Investment banking (and other high profile) careers

ULF AXELSON and PHILIP BOND[*]

October 23, 2009

ABSTRACT

We analyze a general equilibrium labor market model where moral hazard problems are a key concern. We show that variation in moral hazard across types of jobs explains contract terms, work patterns over time, and promotion structures. We explain why high-profile jobs such as investment banking pay more and give higher utility to the employee than other jobs, even if employees have no skill advantage. These jobs also have high firing rates, and inefficiently long hours. Because dynamic incentives are especially important in high-profile industries, they are hard to enter late in a worker's career. Therefore, agents who are unlucky early on, either because they do not land a high-profile job or because they lose a high-profile job, suffer life-long disadvantages in the labor market. We also derive two versions of talent misallocation: High profile employers like investment banks may lure workers whose talent would be more valuable elsewhere, and may reject "over qualified" job applicants – smart workers may be "too hard to manage," because their high outside options make them respond less to firing incentives. Finally, we show that moral hazard problems increase in good times for critical sectors in the economy, leading to both higher pay and higher failure rates.

JEL codes: E24, G24, J31, J33, J41, M51, M52

Keywords: Investment Banking, Compensation Contracts

Jobs differ widely in terms of salary, job security, work hours and conditions, and promotion possibilities. High-profile jobs, such as investment banking, management consulting, jobs with top law firms, or prestigious medical specialities at top hospitals feature very high pay, especially in case of promotion. On the other hand, job security is low, work hours are long, and the initial "dog years" are often characterized by extremely hard work. Despite the tough work conditions, these jobs are considered very attractive. Furthermore, evidence suggests that they have to be entered relatively early in the career; switching to a high-profile job later in life is extremely hard.

Investment banking provides perhaps the quintessential example of a high-profile industry (at least until recently), and throughout the paper we use it as our leading example of a high-profile job. Investment bankers are (in)famously highly paid. For example, Oyer (2008) shows that MBA students from Stanford who entered investment banking had salaries that were around three times higher than other students after six to ten years. However, the high compensation comes at a price: the risk of getting fired is very high, and work hours are notoriously long. In the beginning of an investment banking career at a top firm it is not uncommon to work 100 hour weeks, and much of this time is spent on rather menial tasks such as gathering data and preparing power point presentations.

That tough work conditions go together with high pay may seem natural. It is not surprising, for instance, that oil-rig workers or miners are highly compensated, given the intrinsically high-risk nature of their jobs. Pay is simply a compensating differential. What makes investment banking different, and more challenging to explain, is that many of the unappealing aspects of the work are not intrinsic but rather chosen by the employer. This is true for both work hours and firing probabilities, and some of these choices seem –at least at first glance– inefficient. For example, instead of having a highly qualified MBA graduate work 100 hours per week on menial tasks, the employer could hire one more secretary to do the simpler tasks, have the MBA graduate work less, lower salaries somewhat, and possibly make everyone better off.

Furthermore, the compensating differential explanation for high pay is incompatible with the fact that almost all MBAs choose the investment banking job over all other jobs given the choice. It also seems doubtful that the large pay difference between investment banking

1

and other jobs is a pure skill premium.[1] Oyer (2008), using macroeconomic conditions at the time of graduation as an instrument for the probability of entering investment banking, shows that an MBA student who enters investment banking has an expected lifetime income that is $1.5 million to $5 million higher in present value terms than an equally skilled student who does not. In large part, the big difference is attributable to the fact that a student who misses the opportunity to enter investment banking upon graduation due to a downturn on Wall Street is very unlikely to be able to land a job in investment banking later on, even if the economy subsequently recovers. This striking path dependency in careers and life-time income is hard to explain with reasonable switching costs, and has unnerving implications for students graduating in recessions.

Just as high-profile jobs seem to pair high compensation with tough work conditions, and are hard for older workers to enter, other jobs feature lower compensation, easier hours, and more generous retention policies; can be entered at any stage in an individual's career; but are viewed as less desirable.

In this paper we develop a model based on a single standard friction, moral hazard, that can account for the combination of characteristics in high-profile jobs such as investment banking, the absence of those same characteristics in other jobs, and the allocation of workers across these jobs over their careers and over the business cycle. The basic building block of our model is a standard dynamic moral hazard model (Rogerson 1985; Becker and Stigler 1974; Lazear 1981). In order to discuss differentiated jobs, we allow for heterogeneity across employers in the degree of moral hazard, and in order to discuss career dynamics we allow for heterogeneity in worker age (and in extensions, heterogeneity in observable skill). An important aspect of our analysis is one-sided commitment: employees are free to leave their employer whenever they want, and their outside option is endogenously determined in equilibrium. In this respect we add to the vast literature on dynamic contracting, in which there are very few papers with this characteristic.[2]

---

[1]Less highly paid professionals such as business school professors (a category to which both authors belong) will surely arrive at this conclusion by candid self assessment.

[2]Phelan (1995) studies insurance contracts when agents can walk away from a contract with one principal and sign a new contract with another principal. Both principals and agents are homogenous in Phelan's model. Krueger and Uhlig (2006) allow for some heterogeneity, though not with respect to agents' contracting horizons, and not with respect to principals. Both papers consider economies with unobserved endowments,

In addition to explaining cross-sectional differences in job characteristics, the equilibrium nature of our model lets us consider two oft-asked questions of high-profile jobs in general and investment banking in particular. First, do the right people become investment bankers? Second, did they work conscientiously enough in the recent financial expansion, or are they in part to blame for the subsequent financial crisis? With respect to the first question, as we explain below, our model predicts that some people whose talents would be better used elsewhere become investment bankers ("talent lured"); but on the other hand, investment banks may shy away from hiring some very talented people on the grounds that they are too hard to manage/incentivize ("talent scorned"). With respect to the second question, our model predicts that while employees in less-prestigious sectors work harder and more carefully in economic expansions, the opposite is true in high-profile jobs such as investment banking. Instead, these jobs are characterized by a simultaneous increase in compensation (much of it in the form of bonuses), decrease in effort, and increase in mistakes.

In more detail, our model features a continuum of different types of job, each of which is afflicted by a standard moral hazard problem: an employee either succeeds or fails, and the success probability is increasing in the employee's non-contractible effort. When the employee succeeds, output is produced: for example, a successful merger is accomplished.

It is worth stressing that the only friction in our model is a standard moral hazard problem. This assumption both aids tractability, and allows us to isolate the effects of a single feature of the economic environment. In the conclusion we make some conjectures about how our results would change if skill were also unobservable.[3]

Different types of job are differentiated by the cost to the employer of employee failure. Employee effort is especially valuable in jobs with a high cost of failure, and so in equilibrium moral hazard is especially pronounced in these jobs. These high moral hazard jobs turn out to be "high-profile" in the sense that they are the most attractive to workers, and

---

as opposed to unobserved actions (moral hazard). Closer to us is a contemporaneous paper by Tsuyuhara (2009), who studies dynamic contracts in an economy with moral hazard; however, his focus is very different from ours.

[3]MacDonald (1988a) and Tervio (2009) study models with unobservable skill, but no effort decision, in which skill matters in just one industry. Both features render them silent with respect to many of our main results. Gibbons and Waldman (1999) allow for exogenous human capital acquisition. Like us, MacDonald (1988b) allows for multiple differentiated industries (but still no effort decision); in contrast to our framework, his industries are symmetric, so there is no notion of good and bad jobs in his model.

commensurately hard to get. This is because employers surrender rent to employees in these jobs to ameliorate the moral hazard problem. For the same reason, employers prefer to hire workers with longer career horizons (i.e., younger workers) for these jobs, since the possibility of using promotion, firing and other dynamic incentive devices also ameliorates the moral hazard problem.

It is impossible for all employers to hire young workers, and so the reservation utilities of young and old workers are critical in determining which employers hire which employees. We determine these reservation utilities in equilibrium, and show that employers with moderate and high amounts at stake hire only young workers, and may or may not retain these workers when old. Crucially, the job conditions of young workers differ depending on exactly how much is at stake. Young workers in jobs with only a moderate amount at stake have compensation determined by their equilibrium reservation utility, and are sometimes retained even after failure. In contrast, young workers in jobs with the very most at stake — high-profile jobs — receive surplus over and above their reservation utilities, but work long hours and are always fired when they fail. Because of the high surplus, all young workers wish to obtain high-profile jobs, although only some are able to. If possible, employers try to recoup some of the surplus they surrender in these high-profile jobs by assigning long work hours on mundane activities early on in the career. This is our explanation for the "dog years."

Fired young workers must reenter the labour market as old workers, but can only land low-profile jobs where moral hazard problems are less severe. These more mature, fired workers receive lower compensation both in monetary and utility terms than average workers. In this sense, our model features something that resembles a "stigma of failure," but the result is driven by age and not adverse selection. By the same token, young workers who are not lucky enough to enter a high-profile job such as investment banking are unable to do so later in the career.

Most of the analysis we perform is without any skill difference across workers. When we introduce observable skill differences, we get some surprising results. In particular, our model naturally generates two commonly noted forms of talent misallocation. The first one, which we call "talent lured," is the observation that jobs like investment banking tend to

attract talented workers whose skills might be socially more valuable in other jobs, such as engineers and PhDs. In our model, this type of misallocation follows immediately from the fact that the high surplus earned in high moral hazard industries will make it possible for these industries to outbid other employers for workers even if their talent is wasted in investment banking. The second phenomena, which we call "talent scorned," is the opposite – high profile jobs often reject the most talented applicants on the grounds that they are "difficult" or "hard to manage." This can be rational in our model because talented workers, when fired, have higher outside opportunities, which makes it harder to control them with dynamic incentive schemes.

Finally, we analyze how moral hazard problems vary across the business cycle. When demand in the economy goes up, there are two effects on moral hazard that have a differential impact across industries. The first effect is that for the low-profile segment of the industry hiring only fired old workers, increased demand increases output prices which makes it optimal to incentivize workers to work harder. In that sense, moral hazard problems go down for this segment of the economy. For the highest-profile industries, such as investment banking, the effect is opposite – because fired workers have higher outside opportunities, it is more difficult to incentivize workers. Although principals will partly compensate for this by increasing the level of bonuses, the net effect is still weakened incentives. Thus, moral hazard problems go up, even at the same time as bonuses rise, leading to worse decisions in a critical segment of the economy.

A. Related literature

We have already noted our paper's connection to the dynamic contracting literature. Of particular relevance is Hutchens (1986), who notes that agency considerations make young workers relatively attractive to hire. However, this observation begs the question of whether compensation for young and old workers will adjust so as to leave employers indifferent between the two. In other words, if young workers are so attractive, who hires old workers? One of the contributions of our analysis is to show that the relaxation of agency concerns is especially valuable in jobs where a lot is at stake, and so in equilibrium young workers enter jobs such as investment banking with a lot at stake (high-profile jobs), while old workers

enter jobs with less at stake.

Empirically, Lazear and Moore (1984), Kotlikoff and Gokhale (1992), and Groshen and Krueger (1990) all provide evidence consistent with the use of dynamic labor contracts.[4]

Our paper has some antecedents in the "efficiency wage" literature (Shapiro and Stiglitz 1984), which studies the aggregate effects of one particular dynamic incentive contract (fire if caught shirking, continue at same wage otherwise). The literature focuses primarily on implications for unemployment, although Bulow and Summers (1986) informally discuss an assortment of other predictions. Efficiency wage models are often criticized on the grounds that alternate contracting arrangements might obviate the need for unemployment as an incentive device (Yellen 1984; Carmichael 1985). In particular, it is possible that promising a worker an increasing wage profile would deter the worker from shirking.[5] Relative to the efficiency wage literature, we solve for optimal equilibrium contracts. This enables us to consider employment outcomes on which efficiency wage models are silent, such as promotion, demotion, changes in type of job, and firing.

An important aspect of our model, which we share with the efficiency wage literature, is that some industries/employers reward their employees with more utility than others, even when the employees are otherwise identical. A large empirical literature (see, e.g., Krueger and Summers 1988) provides evidence on the existence of industry- or employer- specific effects in wage determination.[6] Oyer (2008) can be read as providing new evidence.

Finally, a relatively recent literature studies incentive contracting between employers and employees in an equilibrium context: see Moen and Rosen (2006), Edmans *et al* (forthcoming), Baranchuk *et al* (2008), Acharya and Volpin (forthcoming), and Dicks (2009). These papers examine only one period contracts, and so most of the main results in the current paper are not attainable in these models. Additionally, Edmans *et al* specify an incentive problem in which the agent never receives any surplus above his outside option, while

---

[4]See chapter 6 of Cahuc and Zylberberg (2004).

[5]Akerlof and Katz (1989) establish that optimal dynamic contracting cannot by itself eliminate the need for unemployment as an incentive device. MacLeod and Malcomson (1998) observe both that the simple contract of efficiency wage models is a self-enforcing contract (i.e., neither the principal nor the wage wants to renege), and that other outcome-dependent self-enforcing contracts exist in some circumstances. However, they restrict attention to stationary contracts without wage growth over time.

[6]See also papers such as Murphy and Topel (1987) and Abowd et al (1999) on the difficulties of adequately controlling for worker characteristics.

Baranchuk *et al* study a problem with undifferentiated principals,[7] so that there is no sense in which there are good and bad jobs in their model.

B. Paper outline

In Section I, we describe the model. In Section II, we characterize the solution to the static contracting problem. In Sections III, IV and V we characterize the solution to the dynamic contracting problem. Section VI extends the analysis to differentiated tasks within the same industry; differential skill across workers; and comparative statics across the business cycle. In Section VII we formally establish equilibrium existence (as opposed to equilibrium properties). Section VIII concludes.

I. Model

To study the labor market phenomena we are interested in, we need two key elements: Workers of different age, and industries that vary in their degree of moral hazard problems. To this end, we assume a supply $\frac{1}{2}\overline{\lambda}$ of young workers enter the labor market each period, work for two periods, and then exit. Thus, the total supply of workers is $\overline{\lambda}$. Except for age, workers are identical. They all have the same skill, are risk neutral over both consumption and leisure, start out penniless, and have limited liability. (We will analyze a setup where skills differ across workers in Section VI.)

There is a continuum of industries indexed by $k \in \left[\underline{k}, \overline{k}\right]$. A worker employed within an industry a certain period works on a project, which either succeeds or fails. Projects vary across industries in the cost of failure k: the failure payoff is $-k$. We write the success payoff as $g - k$, where g is determined in equilibrium (see below). One way to think about these payoffs is that k is an input cost (e.g., funds provided to a trader) and g is the value of output produced when the project succeeds (e.g., gross value after trading). Alternatively, k is the value destroyed if a project fails (e.g., a takeover fails), and $g - k$ is the value created if a project succeeds (e.g., takeover succeeds).

