30	0.35	0.78	0.95	0.78	1.30	1.56
80	20	0.38	0.75	0.98	0.75	1.35	1.50
90	10	0.39	0.72	0.99	0.72	1.38	1.45
100	0	0.40	-	1.00	-	1.40	-

Table 1: Output and Effort

Population mix (%)		Employment mix (%)		Partnership mix (%)	
Type θ_1	Type θ_2	Type θ_1	Type θ_2	Type θ_1	Type θ_2
10	90	-	100	-	100
20	80	-	100	-	100
30	70	30	70	0	100
40	60	40	60	8	92
50	50	50	50	20	80
60	40	60	40	34	66
70	30	70	30	49	51
80	20	80	20	65	35
90	10	90	10	82	18
100	0	100	0	100	0

Table 2: Employment Mix and Expected Partnership Mix.

Population Mix (%)		Promotion Probability		Information Rent		Payoff to Principal
Type θ_1	Type θ_2	Type θ_1	Type θ_2	Type θ_1	Type θ_2	W
10	90	-	$\frac{1}{n}$	-	0.00	$0.90n$
20	80	-	$\frac{1}{n}$	-	0.00	$0.80n$
30	70	0	$\frac{1.43}{n}$	0.00	0.00	$0.70n$
40	60	$\frac{.21}{n}$	$\frac{1.53}{n}$	0.00	0.10	$0.62n$
50	50	$\frac{.40}{n}$	$\frac{1.60}{n}$	0.00	0.16	$0.56n$
60	40	$\frac{.56}{n}$	$\frac{1.66}{n}$	0.00	0.19	$0.52n$
70	30	$\frac{.70}{n}$	$\frac{1.71}{n}$	0.00	0.21	$0.48n$
80	20	$\frac{.81}{n}$	$\frac{1.75}{n}$	0.00	0.23	$0.45n$
90	10	$\frac{.91}{n}$	$\frac{1.79}{n}$	0.00	0.23	$0.42n$
100	0	$\frac{1}{n}$	—	0.00	-	$0.40n$

Table 3: Promotion Probabilities and Expected Payoffs. (Recall that n is the size of the population of potential employees.)

The properties of this example are illustrated in Tables 1 to 3, for $\tau = 0.6$. Table 1 shows how output, average effort and marginal effort vary as the population mix changes. Table 2 shows how the population mix of types is reflected in the employed population mix and in the population mix of partners (since we have normalized the number of partners to 1, this should be interpreted as the expected number of partners of each type). Table 3 shows how the promotion probabilities, the information rents earned by each type, and payoff to the principal vary.

If the proportion of type θ_1 agents in the population is below the critical level of 30%, then the bottom type is completely excluded and only the top type is employed. These agents are offered a pure rat-race contract (Bardsley [3]) with a zero wage, and an effort level so high (the marginal effort of 2 is well above the efficient level of 1) that the surplus is totally exhausted (the average effort is 1). As a consequence, all ex ante rents are dissipated.

As the proportion of type θ_1 agents rises a threshold is crossed, and it suddenly becomes profitable to employ all of the type θ_1 agents as well as all of the type θ_2 agents. If even one type θ_1 agent is employed it is necessary to pay information rents to all of the type θ_2 agents. Below the threshold there are not enough potential type θ_1 employees to warrant doing this. Above the threshold, once the first type θ_1 agent is employed then these rents must be paid in full, so there is no reason not to employ them all.

As the proportion of type θ_1 agents rises beyond this critical level the population mix is immediately reflected in the employed population, since everybody is employed. However the number who make it through into partnership is quite small. For example, if the population is evenly balanced, with 50% of each type, only 20% of the partnership is of type θ_1 while 80% are of type θ_2 . By way of comparison, under complete information 47% would be of type θ_1 and 53% of type θ_2 . Even if the proportion of type θ_2 agents fell to 10%, they would remain grossly over-represented in the partnership at around 18% (under complete information this would only be 12%). We interpret this as a glass ceiling effect.

The reason for this glass ceiling lies in the extraction of rent. As in any standard adverse selection model, the principal extracts all rent from the bottom type, but there are only a limited number of instruments available with which to do so. In this model wages are immediately driven to 0 for everybody (Theorem 1). Rent cannot be extracted simply by making type θ_1 work harder, as this would not be incentive compatible: the high type would imitate the low type. So the only way to extract rent involves a squeeze on the promotion probability.

