Investment banking careers: An equilibrium theory of overpaid jobs

ULF AXELSON and PHILIP BOND*

October 5, 2011

ABSTRACT

We develop an optimal dynamic contracting theory of overpay for jobs in which moral hazard is a key concern, such as investment banking. Overpaying jobs feature up-or-out contracts and long work hours, yet give more utility to workers than their outside option dictates. Labor markets feature "dynamic segregation," where some workers are put on fast-track careers in overpaying jobs and others have no chance of entering the overpaying segment. Entering the labor market in bad economic times has life-long negative implications for a worker's career both in terms of job placement and contract terms. Moral hazard problems are exacerbated in good economic times, which leads to countercyclical productivity. Finally, workers whose talent would be more valuable elsewhere can be lured into overpaying jobs, while the most talented workers might be unable to land these jobs because they are "too hard to manage".

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

Keywords: Investment Banking, Compensation Contracts

*Ulf Axelson is with the London School of Economics and SIFR, and Philip Bond is with the Carlson School of Management, University of Minnesota. Address correspondence to Ulf Axelson, LSE, Houghton Street, WC2A 2AE London, UK, or e-mail: u.axelson@lse.ac.uk. We want to thank Wouter Dessein, Paolo Fulghieri, Camelia Kuhnen, Alan Morrison, Paul Oyer, Uday Rajan, David Robinson, and seminar audiences at the American Economic Association meetings, BI Oslo, Columbia, Cornell, the European Finance Association meetings, University of Exeter, the Federal Reserve Bank of Philadelphia, University of Frankfurt (Goethe), University of Glasgow, University of Houston, IE, IESE, Jackson Hole Finance Group conference, London School of Economics, MIT, Michigan State University, University of Minnesota, the NBER, Ohio State University, Oxford University, SIFR/SSE, SSE Riga, Temple, and Toulouse IDEI for helpful comments and suggestions. Any errors are our own.

The last few years have seen heated debate about the level of financial sector pay. There is no doubt that financial sector pay is indeed extremely high. Broadly speaking, there are three potential explanations: high pay as a return to skill; high pay as a compensating differential for stressful work conditions; and high pay as overpay—in the sense of being neither a return to skill nor a compensating differential. As we review in more detail below, we think there is substantial evidence against the first two hypotheses. However, a coherent explanation of the third hypothesis—that workers in the financial sector are overpaid relative to their outside options—requires explaining why market entry does not eliminate the pay premium.

In this paper we develop an equilibrium theory of overpay. We build on a strand of an older efficiency wage literature,[1] which points out that a wage premium may exist in one sector of the economy (employed workers), because incentive problems prevent workers from other sectors of the economy (unemployed workers) from bidding these wages down. However, this older literature attracted criticism for its focus on simple wage contracts, and its neglect of the role of dynamic incentives (a criticism broadly know as the "bonding critique").[2] This criticism strikes us as particularly important with respect to the current debate about financial sector pay, because age-compensation profiles are often very steep, consistent with dynamic incentives; and moreover, many policy proposals call for increased use of back-loaded incentive pay.

In this paper we develop a parsimonious dynamic equilibrium model based on the single friction of moral hazard, in which some workers are overpaid relative to other workers, even when firms employ fully optimal dynamic contracts. We further show how this same model matches a variety of empirical observations about both cross-sectional variation of job characteristics, and time-series variation of labor force conditions. All of these predictions hinge crucially on solving for the optimal dynamic contract. For example, our model predicts that overpaid jobs rely heavily on up-or-out promotion, and demand long hours for entry-level workers, often on surprisingly mundane tasks. They are most commonly entered when young, implying that cross-sectional variation in workers' initial employment conditions have long lasting effects. In the time-series, our model predicts that workers who enter the labor force in bad economic times are less likely to get an overpaid job; that even if they do, the overpaid job is worse; and that they work harder, implying countercyclical productivity. We review the empirical evidence supporting these results in the main body of the paper.

[1] See especially Shapiro and Stiglitz (1984). Katz (1986) provides a useful review.

[2] See Katz (1986) for a discussion of the bonding critique.

That overpay persists in a model with optimal dynamic incentives is not a foregone conclusion. The basic rationale for overpay is that when tasks are such that the difference between success and failure is large and effort is unobservable, a profit maximizing firm may find it optimal to use bigger monetary incentives than the outside option of a worker dictates. Dynamic incentives can help to reduce the need for overpay in two ways. First, the firm can backload pay to the end of a worker's career and threaten him with separation in case of failure. Second, tasks can be sequenced such that all workers start out on jobs characterized by relatively low moral hazard, and only gradually get employed on more important tasks as a reward for earlier success. Indeed, the latter is what one would prescribe based on an important insight of contract theory: workers who have built up wealth over time are easier to employ on high moral hazard tasks, because the wealth can be used to acquire an equity stake in the firm and lessen the split between ownership and control that lies at the heart of the moral hazard problem.[3] If all workers in the economy were forced to "work their way up" in this manner, there would be no sense in which some workers are overpaid relative to other workers.