If a worker spends h hours on the project, it succeeds with probability $p(h)$ and fails with

---

[7] Specifically, firm size is chosen endogenously in their model.

7

probability $1 - p(h)$. Workers have a per-period time endowment of H, which they can split between work and leisure, and have linear preferences over leisure. The success probability $p(h)$ is a strictly increasing and strictly concave function with $p'(0) = \infty$ and $p'(H) = 0$.[8] While output (i.e., success or failure) is fully observable, effort is private information to the worker, which leads to a standard moral hazard problem. Analytically, it is slightly easier to express everything in terms of probabilities instead of hours worked: let $\gamma \equiv p^{-1}$, so that the utility cost of a worker achieving success probability p is $\gamma(p)$. The function γ is strictly increasing and strictly convex, with $\gamma'(0) = 0$ and $\gamma'(p(H)) = \infty$.

For the case of financial sectors, the following specific interpretation of the moral hazard problem is worth spelling out. The success payoff g is a target (gross) rate of return. A financial sector worker can meet this target either by working hard and discovering genuinely profitable trading opportunities, or by taking "tail" risk. When tail risk is realized all the input funds k are lost. By working h hours, the amount of tail risk a worker needs to take to achieve his target return is such that the probability of tail risk being realized is $1 - p(h)$.

For a worker entering industry k, let $\overline{x}$ denote his expected compensation per period (i.e., the ratio of total expected compensation in the industry to the number of periods he expects to work in the industry). Likewise, let $\overline{p}$ denote his expected average success probability while working in the industry. The profit created by the worker (net of compensation) is hence $\overline{p}g - \overline{x} - k$. We assume there is free-entry into each industry, with g decreasing in total output. Consequently, in equilibrium we have the zero-profit condition

$$\overline{p}g - \overline{x} - k = 0, \qquad \text{(Zero Profit)}$$

where $\overline{p}$ and $\overline{x}$ are associated with the profit-maximizing contract, which we define formally below. However, the following simple and useful result already follows, using only the fact that if a worker achieves success probability $\overline{p}$ for compensation $\overline{x}$ in some industry k he would be prepared to work for the same terms in any other industry:

---

[8] The assumption that effort has the same effect on success probabilities in all industries is less restrictive than it seems. Variation in the amount at stake k across industries has qualitatively the same effect as variation in the effect of effort on success probability, so we choose to normalize by only considering variation in k.

Lemma 1. *In any equilibrium, $g - k$ is continuous and strictly increasing in k; and $\overline{x}$ and $\overline{p}$ are weakly increasing in k.*

Proof: In Appendix.

These simple equilibrium relations are useful for understanding why moral hazard problems are "bigger" in industries where more is at stake. When the possible loss k is bigger, it must be that the gains are also bigger in equilibrium, or else employers would never enter the industry. Since both gains and losses are bigger, it is more important to incentivize the worker to work hard so that the success probability increases. As it turns out, this can only be done by paying the agent more in case of success, which means that even if he does not increase his work effort he will get a higher expected pay. Hence, if a worker ends up in a high k industry, he will typically get both higher wages and higher utility.[9]

For use below, we also make the following fairly innocuous assumption on the shape of the production function:

Assumption 1. $p\frac{\gamma'''(p)}{\gamma''(p)} > -1$ and $\lim_{p \to 0} p\frac{\gamma'''(p)}{\gamma''(p)} < \infty$.

Economically, the first part of Assumption 1 says that the surplus a worker receives as a result of moral hazard increases at an increasing rate in the desired effort level p.

A. Equilibrium

Rather than inundate the reader with notation upfront, we postpone a formal definition of an equilibrium until Section VII. Loosely speaking, we will require the following conditions to hold in equilibrium: success payoffs $g(k)$ satisfy a zero-profit condition for each industry (see earlier); all workers receive at least their reservation utility; success probabilities are consistent with profit maximization; reservation utilities for young and old workers are such that industries hiring old (respectively, young) workers prefer them to young (respectively,

---

[9]We could have modelled the magnitude of moral hazard problems within an industry in other ways without changing the general message of the paper. For example, instead of varying the money at stake, we could increase the noise between unobservable effort and observable outcome, or we could increase the cost of effort. The important feature, which we will come to further down, is the relative extent to which the incentive constraint rather than the participation constraint binds across industries.

old workers); the expected utility of young workers, if fired, is determined by the market for old workers; and labor markets clear.

We conjecture, and prove in the analysis that follows, that an equilibrium exists in which industries $k < \hat{k}$ hire only old workers, while industries $k > \hat{k}$ hire only young workers. (Industry $\hat{k}$ is indifferent between the two types of worker.)

II. Industries hiring old workers

We start by considering industries $k < \hat{k}$ that hire only old workers. Old workers can be given only one-period contracts, and so the contracting problem itself is very standard. This allows us to specify the equilibrium aspects of our analysis in a very familiar setting. Later, the analysis of industries $k > \hat{k}$ employing young workers makes use of similar ideas, and builds on the solution to the one-period contracting problem.

As a benchmark, note that in the absence of any agency problem in industry k, the marginal benefit of increasing the success probability p would equal the marginal cost of effort. We write $p_{FB}(g)$ for this first-best success probability, defined by

$$\gamma'(p_{FB}(g)) = g. \tag{2.1}$$

Since γ is convex, $p_{FB}(g)$ increases in g, and hence in k.

With moral hazard, the effort level will typically be lower than first best. We now characterize optimal contracts and effort levels in industries that hire old workers. Since old workers only work for one period, a contract is simply a fixed payment w and an extra payment Δ in case of success. Note that we have deliberately left unspecified whether workers are employed directly by consumers of output ("clients"), or whether they are employed by intermediary firms who make zero-profits. Limited liability requires $w \geq 0$ and $\Delta \geq -w$.

If the worker chooses success probability p, the employer's expected profit is $pg - p\Delta - w - k$, while the worker's expected utility is $p\Delta - \gamma(p) + w$. Hence a contract (w, Δ) induces a success probability

$$\gamma'(p) = \Delta. \tag{IC1}$$

10

This says that for the worker to achieve success probability p, he has to be paid a bonus reflecting the marginal cost of effort to avoid shirking. This costs the employer $p\gamma'(p)$ in expectation, which is larger than the total cost of effort $\gamma(p)$ since γ is convex. Hence, the worker may capture some surplus relative to the case where effort is observable. Using relation (IC1), we can write the contract in terms of p and w instead of Δ and w. The worker's expected utility from a contract (p, w) is hence

$$u(p, w) \equiv p\gamma'(p) + w - \gamma(p),$$

where the first two terms are the agent's expected compensation, and the third is his cost of effort. Note that the utility is strictly increasing in both arguments. In particular, increases in effort p raise the agent's utility, and (by Assumption 1) do so at an increasing rate.

Let $v \geq 0$ denote the worker's reservation utility. In equilibrium, this will be determined by the condition that the labor market for old workers clears, while the success payoff g will be determined by the zero profit condition. Taking v and g as given, the contract terms (p, w) in industry k are set to maximize profits subject to the participation constraint $u(p, w) \geq v$ (recall that we have substituted out the incentive constraint), i.e., to solve

$$\max_{w \geq 0, \, p} pg - \gamma(p) - u(p, w) - k \text{ such that } u(p, w) \geq v. \tag{P1}$$

The solution to problem (P1) depends on the size of the outside option v. First, suppose v is so small that the participation constraint is not binding. Then, it is easy to see that it is optimal to not give the agent any fixed pay w, and to set p from the first order condition such that:

$$g = \gamma'(p) + \frac{\partial u(p, 0)}{\partial p} = \gamma'(p) + p\gamma''(p), \tag{1}$$

that is, the employer sets the success probability such that the marginal benefit g equals the marginal cost of effort plus the marginal increase in surplus that is captured by the agent. We call the solution to (1) $p_{SB}(g)$ for the second-best level of p. Note that $p_{SB}(g) \in (0, 1)$, since the right-hand side of (1) is strictly increasing in p, is 0 at $p = 0$, and goes to infinity

as p goes to 1.[10] Also, note that $p_{SB}(g)$ is below the first-best level. More important for our purposes is that $p_{SB}(g)$ strictly increases with the amount at stake k, since g is strictly increasing in k (Lemma 1). This means that the utility of the agent also increases with k. This is why we call high k industries "attractive" or "high moral hazard industries," since the surplus given to agents is typically higher.

The employer will set $w = 0$ and $p = p_{SB}(g)$ as long as the participation constraint is not binding, that is, as long as $u(p_{SB}(g),0) \geq v$. Now suppose $u(p_{SB}(g),0) < v$, so that the participation constraint is binding. There are two ways of increasing the worker's utility to satisfy the participation constraint: either increase p, or increase w. Increasing p is better for the employer as long as $g > \gamma'(p)$, that is, as long as p is below the first-best level. If the promised utility v to the agent is so large that the participation constraint is not satisfied even at the first-best effort level, that is, if $u(p_{FB}(g),0) < v$, it is better to increase agent utility by a fixed payment w instead of increasing the effort p. We collect the solution to the one-period problem in the following lemma:

Lemma 2. *Optimal one-period contracts, success probabilities, and worker utilities, given g and v, are:*

	Fixed wage w	Bonus Δ	p	Worker utility
$v \leq u(p_{SB},0)$	0	$\gamma'(p_{SB})$	p_{SB}	$u(p_{SB},0)$
$u(p_{SB},0) < v \leq u(p_{FB},0)$	0	$\gamma'(p_v)$	p_v	v
$u(p_{FB},0) < v$	$v - u(p_{FB},0)$	$\gamma'(p_{FB})$	p_{FB}	v

where p_v solves $u(p_v,0) = v$.

Earlier, we informally stated a zero-profit condition that must hold in equilibrium. Given Lemma 2, we can now provide a formal version for industries hiring old workers. Define $\pi(v,g,k)$ as the maximal attainable profits in industry k, using only old workers, and for a given reservation utility v and success payoff g. That is, $\pi(v,g,k)$ is the maximized value of problem (P1), and is easily evaluated from Lemma 2.[11] Trivially, π is strictly increasing

---

[10] To see that the right-hand side increases in p, note that $p\gamma''(p)$ increases by Assumption 1, while γ' increases by convexity of γ.

[11] For use below, note that the profit function π is concave in the worker's utility v. The non-trivial step in this claim is to establish concavity when $u(p_{SB},0) < v \leq u(p_{FB},0)$. For such values of v, $\pi(v) =$

in the success payoff g. The zero-profit condition for industries hiring old workers is thus

$$\pi\left(v, g\left(k\right), k\right) = 0. \qquad\qquad\qquad \text{(ZP1)}$$

Observe that for any industry k and reservation utility v there is a unique success payoff $g\left(k\right)$ that satisfies (ZP1).[12]

We have written Lemma 2 to cover arbitrary values of v and g because we use it not only to characterize old-worker contracts, but also as an input to find optimal young-worker contracts. However, the success probability $p_{FB}\left(g\right)$ leads to strictly negative profits if either $k > 0$ or $w > 0$.[13] Consequently, from Lemma 2:

Corollary 1. *In equilibrium, no worker who joins an employer when old receives a fixed wage: $w = 0$ for all such workers.*

In contrast, an employer may pay a fixed wage to young workers it retains when old, provided that it makes enough profits in the first period to offset the losses associated with setting $w > 0$ in the second period.

In addition to a success payoff $g\left(k\right)$, an equilibrium specifies the expected utility of an old worker hired into industry k, which we denote as $u\left(k\right)$, and the success probability attained by old workers, which we denote $p\left(k\right)$. Given $g\left(k\right)$ and reservation utility v, equilibrium values of $u\left(k\right)$ and $p\left(k\right)$ are determined by profit maximization, i.e., by the solution to (P1), and are computed in Lemma 2. Consequently, for a given reservation utility v and cutoff $\hat{k}$ for industries hiring old workers, the utilities and success probabilities of old workers in each of industries $k < \hat{k}$ are uniquely determined.

As noted previously, in equilibrium $p_{SB}\left(g\left(k\right)\right)$ is strictly increasing in k. This implies the following corollary of Lemma 2:

$pg - \gamma\left(p\right) - v$, where $p\gamma'\left(p\right) - \gamma\left(p\right) = v$. Differentiation yields $\frac{d^2\pi}{dv^2} = \frac{d^2p}{dv^2}\left(g - \gamma'\left(p\right)\right) - \gamma''\left(p\right)\left(\frac{dp}{dv}\right)^2$. Since $p < p_{FB}$, $g - \gamma'\left(p\right) > 0$, and so π is concave in v provided $\frac{d^2p}{dv^2} < 0$. Evaluating, $p\gamma''\left(p\right)\frac{d^2p}{dv^2} + \left(p\gamma'''\left(p\right) + \gamma''\left(p\right)\right)\left(\frac{dp}{dv}\right)^2 = 0$. By Assumption 1, $p\gamma'''\left(p\right) + \gamma''\left(p\right) > 0$. Hence $\frac{d^2p}{dv^2} < 0$, establishing that π is concave in v.

[12]$\pi\left(v, g, k\right)$ is continuous in g; is weakly negative at $g = 0$; and grows unboundedly as $g \to \infty$.

[13]This follows from substituting $p = p_{FB}\left(g\right)$ and (IC1) into the profit expression $pg - p\Delta - w - k$.

Corollary 2. *In equilibrium, if old workers in industry $k_1 \in \left[\underline{k}, \hat{k}\right]$ receive more than their reservation utility $(u(k_1) > v)$, the same is true for old workers in any industry $k_2 \in \left[k_1, \hat{k}\right]$ with more at stake $(u(k_2) > v)$.*

The economics behind Corollary 2 is that the gain from paying workers above their reservation wage — namely, the increase in worker effort it allows — is more valuable in industries with more at stake, and higher success payoffs $g(k)$.

In light of Corollary 2, define $k^- \equiv \sup \{k : u(k) = v\}$: so workers in all industries below k^- receive exactly their reservation utility v, while workers in all industries between k^- and $\hat{k}$ receive rent strictly in excess of their reservation utility v. Moreover, since g is strictly increasing, Lemma 2 implies that both effort $p(k)$ and utility $u(k)$ are *strictly* increasing over the range of industries $[k^-, \hat{k})$. Conversely, since no old worker receives a fixed wage (Corollary 1), and since worker utility is constant over industries $[\underline{k}, k^-]$, effort in these industries must also be constant and equal to $p(k^-)$.