It is interesting to note that type θ_2 unambiguously benefits from the adverse selection effect and from the presence of type θ_1 agents in the population. This is a model where the suppression of information, and the adverse selection that then results, is Pareto improving for the agents relative to the (highly inefficient) complete information rat-race outcome.⁹ Under complete information all rents are dissipated and all agents receive an ex ante expected utility of 0. Under adverse selection some agents earn positive rents. As the proportion of type θ_1

⁹See footnote 8 on the discussion of the efficiency criterion.

agents increases, the type θ_2 agents are unambiguously better off. Their effort level decreases, and their promotion probability goes up. The value of being promoted (the principal's payoff) does decrease, because the average quality of the employment pool is declining, but this effect is outweighed by the other two, and the information rent increases.

5.2 Discrimination

We believe that this model throws light on questions of discrimination. Consider a complete information environment where there are two observable types. One is of lower productivity within the firm. For example, women may have higher expected effort costs due to child-bearing and child-rearing conventions; or minority groups may have language problems or they may have faced educational discrimination. As in the Section 1, for clarity of exposition we will refer to the two types as “women” and “men.” In a complete information environment we would expect to see full employment with discriminatory contracts.

If discrimination is impossible, either by law or by social convention, then these effectively become hidden types – no contract can be conditioned on type. Thus we model non-discriminatory behavior as the suppression of information. We then see some interesting possibilities. In the first place, there may be segregated firms which employ only “men,” even though there are potentially productive “women” in the population. So there is a threshold effect. Until the number of “women” in the population (in our example this would be the number of “women” law graduates entering the employment market) reaches a critical level, none are hired. Once the threshold is crossed, they are hired but they face a promotion barrier. There is an obvious coordination problem that may discourage any “women” from investing in legal qualifications.

Secondly, even if both types are employed we expect to see systematic differences in promotion rates. Very few “women” are promoted to the top of the organization. It should be emphasized that this is not a result of explicit discrimination. In fact, the employment arrangements and promotion policies are completely neutral and blind to the type of the agent; only the work that they do and the output that they produce are relevant. The apparently discriminatory outcomes are the result of self selection induced by the structure of the contracts.

In the mixed firm the presence of “women” mitigates the high rat-race effort levels. As a consequence, “men” now earn positive information rents. Thus in this model non-discrimination is welfare improving for the agents (relative to the highly inefficient complete information rat-race, which extracts all rents from everyone). The “women” are no worse off, while “men” are better off. It is curious that it is the top type that benefits from non-discriminatory hiring and promotion.

5.3 Conclusion

In conclusion, the model is consistent with two quite different types of firms: one which hires only men, and one which hires both men and women indiscriminately but which places promotion barriers in front of women. Employment in these firms is characterized by the interaction of a “rat-race” for men and a “glass ceiling” for women. The model also suggests that, because of the threshold effect and the coordination problem faced by women trying to break into the industry, the single-type structure in which only men are hired may be quite persistent. This would seem to be consistent with the historical facts (see Galanter and Palay [11], Spurr [18]). The large law firm was for a long time an exclusively male domain; now roughly equal numbers of men and women are hired, but relatively few women reach partnership.

These effects can be seen most clearly in a pure overlapping generations firm, such as the traditional law partnership, but we conjecture that they will operate, to a reduced degree, in any firm with an internal management ladder and processes that place some weight on internal management succession.

A Proof of Theorem 2

The conditions on the optimal contract may be summarized

$$n_1\sigma_1 \min(\rho_1, W) + n_2\sigma_2 \min(\rho_2, W) = W \quad (11a)$$

$$n_1\sigma_1 (b_1 - \rho_1) + n_2\sigma_2 (b_2 - \rho_2) \geq 0 \quad (11b)$$

$$\sigma_1\rho_1 - \sigma_1 e(b_1, \theta_1) \geq 0 \quad (11c)$$

$$\sigma_2\rho_2 - \sigma_2 e(b_2, \theta_2) \geq 0 \quad (11d)$$

$$\sigma_1\rho_1 - \sigma_1 e(b_1, \theta_1) \geq \sigma_2\rho_2 - \sigma_2 e(b_2, \theta_1) \quad (11e)$$

$$\sigma_2\rho_2 - \sigma_2 e(b_2, \theta_2) \geq \sigma_1\rho_1 - \sigma_1 e(b_1, \theta_2) \quad (11f)$$

$$0 \leq \sigma_i \leq 1 \quad (11g)$$

$$b_i \geq 0 \quad (11h)$$

$$\rho_i \geq 0 \quad (11i)$$

We know in addition that either $\sigma_1 = 1$ and $\rho_1 < W$, or $\sigma_2 = 1$ and $\rho_2 < W$, or both. We note that the result is clear if $\sigma_1 = 0$, in which case the optimal contract is given by $b_1 = \rho_1 = 0$, $b_2 = \rho_2 = b$, where $b = e(b_2, \theta_2)$, and $W = n_2 b_2$.