Our key result is to show that when moral hazard problems are severe enough, putting all workers on the same job ladder is suboptimal. Instead, some workers will be singled out for fast-track careers that feature high moral hazard tasks even early on, and these workers are indeed overpaid. Workers who are not lucky enough to be placed on the fast track when young will never get the chance to work on overpaying tasks. We denote this phenomenon "dynamic segregation" of the labor market. Loosely speaking, dynamic segregation reflects a second insight of contracting theory: the prospect of wealth accumulation in the future can be used to ameliorate moral hazard problems in the present.[4] Workers on fast-track careers expect lucrative job placements in case of success, which in turn makes it easier to motivate them to perform difficult tasks early on in their career.

Our basic model has no aggregate uncertainty, and accounts for the existence and characteristics of overpaid jobs. We next examine the effects of aggregate shocks, which allows us to develop time-series implications for job allocation, contract characteristics, and firm productivity. Our model delivers two types of cohort effects, both of which have considerable support in the empirical labor market literature.[5] First, entering the labor market in bad times leads to worse job placement on

[3] The idea that wealth possessed by the agent ameliorates the moral hazard problem dates back at least as far as Jensen and Meckling (1976). Recent papers that explicitly model the reduction in inefficiency associated with the dynamic accumulation of wealth by the agent include DeMarzo and Fishman (2007) and Biais et al (2007).

[4] Early observations of this point include Becker and Stigler (1974), Akerlof and Katz (1989), and Lazear (1981).

[5] See Oyer (2008), Kahn (2010), Baker, Gibbs, and Holmström (1994), and Beaudry and DiNardo (1991).

average because there are fewer overpaid jobs available, and this has life-long effects on a worker's career because of dynamic segregation. Second, even if an entering worker does land one of the few good jobs available in bad times, this job will pay less and the worker's future wages will also be depressed relative to workers who entered the economy in good times. Our model further predicts that productivity in good jobs is countercyclical, for two reasons. First, in bad times a higher fraction of a firm's workers are old, and old workers in our model are (endogenously) more productive. Second, in bad times the threat of being fired is more powerful, leading to greater effort. In contrast, in good times workers are more reckless because they are confident that they will "land on their feet," a prediction that accords well with anecdotal accounts of the recent financial boom.

As an extension, we also analyze how observable differences in talent affect job placement. 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 overpaying firms can outbid other employers for workers even if their talent is wasted in investment banking. The second phenomenon, which we call "talent scorned," is the opposite—overpaying jobs often reject the most talented applicants on the grounds that they are "difficult" or "hard to manage." In our model, this effect arises because talented workers, when fired, have higher outside opportunities.

Finally, a contribution of a more technical nature is to prove existence of equilibrium in an economy with overpay. As we explain in Section VI, the same features of our model that imply overpay also imply that the excess demand correspondence of the economy may fail to be upper-hemi continuous in prices, which considerably complicates the existence proof.

As stated in our opening paragraph, we describe a worker as overpaid if his pay represents neither a return to skill nor a compensating differential, i.e., if his expected utility exceeds that of another worker with identical skills. It is worth highlighting that under this definition the existence of overpaid workers is not necessarily socially inefficient. In particular, since contracts are set optimally in our model, shareholders would not gain by reducing the amount paid to workers. In this, our notion of overpay is very different from the criticisms of executive pay advanced by, for example, Bebchuk and Fried (2004). Also, our model does not imply that the financial sector as a whole is too large, as suggested by, for example, Murphy, Shleifer and Vishny (1991), or more recently by Philippon (2010), or Bolton, Santos and Scheinkman (2011).