Summarizing the above analysis:

Proposition 1. *In any equilibrium in which industries $k < \hat{k}$ hire only old workers, there exists some k^- such that workers in industries $[\underline{k}, k^-]$ receive exactly their reservation utility v and work a constant amount $p(k^-)$; while workers in industries in $\left(k^-, \hat{k}\right)$ receive strictly more than their reservation utility v, and work strictly harder than $p(k^-)$. Moreover, both utility and effort are strictly increasing over $\left(k^-, \hat{k}\right)$.*

Proposition 1 illustrates a few of the general properties of contracts that we stress in the paper. (See also the part of Figure 1 on page 23 related to industries below $\hat{k}$.) The incentive constraint binds for employers with bigger moral hazard problems (money at stake), while the participation constraint binds for employers with lower moral hazard problems. It is better to end up in one of the "high-profile" industries, since they give workers higher utility. On the other hand, you work more in these industries, but this is not enough to outweigh the higher pay. Hence, the labor market is a lottery, with some workers being luckier than others.

Thus far we have taken the reservation utility v of old workers as given. In equilibrium, it will be determined by labor-market clearing: demand for old workers must equal supply.

Let $\lambda_o(k)$ denote the number of old workers employed in industry k. Total output (i.e., total successes) in industry k is hence $\lambda_o(k) p(k)$. As noted previously, we assume that the value of success in each industry k is decreasing in total output: let $\zeta_k(x)$ be the value of success in industry k if total output in the industry is x. For example, if industry k produces a product, ζ_k is simply the inverse demand curve. Alternatively, if industry k is engaged in trading financial securities, ζ_k reflects the idea that profits decline as a trading strategy becomes more popular. Regardless of the interpretation, the number of old workers in industry k is determined by the equilibrium relation

$$g(k) = \zeta_k(\lambda_o(k) p(k)). \tag{LD1}$$

The reservation utility v itself is then determined by the requirement that the total demand for old workers in industries $[\underline{k}, \hat{k})$ matches total supply. However, the supply of old workers is determined by the promotion policies employed by employers $k > \hat{k}$ hiring young workers, to which we turn next.

III. Dynamic Contracts

We now turn to industries $k > \hat{k}$, where (according to our conjecture) young workers are hired. A contract for a young worker consists of a triple (v_s, v_f, f), where v_s is the continuation utility promised to the worker in case of success, v_f is the continuation utility promised to the worker in case he fails but is retained, and f is the firing probability in case of failure.[14] The continuation utilities are paid in period 2 by giving the worker a one-period contract (p_s, w_s) after success and a contract $(p_f, 0)$ after failure, where the optimal contract terms

---

[14]Because we allow for the worker to be randomly fired (i.e., we allow for $f \in (0, 1)$ as well as $f = 0$ and $f = 1$), the reader may wonder if randomization would help elsewhere in the contract. The answer is no. For one period contracts, footnote 11 establishes that profits are concave in the utility promised to the worker, implying there would be no gain from randomization. For dynamic contracts, on the one hand randomization cannot help if the worker's participation constraint is non-binding, since in this case the firm should simply pick the contract (v_s, v_f, f) that maximizes profits. But on the other hand, nor can randomization help if the worker's participation constraint is binding. To see this, suppose to the contrary that the optimal contract entails randomization. Let (v_s, v_f, f) and $\left(\tilde{v}_s, \tilde{v}_f, \tilde{f}\right)$ be two points in the support. For randomization to be useful, the agent's utility level must differ between these two points. But then the worker's participation constraint is violated with positive probability.

are solved for in Lemma 2 given v_s and v_f.[15]

The utility a young worker receives from a contract (v_s, v_f, f) depends on his expected utility if fired after failure. The young worker has now become an old worker, and so must look for a new job in industries that hire old workers, that is, industries $k < \hat{k}$ in our conjectured equilibrium. Let F be the average utility offered by industries $k < \hat{k}$, weighted by the number of old workers hired by each industry: formally,

$$
F = \frac{\int_{\underline{k}}^{\hat{k}} \lambda_o(k) u(k) \, dk}{\int_{\underline{k}}^{\hat{k}} \lambda_o(k) \, dk}. \tag{3.1}
$$

Given F, a young worker's utility from a contract (v_s, v_f, f) if he exerts effort p in the first period is $pv_s + (1-p)((1-f)v_f + fF) - \gamma(p)$. Consequently, in the first period the young worker chooses effort given by

$$
\gamma'(p) = v_s - ((1-f)v_f + fF). \tag{IC2}
$$

We assume that contracts have to be renegotiation proof, so that they lie on the Pareto frontier in period 2. Formally, write $\underline{u}(g)$ as the utility the worker gets in the solution to the one-period contracting problem (P1) if we set $v = 0$ (so that the participation constraint does not bind): the renegotiation constraint is that $v_s, v_f \geq \underline{u}(g)$. We also need to ensure that the young worker never wants to voluntarily quit after one period and try his luck in the old-worker market, i.e., $v_s, v_f \geq F$. Between them, the no-renegotiation and no-quitting constraints restrict the extent to which a worker who fails can be punished. The worst punishment that can be inflicted on a worker is to fire him, in which case he is free to take a new job. As discussed in the introduction, one of the contributions of our paper is to determine his outside option if fired, F, in equilibrium.

---

[15] In equilibrium, no employer would ever make a non-contingent payment at the end of period 1 (see Lemma A1 in appendix). Likewise, no employer would give severance pay, i.e., pay the worker if he is unsuccesful and fired. Finally, firing a worker if he is succesful in period 1 is suboptimal.

The contracting problem for employers of young workers to solve is hence

$$\max_{\substack{p,f \\ v_s,v_f \geq \max\{\underline{u}(g),F\}}} p\left(g + \pi\left(v_s, g, k\right)\right) + (1-p)(1-f)\pi\left(v_f, g, k\right) - k \tag{P2}$$

subject to the incentive constraint (IC2) and the participation constraint

$$pv_s + (1-p)\left((1-f)v_f + fF\right) - \gamma\left(p\right) \geq V,$$

where V is the young agent's reservation utility, determined in equilibrium below. Parallel to the old worker case, define $\Pi\left(V, F, g, k\right)$ as the maximal attainable profits in industry k, using only young workers, and for a given reservation utility V, outside option if fired F, and success payoff g. Again parallel to the old worker case, the equilibrium success payoff $g\left(k\right)$ in industries employing young workers must satisfy the zero-profit condition[16]

$$\Pi\left(V, F, g\left(k\right), k\right) = 0. \tag{ZP2}$$

IV. The advantage of young workers relative to old workers

We now describe the advantages of using dynamic contracts, and how this translates into the allocation of young and old workers over industries. Note that a repeated one-period contract can always be implemented with a dynamic contract. Hence, holding g and the utility of the worker constant, profits must be at least weakly higher than with a one-period contract. We first show that when an employer hires a young worker, the repeated one-period contract is (basically) *never* optimal:

Lemma 3. *In any profit-maximizing contract for young workers, the worker either exerts strictly more effort in period 2 after he succeeds in period 1 than in period 1, an outcome unattainable using a repeated one-period contract; or else he exerts at least first-best effort in both instances.*

---

[16] $\Pi\left(V, F, g, k\right)$ is continuous in g; is weakly negative at $g = 0$; and grows unboundedly as $g \to \infty$. Hence there is a unique success payoff $g\left(k\right)$ that satisfies (ZP).

Proof: In Appendix.

The economics behind Lemma 3 is familiar from the dynamic contracting literature. In period 1 the agent is motivated to work by the prospect that he will get to continue working on the moral hazard task in period 2 if he succeeds. The surplus captured by the agent from his work in period 2 can be used as a reward for his period 1 work, which alleviates moral hazard problems relative to the one-period solution.[17]

Lemma 3 implies that an employer would be willing to offer a young worker an employment contract that gives the worker strictly higher per-period utility than the employer would be willing to offer a newly hired old worker. In equilibrium as employers compete for workers, this means that young workers' reservation utility must be strictly higher than what even the best compensated newly hired old worker can get:

Corollary 3. *In equilibrium, the per-period reservation utility of young workers, $V/2$, must be strictly greater than the utility of any old worker newly employed in any industry.*

To prove Corollary 3, suppose to the contrary that there is some industry k that is willing to give a newly hired old worker an employment contract that gives utility $V/2$ or higher. This industry could switch to employing young workers using a repeated version of the contract they offered the old worker and continue to break even. Lemma 3 then implies that employers in industry k could make strictly greater profits by switching to the profit-maximizing contract for young workers, contradicting the hypothesis that they would be willing to offer an old worker $V/2$.

Corollary 3 highlights the disadvantages of old workers seeking employment. Because these workers cannot commit to a dynamic contract, employers know that they will be harder to incentivize — colloquially, old workers are "jaded" and "hard to motivate." This means that they will be strictly discriminated against in the labor market. In our model, old workers seeking employment are workers who have been fired from their previous job because

---

[17]Note that Lemma 3 shows that effort typically goes *up* over the career as the worker gets promoted. We would like to downplay the actual increase in work hours at this stage; rather, we want to stress the fact that promotion leads to more work on the important task, which in turn gives the worker high rents. We will see in Section VI that when there are more tasks to be performed within an organization, the worker will typically work longer hours early on in the career but on more "menial" (lower moral hazard) tasks. As he gets promoted, he works less hours, but all on the important task.

of sub-par performance. Hence, the result that fired old workers have a hard time finding jobs at good terms resembles a stigma of failure. Note, however, that this has nothing to do with learning about worker type – it is purely an age effect.

We are now ready to establish another, related benefit of being young – the fact that young workers will end up in higher profile industries than old workers seeking new employment. As we will see, this implies that young workers not only have a higher reservation utility, they also have a chance to get jobs in industries that pay strictly more than the reservation utility. On our way to establishing the sorting of workers across industries, we first show that young workers work harder than old workers would have done in the same industry. In order to do this, denote by $p_o(k)$ the effort industry k would use if they hired old workers, given g. Note that this definition covers all industries, including those that in fact only hire young workers in equilibrium. Likewise, let $p_y(k)$ denote the average (over their employment in industry k) effort of young workers under the profit-maximizing contract for young workers. We then have the following result:

Lemma 4. *In any industry k that can make weakly positive profits from young workers, $p_y(k) > p_o(k)$ if $p_o(k) < p_{FB}(g)$ and $V/2 \geq F \geq v$.*

Proof: In Appendix.

This result further accentuates the benefit of hiring young workers – they can be made to work harder over their career with the employer relative to a newly hired old worker. In addition to being of independent interest, Lemma 4 also helps in delivering the following central result on the sorting of workers across industries:

Proposition 2. *In any candidate equilibrium, there must be a $\hat{k} < \overline{k}$ such that industries $k < \hat{k}$ hire old workers and industries $k > \hat{k}$ hire young workers. Furthermore, if in such a candidate equilibrium $g(k)$ is set such that the zero-profit condition holds for the postulated allocation of workers, all industries $k \neq \hat{k}$ are worse off by switching workers (strictly so for $k > \hat{k}$), while industry $\hat{k}$ is indifferent between old and young workers.*

Proposition 2 is deliberately carefully stated since we have not yet established the existence of an equilibrium; indeed, the proposition is a critical stepping stone in proving that

such an equilibrium exists. What it says is that if there is an equilibrium, it must be such that low moral hazard industries hire old workers and high moral hazard industries hire young workers, and no employer would like to switch workers if the zero profit condition holds. It remains to be shown that reservation utilities V, v, and F can be found such that the labor market clears and the product market clears.

We sketch the proof of the proposition here and leave the details to the appendix (in particular, in the main text we assume that $p(\cdot)$ is continuous and g has at least one-sided derivatives). First, note that the local change in maximal attainable profits from a young worker, $\pi(v, g(k), k)$, is

$$d\pi(v, g(k), k) = p_o(k)\, dg(k) - dk. \tag{4.1}$$

This is an envelope-argument: employers in both industries k and $k + dk$ pick contracts to maximize profits, and so the change in maximal profits is determined by the change in parameters outside employers' control, namely k and $g(k)$. Similarly, there exists a positive-valued function $n(\cdot)$ such that "normalized" young worker profits vary with k according to

$$d\left(\frac{\Pi(k)}{n(k)}\right) = p_y(k)\, dg(k) - dk. \tag{4.2}$$

Comparing (4.1) and (4.2), young workers and high stakes (high k and $g(k)$) are *strategic complements* by Lemma 4. That is, young workers are more valuable in high-k industries. Employers with more at stake have more need to incentivize workers to work harder, and hence benefit more from the reduction in moral hazard that dynamic contracts provide. In equilibrium, they can therefore outbid employers with less at stake for the young workers. This is the key economic idea behind Proposition 2.

Slightly more formally, consider an industry $\hat{k}$ in which zero-profits are attainable using old workers, and also using young workers. We know $p_y\left(\hat{k}\right) > p_o\left(\hat{k}\right)$ from Lemma 4. So at industry $\hat{k}$, normalized young worker profits Π/n must cross old-worker profits π from below. In other words, normalized young worker profits Π/n exceed (are below) old-worker profits π in industries above $\hat{k}$ (below $\hat{k}$). Because in our conjectured equilibrium young worker profits are zero above $\hat{k}$, and old worker profits are zero below $\hat{k}$ (i.e., (ZP1) and (ZP2) hold),

20

this in turn implies that industries below $\hat{k}$ would lose money hiring young workers ($\Pi < 0$), while industries above $\hat{k}$ would lose money hiring old workers ($\pi < 0$).

V. Industries hiring young workers

Thus far, we have shown that industries with a lot at stake (i.e., $k \geq \hat{k}$) hire young workers, and that even the least rewarded young worker is better off than the luckiest old worker. We next consider how the terms on which young workers are employed differ across industries $\left[\hat{k}, \bar{k}\right]$. First, we show that increasing the probability f of firing after failure always increases profits if the participation constraint of the worker is ignored:

Lemma 5. *Either young workers are always fired after period 1 failure ($f = 1$) or else receive exactly their reservation utility V.*

This follows from the fact that in equilibrium, the employer has negative profits in the second period: $\pi(v, g, k) < 0$ for $k > \hat{k}$ from Proposition 2 and the zero profit condition. The only reason to retain a worker in the second period is either as a reward for success after the first period, or to satisfy the ex ante participation constraint.