The result in the case that $\sigma_1 > 0$ will be established by way of a number of Lemmas in the remainder of this appendix. We now impose the assumption that $\sigma_1 > 0$, (which, by Lemma 3, implies that $\sigma_2 > 0$ as well).

Lemma 5 *The IR₂ constraint (11d) is slack.*

Proof. This follows from Lemma 3. ■

Lemma 6 *The IC₂₁ constraint (11f) binds.*

Proof. Otherwise we could increase b_2 without violating any of the constraints (11a) to (11i). This relaxes the budget constraint (11b), yielding a surplus that could be distributed in an incentive compatible manner (see Lemma 2). ■

Lemma 7 *A pooling contract can never be optimal. In fact, any contract with $\sigma_1 = \sigma_2$, $b_1 = b_2$ and $\rho_1 = \rho_2$ cannot be optimal.*

Proof. In any pooling contract we must have $\sigma_1 = \sigma_2 = 1$, $\rho_1 = \rho_2 < W$, $b_1 = b_2$ and $W = n_1\rho_1 + n_2\rho_2$. We will not at this stage impose all these constraints.

Let us look for an optimal contract under the assumptions $\sigma_1 = \sigma_2 = 1$, $\rho_1 < W$, $\rho_2 < W$. By Theorem 1, $w_1 = w_2 = 0$ and $\pi_i < 1$ for each i . It will be sufficient to show that in any such contract $b_1 \neq b_2$. The Lagrangean is

$$\begin{aligned} \mathcal{L} = & (1 - \alpha)(n_1\rho_1 + n_2\rho_2) + \alpha(n_1b_1 + n_2b_2) \\ & + \beta(\rho_1 - e(b_1, \theta_1)) \\ & + \delta_1(\rho_1 - e(b_1, \theta_1) - \rho_2 + e(b_2, \theta_1)) \\ & + \delta_2(\rho_2 - e(b_2, \theta_2) - \rho_1 + e(b_1, \theta_2)), \end{aligned}$$

yielding first order conditions

$$\begin{aligned}(1 - \alpha) n_1 + \beta + \delta_1 - \delta_2 &= 0 \\ (1 - \alpha) n_2 - \delta_1 + \delta_2 &= 0 \\ \alpha n_1 - (\beta + \delta_1) e_b(b_1, \theta_1) + \delta_2 e_b(b_1, \theta_2) &= 0 \\ \alpha n_2 + \delta_1 e_b(b_2, \theta_1) - \delta_2 e_b(b_2, \theta_2) &= 0.\end{aligned}$$

The multipliers α , β , δ_1 and δ_2 are non-negative.

Let us now impose the assumption $b_1 = b_2 = b$. The *IC* constraints (11e), (11f) imply that $\rho_1 = \rho_2 = \rho$, and the budget constraint (11b) implies that $\rho = b$. The first order conditions become

$$(1 - \alpha) n_1 + \beta + \delta_1 - \delta_2 = 0 \quad (12a)$$

$$(1 - \alpha) n_2 - \delta_1 + \delta_2 = 0 \quad (12b)$$

$$\alpha n_1 - (\beta + \delta_1) e_b(b, \theta_1) + \delta_2 e_b(b, \theta_2) = 0 \quad (12c)$$

$$\alpha n_2 + \delta_1 e_b(b, \theta_1) - \delta_2 e_b(b, \theta_2) = 0. \quad (12d)$$

Adding these together in pairs,

$$\begin{aligned}\beta &= (\alpha - 1) n \\ \alpha n &= \beta e_b(b, \theta_1).\end{aligned}$$

Since $\beta \geq 0$ we deduce from these equations that $\beta > 0$, $\alpha > 1$, $e_b(b, \theta_1) > 1$, and that

$$\begin{aligned}\alpha &= \frac{e_b(b, \theta_1)}{e_b(b, \theta_1) - 1} \\ \beta &= \frac{n}{e_b(b, \theta_1) - 1}.\end{aligned}$$

Substituting in (12b) and (12d),

$$\begin{aligned}\delta_2 - \delta_1 &= \frac{n_2}{e_b(b, \theta_1) - 1} \\ \delta_2 e_b(b, \theta_2) - \delta_1 e_b(b, \theta_1) &= \frac{n_2 e_b(b, \theta_1)}{e_b(b, \theta_1) - 1}.\end{aligned}$$

Thus

$$\delta_2 \frac{e_b(b, \theta_2)}{e_b(b, \theta_1)} - \delta_1 = \frac{n_2}{e_b(b, \theta_1) - 1} = \delta_2 - \delta_1 > 0.$$

Thus $\delta_2 e(b, \theta_2) = \delta_2 e(b, \theta_1)$. But $\delta_2 \neq 0$, since $\delta_1 \geq 0$, so $e(b, \theta_2) = e(b, \theta_1)$, which implies that $b = 0$. This is a contradiction, as it would imply that $W = 0$. ■

Lemma 8 *The IC_{12} constraint (11e) does not bind.*

Proof. We label the types θ_i and θ_j in such a way that $b_i \leq b_j$. If there is more than one contract that satisfies (11a) to (11i) and that yields the same payoff to the principal, we assume that a contract has been chosen such that b_j is minimal.

