Recursive Contracts

Peter Bardsley

The University of Melbourne.

June 2001

Abstract

We model career design as a recursive contract design problem in an overlapping generations firm. Agents live two periods. In period 1 they may be hired as employees, paid a wage, and produce output. In period 2 they may be promoted to become joint owners (partners) of the firm, producing no output directly, but setting the rules and receiving the residual income. Professional partnerships, such as the traditional law firm, are often organized like this.

Employees are motivated not only by the wage but by the possibility of promotion to the partnership, and the opportunity to set the rules in the next period; the reward structure is thus recursive. The contracts that emerge in this environment are always inefficient. In many circumstances the inefficiency takes the form of "rat-race" contracts that specify very low wages and an inefficiently high level of effort.

This conclusion seems to be robust to a range of variations in the environment.

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

1 Introduction

Firms are ongoing organizations. Individuals join them when they are young, hoping to make a career. As their career develops, they hope to move to positions of greater responsibility and autonomy, to gain greater rewards, and to exert some influence over the direction of the firm.

This process can be observed in all 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 (or at least to contribute to) management decisions, unrestricted by external owners. In these firms the employment contract is recursive: the reward for one cohort derives, at least in part, from the right to choose the contract under which the next cohort will work. This leads to a dynamic programming problem in contract design.

Career design and dynamic (but not recursive) contracts have been studied quite extensively; see for example Holmstrom [10], Gibbons and Murphy [8], and Meyer and Vickers [12]. Work norms and career design in an overlapping generations firm have been discussed by Cremer [4] and, in the context of law firms, by Carr and Matthewson [3], Gilson and Mnookin [9], Galanter and Palay [7], O'Flaherty and Siow [13], Ferrall [5], and Landers, Rebitzer and Taylor [11]. Staughton and Talmor [14] have recently considered managerial compensation as a mechanism design problem, but not in a recursive context.

One theme of this literature is the "up or out" nature of typical contracts, and the emphasis on low wages and deferred rewards that are paid only to the successful. Another is the pervasiveness of what appear to be inefficiently high work levels, or "rat-race" contracts (see also Akerlof [1]).

In this paper we will focus on the nature of recursive contracts, which we view as a contract design problem in an overlapping generations environment. We consider here the simplest case, with full information and identical agents. This will allow us to isolate the recursive contracting structure from any effects of imperfect information. The results are thus free from any adverse selection or moral hazard effects. In particular, the source of the rat-race that we will find is distinct from that identified by Landers, Rebitzer and Taylor [11] (which is driven by unobservable shirking within partnerships). Recursive contracts under adverse selection are discussed by Bardsley and Sherstyuk [2].

In a recursive environment, there is an ambivalence in the attitude of agents. On the one hand, as in any principal-agent relationship, the agent prefers a soft contract with low effort and high rewards. But on the other hand, the agent may well become the principal in the next period, and can see the virtues of a harsher contract, especially if this or a similar contract can be imposed on the next generation of agents. We find that, as a result of this ambivalence, in many circumstances principals will design and agents will accept inefficient contracts that specify low wages and an inefficiently high level of effort. This conclusion seems to be robust to a range of variations in the environment.

The structure of the paper is as follows. In Section 2 we set out the basic, risk-neutral model, under both finite horizon and steady state assumptions. One of the implications of risk neutrality is that the firm may grow very large, even though this may impose a highly risky contract on the agents. In Section 3 we explore the implications of risk aversion, and also the effect of congestion costs associated with increasing the principal's span of control. Finally, in Section 4 we discuss some implications of the results.

2 The Basic Model

There are n agents born each period, and each agent lives for two periods. Utility is separable between periods and there is no discounting; all agents are risk neutral.

There is a single firm. An agent who is born at date t is, with probability σ_t, offered an employment contract (w_t, π_t, b_t) with the firm. This contract specifies a wage w_t, a promotion probability[1] π_t, and an output level b_t. 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. Partners exert no effort, and they share the partnership profits equally. Thus they have no difficulty in agreeing on the objectives of the firm.

To simplify notation, write $w = w_t$ for the current period wage, $w_+ = w_{t+1}$ for the next period's wage, and $w_- = w_{t-1}$ for the previous wage. A similar notational convention will be used for other variables.

Let

$$W = n\sigma (b - w) \tag{1}$$

be the current partnership profits (in period t). Let

$$m = n\sigma_- \pi_- \tag{2}$$

be the number of partners in the firm in period t. Agents will accept the employment contract provided that it satisfies the individual rationality constraint

$$w + \pi \left(\frac{W_+}{m_+} \right) \geq e(b), \tag{3}$$

where $e(b)$ is the cost of exerting effort. We make the following assumptions about the effort function:

Assumption 1 *The effort function is smooth, increasing, convex, and* $e(0) = 0$.

Assumption 2 *There exists an output level* b *such that* $b > e(b)$; *(otherwise the optimal production level will always be* 0*).*

Since the partners are identical and share equally, they will choose (σ, w, π, b) to maximize the total partnership profit (1), subject to the participation constraint (3) and the feasibility constraints $w \geq 0$, $b \geq 0$, $0 \leq \sigma \leq 1$, and $0 \leq \pi \leq 1$. After eliminating w and π the constraints can be rewritten $\frac{W}{n} \leq \sigma b$, $\frac{W}{n} \leq \sigma (b - e(b)) + \frac{W_+}{n}$, $b \geq 0$, and $\frac{m_+}{n} \leq \sigma \leq 1$; so the principals' problem is to maximize

$$\frac{W}{n} = \min \left[\sigma b, \sigma (b - e(b)) + \frac{W_+}{n} \right]$$

subject to $b \geq 0$, and $\frac{m_+}{n} \leq \sigma \leq 1$.

We note that a solution to this maximization problem always exists, since σ is restricted to a compact interval, and $\sigma (b - e(b)) + \frac{W_+}{n}$ is coercive in b (it is negative outside a compact interval, so we can restrict b to this interval). It is also clear that the individual rationality constraint always binds, for if it did not then $\frac{W}{n}$ could be made arbitrarily large, violating the existence conclusion.