We next characterize the profit-maximizing contract when the worker's participation constraint is absent and attention is restricted to *firing* contracts, i.e., $f = 1$. The only contract term to consider is the utility reward after success, v_s: write $p(v_s)$ for the associated period 1 success probability, where from the incentive constraint (IC2),

$$v_s - F = \gamma'(p(v_s)).$$

Formally, the profit-maximizing firing contract solves

$$\max_{v_s \geq u(p_{SB}(g), 0)} p(v_s)(g + \pi(v_s, g, k)). \tag{5.1}$$

Let $U(g, k)$ denote the worker's utility when employed under the profit-maximizing firing contract, i.e.,

$$U(g, k) = p(v_s) v_s + (1 - p(v_s)) F - \gamma(p(v_s)),$$

where v_s is the maximizing value of (5.1).[18] The profit-maximizing firing contract is impor-tant because, in equilibrium, it is used in the highest-k industries. More precisely, it is used in industries above k^+, where k^+ is defined by[19]

$$U\left(g^+, k^+\right) = V,$$

where g^+ is such that industry k^+ makes zero profits using the profit-maximizing contract.

Proposition 3. *In equilibrium, the expected utility of young workers in industries $k \in \left[\hat{k}, k^+\right)$ is V, where $\frac{V}{2}$ is strictly higher than the utility of the luckiest fired worker. Both the expected utility of young workers and their average success probability are strictly increasing in k over $\left[k^+, \bar{k}\right]$. Workers in these industries are always fired after failure: $f = 1$.*

The per-period utility of young workers is higher than that of old workers. There are two reasons. First, they can be incentivized using the promise of promotion, and in equilibrium they retain some of the extra surplus this delivers (Corollary 3). Second, *because* they are easier to incentivize, in equilibrium they are hired by the industries with the greatest moral hazard problems, which means they end up with additional surplus in industries $k \geq k^+$ (Proposition 3).

Finally, we consider how the optimal contract varies across the young worker industries $k > \hat{k}$. From Lemma 1, we already know that average work and wages increase (weakly) with k, and from Proposition 3 we know these relations are strict for $k > k^+$. For industries between $\hat{k}$ and k^+, worker utility is constant at V. In contrast to workers in the "top" industries $k > k^+$, workers in these intermediate industries may be retained even after period 1 failure:

---

[18]Differentiating (5.1), a small change in v_s affects employer profits by

$$\frac{g + \pi\left(v_s, g, k\right)}{\gamma''\left(p\left(v_s\right)\right)} + p\left(v_s\right)\pi_v\left(v_s, g, k\right). \tag{5.2}$$

Lemma 2 and Assumption 1 imply that if (5.2) is weakly negative for some v_s, it is strictly negative for all higher values. Hence the objective (5.1) is single-peaked in v_s, and has a unique maximizer.
[19]The proof of Proposition 3 ensures that k^+ exists and is unique.

Lemma 6. *For $\hat{k}$ small enough, we have $f < 1$ for the marginal industry hiring young workers.*

Proof: In Appendix.

The reader should recall that workers in our model are risk neutral. Given this, Lemma 6's conclusion that workers who fail on the first task may nonetheless be retained is perhaps surprising. As the lemma suggests, retention is most easily explained in industries close to the marginal young worker industry $\hat{k}$. For these industries, there exists a one period contract that firms can give to retained old workers that gives the firm roughly zero profits, and gives the worker a utility v_f strictly above the utility F of a fired worker who must look for a new job when old. Consequently, at almost no direct cost, a firm can increase a worker's utility if he fails in period 1; the firm can then strictly increase its profits by decreasing the compensation given to a worker if he succeeds. However, retention does have an indirect cost, namely that it reduces a worker's effort in his first period. Although in general the net effect of the direct gain and the indirect cost is ambiguous, the lemma shows that when $\hat{k}$ is small enough, the indirect cost is likewise small and hence the net effect is positive.

For any industry in which retention after failure ($f < 1$) does take place, the following result characterizes the worker's effort in the second period:

Lemma 7. *Young workers work strictly more after success than if they are retained after failure.*

Proof: In Appendix.

FIGURE 1 ABOUT HERE

These two lemmas together with Proposition 1, Proposition 3, and Lemma 3 illustrate how work, job security, and promotion structure vary across industries depending on the

degree of moral hazard. In the industries with the least at stake, $k \leq k^-$, work hours are limited and constant across industries, while in low- to intermediate-k industries ($k^- < k < \hat{k}$) work and pay increase strictly with the amount at stake, as does the surplus to workers. All these industries hire only old workers and so do not use any dynamic incentive contracts. For industries with intermediate to high moral hazard problems ($\hat{k} < k < k^+$), dynamic incentive contracts are used in which after success workers get promoted to positions with more job responsibility, while after failure workers are either fired or get demoted to positions with less job responsibility. Employees in these industries work harder than in industries hiring old workers. For really high-profile jobs ($k \geq k^+$) work and pay is strictly increasing in k and workers are always fired after failure, so job security is the lowest.

Figure 1 summarizes how the surplus of workers varies over industries in equilibrium. Old fired workers are excluded from the high-profile labor market, and earn strictly lower rents than even the unluckiest young workers. Labor markets for both old and young workers are lotteries, where the lucky workers end up in high moral hazard sectors and earn higher rents. Our equilibrium features both regions in which high pay is a compensating differential for worse work conditions (industries where $\hat{k} < k < k^+$) and regions in which pay is more than a compensating differential (industries where $k^- < k < \hat{k}$ or $k \geq k^+$). In general, though, the most attractive jobs feature the longest work hours and the lowest job security.

Before formally establishing equilibrium existence in Section VII, we analyze a number of important labor market phenomena that our model allows us to study.

VI. Labor market equilibrium: features and extensions

A. Lucky cohorts: temporary industry shocks have life-long effects

Oyer (2008) shows that temporary shocks to Wall Street that affect the number of workers hired in a year have big and life-long effects on the careers of the MBA students who are on the margin of getting hired by an investment bank. Relative to an MBA student who gets an investment banking job, an otherwise identical student who doesn't because he is unlucky enough to graduate in a year when Wall Street is down has a loss of life-time income of up to 5 million dollars in present value terms. He is also very unlikely to enter investment

banking later in life, even if Wall Street is booming. Oyer finds it hard to explain this with differences in skill or preferences. Instead, there seem to be a large element of randomness in who ends up on Wall Street and who does not. Oyer tentatively suggests that the difference in income is not a skill premium but rather a compensating differential for the hours, risk, travel, and other factors that go with working on Wall Street.

Our model provides an explanation for the wage differential, the importance of initial conditions, and the stickiness of careers documented by Oyer, without appeal to either skill differences, development of specific human capital, or other switching costs. Imagine a temporary shock in the demand function for services in the top moral hazard industry $(k = \bar{k})$ in our model, which leads to one less worker being hired. This worker, who instead ends up in a random industry in $\left[\hat{k}, \bar{k}\right)$, can expect a significantly lower life-time income. Furthermore, his chance to get into a higher-profile industry is gone – as he gets older, he will either stay in his industry or move to a lower k industry. This is because he grows relatively unattractive to high moral hazard industries as he ages, because he becomes harder to incentivize.

Consistent with Oyer's findings, this worker also avoids the long hours and risks associated with the top moral hazard industry (where the firing probability is one in case of failure). However, it is not the case that the high pay is set as a compensating differential for the gruelling work conditions. Instead, the causation goes the other way – the fact that rents are so high leads employers to create work conditions that partly eat up some of those rents (see the following subsection). The job is still attractive, though – not only is life-time income substantially higher, but life-time *utility* is as well.

Finally, we comment briefly on the quantitative interpretation of our analysis. For expositional ease we have examined only two-period dynamic contracts, and so one period in our model lasts 15-20 years — much longer than the length of typical recessions. We would obtain similar results from a model with more periods, where a period is 3-5 years (the typical length of a business cycle). However, the skeptical reader might then wonder whether an age difference of one period has a quantitatively relevant effect on employers' hiring decisions. In this respect, we wish to make three related points. First, our model is intended to apply to relatively homogenous sectors of the labor market, in which skill

differences among workers are relatively small (in our basic model, they are nonexistent). This seems a reasonable description of pools of graduates from top universities, and implies that even small differences in the efficiency of contracting can play a decisive role in hiring conditions. Second, to the extent to which workers are not homogenous, and instead differ in skill, we show below that it is not necessarily the case that employers want to employ the highest skill workers. In this sense, skill may be a less decisive factor in hiring decisions than sometimes believed (again, within pools of already relatively homogenous applicants). Third, while Oyer suggests that direct switching costs and human capital accumulation are unlikely to be significant factors, dynamic contracts postpone some of the rewards until later periods, and generate potentially important endogenous switching costs (this point is made elsewhere in the literature).

B. Multiple tasks: Dog years and promotion

Lemma 3 shows that promotion leads to more work - p goes up. This is attractive to the worker, as he earns higher rents when he works more. One should keep in mind that the extra work is on an *important* task, that is, one where the marginal productivity of labor is very high and the moral hazard rents are correspondingly high.

Now imagine that there is an extra task, which we call the *menial* task, that can also be performed in the organization. For example, this could involve gathering data, preparing spreadsheets, copying papers, or fetching burgers for more senior employees. The menial task is also easily monitored: the employer can simply stipulate how much of the menial task it wants a worker to do.

Proceeding a little more formally, we take the equilibrium of the economy without menial tasks, and then introduce menial tasks to a null set of industries (this allows us to hold the overall structure of the equilibrium unchanged). The marginal product of a worker in industry k with a success probability p^* is $p'\left(\gamma\left(p^*\right)\right)g\left(k\right)$.[20] To ensure that the menial task is truly menial, we assume that if a worker spends time m on the menial task he produces εm, where ε is small enough to ensure that a worker's marginal product in the menial task

---

[20]Recall that $p\left(h\right)$ is the success probability if a worker works h.

is below his marginal product in the important task in any industry and in any period. A worker can work on both the menial and important tasks: his total hours worked is $\gamma(p) + m$, which must be less than H, his total time endowment.

The following result is then immediate:

Proposition 4. *Consider an industry k in which the menial task is available. If the worker's participation constraint binds, the menial task is never used. Likewise, retained young workers never perform the menial task. However, young workers in industries above k^+ and fired old workers in industries above k^- perform the menial task, up to the point where either their participation constraint or time endowment constraint binds. Total hours worked $(m + \gamma(p))$ is increasing in k across industries with the menial task.*

We want to stress two features of this result. First, for industries employing young workers, the menial task is only used in the early stage of the career. If the worker is promoted, he is assigned only to important tasks. The reason is that in the second period, the worker must be promised some surplus to motivate work in the first period, so extracting surplus from the worker in the second period is counterproductive. This explanation for why workers graduate to more important tasks is distinct from theories based on learning or screening of talent, since those elements are absent in our model.

Second, since the menial task is used as an inefficient surplus extraction mechanism, its use will be concentrated in high-profile industries. This is our "dog years" result: In high-profile industries, there are typically very long hours early on in the career, much of which is spent on less prestigious tasks. This can be a second best solution even when work hours are inefficiently long, and even when the menial task can be performed better or cheaper with less qualified workers.

C. Distortions in the allocation of talent

We now add (observable) heterogeneity in talent between workers to study how talent is allocated across industries. Our aim is to explain two opposite versions of misallocation of talent, which we call the "talent lured" and the "talent scorned" phenomena. First, it is often argued that high prestige jobs draw in people whose talents would be more appropriately

used elsewhere. For example, many people bemoan the large number of engineers and scientists who have been "lured" into investment banking. Second, misallocation may go in the opposite direction – the most talented people do not always land the most prestigious jobs, even when they prefer those jobs, and even if their skill in a first-best world would be more valuable in those jobs. Colloquially, very skilled people are sometimes viewed as being "too difficult to handle" or "too hard-to-manage."

Our model lends some support to both to the "talent lured" and the "talent scorned" views, as we next explain.

C.1. Talent lured

To keep as close as possible to our basic model, we introduce differences in talent by assuming that only a null set of workers have higher skills, while the remaining workers are homogenous as before. This assumption ensures that the basic structure of the equilibrium remains unchanged. Specifically, suppose that a null set of workers are unusually suited for working in some select industry k_1 in that they have a cost $c_1 \gamma (p)$ of achieving success probability p, where $c_1 < 1$. Also, suppose this is not one of the highest prestige industries – for example, suppose k_1 is lower than k^+ (the top industry for which the participation constraint of young workers still binds), but higher than $\hat{k}$ (the marginal industry hiring old workers). Furthermore, suppose these workers are also slightly more efficient in the most prestigious industry $\bar{k}$ in that they have a cost $c_2 \gamma (p)$ of achieving success probability p, where $c_1 < c_2 < 1$. We assume c_2 is close enough to 1 so that the marginal productivity of these talented workers is higher in industry k_1, even given the higher price of output in industry k_2. For simplicity, assume that their cost of effort in other industries is equal to that of other workers.

The maximization of social welfare (measured as the sum of output and worker utility) dictates that these workers should be employed in industry k_1 where their extra talent has the biggest effect. It is then easy to see that if the talent advantage relative to regular workers is not too large, then this is not the outcome. Industry $\bar{k}$ employers reward their workers with so much surplus that they will outbid industry k_1 employers for these workers. This feature of our model is very much in line with popular impressions of investment banks

hiring away talented scientists from research careers. Moreover, note that this prediction emerges only because workers in industries above $\hat{k}$ endogenously receive more than their reservation utility.

C.2. Talent scorned

The "talent lured" effect is inherently static, as it relies only on some industries giving more than the reservation utility of workers. The "talent scorned" effect, on the other hand, depends crucially on having a dynamic general equilibrium model as it works through the outside option of fired workers. Essentially, the idea is that a talented worker has higher outside options when fired and therefore responds less to dynamic incentives.

To formalize this idea, assume now that a null set of workers have a cost $\tilde{c}\gamma(p)$ of achieving success probability p in any industry, where $\tilde{c} < 1$. Because output is most valuable in the most prestigious industry $\bar{k}$, social efficiency dictates that these talented workers should all enter this industry when young. However, this is not the equilibrium outcome when $1 - \tilde{c}$ is sufficiently small, as we now show.