Let us assume that the constraints IC_{12} and IC_{21} both bind. The constraints IR_1 , IC_{12} and IC_{21} can then be written (we already know that $\sigma_i, \sigma_j \neq 0$)

$$\rho_i = h(b_i) + \frac{\Delta}{\sigma_i} \quad (13)$$

$$\rho_j = h(b_j) + \frac{\Delta}{\sigma_j} \quad (14)$$

$$\sigma_i \gamma(b_i) = \sigma_j \gamma(b_j), \quad (15)$$

where for convenience we write

$$\begin{aligned} h(b) &= e(b, \theta_1) \\ \gamma(b) &= e(b, \theta_1) - e(b, \theta_2). \end{aligned}$$

These are respectively the effort cost of type θ_1 and the cost advantage of type θ_2 over type θ_1 . The sorting condition (3a) implies that $\gamma(b)$ is strictly increasing. The constant $\Delta \geq 0$ measures the information rent of type θ_1 ; if the IR_1 constraint binds then $\Delta = 0$, but at this stage we do not know whether this is so. The budget constraint (11b) becomes:

$$n_i \sigma_i (b_i - h(b_i)) + n_j \sigma_j (b_j - h(b_j)) \geq n \Delta, \quad (16)$$

where $n = n_i + n_j$.

We may assume that $b_i < b_j$, for if $b_i = b_j$ then, by (13) to (15), we have $\sigma_i = \sigma_j$ and $\rho_i = \rho_j$, which is excluded by Lemma 7.

Since $\gamma(b)$ is strictly increasing, we note, from (15), that $\sigma_i > \sigma_j$. So we know by Lemma 4 that in fact $\sigma_i = 1$ and $\sigma_j < 1$. It then follows, from (13) and (14), that $\rho_i < \rho_j$. Also note that, by Theorem 1, $\sigma_i = 1$ implies $\rho_i < W$.

To show that the above cannot be optimal, we now consider the following contract. For type θ_i the contract (σ_i, b_i, ρ_i) is unchanged. For type θ_j we set $(\sigma_j, b_j, \rho_j) = (\sigma(b), b, \rho(b))$, where the functions $\sigma(b)$ and $\rho(b)$ are defined implicitly by

$$\begin{aligned} \sigma_i \gamma(b_i) &= \sigma(b) \gamma(b) \\ \rho(b) &= h(b) + \frac{\Delta}{\sigma(b)}. \end{aligned}$$

The constant Δ is held fixed. We set

$$\begin{aligned} B(b) &= n_i \sigma_i (b_i - h(b_i)) + n_j \sigma(b) (b - h(b)) - n \Delta \\ W(b) &= \begin{cases} \frac{n_i \sigma_i h(b_i) + n_i \Delta}{1 - n_j \sigma(b)} & \text{if } \rho(b) > W, \\ n_i \sigma_i h(b_i) + n_j \sigma(b) h(b) + n \Delta & \text{if } \rho(b) \leq W. \end{cases} \end{aligned}$$

These are respectively the budget surplus and the payoff to the principal under the modified contract.

Note that $\sigma(b)$ is decreasing in b by definition; it follows that $\rho(b)$ is increasing in b . We claim that for $b < b_j$ and sufficiently close to b_j this new contract meets all the constraints but yields an equal or higher payoff to the principal. It is clear by definition that the constraints IR_1 , IC_{12} and IC_{21} are met. We know that the IR_2 constraint is strictly slack, so by continuity it is satisfied for b close to b_j . We now consider two cases.

Case 1: $\rho_j > W$

We will show that the budget surplus strictly improves under the new contract, and the principal's share improves weakly. By redistributing some of this surplus (as in the proof of Lemma 2) the principal can be made better off.