[1] Explicit randomization in promotion is unlikely to be seen, but in practice the uncertain nature of the business environment does the job just as well. When promotion time comes round nobody is sure exactly who will be promoted, as it depends on a variety of uncertain factors. But everybody has a good idea of who is likely to get up and who is not; and sometimes there are surprises.

2.1 Preliminary Remarks

We show first that, under these assumptions, all agents will be employed.

Lemma 1 (full employment) $\sigma = 1$

Proof. We consider separately three cases, showing in each case that W is increasing in σ.

If $W = n\sigma b > n\sigma(b - e(b)) + W_+$ then W is clearly increasing in σ.

If $W = n\sigma(b - e(b)) + W_+ > n\sigma b$, then b maximizes $(b - e(b))$. But by Assumption 2 this function takes non-negative values. Thus $(b - e(b)) > 0$. It is also clear that this optimal value of b does not depend on σ. Thus W is increasing in σ.

Finally, if $W = n\sigma(b - e(b)) + W_+ = n\sigma b$, then we can write $b = b(\sigma)$ and $W = W(\sigma)$ as functions of σ, and we have $W(\sigma) = n\sigma b$ and $W_+ = n\sigma e(b(\sigma))$. Writing the derivative as a dot, $\dot{W} = n\sigma\dot{b} + n\dot{\sigma}b$ and $0 = n\sigma e'(b)\dot{b} + n\dot{\sigma}e(b)$, so $\dot{W} = n\dot{\sigma}b\left(1 - \frac{e(b)}{be'(b)}\right)$. So W is increasing in σ provided that the effort elasticity $\frac{be'(b)}{e(b)} > 1$. That is to say, provided that the marginal effort exceeds the average effort; but this follows from Assumption 1. ■

Bardsley and Sherstyuk [2] show that if there are hidden types then this result may no longer be true. Adverse selection and information effects may endogenously limit the size of the firm.

Corollary 1 *The contract is invariant to the size of the partnership.*

Proof. Since the constraint $\frac{m_+}{n} \leq \sigma$ does not bind, the contract is independent of the value of m_+. ■

The intuition is straight forward. Firm profits depend only on the actions of the agents. For a given future level of profit W_+, the profit per partner is $\frac{W_+}{m_+}$, and the probability of promotion is $\frac{m_+}{n\sigma}$, so the expected future profit per employee is $\frac{W_+}{n\sigma}$. Doubling the number of partners halves the profitability of partnership but doubles the probability of becoming partner. To risk neutral agents this is the same.

Thus, under the assumptions made here, it is always profitable to employ all agents, but the size of the partnership is indeterminate. There may be a single principal ($m = 1$), with the probability of promotion being $\frac{1}{n}$, or there may be many principals ($m = n$), with the probability of promotion being 1.

We can also interpret an m partner firm as an amalgam of m single principal firms, with the span of control of each principal being $\frac{n}{m}$. Given this interpretation, the partnership size indeterminacy is quite natural: since the principals do nothing productive (apart from designing the contract) the span of control is immaterial. So without loss of generality, we may set $m_t = 1$ for all t, and we may assume that there is a single principal. The results may then be reinterpreted in the multi-principal case by regarding n as the span of control of a representative principal[2].

[2] In a model where the partners exert some effort, partnership size might be limited by free riding between partners (See Landers, Rebitzer and Taylor [11] for a model that focuses on

2.2 Efficiency

We now show that recursive contracts are always inefficient. The concept of efficiency that we will use is ex-ante Pareto efficiency. A contract is efficient if no generation can be made better off without reducing the welfare of another generation. Using the notation set out above, the ex-ante expected utility of the current generation is (after eliminating w and π)

$$\begin{aligned} u &= \sigma \left(w + \pi W^{+} - e(b) \right) \\ &= \sigma \left(b - e(b) \right) + \frac{\left(W^{+} - W \right)}{n}. \end{aligned}$$

It is worth noting explicitly that the utility of the current principal plays no direct part in the efficiency criterion because his or her utility has already been accounted for in the utility of agents of the previous generation. The feasibility constraints on the contract are $b \geq 0$, $0 \leq \sigma \leq 1$, and $0 \leq W \leq n\sigma b$.

We first establish some baseline results that can be used to assess efficiency. It is natural to conjecture that a feasible contract (σ_t, b_t, W_t) will be efficient provided that there is full employment ($\sigma_t = 1$), that there is efficiency in production in every period ($e'(b_t) = 1$). We will confirm this in the two cases that are of interest to us. We consider first the case where it is known that the firm will cease to exist after a certain date T.

Lemma 2 (finite horizon efficiency criterion) *Assume that the firm will cease to exist after a finite date T, that $\sigma_t = 1$ for all $t \leq T$, that $e'(b_t) = 1$ for all $t \leq T$, and that the contract is feasible. Then it is efficient.*

Proof. Assume that there exists a Pareto improvement $(\Delta\sigma, \Delta b, \Delta W)$. Then $\Delta u = \Delta\left[\sigma\left(b - e(b)\right)\right] + \frac{\left(\Delta W^{+} - \Delta W\right)}{n} \geq 0$, with a strict inequality holding for some generation. But (σ, b) already maximizes $\sigma\left(b - e(b)\right)$ period by period, so $\Delta\left[\sigma\left(b - e(b)\right)\right] \leq 0$. Thus $\Delta W^{+} \geq \Delta W$. So the change ΔW is non-decreasing through time. But after the terminal date $\Delta W = 0$, since then W is constrained to be zero. Thus $\Delta W = 0$ at all dates. It then follows that $\Delta u = \Delta\left[\sigma\left(b - e(b)\right)\right] \leq 0$, so the variation cannot have been a Pareto improvement. ■

We note in passing that the condition in this Lemma is sufficient but not necessary. One can find optimal allocations that do not satisfy this criterion.

We now consider stationary allocations, imposing the constraint $\sigma^{+} = \sigma$, $b^{+} = b$, $W^{+} = W$. The ex-ante utility of a representative generation is then $u = \sigma\left(b - e(b)\right)$.