When a talented worker is unlucky when young and is fired, he is employed in industry $\hat{k}$, the most prestigious of the industries hiring old workers. Consequently, his expected utility after being fired is $u(\hat{k})$, whereas the payoff of a regular worker when fired is $F < u(\hat{k})$. Therefore, a talented worker is harder to incentivize ("difficult," or "hard-to-manage"), precisely because his talent gives him a better outcome if he fails. This effect dominates the direct effect of talent whenever $1 - \tilde{c}$ is sufficiently small, and ensures that industry $\bar{k}$ employers will not hire the most talented workers. Note that this is truly talent scorned, because the talented worker would prefer to get the prestigious job and enjoy the high surplus.

D. Moral hazard in booms and busts

Our model has implications for how moral hazard problems vary over the business cycle. Consider the effect of a "boom" that raises the demand for output in all industries. The standard effect of this demand increase is to raise v, the utility of old workers in industries between $\underline{k}$ and k^-, in order to ensure labor market clearing. In equilibrium, this is associated with an increase in work in these industries, so that success probabilities rise.

At the same time, however, the increase in demand has no *direct* effect on contract terms or effort in industries where a worker's participation constraint is non-binding, namely industries between k^- and $\hat{k}$, and above k^+. (Of course, the increase in demand does affect the *number* of workers employed in these sectors.) But the demand increase does have an *indirect* effect on the most prestigious industries, i.e., those above k^+: since the utility of old workers is now higher in industries below k^-, the utility of a fired worker, F, is now higher. This makes young workers harder to incentivize, and simultaneously increases their pay[21] (much of it in the form of bonuses) and the price of output in the most prestigious sectors, while decreasing effort and success probabilities. Consequently, if higher failure rates in the most prestigious industries cause a boom to end, booms may contain the seeds of their own destruction. This is a potential explanation for the financial crisis starting in 2007, which was preceded by an unprecedented boom in the financial industry.

VII. Equilibrium existence

Our analysis characterizes properties of an equilibrium in which young and old workers are hired by industries below and above $\hat{k}$, respectively. Conditional on existence, Proposition 2 shows that any equilibrium must be of this form. Before concluding, we tie-up a (large) loose end: we show that an equilibrium actually exists.

We start by formalizing the definition of an equilibrium for our economy. In addition to notation already defined, for industries $k > \hat{k}$ hiring young workers, let $n(k)$ be the expected number of tasks a young worker expects to perform: formally, if p_1 is the worker's first period effort in industry k, then $n(k) = 1 + p_1 + (1 - p_1)(1 - f(k))$. Observe that the worker's unconditional probability of being retained is simply $n(k) - 1$. For industries $k < \hat{k}$ hiring only old workers, we write $n(k) = 1$.

An equilibrium is a set

$$\left\{ \hat{k}, v, V, F, g\left(\cdot\right), u\left(\cdot\right), n\left(\cdot\right), p\left(\cdot\right), r\left(\cdot\right), \lambda_o\left(\cdot\right), \lambda_y\left(\cdot\right) \right\}$$

---

[21]To see this, note that an increase in F reduces $p(v_s)$ and hence $p\gamma''(p(v_s))$ (by Assumption 1). The equilibrium success payoff g is also increased. From the derivative expression (5.2) in footnote 18, it follows that the profit-maximizing v_s is greater.

satisfying the following conditions, all of which we have already discussed (at least informally):

No-entry and no-poaching: (ZP1) and (ZP2) hold for $k < \hat{k}$ and $k > \hat{k}$ respectively. Moreover, no industry above $\hat{k}$ wants to hire old workers, and no industry below $\hat{k}$ wants to hire young workers.

Profit maximization: For $k < \hat{k}$, $u(k)$ and $p(k)$ are determined by the profit-maximizing contract for old workers. For $k > \hat{k}$, $n(k)$, $u(k)$ and $p(k)$ are determined by the profit-maximizing contract for young workers.

Consistency of v, V, F: There are industries k_o and k_y employing old and young workers, respectively, such that that $u(k_o) = v$ and $u(k_y) = V$; and F is determined by (3.1).

Labour market clearing: (LD1) holds for $k < \hat{k}$; analogously,

$$g(k) = \zeta_k (n(k) p(k) \lambda_y(k)) \qquad \text{(LD2)}$$

for $k > \hat{k}$. For $k > \hat{k}$, demand for old workers is determined by retention policies, $\lambda_o(k) = (n(k) - 1) \lambda_y(k)$. Aggregate demand for old and young workers matches aggregate supply of old and young workers, $\int_{\underline{k}}^{\bar{k}} \lambda_o(k)\, dk = \frac{1}{2}\overline{\lambda}$ and $\int_{\hat{k}}^{\overline{k}} \lambda_y(k)\, dk = \frac{1}{2}\overline{\lambda}$.

To establish the existence of an equilibrium, we proceed as follows.

First, as discussed in Section II, for a given $\hat{k}$ and v there are unique specifications for $k < \hat{k}$ of $u(\cdot)$, $p(\cdot)$, $g(\cdot)$, $\lambda_o(\cdot)$ and F that satisfy the equilibrium conditions.

Second, given $\hat{k}$, v and F, define young worker reservation utility by the requirement that industry $\hat{k}$ employers are indifferent between hiring young and old workers, $\Pi\left(V, F, g\left(\hat{k}\right), \hat{k}\right) = \pi\left(v, g\left(\hat{k}\right), \hat{k}\right) = 0$.

Third, given V and F, use (ZP2) to determine the success payoff $g(k)$ for industries $k > \hat{k}$. Given V, F, $g(k)$, the quantities $u(\cdot), n(\cdot), p(\cdot), r(\cdot)$ follow from profit-maximization, and $\lambda_y(k)$ follows from (LD2).

Proposition 2 ensures that the "no-poaching" condition is satisfied. Consequently, we are left with just two parameters to determine, $\hat{k}$ and v, and just the two aggregate labor market clearing conditions to check.

In the Appendix, we establish that it is possible to choose $\hat{k}$ and v to satisfy these two

remaining conditions, and thereby establish equilibrium existence.

In brief, the main problem handled in the Appendix is that changes in $\hat{k}$ and v potentially affect demand for old and young workers non-monotonically. For example, increasing v reduces the demand for old workers in low-k industries; but it also increases F and hence changes the probability that they are fired from industries above $\hat{k}$ in the first place. Similarly, raising $\hat{k}$ both increases the demand for old workers in industries below $\hat{k}$, and raises F, again changing the firing probability in industries above $\hat{k}$. We circumvent these difficulties by establishing a fixed-point theorem which relies only on the excess demand for labor being strictly positive (respectively, negative) whenever the reservation utility of old workers is sufficiently low (respectively, high).

VIII. Conclusion

We have analyzed a general equilibrium labor market model that we think applies particularly well to workers in jobs where the exact link between effort and output is hard to measure. Although we have cast this within an effort model, we think the principles apply to other types of moral hazard as well, such as stealing. We think these problems are especially relevant for the types of jobs sought by MBA students, such as consultancy, investment banking, or general management. We have explained several features of wages, career paths, and contracts in these types of jobs, and how these features covary with the attractiveness of the job. In particular, jobs characterized by higher moral hazard problems where more value is at stake, such as investment banking, will have longer work hours, steeper career paths, higher risk of firing, but also higher compensation. They are also more attractive because in spite of the gruelling work conditions, they give workers higher utility, even when there are no skill differentials between workers.

We have also shown the value of being young for landing high-profile jobs. Being young makes it possible to use future work as collateral, which makes it easier to incentivize the worker early on. Young workers are therefore especially attractive to high moral hazard industries, and if a worker fails to get a job in such an industry early on he will have a very hard time entering later.

When we introduce observable skill differences into the model, we can explain two much discussed phenomena of talent misallocation – the "talent lured" and the "talent scorned" effects. "Talent lured" is when workers whose talents would be socially more valuable in lower moral hazard industries get poached by high moral hazard employers, such as investment banks. This effect appears in our model because the high surplus offered by prestigious industries makes it possible for them to outbid for talent even when that talent is better used somewhere else. "Talent scorned" is when intrinsically more skilled workers do not land the highest profile jobs, even though they apply for them. This happens in our model because more talented workers, when fired, are assured of landing in the best possible job in the segment of the economy that does not use dynamic incentive schemes. This higher outside opportunity implies that firing is not as much of a threat to these workers. Therefore, they become hard to manage for an employer that relies heavily on firing incentives for controlling moral hazard.

Finally, we show that moral hazard problems vary in an interesting way over the business cycle. In a boom, as competition for workers increases, industries in which moral hazard problems are relatively small to start with will have to increase the utility they give to workers. This utility is best used to have the workers work harder, which makes them more successful at their tasks – in this sense, moral hazard goes down for this segment of the economy. In contrast, employers with the highest moral hazard problems who use dynamic incentive schemes to control workers do not have to increase worker utility to successfully compete for workers. The increased outside opportunities of fired workers leads to lower incentives in these industries – moral hazard is increased – even at the same time as bonuses increase.

In addition to high-profile jobs such as investment banking, our analysis also has possible implications for the market for CEOs. Under this interpretation, high-profile jobs correspond to CEO-track careers: individuals who succeed in a senior management position at a large firm may be promoted to CEO, while those who do less well are either demoted, or else leave to take a management job with a smaller firm. That said, in other respects our model is an imperfect fit for the CEO market. In particular, in our model an employer can double his profits by hiring twice as many workers, and all jobs are the same: both features are hard to

interpret in the context of senior management positions. Because of this, we leave a fuller exploration of our general framework's implications for CEOs for future research.

As we emphasized in the introduction, the only friction in our model is a standard moral hazard problem. This assumption both aids tractability, and allows us to isolate the effects of a single feature of the economic environment. Nonetheless, it is worth considering how our results would change if workers *also* differed in unobservable skill levels in the manner of Holmström (1999).[22] We conjecture that this extension of our model would largely reinforce our existing results. In the second period, firms with a lot at stake (high k) would seek to employ workers with high perceived abilities. Holmström's "career concerns" effect would then give workers an incentive to work hard in the first period so as to obtain a job in a high-stakes industry, which would be highly compensated for exactly the same reasons as in our existing model. Consequently, high-stakes firms would still exhibit a preference of employing workers at the start of their careers, for parallel reasons to those in our existing model, namely that the dynamic aspect of the agency problem makes young workers easier to motivate than older workers. Again, we leave a fuller exploration of this extension of our model for future research.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, 1999, High Wage Workers and High Wage Firms, *Econometrica* 67, 251-333.

Acemoglu, Daron, 2001, Good jobs versus bad jobs, *Journal of Labor Economics* 19, 1–22.

Acemoglu, Daron, and Robert Shimer, 2000, Wage and technology dispersion, *The Review of Economic Studies* 67, 585–607.

Acharya, Viral V., and Paolo F. Volpin, forthcoming, Corporate governance externalities, *Review of Finance.*

Akerlof, George A. and Lawrence F. Katz, 1989, Workers' trust funds and the logic of wage profiles, *The Quarterly Journal of Economics* 104, 525–536.

Baranchuk, Nina, Glenn MacDonald, and Jun Yang, 2008, The economics of super managers,

---

[22]We are unaware of any paper that embeds Holmström's framework in an equilibrium model.

Indiana University Working Paper.

Becker, Gary S., and George J. Stigler, 1974, Law Enforcement, Malfeasance, and Compensation of Enforcers, *Journal of Legal Studies*, 3, 1-18.

Bulow, Jeremy I., and Lawrence H. Summers, 1986, A theory of dual labor markets with application to industrial policy, discrimination, and Keynesian unemployment, *Journal of Labor Economics* 4, 376–414.

Cahuc, Pierre, and André Zylberberg, 2004, Labor Economics, MIT Press.

Carmichael, Lorne, 1985, Can Unemployment Be Involuntary? Comment, *American Economic Review 75*, 1213-1214.

Dicks, David L., 2009, Executive Compensation, Incentives, and the Role for Corporate Governance Regulation, *University of North Carolina Working Paper.*

Edmans, Alex, Xavier Gabaix, and Augustin Landier, forthcoming, A multiplicative model of optimal CEO incentives in market equilibrium, *Review of Financial Studies.*

Gayle , George-Levi, and Robert A. Miller, forthcoming, Has moral hazard become a more important factor in managerial compensation?, *American Economic Review.*

Gibbons, Robert,and Michael Waldman, 1999, A Theory of Wage and Promotion Dynamics inside Firms, *Quarterly Journal of Economics* 114, 1321-1358.

Groshen, Erica L., and Alan B. Krueger, 1990, The Structure of Supervision and Pay in Hospitals, *Industrial and Labor Relations Review*, 43, 134S-146S.

Hutchens, Robert, 1986, Delayed payment contracts and a firm's propensity to hire older workers, *Journal of Labor Economics* 4, 439–457.

Kotlikoff Laurence J., and Jagadeesh Gokhale, 1992, Estimating a Firm's Age-Productivity Profile Using the Present Value of Workers' Earnings, *Quarterly Journal of Economics*, 107, 1215-1242.

Krueger, Alan B., and Lawrence H. Summers, 1988, Efficiency wages and the inter-industry wage structure, *Econometrica* 56, 259–293.

Krueger, Dirk and Harald Uhlig, 2006, Competitive risk sharing contracts with one-sided commitment, *Journal of Monetary Economics*, 53, 1661–1691.

Lazear, Edward P., 1981, Agency, earnings profiles, productivity, and hours restrictions, *The American Economic Review* 71, 606–620.

Lazear, Edward, and Robert L. Moore, 1984, Incentives, productivity and labor contracts, *Quarterly Journal of Economics*, 99, 275-296.

MacDonald, Glenn M., 1988a, The economics of rising stars, *American Economic Review* 78, 155-166

MacDonald, Glenn M., 1988b, Job mobility in market equilibrium, *Review of Economic Studies* 55, 153-168.

MacLeod, W. Bentley, and James M. Malcomson (1998), Motivation and markets, *American Economic Review* 88, 388–411.

Medoff , James L., and Katharine G. Abraham, 1980, Experience, performance, and earnings, *The Quarterly Journal of Economics* 95, 703–736.

Moen, Espen R., and Åsa Rosén, 2007, Incentives in competitive search equilibrium, *Stockholm University working paper.*

Moen, Espen R., and Åsa Rosén, 2006, Deferred compensation and turnover, *Stockholm University working paper.*

Moen, Espen R., and Åsa Rosén, 2006, Equilibrium incentive contracts and efficiency wages, *Journal of European Economic* Association 4, pages 1165-1192.

Murphy, Kevin M., and Robert H. Topel, 1987, "Unemployment, Risk, and Earnings: Testing for Equalizing Wage Differences in the Labor Market," in Unemployment and the Structure of the Labor Market, ed. by Kevin Lang and Jonathan S. Leonard. New York: Basil Blackwell.