To show that the budget constraint is satisfied if $b < b_j$ it is sufficient to show that $B(b)$ is a decreasing function of b . Apart from a constant term, $B(b)$ is just

$$\begin{aligned}\sigma(b)(b - h(b)) &= A \frac{b - h(b)}{\gamma(b)} \\ &= -A \frac{e(b, \theta_1) - b}{e(b, \theta_1) - e(b, \theta_2)}\end{aligned}$$

where $A = \sigma_i(e(b_i, \theta_1) - e(b_i, \theta_2))$ is a positive constant that does not depend on b . Thus it is sufficient to check that $\frac{e(b, \theta_1) - b}{e(b, \theta_1) - e(b, \theta_2)}$ is increasing in b . But this is Assumption 3, condition (3c).

We must also show that $W(b)$ is decreasing in b . We know that $\rho_i < W$ (Theorem 1). Since $\rho(b)$ is increasing in b , for b smaller than but close to b_j , we have $\rho_i < W < \rho(b) < \rho_j$. So $W(b)$ is defined by

$$W(b) = \frac{n_i \sigma_i h(b_i) + n_i \Delta}{1 - n_j \sigma(b)}$$

Then $W(b)$ is decreasing in b , since $\sigma(b)$ is decreasing in b .

Case 2: $\rho_j \leq W$

First note, by Assumption 2, condition (2a), and by Assumption 3, conditions (3a) and (3b), that $\gamma(0) = 0$, $\gamma'(b) > 0$, and $\gamma''(b) \geq 0$. Thus $0 = \gamma(0) \geq \gamma(b) - b\gamma'(b)$, which implies that $\frac{b}{\gamma(b)}$ is weakly decreasing in b . Thus the total surplus

$$\begin{aligned}B(b) + W(b) &= n_i \sigma_i b_i + n_j \sigma(b) b \\ &= n_i \sigma_i b_i + n_j \sigma_j \gamma(b_j) \frac{b}{\gamma(b)}\end{aligned}$$

is weakly decreasing in b . As b is reduced a little below b_j the total surplus $B(b) + W(b)$ will not decrease, and it may increase.

It is not immediately clear that such a change is beneficial to the principal, since $W(b)$ could decrease even if $B(b) + W(b)$ increases. However, $\rho_i < W$ and $\rho(b) < \rho_j \leq W$ so by Remark 2 the principal can redesign the contract to appropriate fully any marginal increase in the budget surplus $B(b)$. After appropriating this surplus the principal is no worse off, and may be better off, but the output level of the high output (type θ_j) agents has been reduced. This contradicts the minimality of b_j . ■

Corollary 2 *The constraint IR_1 binds.*

Proof. We have shown that the constraints IR_2 constraint (11d) and the IC_{12} constraint (11e) do not bind. If IR_1 does not bind then we can increase b_1 , holding all other parameters of the contract fixed, without violating any incentive compatibility or individual rationality constraints. But this relaxes the budget constraint, yielding a surplus that can be distributed as in Lemma 2. ■

B Proof of Theorem 3(1)

The following lemma shows that, in an economy with an arbitrary number $I \geq 1$ of agent types, an optimal contract under complete information is characterized by full employment. Result (1) of Theorem 3 then follows.

It is convenient to return to the original Problem (P) with the constraints given by (1a)-(1h).

Lemma 9 (full employment) *In the absence of incentive constraints, all agents are employed: $\sigma_i = 1$ of all i .*

Proof. Consider Problem (P) without the constraints (1c). It is clear that the individual rationality constraints (1b) bind under an optimal contract, for otherwise the agents could be required to work harder, and the principal's payoff W could be increased. We now show that in an optimal contract, if $\sigma_i > 0$, then $\sigma_i = 1$, for all $i \leq I$. Suppose not, i.e., there is type θ_i which is offered a contract $(\sigma_i, b_i, w_i, \pi_i)$ with $0 < \sigma_i < 1$. Now consider an alternative contract, in which all other types' contracts do not change, and type θ_i is offered a modified contract $(\tilde{\sigma}_i, \tilde{b}_i, \tilde{w}_i, \tilde{\pi}_i)$, with $\tilde{\sigma}_i = 1$, $\tilde{b}_i = \sigma_i b_i$, $\tilde{w}_i = \sigma_i w_i$, and $\tilde{\pi}_i = \sigma_i \pi_i$. Since, by Lemma 2, the budget constraint (1a) was binding before the modification, we observe that the payoff W to the principal remains unchanged under the modified contract; all feasibility constraints and the stationarity constraint (1h) also hold. Checking the individual rationality constraint:

$$\tilde{\sigma}_i \left(\tilde{w}_i + \tilde{\pi}_i W - e \left(\tilde{b}_i, \theta_i \right) \right) = \sigma_i (w_i + \pi_i W) - e(\sigma_i b_i, \theta_i) > \sigma_i (w_i + \pi_i W - e(b_i, \theta_i)),$$

where the inequality follows from the convexity of effort assumption (2c). This shows that the IR_i constraint is slack and hence this contract can be improved upon in a feasible manner. Thus, the original contract could not be optimal.