Lemma 3 (infinite horizon efficiency criterion) *Consider the stationary contract (σ, b, W). If $\sigma = 1$, $e'(b) = 1$ and $0 \leq W \leq nb$ then the contract is efficient subject to the stationarity constraint.*

---

free riding).

Proof. Under stationarity all generations are treated identically, so a Pareto optimal allocation is simply one that maximizes $\sigma(b - e(b))$ subject to the feasibility constraints $b \geq 0$, $0 \leq \sigma \leq 1$, and $0 \leq W \leq n\sigma b$. From the convexity of $\sigma(b - e(b))$ it follows that this will happen when the constraint $0 \leq W \leq n\sigma b$ does not bind, $\sigma = 1$, and $b - e(b)$ is maximized. ■

We will refer to the contract $(\sigma = 1,\ W = 0,\ e'(b) = 1)$, with full employment, no intergenerational transfers, and efficient production in each period, as the infinitely repeated myopic contract.

Corollary 2 *Under either the finite horizon assumption, or the constraint of stationarity, the infinitely repeated myopic contract is efficient. Furthermore, it yields every agent a strictly positive ex-ante expected utility.*

Corollary 3 *The recursive contract will always be inefficient.*

Proof. Since the individual rationality constraint always binds, the ex-ante utility of every generation is zero. But by the preceding corollary, there exists a feasible contract that yields to every generation a positive utility. ■

The intuition is straight forward. The current principal has monopoly power to impose a contract on the next generation, and naturally chooses to extract all the surplus. However the previous principal was in the same position. Going back through time, everybody loses out.

2.3 Finite Horizon Contracts

We now show how the inefficiency of recursive contracts may take the form of a "rat-race," with very low (in fact zero) wages, and excessively high effort levels.

The best contract for the principal to offer at time t depends on the expected partnership profit W_+ at time $t+1$. If agents are optimistic about the future of the firm then deferred payment is attractive; but this may not be so if the firm is declining. Here we consider the finite horizon case and compute the optimal contract by backward induction.

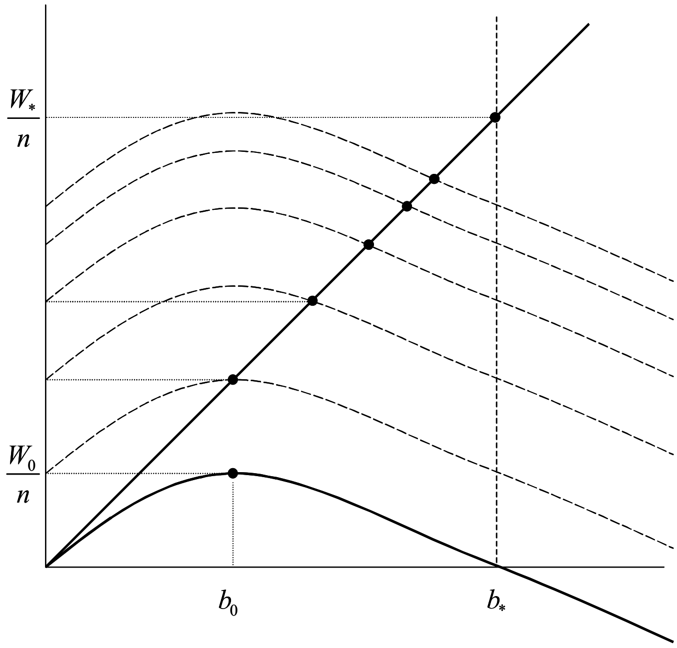

Figure 1. Finite Horizon Contracts

Consider the function $\phi(b) = b - e(b)$, which is shown in Figure 1. The assumptions on the effort curve imply that $\phi(b)$ is strictly concave, that its graph passes through the origin with a slope that is strictly positive but less than 1, and that it is eventually negative. Thus it crosses the axis at a point b$^* > 0$, and it reaches a maximum at a point b_0 such that $0 < b_0 < b^*$. The output level b_0, where marginal effort equals 1, is the efficient output level that would be chosen by a conventional profit maximizing firm. For simplicity of exposition we will call b_0 the *efficient point*[3]. We will call b_*, where average effort equals 1, the *rat-race point*.

The principal's problem is to maximize $\frac{W}{n}$ subject to the constraints $\frac{W}{n} \leq b$ (the non-negativity constraint on w), and $\frac{W}{n} \leq \phi(b) + \frac{W_+}{n}$ (the individual rationality constraint). Figure 1 shows graphs of these constraints. Notice $\phi(b) + \frac{W_+}{n}$ shifts vertically up or down as W^+ varies, so its maximum remains fixed at b_0. We also note that the functions b and $\phi(b) + \frac{W_+}{n}$ cross at the point where $b = e^{-1}\left(\frac{W_+}{n}\right)$, or equivalently where $\frac{W_+}{n} = e(b_0)$. There are thus two regimes, depending upon the size of the future partnership profit W^+.

[3] If $b = b_0$ in all periods then the contract is efficient in the sense of Section 2.2. However we note that the converse is not true. There may exist Pareto optimal contracts in which $b > b_0$ at some dates, due to binding of the $w \geq 0$ constraint.

Regime 1: $\frac{W^+}{n} < e(b_0)$**.** In this regime the constraint $\frac{W}{n} \leq b$ does not bind, so the principal chooses b to maximize $\phi(b)$. Output b is set at the efficient level b_0, and the wage paid is $w = b_0 - \frac{W}{n} = e(b_0) - \frac{W_+}{n}$, which is strictly positive. The partnership profit is $W = n(b_0 - e(b_0)) + W_+ > W_+$. Thus partnership profits are declining through time, which has the consequence that once the firm enters this regime it will not leave it. The wage paid to the agents is increasing through time. The output level $b = b_0$ does not change, and remains fixed at the efficient level.