Oyer, Paul, 2008, The making of an investment banker: Stock market shocks, career choice, and life-time income, *Journal of Finance* 63 2601-2628.

Phelan, Christopher, 1995, Repeated moral hazard and one-sided-commitment, *Journal of Economic Theory*, 66, 488-506.

Rogerson, William P., 1985, Repeated moral hazard, *Econometrica*, 53, 69-76.

Shapiro, Carl, and Joseph E. Stiglitz, 1984, Equilibrium unemployment as a worker discipline device, *The American Economic Review* 74, 433–444.

Terviö, Marko, 2009, Superstars and mediocrities: market failure in the discovery of talent, *Review of Economic Studies* 76, 829–850.

Tsuyuhara, Kunio, 2009, Repeated Moral Hazard with Worker Mobility via Directed On-

the-Job Search, *Univeristy of Toronto Working Paper.*

Walsh, Frank , 1999, A multisector model of efficiency wages, *Journal of Labor Economics* 17, 351–376.

Yellen, Janet L., 1984, Efficiency Wage Models of Unemployment, *American Economic Review* 74, 200-205.

Appendix. Proofs omitted from main text

Proof of Lemma 1:

$g(k) - k$ *is strictly increasing:* Zero profits in industry k_1 implies

$$\bar{p}(k_1) g(k_1) - \overline{x}(k_1) - k_1 = 0. \tag{1.1}$$

Hence for any k_2,

$$
\begin{aligned}
\bar{p}(k_2) g(k_1) - \overline{x}(k_2) - k_1 &\leq 0 \\
&= \bar{p}(k_1) g(k_1) - \overline{x}(k_1) - k_1 \\
&= \bar{p}(k_2) g(k_2) - \overline{x}(k_2) - k_2 \tag{1.2}
\end{aligned}
$$

where the inequality follows from the fact that, in equilibrium, an employer in k_1 cannot make strictly positive profits using the contractual arrangements of industry k_2. Hence

$$\bar{p}(k_2) (g(k_2) - g(k_1)) \geq k_2 - k_1. \tag{1.3}$$

Inequality (1.3) can be rewritten as

$$\bar{p}(k_2) ((g(k_2) - k_2) - (g(k_1) - k_1)) \geq (1 - \bar{p}(k_2)) (k_2 - k_1),$$

which immediately implies that $g(k) - k$ is strictly increasing in k (from the conditions on γ, the equilibrium success probability always lies in $(0,1)$).

$\bar{p}(k)$ *and* $\overline{x}(k)$ *weakly increasing:* From (1.2),

$$(\bar{p}(k_2) - \bar{p}(k_1)) g(k_1) \leq \overline{x}(k_2) - \overline{x}(k_1).$$

Interchanging k_1 and k_2 yields

$$(\bar{p}(k_2) - \bar{p}(k_1)) g(k_1) \leq \overline{x}(k_2) - \overline{x}(k_1) \leq (\bar{p}(k_2) - \bar{p}(k_1)) g(k_2), \tag{1.4}$$

and hence

$$0 \leq (\bar{p}(k_2) - \bar{p}(k_1)) (g(k_2) - g(k_1)).$$

If $k_2 > k_1$, then $g(k_2) > g(k_1)$ from above, and so $\bar{p}(k_2) \geq \bar{p}(k_1)$. Inequality (1.4) then implies $\overline{x}(k_2) \geq \overline{x}(k_1)$.

Continuity of g: Fix $k \in \left[\underline{k}, \bar{k}\right]$, and any $\delta > 0$. We show there exists some $\varepsilon > 0$ such that if $\left|\tilde{k} - k\right| \leq \varepsilon$, $\left|g\left(\tilde{k}\right) - g(k)\right| \leq \delta$. Choose $\varepsilon > 0$ such that $\varepsilon < \bar{p}\left(\frac{k+k}{2}\right)\delta$ and $k - \varepsilon > \frac{k+k}{2}$. If $\tilde{k} < k$ then observe that (1.3) (with $k_1 = k$ and $k_2 = \tilde{k}$) implies $g(k) - g\left(\tilde{k}\right) \leq \left(k - \tilde{k}\right)/\bar{p}\left(\tilde{k}\right) \leq \delta\bar{p}\left(\frac{k+k}{2}\right)/\bar{p}\left(\tilde{k}\right) \leq \delta$, since $\tilde{k} > \frac{k+k}{2}$ and p is weakly increasing. If instead $\tilde{k} > k$ then observe that (1.3) (with $k_1 = \tilde{k}$ and $k_2 = k$) implies $g\left(\tilde{k}\right) - g(k) \leq \left(\tilde{k} - k\right)/\bar{p}(k) \leq \delta\bar{p}\left(\frac{k+k}{2}\right)/\bar{p}(k) \leq \delta$. ∎

Proof of Lemma 3: Let p and $p(v_s)$ respectively denote effort in period 1 and effort in period 2 after period 1 success. First, if $u(p_{FB}(g), 0) > v_s$, then from Lemma 2, $p(v_s)$ is given implicitly by

$$v_s = p(v_s)\gamma'(p(v_s)) - \gamma(p(v_s)),$$

or equivalently,

$$\frac{v_s + \gamma(p(v_s))}{p(v_s)} = \gamma'(p(v_s)). \tag{1.5}$$

Certainly

$$v_s - ((1-f)v_f + fF) < \frac{v_s + \gamma(p(v_s))}{p(v_s)}.$$

Substituting (1.5) and (IC2) into this last inequality implies $\gamma'(p) < \gamma'(p(v_s))$, and hence $p < p(v_s)$. Second, if instead $u(p_{FB}(g), 0) \leq v_s$, then from Lemma 2, $p(v_s) = p_{FB}(g)$, the

first-best level. For this case, the result in the Lemma is immediate. ■

Proof of Lemma 4: If $p_o(k) = p_{SB}(g)$ the result is immediate: the young worker's success probability must be at least $p_{SB}(g)$ in the first period and in the second period after failure, and by Lemma 3 the success probability in the second period after success is strictly greater than $p_{SB}(g)$.

The remainder of the proof deals with the case $p_o(k) > p_{SB}(g)$. From Lemma 2, the old worker receives his reservation utility v. For the young worker, consider any profit maximizing contract. Let f be the firing probability after failure, and p be the first period success probability. So the young worker expects to work on $2 - (1 - p)f$ tasks. By the convexity of effort costs, the young worker's total expected cost of effort is at least $(2 - (1 - p)f)\gamma(p_y(k))$. Denote his actual effort cost per task by $\widetilde{\gamma}(p_y(k)) > \gamma(p_y(k))$, with strict inequality from the convexity of effort and from Lemma 3, as long as effort is not first best in all states. The employer must pay the young worker

$$
\begin{aligned}
& V - (1 - p)fF + (2 - (1 - p)f)\widetilde{\gamma}(p_y(k)) \\
= & (2 - (1 - p)f)\left(\frac{\widetilde{V}}{2} + \widetilde{\gamma}(p_y(k))\right),
\end{aligned}
$$

where
$$
\widetilde{V} \equiv 2\frac{V - (1 - p)fF}{(2 - (1 - p)f)}
$$

Note that since $V \geq 2F$, we have $\widetilde{V} \geq V$. The profit from this contract is

$$
(2 - (1 - p)f)\left(p_y(k)g - \frac{\widetilde{V}}{2} - \widetilde{\gamma}(p_y(k)) - k\right).
$$

Suppose contrary to the claim in the lemma that at the profit maximizing contract, $p_y(k) \leq p_o(k)$. By the assumption that the industry makes non-negative profits from the optimal contract, we then have

$$
p_y(k)g - \frac{\widetilde{V}}{2} - \widetilde{\gamma}(p_y(k)) - k \geq 0.
$$

Suppose the contract is replaced by a repeated one-period contract with work $p_y(k)$ each

period. This gives profits

$$2\left(p_y\left(k\right)g - \frac{V}{2} - \gamma\left(p_y\left(k\right)\right) - k\right) > \left(2 - \left(1 - p\right)f\right)\left(p_y\left(k\right)g - \frac{\widetilde{V}}{2} - \widetilde{\gamma}\left(p_y\left(k\right)\right) - k\right).$$

This is a feasible contract since $\frac{V}{2} \geq v$. The strict inequality above follows since $p_y\left(k\right) < p_{FB}\left(g\right)$, and hence $\gamma\left(p_y\left(k\right)\right) < \widetilde{\gamma}\left(p_y\left(k\right)\right)$; and moreover, $V \leq \widetilde{V}$, and $2 \geq 2 - \left(1 - p\right)f$. Since this is a strict improvement, the contract with $p_y\left(k\right) \leq p_o\left(k\right)$ cannot have been optimal in the first place, completing the proof of the Lemma. ∎

Proof of Proposition 2:

As a preliminary, note that the profits in industry k_1 from employing an old worker optimally exceed those attained employing the old worker on industry $k_2 \neq k_1$ terms. Formally, let $\overline{x}_o\left(k\right)$ be the expected compensation paid to an old worker in industry k using the profit maximizing contract. So

$$
\begin{aligned}
\pi\left(v, g\left(k_1\right), k_1\right) &\geq p_o\left(k_2\right)g\left(k_1\right) - \overline{x}_o\left(k_2\right) - k_1 \\
&= p_o\left(k_2\right)g\left(k_2\right) - \overline{x}_o\left(k_2\right) - k_2 \\
&\quad - \left(p_o\left(k_2\right)\left(g\left(k_2\right) - g\left(k_1\right)\right) - \left(k_2 - k_1\right)\right),
\end{aligned}
$$

and hence

$$\pi\left(v, g\left(k_2\right), k_2\right) - \pi\left(v, g\left(k_1\right), k_1\right) \leq p_o\left(k_2\right)\left(g\left(k_2\right) - g\left(k_1\right)\right) - \left(k_2 - k_1\right). \tag{1.6}$$

Interchanging k_2 and k_1 gives

$$\pi\left(v, g\left(k_2\right), k_2\right) - \pi\left(v, g\left(k_1\right), k_1\right) \geq p_o\left(k_1\right)\left(g\left(k_2\right) - g\left(k_1\right)\right) - \left(k_2 - k_1\right). \tag{1.7}$$

For young workers, let $n\left(k\right)$ denote the expected number of tasks a young worker will perform

in industry k. Then analogous arguments imply

$$\Pi\left(V,F,g\left(k_{2}\right),k_{2}\right)-\Pi\left(V,F,g\left(k_{1}\right),k_{1}\right) \quad \leq \quad n\left(k_{2}\right)p_{y}\left(k_{2}\right)\left(g\left(k_{2}\right)-g\left(k_{1}\right)\right)-\left(k_{2}-k_{1}\right) \quad (1.8)$$

$$\Pi\left(V,F,g\left(k_{2}\right),k_{2}\right)-\Pi\left(V,F,g\left(k_{1}\right),k_{1}\right) \quad \geq \quad n\left(k_{1}\right)p_{y}\left(k_{1}\right)\left(g\left(k_{2}\right)-g\left(k_{1}\right)\right)-\left(k_{2}-k_{1}\right) \quad (1.9)$$

Claim 1: $\pi\left(v,g\left(k\right),k\right)<0$ for all $k>\hat{k}$.

Proof of Claim 1: To establish Claim 1, we show that if $\pi\left(v,g\left(k_{1}\right),k_{1}\right)\leq 0$ for some $k_{1}\geq\hat{k}$, then $\pi\left(v,g\left(k_{2}\right),k_{2}\right)<\pi\left(v,g\left(k_{1}\right),k_{1}\right)$ for all $k_{2}>k_{1}$ sufficiently close to k_{1}. Consider any $k_{2}>k_{1}\geq\hat{k}$. Inequality (1.9), together with the hypothesis $\Pi\left(V,F,g\left(k_{1}\right),k_{1}\right)=\Pi\left(V,F,g\left(k_{2}\right),k_{2}\right)=0$, implies

$$p_{y}\left(k_{1}\right)\left(g\left(k_{2}\right)-g\left(k_{1}\right)\right)-\left(k_{2}-k_{1}\right)\leq 0.$$

Since g is strictly increasing in k (Lemma 1), inequality (1.6) implies that for any $k_{2}>k_{1}\geq\hat{k}$ such that $p_{o}\left(k_{2}\right)<p_{y}\left(k_{1}\right)$,

$$\pi\left(v,g\left(k_{2}\right),k_{2}\right)<\pi\left(v,g\left(k_{1}\right),k_{1}\right).$$

By Lemma 4, $p_{o}\left(k_{1}\right)<p_{y}\left(k_{1}\right)$.[23] Since $p_{o}\left(k\right)$ is continuous in k (recall g is continuous in k), $\pi\left(v,g\left(k_{2}\right),k_{2}\right)<\pi\left(v,g\left(k_{1}\right),k_{1}\right)$ for all $k_{2}>k_{1}$ sufficiently close to k_{1}.

Claim 2: $\Pi\left(V,F,g\left(k\right),k\right)\leq 0$ for all $k\leq\hat{k}$.

Proof of Claim 2:

Suppose to the contrary that there exists $k_{1}<\hat{k}$ such that an employer can make strictly positive profits employing a young worker, i.e., $\Pi\left(V,F,g\left(k_{1}\right),k_{1}\right)>0$. Since $\Pi\left(V,F,g\left(\hat{k}\right),\hat{k}\right)=0$, and $\Pi\left(V,F,g,k\right)$ is continuous in k and g, and g is continuous in k (see Lemma 1), it follows that there exists $k_{2}\in(k_{1},\hat{k}]$ such that $\Pi\left(V,F,g\left(k\right),k\right)\geq 0$ for all $k\in[k_{1},k_{2}]$, with equality at k_{2}. For any $k\in[k_{1},k_{2}]$, inequality (1.9) implies

$$\Pi\left(V,F,g\left(k_{2}\right),k_{2}\right)\geq\Pi\left(V,F,g\left(k\right),k\right)+\left(k_{2}-k\right)\left(p_{y}\left(k\right)\frac{g\left(k_{2}\right)-g\left(k\right)}{k_{2}-k}-1\right).$$

---

[23] Since $\pi\left(v,g\left(\hat{k}\right),\hat{k}\right)=0$, we know from the discussion following Lemma 2 that $v<u\left(p_{FB}\left(g\left(\hat{k}\right)\right),0\right)$. Since g is strictly increasing (Lemma 1), Lemma 2 implies $p_{o}\left(k_{1}\right)<p_{FB}\left(g\left(k_{1}\right)\right)$. So Lemma 4 applies.