It is left to show that in an optimal contract $\sigma_i > 0$ for all i . Suppose not, that is, $\sigma_i = 0$ for some type θ_i . This type is potentially productive (Assumption 4). Consider a new contract in which this type works productively and is compensated solely through a wage $w_i = e(b_i, \theta_i)$; the surplus goes to the principal. This relaxes the individual rationality constraint for the other types, so the new contract is feasible and it is better for the principal. ■

C Proof of Theorem 4(1)

The result follows from the following two lemmas:

Lemma 10 (*Full employment at the top*) *In an optimal contract under adverse selection, $\sigma_2 = 1$.*

Proof. Consider the original Problem (P) with the constraints given by (1a)-(1h). Let $I = 2$. We know, from Theorem 2, that IR_1 , IC_{21} bind, and IR_2 , IC_{12} are strictly slack. If $\sigma_1 = 0$ under an optimal contract, then we are in the one-type world, and the result follows trivially. Now suppose that $\sigma_1 > 0$, and $\sigma_2 < 1$. To prove that this contract cannot be optimal, we show that we can improve on the principal's payoff by substituting a small number of type θ_1 agents by type θ_2 agents, without violating any of the constraints. Pick $\Delta\sigma > 0$ small, with $\Delta\sigma \leq \min\{\sigma_1, \frac{n_2}{n_1}(1 - \sigma_2)\}$, and consider the following modified contract: $\tilde{\sigma}_1 = \sigma_1 - \Delta\sigma$, $\tilde{\sigma}_2 = \sigma_2 + \frac{n_1}{n_2}\Delta\sigma$, $(\tilde{b}_1, \tilde{w}_1, \tilde{\pi}_1) = (b_1, w_1, \pi_1)$, and $(\tilde{b}_2, \tilde{w}_2, \tilde{\pi}_2) = \alpha(b_2, w_2, \pi_2) + (1 - \alpha)(b_1, w_1, \pi_1)$, where $\alpha \equiv \frac{\sigma_2 n_2}{\sigma_2 n_2 + \Delta\sigma n_1}$. It is easy to check (as in the proof of Lemma 9) that principal's payoff W remains the same as in the original contract, and the budget constraint, all feasibility constraints and the stationarity constraint still hold. Next notice that, since the effort function is convex and decreasing in type, this contract increases the utility of type θ_2 agents:

$$\begin{aligned} & \tilde{\sigma}_2(\tilde{w}_2 + \tilde{\pi}_2 W - e(\tilde{b}_2; \theta_2)) > \\ & > \tilde{\sigma}_2\left(\alpha \frac{\sigma_2(w_2 + \pi_2 W - e(b_2; \theta_2))}{\sigma_2} + (1 - \alpha) \frac{\sigma_1(w_1 + \pi_1 W - e(b_1; \theta_1))}{\sigma_1}\right) = \\ & = \sigma_2(w_2 + \pi_2 W - e(b_2; \theta_2)) + \frac{\Delta\sigma n_1}{\sigma_1 n_2} \sigma_1(w_1 + \pi_1 W - e(b_1; \theta_1)) > \\ & > \sigma_2(w_2 + \pi_2 W - e(b_2; \theta_2)) \end{aligned}$$

The utility of type θ_1 agents is unchanged. Hence IC_{21} is now slack, and IR_1 , IR_2 still hold. The type θ_2 contract becomes more attractive to type θ_1 after the modification, but since IC_{12} was slack under the original contract, by continuity it will still hold under the modified contract for $\Delta\sigma$ small enough. Since IC_{21} is now slack, this contract can be further improved upon to increase W (as in the proofs of lemma 6). Thus the original contract could not be optimal. ■

Lemma 11 *In an optimal contract under adverse selection, either $\sigma_1 = 1$ or $\sigma_1 = 0$.*

Proof. Consider the Problem (P''), with the constraints given by (8a) to (8g). By Theorem 2, we know that IR_1, IC_{21} bind. We can use these binding constraints to get rid of ρ_i 's:

$$\begin{aligned}\rho_1 &= e(b_1, \theta_1), \\ \rho_2 &= e(b_2, \theta_2) + \frac{\sigma_1}{\sigma_2} \gamma(b_1),\end{aligned}$$

where $\gamma(b) \equiv e(b, \theta_1) - e(b, \theta_2)$. We note that, by assumptions (3a) and (3b), $\gamma(b)$ is positive for $b > 0$, increasing and convex.