Regime 2: $\frac{W^+}{n} \geq e(b_0)$**.** In this regime both the constraints bind, so the contract is pushed away from the efficient production point b_0. Since the constraint $\frac{W}{n} \leq b$ binds, $w = 0$ and the wage is fixed at zero. In this regime all rewards are deferred and agents are compensated for their effort only by the expectation of promotion. The output level is $b = e^{-1}\left(\frac{W_+}{n}\right) \geq b_0$, so output is inefficiently high. Since the wage is zero, $\frac{W}{n} = b = e^{-1}\left(\frac{W_+}{n}\right)$ so $\frac{W_+}{n} = e\left(\frac{W}{n}\right)$ and $b_+ = e(b)$. This allows us to explore the dynamics of the contract. If $b = b_*$ then $b^+ = e(b_*) = b_* = b$, so the contract is stationary. If $b > b_*$ then $b_+ > b$, so output and partnership profits are growing. If $b < b_*$ then $b_+ < b$, so output and partnership profits are declining; this is the only type of trajectory that is consistent with the terminal conditions.

In both regimes we have, by the results of Section 2.1, full employment. So $\sigma = 1$ and the promotion probability $\pi = \frac{1}{n}$ is constant through time and across both regimes.

We can now get a clear idea of the sequence of contracts by proceeding backwards from the terminal period. Let the terminal horizon be at time 0. Since the firm no longer exists after this date, $W_+ = 0$ and we are in regime 1, with b_0 at the efficient level, $\frac{W_0}{n} = b_0 - e(b_0) > 0$, $w_0 = e(b_0) > 0$ and $\pi_0 = \frac{1}{n}$; in fact the value of π_0 is irrelevant in this terminal period.

In the previous period the individual rationality constraint shifts upwards. Let us assume (as in Figure 1) that we are still in regime 1; this will be so provided that $\frac{W_0}{n} = b_0 - e(b_0) < e(b_0)$, that is to say $\frac{e(b_0)}{b_0} > \frac{1}{2}$. Then output remains at the efficient value, the partnership profit increases to $\frac{W_{-1}}{n} = (b_0 - e(b_0)) + \frac{W_0}{n} = 2(b_0 - e(b_0))$, and the wage falls to $w_{-1} = e(b_0) - \frac{W_0}{n} = 2e(b_0) - b_0$.

Moving backwards in time, partnership profit $\frac{W}{n}$ increases until we enter the second regime, where the constraint $\frac{W}{n} \leq b$ binds and the wage falls to zero. The output b rises above the efficient level, tending in the distant past to the "rat-race" level b_*. Partnership profit also rises as we move back in time. Agents exert an inefficiently high level of effort, they receive a zero wage, and they are compensated purely by the expectation that they may be promoted.

As noted in Corollary 3, this contract is inefficient. It is instructive to consider exactly why the principal chooses such an inefficient outcome[4]. Let us start first with the myopic efficient contract of Corollary 2, which is just the terminal contract repeated infinitely often. Production is always at the efficient level b_0, the agent receives the whole of his or her output as wage, and there are no intergenerational transfers. Beginning from this baseline, each principal would like to extract some rent. There are two instruments that can be used, w and b. The most straight-forward way to proceed is to reduce the wage w; by doing so some of the surplus can be extracted without upsetting production efficiency[5]. However there is a limit to what can be done, because of the non-negativity constraint on wages[6]. But the principal is concerned not with ex-ante efficiency but with the ex-post pay-off, conditional on having been promoted. Having no other instrument to use, he or she is quite happy to push up effort levels even at the cost of production efficiency.

We summarize as follows.

Proposition 1 *Assume that the firm will cease to exist after a finite date T. Let N be the largest integer such that $\frac{\mathbf{N}-1}{\mathbf{N}} < \frac{e(b_0)}{b_0}$. Then*

1. *In all except the last* $\mathbf{N}$ *periods the wage is fixed at zero and effort is inefficiently high. In the distant past the output level b tends to the "rat-race" level b_*; it declines through time towards the efficient level b_0.*
2. *In the final* $\mathbf{N}$ *periods output is fixed at the efficient level b_0, and the wage w is positive and increasing through time.*
3. *The contract is inefficient, yielding an ex-ante expected utility of zero to every generation.*

2.4 Steady State Contracts

As the terminal horizon recedes into the future the finite horizon contract approaches a steady state. So we now consider steady state contracts, under the assumption that the firm may cease to exist with probability α at any time. This can also be considered as a model of an infinitely lived firm with discounting.

The contract design problem is now to maximize $W = n\,(b - w)$ subject to the constraints

---

[4] I would like to thank the members of the theory group at the University of Sydney for a very lively seminar and dinner where this point was clarified.

[5] We do not consider here the question of allocative inefficiency. See the next section, where allocative effects are seen more easily.

[6] Without a constraint that imposes a lower bound on wages no equilibrium would exist. The model becomes a Ponzi game: the current generation can extract any surplus that they wish, passing the cost back, generation by generation, into the distant past.

$$n\pi = 1 \tag{4}$$
$$w + (1-\alpha)\,\pi W \geq e\,(b) \tag{5}$$
$$w \geq 0 \tag{6}$$
$$0 \leq \pi \leq 1. \tag{7}$$

We have imposed the stationarity constraints $W = W_+$, $w = w_+$, $\pi = \pi_+$, set $\sigma = 1$ and normalized the number of partners to 1; the justification for this is similar to that set out previously.