The profits $\Pi\left(V, F, g\left(k\right), k\right)$ are weakly positive. As k approaches k_2 from below, the term $p_y\left(k\right)\frac{g(k_2)-g(k)}{k_2-k} - 1$ converges to $p_y\left(k\right)\frac{d^-}{dk}g\left(k\right) - 1$. From (1.6) and (1.7), and the hypothesis that $\pi\left(v, g\left(\cdot\right), \cdot\right)$ is zero below $\hat{k}$, $\frac{d^-}{dk}g\left(k\right) = 1/p_o\left(k\right)$. From Lemma 4, $p_y\left(k\right) > p_o\left(k\right)$ for all $k \in [k_1, k_2]$, and moreover, it is easy to see from the proof of that result that $p_y\left(k\right)$ is bounded away from $p_o\left(k\right)$. Hence for all $k < k_2$ sufficiently close to k_2, $\Pi\left(V, F, g\left(k_2\right), k_2\right) > 0$, giving a contradiction and completing the proof. ∎

Proof of Lemma 5: Consider a contract $\left(v_s, v_f, f\right)$, where $f < 1$, and let p be the associated probability of success in the first period. By hypothesis, $\pi\left(v_f, g\left(k\right), k\right) \le 0$ (see main text). Hence the employer's payoff after first period success must be enough to cover k, i.e., $p\left(g\left(k\right) + \pi\left(v_s, g\left(k\right), k\right)\right) \ge k > 0$. Consider the effect of a small increase in the firing probability f. Differentiating, the effect on the employer's profits is

$$\frac{dp}{df}\left(g\left(k\right) + \pi\left(v_s, g\left(k\right), k\right) - \left(1 - f\right)\pi\left(v_f, g\left(k\right), k\right)\right) - \left(1 - p\right)\pi\left(v_f, g\left(k\right), k\right). \quad (1.10)$$

We know v_f is at least $u\left(p_{SB}\left(g\left(k\right)\right), 0\right)$, which in turn is strictly greater than $u\left(\hat{k}\right)$ and hence strictly greater than F. Hence $\frac{dp}{df} > 0$, and so expression (1.10) is strictly positive, implying the result. ∎

Proof of Proposition 3:

The proof of Proposition 3 uses the following notation: For any $k \in \left[\hat{k}, \bar{k}\right]$, let $\hat{g}\left(k\right)$ be the price such that the profit-maximizing firing contract gives exactly zero profits. It also uses the following result:

Lemma 8. *Fix an industry $k \ge \hat{k}$, a level of worker utility $u > V$, and a price $g = \hat{g}\left(k\right)$ (i.e., profit-maximizing firing contract gives zero profits). Then g and u together satisfy the industry-k equilibrium condition if and only if $u = U\left(g, k\right)$.*

Proof of Lemma 8:

"IF": Any firing contract other than the profit-maximizing firing contract gives strictly negative profits. Suppose that, contrary to the claimed result, g and $u = U\left(g, k\right)$ fail the industry-k equilibrium condition. So there must exist a contract $\left(v_s, v_f, f\right)$ with $f < 1$

42

that generates non-negative profits. By Lemma 5, this implies there is a firing contract generating strictly positive profits, giving a contradiction.

"ONLY IF": Suppose to the contrary that g and u together satisfy the industry-k equilibrium condition, but that $u \neq U(g, k)$. If $u < U(g, k)$ the industry-k equilibrium conditions are certainly violated, since it is possible to increase worker utility and still generate non-negative profits. If $u > U(g, k)$ then there is a contract that delivers non-negative profits and utility u. By the stated conditions, this contract cannot be a firing contract, and so must feature $f < 1$. But then Lemma 5 implies there exists a contract that gives non-negative profits, and delivers worker utility strictly above V. The contradiction completes the proof. ∎

Given Lemma 8, it is sufficient to show that $U(\hat{g}(k), k)$ is continuous and strictly increasing. This is equivalent to showing the maximizing value v_s of (5.1) is continuos and strictly increasing. The maximizing value is determined by the first-order condition of setting (5.2) to equal zero (see footnote 18). The price $\hat{g}(k)$ is defined so $p(v_s)(\hat{g}(k) + \pi(v_s, \hat{g}(k), k)) = k$, where v_s is the maximizing value of (5.1). Substituting, the maximizing value v_s of (5.1) solves

$$\frac{k}{p(v_s)^2 \gamma''(p(v_s))} + \pi_v(v_s, g, k) = 0.$$

The solution is continuous and strictly increasing in k, completing the proof. ∎

Proof of Lemma 6: Suppose $f = 1$ for the marginal case. The problem of the employer is, taking the price $\underline{g}$ as given from the old worker marginal sector:

$$\underline{g} = \gamma'(\underline{p}) + \underline{p}\gamma''(\underline{p})$$

where $\underline{p}$ is defined by

$$k = \underline{p}^2\gamma''(\underline{p}).$$

The employer's problem is:

$$\max_{v_s} p\left(v_s\right) \left(\underline{g} + \pi\left(v_s, \underline{g}\right)\right) - k$$

such that agent participation condition holds condition holds:

$$p\left(v_s\right) v_s + \left(1 - p\left(v_s\right)\right) F - \gamma\left(p\left(v_s\right)\right) \geq 2\underline{v},$$

where

$$\underline{v} = \underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right),$$

and the incentive condition holds:

$$v_s - F = \gamma'\left(p\left(v_s\right)\right)$$

The zero profit condition is

$$p\left(v_s\right) \left(\underline{g} + \pi\left(v_s, \underline{g}\right)\right) = k.$$

Now, consider a local change in f, with v_s also changed to hold worker utility constant. Worker utility given by:

$$pv_s + \left(1 - p\right)\left(1 - f\right) v_f + \left(1 - p\right) f F - \gamma\left(p\right),$$

where:

$$v_s - \left(1 - f\right) v_f - fF = \gamma'\left(p\right),$$

so we have:

$$\frac{dv_s}{df} = \frac{1 - p}{p}\left(v_f - F\right).$$

The derivative of profit with respect to f (moving v_s to hold worker utility constant) is

hence:

$$\frac{dp}{df}\left(g+\pi\left(v_s,g\right)-\left(1-f\right)\pi\left(v_f,g\right)\right)+p\pi_v\left(v_s,g\right)\frac{1-p}{p}\left(v_f-F\right)-\left(1-p\right)\pi\left(v_f,g\right).$$

At $f=1$, substituting in the zero-profit condition, together with $\pi_v\left(v_s,g\right)=-1$ (we show that this holds below) gives

$$\frac{dp}{df}\frac{k}{p}-\left(1-p\right)\left(v_f-F\right)-\left(1-p\right)\pi\left(v_f,g\right).$$

Let k_0 denote the marginal case. For this case, there is a v_f such that $\pi\left(v_f,g\right)=0$. Note that $\frac{dp}{df}$ is given by:

$$\frac{dp}{df}=\frac{\frac{dv_s}{df}+v_f-F}{\gamma''\left(p\right)}=\frac{v_f-F}{\gamma''\left(p\right)p},$$

and since $\frac{dp}{df}>0$, we would like to show:

$$k_0<\gamma''\left(p\right)p^2\left(1-p\right).$$

We know that:

$$k_0=\underline{p}^2\gamma''\left(\underline{p}\right),$$

so we want to show:

$$\underline{p}^2\gamma''\left(\underline{p}\right)<p^2\gamma''\left(p\right)\left(1-p\right),$$

or:

$$\frac{\underline{p}^2\gamma''\left(\underline{p}\right)}{p^2\gamma''\left(p\right)}<1-p.$$

Suppose $p\to0$, which is the case when $k\to0$. .Then, it is enough to show that $\underline{p}/p\leq\frac{1}{1+\lambda}$ for some $\lambda>0$. At $f=1$, the worker's utility is:

$$pv_s+\left(1-p\right)F-\gamma\left(p\right)$$
$$=\ p\left(v_s-F\right)-\gamma\left(p\right)+F$$
$$=\ p\gamma'\left(p\right)-\gamma\left(p\right)+F.$$

45

This has to exceed:

$$2\underline{v} = 2\left(\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)\right),$$

so:

$$p\gamma'\left(p\right) - \gamma\left(p\right) + F \geq 2\left(\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)\right).$$

Suppose $F \leq (1 - \lambda)\underline{v}$. Then, we have to have

$$\left(p\gamma'\left(p\right) - \gamma\left(p\right)\right) \geq (1 + \lambda)\left(\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)\right),$$

i.e.,

$$\frac{\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)}{p\gamma'\left(p\right) - \gamma\left(p\right)} \leq \frac{1}{1 + \lambda}.$$

We want to show that as $k \to 0$,

$$\frac{\underline{p}^2\gamma''\left(\underline{p}\right)}{p^2\gamma''\left(p\right)} < 1 - p.$$

It is sufficient to show that

$$\frac{\underline{p}^2\gamma''\left(\underline{p}\right)}{p^2\gamma''\left(p\right)} \leq \left(1 + \frac{\lambda}{2}\right)\frac{\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)}{p\gamma'\left(p\right) - \gamma\left(p\right)},$$

i.e., that

$$\frac{\underline{p}^2\gamma''\left(\underline{p}\right)}{\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)} \leq \left(1 + \frac{\lambda}{2}\right)\frac{p^2\gamma''\left(p\right)}{p\gamma'\left(p\right) - \gamma\left(p\right)}.$$

The limit of both $\frac{p^2\gamma''(p)}{p\gamma'(p) - \gamma(p)}$ and $\frac{\underline{p}^2\gamma''\left(\underline{p}\right)}{\underline{p}\gamma'\left(\underline{p}\right) - \gamma\left(\underline{p}\right)}$ is

$$2 + \lim_{p \to 0}\frac{p\gamma'''}{\gamma''}.$$

So the result follows provided that $\lim_{p \to 0}\frac{p\gamma'''}{\gamma''}$ is finite, which we assume.

We also have to show that as k goes to zero, it is indeed true that $\pi_v\left(v_s, g\right) = -1$ if $f = 1$. This amounts to showing that p_g as defined by:

$$\gamma'\left(\underline{p}\right) + \underline{p}\gamma''\left(\underline{p}\right) = \gamma'\left(p_g\right),$$

is smaller than p_v as defined by:

$$p_v \gamma'\left(p_v\right) - \gamma\left(p_v\right) = v_s,$$

where:

$$v_s - F = \gamma'\left(p\left(v_s\right)\right).$$

So, in other words, p_v is defined by:

$$p_v \gamma'\left(p_v\right) - \gamma\left(p_v\right) = \gamma'\left(p\right) + F.$$

We also know that $p > \underline{p}$. So we have:

$$
\begin{aligned}
\gamma'\left(p_g\right) &= \gamma'\left(\underline{p}\right) + \underline{p}\gamma''\left(\underline{p}\right), \\
\gamma'\left(p_v\right) &= \frac{\gamma'\left(p\right) + \gamma\left(p_v\right) + F}{p_v}.
\end{aligned}
$$

Close to 0, we assume that $\gamma''\left(\underline{p}\right)$ is bounded away from zero and bounded, and we assume that γ' goes to zero. Dividing one with the other we have

$$\frac{\gamma'\left(p_g\right)}{\gamma'\left(p_v\right)} = p_v \frac{\gamma'\left(\underline{p}\right) + \underline{p}\gamma''\left(\underline{p}\right)}{\gamma'\left(p\right) + \gamma\left(p_v\right) + F}.$$

Suppose contrary to the claim that $p_g > p_v$ so that $\frac{\gamma'(p_g)}{\gamma'(p_v)} > 1$. We then must have:

$$\frac{\gamma'\left(\underline{p}\right) + \underline{p}\gamma''\left(\underline{p}\right)}{\gamma'\left(p\right) + \gamma\left(p_v\right) + F} \rightarrow \infty,$$

since $p_v \rightarrow 0$ as $k \rightarrow 0$. But we know $\underline{p} < p < p_v$, so we have:

$$\frac{\gamma'\left(\underline{p}\right) + \underline{p}\gamma''\left(\underline{p}\right)}{\gamma'\left(p\right) + \gamma\left(p_v\right) + F} < \frac{\gamma'\left(p\right) + p\gamma''\left(p\right)}{\gamma'\left(p\right) + \gamma\left(p_v\right) + F} < 1 + p\frac{\gamma''\left(p\right)}{\gamma'\left(p\right)}.$$

Since we have assumed that $p\frac{\gamma''(p)}{\gamma'(p)}$ is bounded, the result follows. ∎

The following lemma shows that there is no fixed payment in period 1 in the dynamic problem:

LEMMA A1: *In equilibrium, the fixed payment w in the first period must be zero in industries hiring only young workers.*

Proof: First, if the participation constraint is not binding, it is obvious that $w = 0$. Suppose instead that the participation constraint is binding and $w > 0$, and consider a perturbation of the contract in which w is changed by an infinitesimal amount $dw < 0$ while v_s is changed by $dv_s = -dw/p$ (where p is the agent's optimal first-period effort under the original contract). By the envelope theorem this perturbation leaves the agent's utility unchanged, while it increases profits by

$$dp \left(g + \pi \left(v_s \right) - \left(1 - f \right) \pi \left(v_f \right) \right) + p\pi_v \left(v_s \right) dv_s - dw$$
$$= dp \left(g + \pi \left(v_s \right) - \left(1 - f \right) \pi \left(v_f \right) \right) - dw \left(\pi_v \left(v_s \right) + 1 \right),$$

where $dp > 0$ is the change in first-period effort associated with the perturbation. From the equilibrium zero-profit condition,

$$g + \pi \left(v_s \right) - \left(1 - f \right) \pi \left(v_f \right) = \frac{1}{p} \left(k - \left(1 - f \right) \pi \left(v_f \right) \right).$$

Since $\pi \left(v_f \right) \leq 0$ in an industry employing young workers, and $\pi_v \left(v_s \right) + 1 \geq 0$ from Lemma 2, it follows that the change in profits is strictly positive. So a contract with $w > 0$ for young workers will never be used in equilibrium. ■

Proof of Lemma 7: By Lemma 2, it is sufficient to show $v_s > v_f$.