By Lemma 10, $\sigma_2 = 1$, and hence, by Theorem 1, $\rho_2 < W$. Thus we can rewrite the principal's problem as to maximize W subject to:

Assume that $0 < \sigma_1 < 1$. By Remark 1 we have $b_1 > 0, b_2 > 0$.

First consider the case $\rho_1 \leq W$, which implies that $\min(e(b_1, \theta_1), W) = e(b_1, \theta_1)$. Then we can rewrite the problem in terms of virtual effort functions:

$$\begin{aligned}\varepsilon_1(b) &= e(b, \theta_1) + \frac{n_2}{n_1} \gamma(b) \\ \varepsilon_2(b) &= e(b, \theta_2),\end{aligned}$$

where $\gamma(b)$ is as defined above. Note that the virtual effort curves satisfy all of the Assumptions 2 made on the effort curves.

The problem's constraints are

$$n_1 \sigma_1 \varepsilon_1(b_1) + n_2 \varepsilon_2(b_2) = W \quad (17a)$$

$$n_1 \sigma_1 (b_1 - \varepsilon_1(b_1)) + n_2 (b_2 - \varepsilon_2(b_2)) \geq 0 \quad (17b)$$

$$0 \leq \sigma_1 \leq 1 \quad (17c)$$

$$b_i \geq 0. \quad (17d)$$

By our supposition, constraints (17c) and (17d) do not bind. The Lagrangean is

$$\begin{aligned}\mathcal{L} &= n_1 \sigma_1 \varepsilon_1(b_1) + n_2 \varepsilon_2(b_2) \\ &+ \beta (n_1 \sigma_1 b_1 + n_2 b_2 - n_1 \sigma_1 \varepsilon_1(b_1) - n_2 \varepsilon_2(b_2)) \\ &= (1 - \beta) (n_1 \sigma_1 \varepsilon_1(b_1) + n_2 \varepsilon_2(b_2)) + \beta (n_1 \sigma_1 b_1 + n_2 b_2)\end{aligned}$$

with $\beta \geq 0$. Then

$$\begin{aligned}\mathcal{L}_{b_1} &= (1 - \beta) n_1 \sigma_1 \varepsilon_1'(b_1) + \beta n_1 \sigma_1 = 0 \\ \mathcal{L}_{\sigma_1} &= (1 - \beta) n_1 \varepsilon_1(b_1) + \beta n_1 b_1 = 0.\end{aligned}$$

It is clear that $\beta \neq 0, 1$. So $b_1 \varepsilon_1'(b_1) = \varepsilon_1(b_1)$, or marginal virtual cost equals average virtual cost. But this is impossible if $b_1 > 0$, given the virtual effort function is increasing, convex and $\varepsilon_1(0) = 0$.

Now consider the case $\rho_1 > W$, which implies that $W = n_1\sigma_1W + n_2\rho_2$. The Lagrangean may be written

$$\begin{aligned}\mathcal{L} &= W \\ &+ \alpha (W - n_1\sigma_1W - n_2\sigma_1\gamma(b_1) - n_2e(b_2, \theta_2)) \\ &+ \beta (n_1\sigma_1b_1 + n_2b_2 - n_1\sigma_1e(b_1, \theta_1) - n_2\sigma_1\gamma(b_1) - n_2e(b_2, \theta_2))\end{aligned}$$

with $\beta \geq 0$. Then

$$\mathcal{L}_W = 1 + \alpha(1 - n_1\sigma_1) = 0 \quad (18)$$

$$\mathcal{L}_{b_1} = \beta n_1\sigma_1(1 - e_b(b_1, \theta_1)) - (\alpha + \beta)n_2\sigma_1\gamma'(b_1) = 0 \quad (19)$$

$$\mathcal{L}_{b_2} = n_2(\beta - (\alpha + \beta)e_b(b_2, \theta_2)) = 0 \quad (20)$$

$$\mathcal{L}_{\sigma_1} = n_1(\beta(b_1 - e(b_1, \theta_1)) - \alpha W) - (\alpha + \beta)n_2\gamma(b_1) = 0 \quad (21)$$