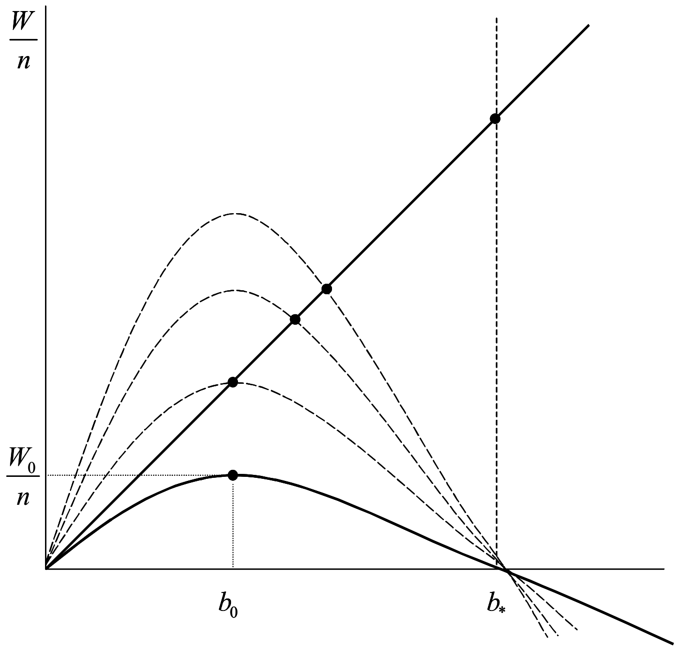

Figure 2. Steady State Contracts

Eliminating π, w, and once again writing $\phi\,(b) = b - e\,(b)$, we find that

$$\frac{W}{n} = \max_{b \geq 0} \min \left[b, \frac{\phi\,(b)}{\alpha} \right]$$

in the case $\alpha > 0$, and to

$$W = \max\left\{W : W \leq b, \phi\,(b) \geq 0\right\}$$

in the case $\alpha = 0$. This last case leads immediately to the pure rat-race contract $b = b_*$, $w = 0$, $\pi = \frac{1}{n}$, and requires no further discussion.

Once again, there are two regimes, depending on how confident agents can be about the future.

Regime 1: $\alpha > \frac{\phi(b_0)}{b_0}$. In this regime in this regime the risk of failure is high. The constraint $\frac{W}{\alpha} \leq b$ does not bind, so the principal chooses b to maximize $\phi(b)$. Thus b is set at the efficient level b_0, and a strictly positive wage $w = b_0 - \frac{\phi(b_0)}{\alpha}$ is paid.

Regime 2: $0 < \alpha \leq \frac{\phi(b_0)}{b_0}$. In this regime it is very likely that the firm will continue into the next period, and the constraint $\frac{W}{\alpha} \leq b$ binds. The wage is zero, with the agents compensated only through the expectation of promotion. Output b is inefficiently high, being characterized by the average effort condition $\frac{e(b)}{b} = 1 - \alpha > \frac{e(b_0)}{b_0}$. As $\alpha \to 0$ the contract tends to the pure rat-race contract with $b = b_*$.

In both regimes there is full employment and the promotion probability is $\pi = \frac{1}{n}$. Once again, the outcome is inefficient.

As before, it is instructive to see why the principal is prepared to choose such an inefficient outcome. Consider first the efficient contract $b = b_0$, $w = b$, $W = 0$. The principal can improve his or her pay-off by reducing the wage w, keeping output at $b = b_0$. This does not disturb production efficiency, but it does lead to allocative inefficiency, at least provided that $\alpha > 0$. To see this, note that the ex-ante welfare of a representative generation can be written

$$u = b - e(b) - \alpha \pi W.$$

Reducing the wage shifts income forward between generations, and incurs the cost of doing so. In the current model, this cost arises because of the risk that the firm might go out of business before the transfer is realized in the next period. The non-negativity constraint on w means that reducing the wage alone is not enough to extract all the surplus. So, as discussed in the finite horizon model, the principal will also push up the effort level beyond b_0.

We summarize as follows.

Proposition 2 *Let α be the failure probability.*

1. *If $\alpha \leq \frac{\phi(b_0)}{b_0}$ then the wage w is zero, and output b is inefficiently high, tending to the pure rat-race level b_* as α tends toward zero.*

2. *If $\alpha > \frac{\phi(b_0)}{b_0}$ then the output level $b = b_0$ is efficient and a positive wage is paid; the wage is increasing in α.*

3. *In all cases the contract is inefficient, there is full employment ($\sigma = 1$) and $\pi = \frac{1}{n}$.*

3 Risk aversion and congestion

In this section we investigate the robustness of the conclusions that we have drawn, by introducing some complicating factors into the basic model.

In both of the models considered so far, there is full employment ($\sigma = 1$ and $\pi = \frac{1}{n}$) and the firm can grow to be very large. One consequence is that the promotion probability of any one individual may become very small, although the ultimate reward may be great. Another is that the span of control $\frac{1}{\pi}$ may become very large. In this section we investigate the effect of risk aversion, and of congestion or span of control costs. We are interested in whether these effects may provide an endogenous limit on the size of the firm, and whether they may mitigate the "rat-race" nature of contracts.

3.1 Risk Aversion

Introducing risk aversion into the models as already specified has a very simple consequence. There is no intrinsic uncertainty in these models; all uncertainty is introduced by the contract through random promotion. As discussed in Section 2, the size of the partnership is indeterminate in these risk neutral models. Uncertainty can be completely eliminated by setting $m = n$ and promoting everybody. That is to say, instead of giving all of the deferred reward to one individual we share it equally across the whole cohort.

To eliminate this uninteresting case, it is necessary to make an assumption that limits the size of the partnership. There are many ways to justify such an assumption. For example, if partners were required to exert some effort, then equal sharing will induce free riding, which may limit the size of the partnership. Here we will simply assume that there is an upper bound on the size of the partnership, and set that bound equal to 1.

We will also make the simplest possible assumption about risk preferences, assuming constant absolute risk aversion. Without loss of generality we may assume that $u(0) = 0$ and $u'(0) = 1$, so that the utility function is a quadratic of the form $u(x) = x - Ax^2$ with $A = -\frac{u''(x)}{2u'(x)} > 0$. We consider steady state contracts, setting the failure rate $\alpha = 0$; so the firm will with certainty continue to exist forever.