Suppose to the contrary that $v_s \leq v_f$. We want to show that this leads to a contradiction in that we can then increase v_s and decrease v_f and strictly increase profits without lowering agent utility. For this perturbation to be possible, we need that v_f is strictly bigger than the lower bound $\max \left(F, u \left(p_{SB} \left(g \right), 0 \right) \right)$, so that lowering v_f. is possible. To show that, suppose to the contrary that $v_s = v_f = \max \left(F, u \left(p_{SB} \left(g \right), 0 \right) \right) = F$. Then, since v_s is then also equal to F, the agent gets less than $2F$ in utility, which contradicts Corollary 3. Next, suppose that $v_s = v_f = \max \left(F, u \left(p_{SB} \left(g \right), 0 \right) \right) = u \left(p_{SB} \left(g \right), 0 \right)$. If this is an optimal contract, profits

are given by

$$p\left(g + \pi\left(v_s, g, k\right)\right) + \left(1 - p\right)\left(1 - f\right)\pi\left(v_f, g, k\right) - k = 0.$$

Note that since $\pi\left(v_f, g, k\right) < 0$, this means that $p\left(g + \pi\left(v_s, g, k\right)\right) > 0$. We show that increasing v_s increases profits strictly. The derivative of profits with respect to v_s is given by

$$\frac{dp}{dv_s}\left(g + \pi\left(v_s, g, k\right)\right) + p\pi_v\left(v_s, g, k\right).$$

Since $v_s = u\left(p_{SB}\left(g\right), 0\right)$ we have $\pi_v\left(v_s, g, k\right) = 0$, and since $\frac{dp}{dv_s} > 0$ and $g + \pi\left(v_s, g, k\right) > 0$, the derivative is strictly positive, and hence $v_s = v_f = u\left(p_{SB}\left(g\right), 0\right)$ cannot be the profit maximizing contract. Hence, we must have $v_f > \max\left(F, u\left(p_{SB}\left(g\right), 0\right)\right)$ if $v_s \leq v_f$. Suppose this is the case. Then, an increase of v_s and simultaneous decrease of v_f that keeps agent utility constant has

$$\frac{\partial v_f}{\partial v_s} = -\frac{p}{\left(1 - f\right)\left(1 - p\right)}.$$

Such a perturbation changes the profits by

$$\frac{\partial p}{\partial v_s}\left(g + \pi\left(v_s\right) - \left(1 - f\right)\pi\left(v_f\right)\right) + p\left(\frac{\partial \pi\left(v_s\right)}{\partial v_s} - \frac{\partial \pi\left(v_f\right)}{\partial v_f}\right).$$

This is strictly positive, since $\frac{\partial p}{\partial v_s} > 0$, since $g + \pi\left(v_s\right) > 0$ from (ZP2), since $\pi\left(v_f\right) \leq 0$, and since

$$\frac{\partial \pi\left(v_s\right)}{\partial v_s} - \frac{\partial \pi\left(v_f\right)}{\partial v_f} \geq 0.$$

This last inequality holds since $\pi\left(v\right)$ is decreasing and concave in v. Hence the perturbation leads to strictly higher profits, giving a contradiction and completing the proof. ∎

Appendix. Equilibrium existence

In the main text we have seen how to construct a candidate equilibrium given a specification of v and $\hat{k}$. It remains only to choose v and $\hat{k}$ so that labor markets clear.

It is slightly easier to change variables and work with k^- instead of v. Recall that k^- is the highest industry employing old workers in which old workers receive exactly their reservation utility v. Moreover, in industry k^- the contract that maximizes profits ignoring

the worker's participation constraint nonetheless gives the worker exactly his reservation utility v. Algebraically, this implies that[24]

$$v = u\left(p_{SB}^-, 0\right) \text{ where } \left(p_{SB}^-\right)^2 \gamma''\left(p_{SB}^-\right) = k^-. \tag{2.1}$$

So there is a one-to-one mapping between reservation utility v and the industry k^- that divides old workers into those who receive more than v and those who do not.

Define the excess demand for old and young workers by

$$
\begin{aligned}
\Lambda_o\left(k^-, \hat{k}\right) &= \int_0^{\bar{k}} \lambda_o\left(k\right) dk - \frac{\lambda}{2} \\
\Lambda_y\left(k^-, \hat{k}\right) &= \int_{\hat{k}}^{\bar{k}} \lambda_y\left(k\right) dk - \frac{\lambda}{2}.
\end{aligned}
$$

Both Λ_o and Λ_y are continuous functions of k^- and $\hat{k}$.

Lemma 9. *For all $\underline{k}$ sufficiently small, if $k^- = \underline{k}$ there is excess demand for labor for all $\hat{k}$, i.e., $\Lambda_o\left(k^-, \hat{k}\right) + \Lambda_y\left(k^-, \hat{k}\right) > 0$ for all $\hat{k} \in \left[\underline{k}, \bar{k}\right]$.*

We assume $\underline{k}$ is low enough for the conclusion of Lemma 9 to hold. In addition, we assume that if the reservation utility of old workers (and hence young workers also) is high, then total labor demand is less than total labor supply $\bar{\lambda}$, regardless of whether employers hire young or old workers. Formally:

Assumption 2. *If $k^- = \bar{k}$, total demand for workers is less than total labor supply $\bar{\lambda}$, i.e., $\Lambda_o\left(k^-, \hat{k}\right) + \Lambda_y\left(k^-, \hat{k}\right) < 0$ for all $\hat{k} \in \left[\underline{k}, \bar{k}\right]$.*

We establish:

Proposition 5. *Under the stated assumptions, there exist k^-, $\hat{k} \in \left(\underline{k}, \bar{k}\right)$ such that labor markets clear, i.e., $\Lambda_o\left(k^-, \hat{k}\right) = \Lambda_y\left(k^-, \hat{k}\right) = 0$.*

---

[24]In detail, in industry k^-, the following three identities hold: $u\left(p_{SB}, 0\right) = v$; $g = p_{SB}\gamma''\left(p_{SB}\right) + \gamma'\left(p_{SB}\right)$ (the definition of p_{SB}); and $p_{SB}g - p_{SB}\gamma'\left(p_{SB}\right) = k^-$ (the zero-profit condition). The second and third of these imply $p_{SB}^2\gamma''\left(p_{SB}\right) = k^-$.

Lemma 9 and Assumption 2 say that

$$\Lambda_o\left(k^-,\hat{k}\right) + \Lambda_y\left(k^-,\hat{k}\right) \; > \; 0 \text{ if } k^- = \underline{k}, \text{ all } \hat{k} \in \left[\underline{k},\bar{k}\right]$$
$$\Lambda_o\left(k^-,\hat{k}\right) + \Lambda_y\left(k^-,\hat{k}\right) \; < \; 0 \text{ if } k^- = \bar{k}, \text{ all } \hat{k} \in \left[\underline{k},\bar{k}\right]$$

Moreover, if all industries hire young workers $(\hat{k} = \underline{k})$ then certainly excess demand for young workers exceeds excess demand for old workers,

$$\Lambda_y\left(k^-,\hat{k}\right) \geq \Lambda_o\left(k^-,\hat{k}\right) \text{ if } \hat{k} = \underline{k}, \text{ all } k^- \in \left[\underline{k},\bar{k}\right],$$

while if no industry hires young workers $(\hat{k} = \bar{k})$ the reverse is true,

$$\Lambda_y\left(k^-,\hat{k}\right) < \Lambda_o\left(k^-,\hat{k}\right) \text{ if } \hat{k} = \bar{k}, \text{ all } k^- \in \left[\underline{k},\bar{k}\right].$$

The labor markets for both young and old workers clear if $\Lambda_o\left(k^-,\hat{k}\right) + \Lambda_y\left(k^-,\hat{k}\right) = 0$ and $\Lambda_y\left(k^-,\hat{k}\right) - \Lambda_o\left(k^-,\hat{k}\right) = 0$. The existence of k^- and $\hat{k}$ that clear the markets is a direct consequence of the following fixed-point result, which we prove using Brouwer's fixed-point theorem:

Lemma 10. *Let $f : [0,1]^2 \to \Re^2$ be a continuous function with continuous first derivatives, such that for $i = 1,2$, $f^i(x) \leq 0$ if $x_i = 0$ and $f^i(x) \geq 0$ if $x_i = 1$. Then there exists $x \in [0,1]^2$ such that $f(x) = 0$.*[25]

A. Proofs of Lemmas 9 and 10

Proof of Lemma 9: For any $\hat{k} \in \left[0,\bar{k}\right]$, consider industry $\underline{k}$ as we let $\underline{k}$ approach 0, holding $k^- = \underline{k}$.

To evaluate this limit, we need to be specific on how to handle cases with $\hat{k} < \underline{k}$. For such cases, first define v using (2.1). Then, define a price $g\left(\hat{k}\right)$ for industry $\hat{k}$ using the worker utility condition, $p\left(\hat{k}\right)\gamma'\left(p\left(\hat{k}\right)\right) - \gamma\left(p\left(\hat{k}\right)\right) = v$, and the zero-profit condition, $p\left(\hat{k}\right)g\left(\hat{k}\right) - \gamma'\left(p\left(\hat{k}\right)\right)\hat{k} - \hat{k} = 0$. Let $F = v$. Finally, let V be given by

---

[25] *Lemma 10 extends in the obvious way to functions from $[0,1]^n$ to $\Re^n$.*

$$\Pi\left(V, F, g\left(\hat{k}\right), \hat{k}\right) = \pi\left(v, g\left(\hat{k}\right), \hat{k}\right) = 0.$$

It follows that for any $\hat{k}$, as $\underline{k} \to 0$, the output price $g(\underline{k})$ approaches 0. Consequently, the total amount produced in industry $\underline{k}$ must grow without bound as $\underline{k} \to 0$ (recall that $\zeta(x) \to 0$ as $x \to \infty$), and hence the labor employed in industry $\underline{k}$ must also grow without bound. Because Λ_o and Λ_y are continuous, and $[0, \bar{k}]$ is compact, it follows that there exists some $\underline{k} > 0$ such that $\Lambda_o\left(k^-, \hat{k}\right) + \Lambda_y\left(k^-, \hat{k}\right) > 0$ if $k^- = \underline{k}$, for all $\hat{k} \in [0, \bar{k}]$. This completes the proof. ■

Proof of Lemma 10: For any pair of strictly positive constants α_1 and α_2, define a new function h over $[1,2]^2$ by $h^i(x) = x_i \exp\left(-\alpha_i f^i(x_1 - 1, x_2 - 1)\right)$ for $i = 1, 2$.

By the construction of h, if $x \in [1,2]^2$ is such that $h(x) = x$ then $f(x_1 - 1, x_2 - 1) = 0$. Consequently, the result follows immediately from Brouwer's fixed point theorem provided that $h\left([1,2]^2\right) \subset [1,2]^2$. We show that it is possible to choose α_1 and α_2 so that this is indeed the case.

We deal with the first dimension of the function, h^1; the second dimension is handled identically. Observe that for any $x_2 \in [1,2]$, $h^1(1, x_2) \geq 1$ and $h^1(2, x_2) \leq 2$. Moreover,

$$\frac{\partial h^1}{\partial x_1} = \exp\left(-\alpha_1 f^1(x_1 - 1, x_2 - 1)\right)\left(1 - \alpha_1 x_1 \frac{\partial}{\partial x_1} f^1(x_1 - 1, x_2 - 1)\right).$$

Since $\frac{\partial f^1(x)}{\partial x_1}$ is continuous over $[0,1]^2$, it is bounded, and so there exists some $\alpha_1 > 0$ such that $\frac{\partial h^1}{\partial x_1}$ is strictly positive over all $[1,2]^2$. Hence for this choice of α_1, $h^1(x) \in [1,2]$ for all $x \in [1,2]^2$, completing the proof. ■
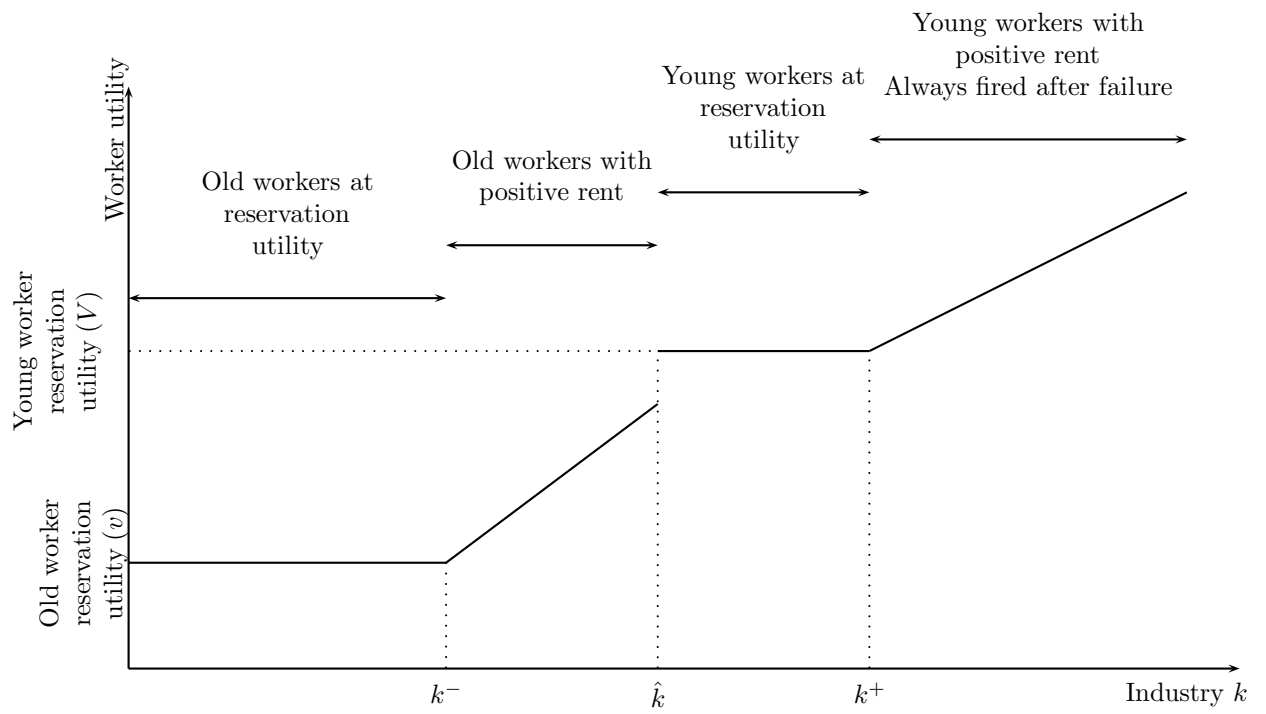
Figure 1: Equilibrium structure