We note first that, from (18), $\alpha \neq 0$ and hence, from (20), $\beta \neq 0$. Further, from (20), we obtain that since $\beta > 0$, then $(\alpha + \beta) > 0$. Next we sign α . Since $W = n_1\sigma_1W + n_2\rho_2$, we have $W(1 - n_1\sigma_1) = n_2\sigma_1\gamma(b_1) + n_2e(b_2, \theta_2)$. Multiplying by α and using (18), $-W = \alpha(n_2\sigma_1\gamma(b_1) + n_2e(b_2, \theta_2))$, so $\alpha < 0$. Now we know that $\mathcal{L}_{b_1} = 0$ and, by assumption, $\mathcal{L}_{\sigma_1} = 0$, so $\frac{1}{n_1\beta\sigma_1b_1}(b_1\mathcal{L}_{b_1} - \sigma_1\mathcal{L}_{\sigma_1}) = 0$. Setting $f(b) = e(b, \theta_1) + \frac{\alpha + \beta}{\beta} \frac{n_2}{n_1} \gamma(b)$, this equation can be written as $\frac{f(b_1)}{b_1} - f'(b_1) + \frac{\alpha W}{\beta b_1} = 0$. Thus $f'(b_1) < \frac{f(b_1)}{b_1}$. But $f(0) = 0$ and f is strictly convex, so $f'(b) > \frac{f(b)}{b}$ for $b > 0$, which provides us with a contradiction. ■

References

- [1] Akerlof, George A., The Economics of Caste and of the Rat Race and Other Woeful Tales, in *An Economic Theorist's Book of Tales: Essays that Entertain the Consequences of New Assumptions in Economic Theory*, Cambridge, New York and Sydney: Cambridge University Press, 1984, pages 23-44. Previously published: [1976].
- [2] Albrecht, J., A. Björklund and S. Vroman, Is there a Glass Ceiling in Sweden? SOLE Working Paper, 2001.
- [3] Bardsley, P. Recursive Contracts, University of Melbourne Working Paper, 2001. Electronic version at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=275231
- [4] Beckman, Martin J., Hierarchy vs. Partnership, *Journal of Economic Behavior and Organization*, 5(2), June 1984, pages 237-45.
- [5] Blau, F. D. and L. M. Kahn, Gender Differences in Pay, *Journal of Economic Perspectives*, 14(4), 2000, 75-99.
- [6] Carr, Jack, Mathewson, Frank, The Economics of Law Firms: A Study in the Legal Organization of the Firm, *Journal of Law and Economics*, 33(2), October 1990, pages 307-30.
- [7] Cremer, Jacques, Cooperation in Ongoing Organizations, *Quarterly Journal of Economics*, 101(1), February 1986, pages 33-49.
- [8] Feinberg, Robert M., Paralegals and Associate Lawyers: Substitutability within the Law Firm, 1977-87, *Review of Economics and Statistics*, 76(2), May 1994, pages 367-72.
- [9] Ferrall, Christopher, Promotions and Incentives in Partnerships: Evidence from Major U.S. Law Firms, *Canadian Journal of Economics*, 29(4), November 1996, pages 811-27.
- [10] Fudenberg, Drew, Tirole, Jean, *Game Theory*, Cambridge, Massachusetts, and London: The MIT Press, 1992.
- [11] Galanter, Marc, Palay, Thomas, *Tournament of lawyers: The transformation of the big law firm*, Chicago and London: University of Chicago Press, 1991.
- [12] Gilson, Ronald J., Mnookin, Robert H., The Implicit Contract for Corporate Law Firm Associates: Ex Post Opportunism and Ex Ante Bonding, in Aoki, Masahiko, Gustafsson, Bo, Williamson, Oliver E., eds., *The firm as a nexus of treaties*, Swedish Collegium for Advanced Study in the Social Sciences series, London, Newbury Park, Calif. and New Delhi: Sage, 1990, pages 209-36.

- [13] Laffont, Jean-Jacques, Tirole, Jean, *A Theory of Incentives in Procurement and Regulation*, Cambridge, Massachusetts, and London: The MIT Press, 1993.
- [14] Landers, Renee M., Rebitzer, James B., Taylor, Lowell J., Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms, *American Economic Review*, 86(3), June 1996, pages 329-48.
- [15] O'Flaherty, Brendan, Siow, Aloysius, Up-or-Out Rules in the Market for Lawyers, *Journal of Labor Economics*, 13(4), October 1995, pages 709-35.
- [16] Sicherman, Nachum, and Oded, Galor, A Theory of Career Mobility, *Journal of Political Economy*, 98(1), February 1990, pages 169-92.
- [17] Stoughton, Neal M., Talmor, Eli, Managerial Bargaining Power in the Determination of Compensation Contracts and Corporate Investment, *International Economic Review*, 40(1), February 1999, pages 69-93.
- [18] Spurr, Stephen J. Sex Discrimination in the Legal Profession: A Study of Promotion, *Industrial and Labor Relations Review*, 43(4), April 1990, pages 406-17.
- [19] R. G. Wood, M. E. Corcoran, and P. N. Courant, Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyer's Salaries, *Journal of Labour Economics*, 11(3), 1993, 417-441.