The contract design problem is to choose σ, w, π, b to maximize W subject to

$$\begin{aligned}
\pi\sigma n &= 1 \\
W &= \sigma n (b - w) \\
(1-\pi) u(w) + \pi u(w + W) &\geq e(b) \\
0 \leq \sigma &\leq 1 \\
0 \leq \pi &\leq 1 \\
b &\geq 0 \\
w &\geq 0.
\end{aligned}$$

Eliminating w and σ, and expanding the quadratic utility function u as a Taylor series expansion about b, these constraints can be written

$$\frac{1}{n} \leq \pi \leq 1 \tag{8}$$

$$b \geq 0 \tag{9}$$

$$\pi W \leq b \tag{10}$$

$$A\pi (1-\pi) W^2 \leq \phi(b) . \tag{11}$$

where $\phi(b) = u(b) - e(b)$. We note that ϕ is concave, since u is concave and e is convex. The efficient output level b_0 is now characterized by $\phi'(b_0) = 0$ (this is the output level at which marginal effort equals marginal utility; this is the output level that would be chosen by a profit maximizing firm) and the rat-race level b_* by $\phi(b_*) = 0$. It is clear that the constraint 11 must bind, for otherwise W can be made arbitrarily large be letting $b \to \infty$; this constraint can be used to eliminate W, leading to a final reformulation of the problem as

$$\begin{aligned} &\text{maximize } J(\pi, b) = W^2 = \frac{\phi(b)}{A\pi(1-\pi)} \\ &\text{subject to } \frac{1}{n} \leq \pi \leq \bar{\pi}(b) . \end{aligned}$$

Here we write $\bar{\pi}(b) = \frac{Ab^2}{Ab^2+\phi(b)} = \frac{Ab^2}{b-e(b)}$, and we put aside for the moment the possibility that $\pi = 1$.

See Figure 3, where the graph of $\bar{\pi}(b)$ is shown. Note that $\bar{\pi}(b)$ is monotonic increasing, since $\frac{b\pi'(b)}{\pi(b)} = 2 - \frac{b\phi'(b)}{\phi(b)}$; but $\phi(b) - b\phi'(b) \geq \phi(0) = 0$, since ϕ is concave, so $\frac{b\phi'(b)}{\phi(b)} \leq 1$ and $\frac{b\pi'(b)}{\pi(b)} \geq 1$. We also note that $\bar{\pi}(b_*) = 1$ and (by l'Hôpital's rule) that $\bar{\pi}(0) = 0$. The feasible region is $\left\{(b, \pi) : \frac{1}{n} \leq \pi \leq \bar{\pi}(b)\right\}$; this shrinks to the right as the risk aversion parameter A tends to 0. We also note that along this locus $J(b) = J(b, \bar{\pi}(b)) = \left(\frac{b-e(b)}{Ab}\right)^2 = \left(\frac{b}{\bar{\pi}(b)}\right)^2$, which is a decreasing function of b. In fact, this relationship holds even when $\pi = 1$, so no further separate argument will be needed for that case. For future convenience, let b_1 be the output level such that $\bar{\pi}(b_1) = \frac{1}{n}$.

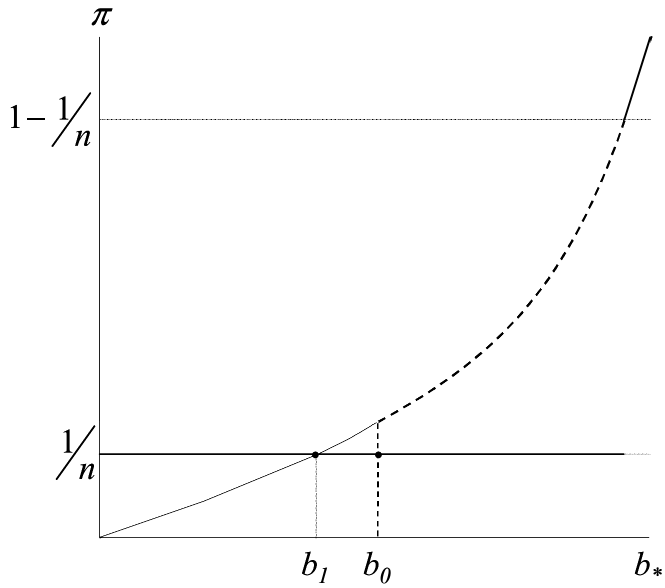

Figure 3. The Optimal Recursive Contract under Risk aversion

We consider separately the optimal choice of π, holding b fixed, and the optimal choice of b, holding π fixed. The objective $\frac{\phi(b)}{\pi(1-\pi)}$ is convex in π and symmetrical about $\pi = \frac{1}{2}$. Thus, for a given b, the optimal choice of π is $\pi = \bar{\pi}(b)$ provided that $\bar{\pi}(b) - \frac{1}{2} \geq \frac{1}{2} - \frac{1}{n}$; otherwise it is $\pi = \frac{1}{n}$. The locus of optimal choices for π is shown as a solid line in the Figure. Conversely, for a given π the optimal choice of b is b_0 provided that it is feasible; that is if $\pi \leq \bar{\pi}(b_0)$. Otherwise it is the smallest value of b such that $\bar{\pi}(b) \geq \pi$. The locus of optimal choices for b is shown as a dashed bold line in the Figure.

It is now clear that risk aversion (at least under the assumption of constant absolute risk aversion) will not limit the size of the firm, and the optimal contract will always be a full employment contract with $\pi = \frac{1}{n}$. For if not, then the optimal contract must lie on the locus $\{b, \bar{\pi}(b)\}$. But $\left(b_1, \frac{1}{n}\right)$ is the smallest feasible contract on this locus, and the objective $J(b)$ is decreasing in b along this locus. Thus the contract $\left(b_1, \frac{1}{n}\right)$ will dominate any contract for which $\pi > \frac{1}{n}$.

There are thus two possible types of contract, depending on whether b_1 is to the left or to the right of b_0 (both of these points vary with the risk aversion parameter A, shifting to the right as A tends to 0). Let us write, for the moment, $\psi(b) = b - e(b)$, let $\tilde{b}$ be the point such that $\frac{\tilde{b}\psi'(\tilde{b})}{\psi(\tilde{b})} = \frac{1}{n}$, and let $\tilde{A}$ be the value of the risk aversion parameter such that $\psi'\left(\tilde{b}\right) = 2\tilde{A}\tilde{b}$. One can check that such a point exists and that $b_1 < b_0$ if and only if $A > \tilde{A}$.

In the first regime (low risk aversion), we have $A < \tilde{A}$ and $b_1 > b_0$. The optimal contract is then $\left(\frac{1}{n}, b_1\right)$. This is a rat-race contract: output and effort are inefficiently high, and a zero wage is paid. As $A \to 0$ it approaches the pure, risk-neutral rat-race contract. In the second regime, we have $A > \tilde{A}$ and $b_1 < b_0$. The optimal contract is then $\left(\frac{1}{n}, b_0\right)$. The output level is efficient, and a positive wage is paid. We summarize in the following Proposition.

Proposition 3 *If the agents display constant absolute risk aversion then the optimal contract still leads to full employment. For high levels of risk aversion the contract implements the efficient output level, and a positive wage is paid. If the coefficient of risk aversion falls below a critical level then effort is inefficiently high and a zero wage is paid. As the coefficient of risk aversion decreases to* 0 *the contract converges to the risk-neutral pure rat-race contract.*

3.2 Congestion Effects.

Another effect that may be expected to limit the size of the firm is congestion, or diseconomies of scale associated with an excessively large span of control when there are many agents to each principle. To explore this effect we assume that the agent's marginal effort depends both on output b and on the the span of control $\frac{1}{\pi}$. For simplicity we assume that the effect of congestion is additive and linear:

$$e(b, \pi) = e(b) + \frac{\gamma b}{\pi},$$

where $\gamma > 0$ is a parameter that indicates the strength of the congestion effect. We normalize the number of principals to 1, and once again we assume risk neutrality.

For for future convenience we write

$$\begin{aligned} \pi_1(b) &= \frac{\gamma b}{\phi(b)} \\ \pi_2(b) &= \frac{\gamma}{\phi'(b)}. \end{aligned}$$

We note (by l'Hôpital's rule) that $\pi_1(0) = \pi_2(0) = \frac{\gamma}{\phi'(0)} > 0$, that both functions are monotone increasing, that $\pi_1(b) \to \infty$ as $b \to b_*$ and that $\pi_2(b) \to \infty$ as $b \to b_0$ (see Figures 4 and 5).

3.2.1 Baseline case

It is useful first of all to establish the efficient baseline. Given complete information, a profit maximizing firm would choose σ, π, w, b to maximize

$$\begin{aligned} J(b,\pi) &= n\sigma\left(b - e(b,\pi)\right) \\ &= \frac{b - e(b) - \frac{\gamma b}{\pi}}{\pi} \\ &= \frac{\phi(b)}{\pi} - \frac{\gamma b}{\pi^2}, \end{aligned}$$

where $\phi(b) - b - e(b)$, subject to $\frac{1}{n} \leq \pi \leq 1$.

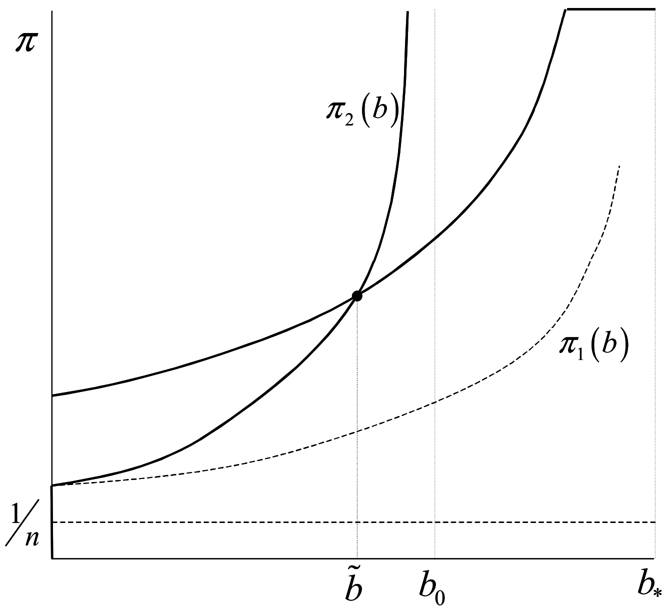

Figure 4. Baseline Contract under Congestion

We note that, for a given $b > 0$, $J(b,\pi)$ has a global maximum (as a function of π) if $\pi = 2\frac{\gamma b}{\phi(b)} = 2\pi_1(b)$. Thus the optimal choice of π, holding b fixed, occurs along the locus $\left\{\left(b, \max\left(\frac{1}{n}, \min\left(1, 2\pi_1(b)\right)\right)\right)\right\}$. For a given value of π, J is maximized as a function of b if $\pi = \frac{\gamma}{\phi'(b)} = \pi_2(b)$. Thus the optimal choice of b, holding π fixed, occurs along the locus $\left\{(b, \pi_2(b)) : \frac{1}{n} \leq \pi_2(b) \leq 1\right\} \cup \left\{(0,\pi) : \frac{1}{n} \leq \pi \leq \pi_2(0)\right\}$. These loci are shown in bold in Figure 4. For simplicity we will assume that n is large enough that the constraint $\frac{1}{n} \leq \pi$ does not bind, and that γ is small enough that the constraint $\pi \leq 1$ does not bind. This means that the congestion effect is strong enough to limit the size of the firm to less than n, but not to reduce it to a firm with only a single employee.

Then the baseline optimal production plan $\left(\tilde{b}, \tilde{\pi}\right)$ occurs in Figure 4 where the two loci cross: it is characterized by the fact that $\frac{\tilde{b}\phi'(\tilde{b})}{\phi(\tilde{b})} = \frac{1}{2}$, and $\tilde{\pi} = \frac{\gamma}{\phi'(\tilde{b})}$.

3.2.2 Recursive case

We now consider the optimal recursive contract. The principal's problem is to maximize W subject to

$$\frac{1}{n} \leq \pi \leq 1$$
$$\pi W \leq b$$
$$\phi(b) \geq \frac{\gamma b}{\pi}.$$

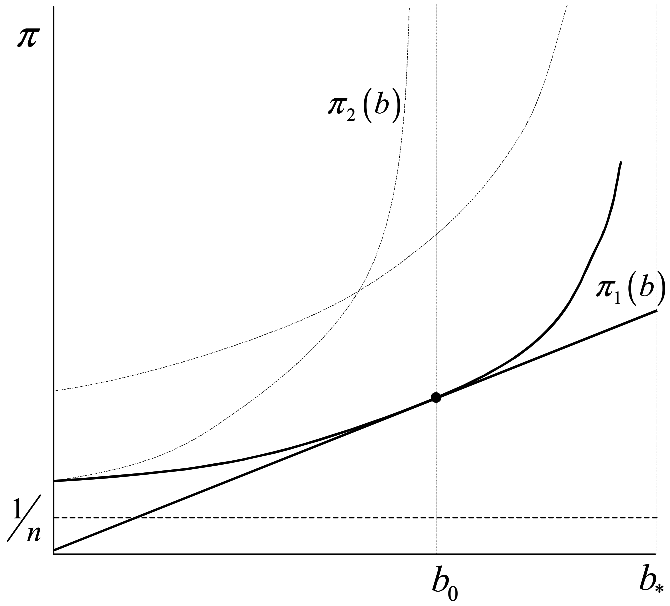

Figure 5. The Recursive Contract under Congestion

Assuming again that n is sufficiently large that the constraint $\frac{1}{n} \leq \pi$ does not bind, this is equivalent to the problem of minimizing $\frac{\pi}{b}$ subject to the constraint $\pi \geq \frac{\gamma b}{\phi(b)} = \pi_1(b)$, and the optimal contract is characterized geometrically by tangency between the curve $\pi_1(b)$ and the ray through the origin (see Figure 5). Thus $\pi_1(b) = b\pi_1'(b)$, which implies that $\phi'(b) = 0$. Hence $b = b_0 > \tilde{b}$; output is above the efficient baseline level. However it is not possible to say, without restricting the effort function $e(b)$, whether the size of the firm $\frac{1}{\pi_0} = \frac{\phi(b_0)}{\gamma b_0}$ is greater than or less than the base-line firm size.

If the constraint $\frac{1}{n} \leq \pi$ binds then the shifts to the right along the $\pi_1(b)$ curve, so the output level is even higher.

In either case the constraint $\pi W \leq b$ binds, so the agent's wage is always zero. We summarize.

Proposition 4 *Assume that the firm is subject to a linear congestion externality of the form* $e(b) + \frac{\gamma b}{\pi}$. *If* $n\gamma \geq \frac{b_0}{\phi(b_0)}$ *(that is, if the population is large or the congestion cost is high) then the firm size is limited to* $\frac{1}{\pi} = \frac{\phi(b_0)}{\gamma b_0} < n$, *and the output level* b_0 *is inefficiently high. If* $n\gamma < \frac{b_0}{\phi(b_0)}$ *then there is full employment, but the output level* b *(which is such that* $\frac{\phi(b)}{b} = n\gamma$*) is even higher, tending to* b_* *as* $\gamma \to 0$. *In all cases the wage is zero, and the agent is compensated only through the prospect of promotion.*

4 Discussion

In the OLG firm agents are motivated not only by current rewards but also by the prospect of becoming principal in the next period, and getting a chance to set the rules next time round. We find that the recursive contracts that emerge in such an environment are always inefficient.

The typical structure of such contracts is illustrated most clearly by the steady state rat-race contracts of Section 2.4. Agents are paid a zero wage, being compensated for their effort only by the probability of promotion. They work inefficiently hard, accepting contracts under which they produce up to the point where all surplus is dissipated and their average product falls to zero. The wage is zero because the agents are risk neutral and there is no discounting, so they do not mind exchanging a current wage for an equivalent future gamble. But the principal clearly prefers to reduce the current wage to zero rather than to leave some surplus to the next generation. The effort level is high because the principal, who designs the contract, cares not about efficiency but only about the ex-post payoff conditional on having been promoted. The agents are willing to accept such a harsh contract because they can impose a similar contract on the next generation.

This conclusion seems to be robust to a range of variations in the environment. In a finite horizon model, rat-race contracts, with a zero wage and an inefficient level of effort, can be sustained up until a finite number of periods before the final horizon. In a stationary model, where the firm may go out of business with a constant probability in any period, rat-race contracts can be sustained if the failure probability is not too great. Rat-race contracts can be sustained (under constant absolute risk aversion) provided that the degree of risk aversion is not too great. They can also be sustained under congestion effects associated with the costs that may be imposed by an increasing span of control.

There are several interesting areas for further investigation. Bardsley and Sherstyuk [2], for example, have investigated a recursive contracting model with adverse selection.

Recursive contracts may occur in the purest form in overlapping generation partnership firms, such as the large commercial law firm, where there is no outside equity, and where ownership and control rest with the partners. However something similar may occur in firms with a more conventional structure. If ownership is divorced from day to day control, then there may be a career structure within the firm where lower managers hope to be promoted to senior management, with the right to exercise control. However the extent of this control, and the contracts which would be designed, would be constrained by right of the owners to appoint from outside the firm. It would be interesting to investigate the recursive contracts that would emerge in this environment.

References

[1] Akerlof, George A., The Economics of Caste and of the Rat Race and Other Woeful Tales, *Quarterly Journal of Economics*, 90(4), 1976, 599-617.

[2] Bardsley, Peter and Katerina Sherstyuk, Career Paths in Partnerships: an OLG agency model, working paper, 2001.

[3] 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.

[4] Cremer, Jacques, Cooperation in Ongoing Organizations, *Quarterly Journal of Economics*, 101(1), February 1986, pages 33-49.

[5] 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.

[6] Farrell, J and S. Scotchmer, Partnerships, *Quarterly Journal of Economics*, May1988, 279-97.

[7] Galanter, Marc, Palay, Thomas, *Tournament of lawyers: The transformation of the big law firm*, Chicago and London: University of Chicago Press, 1991.

[8] Gibbons, Robert, and Kevin J. Murphy, Optimal Incentive Contracts in the Presence of Career concerns: Theory and Evidence, *Journal of Political Economy*, 100(31), 1992, 468-505.

[9] 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.

[10] Holmstrom, B., Mangerial Incentive Schemes — a dynamic Perspective, in *Essays in Economics and Management in Honour of Lars Wahlbeck*, Helsinki: Swenska Handelshögkolan, 1982.

[11] 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.

[12] Meyer, Margaret A, and John Vickers, Performance comparisons and Dynamic Incentives, Journal of Political Economy, 105(3), 1997, 547-581.

[13] 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.

[14] 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.