We conclude with a discussion of why we believe some jobs are truly overpaid, i.e., why high compensation is neither a return to skill nor a compensating differential. In the particular context of finance jobs, Oyer (2008) and Philippon and Reshef (2008) provide evidence against high pay being a return to skill: Philippon and Reshef (2008) control for unobserved worker characteristics using a fixed effect regression, while Oyer instruments for worker characteristics using aggregate economic conditions when an MBA student graduates. More generally, these conclusions are consistent with a large empirical literature arguing that different jobs pay otherwise identical workers different amounts.[6]

High pay as a compensating differential for bad work conditions may seem a plausible explanation at first sight, since investment banking jobs feature notoriously long hours and low job security. However, these onerous work conditions are chosen by the employer rather than being an intrinsic feature of the job (as they are in, for example, mining). Hence one must explain why employers do not make the job more attractive, rather than paying very high amounts to compensate for unattractive job characteristics of their own choosing.[7] Moreover, Philippon and Reshef (2008) control for hours worked, and still find excess pay in the financial sector. Finally, and less formally, the pay differences between finance and other (themselves high-paying) occupations documented by Oyer and others strike us as too large to be easily explained as compensating differentials; and related, students who obtain investment banking jobs act as if they have won the lottery (consistent with our model) rather than as if the high compensation is a compensating differential.[8]

Related literature: As noted, our paper is related to the efficiency wage literature. Relative to this literature, our contribution is to fully evaluate optimal dynamic contracts for finite-lived agents, and to analyze both how a worker's prospects evolve over his career, and how contracts respond to business cycle conditions. In addition, our main interest is in understanding cross-sectional variation in job characteristics, rather than unemployment;[9] hence our model also features multiple tasks, a further distinguishing feature. Separately, the extensive search literature in labor economics also predicts heterogeneity in wages for homogenous workers.[10] In common with the efficiency wage

[6] See, e.g., Krueger and Summers (1988) and Abowd et al (1999).

[7] In our model, unattractive job characteristics such as low job security and long hours emerge endogenously.

[8] Of course, the compensating differential explanation says only that the *marginal* worker is indifferent. We have yet to meet the marginal student who is just indifferent between receiving and not receiving an investment banking offer.

[9] Bulow and Summers (1986) explore some microeconomic predictions of efficiency wage models. Also, much of the *empirical* efficiency wage literature is concerned with examining whether different industries pay otherwise identical workers different amounts (see, e.g., Abowd et al, 1999).

[10] See, e.g., Mortensen (2003).

literature, this literature largely ignores the possibility of dynamic contracting.

Tervio (2009), in a very interesting and related recent paper, explains high income in a model that builds on talent discovery rather than incentive problems. In his setting, overpay arises because young, untried workers who get a chance to work in an industry where talent is important enjoy a free option: If they turn out to be talented, competition between firms drives up their compensation, while if not, they work in the normal sector of the economy. Firms cannot charge for this option when workers have limited wealth. Hence entry into the sector is limited, and compensation for "proved" talent very high. Because Tervio's main focus is the wage and talent distribution of a sector rather than career dynamics, he does not attempt to explain dynamic segregation: In fact, an important assumption in his model is that a worker can only enter the high-paying sector when young. In contrast, endogenizing dynamic segregation is at the heart of our analysis. In terms of applications, while we find his exogenous dynamic segregation assumption realistic for the entertainment business (which is his main example), this assumption seems less realistic for many professional jobs such as banking, where the skills needed for success are less sector-specific. In contrast, incentive problems strike us as of central importance in the financial sector, and are correspondingly central to our analysis.

Paper outline: The paper proceeds as follows. Section I describes the model. Section II derives the structure of equilibrium contracts. Section III derives the dynamic segregation result, along with the characteristics of career paths in overpaying jobs. Section IV studies the effects of demand shocks on careers and incentives. Section V introduces observable talent differences. Section VI deals with equilibrium existence. Section VII concludes.

I Model

To study the labor market phenomena we are interested in, we need two key elements: Workers of different age, and tasks that vary in their degree of moral hazard problems. There is a continuum of workers of measure 1, and we assume a measure $\frac{1}{2}$ of young workers enter the labor market each period, work for two periods, and then exit. 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 analyze an extension where skills differ across workers in Section V.)

There are two tasks denoted as H (the "high stakes" task) and L (the "low stakes" task). A task can either succeed or fail, where the failure cost is what differs across tasks. For task H, the failure cost is $k_H > 0$, while for task L the failure cost is $k_L = 0$. For each task $i \in \{H, L\}$, we write the success payoff as $g_i - k_i$, where g_i is determined in equilibrium (see below). One way to think about these payoffs is that k_i is an input cost (e.g., funds provided to a trader) and g_i is the value, or market price, of output produced when the task succeeds (e.g., gross value after trading). Alternatively, k_i is the value destroyed if a task fails (e.g., a takeover fails), and $g_i - k_i$ is the value created if a task succeeds (e.g., takeover succeeds). We write g for the price vector (g_L, g_H).

A worker can spend time on one task per period. If a worker spends time h on the task, it succeeds with probability $p(h)$ and fails with probability $1 - p(h)$. Hence, we can think of g_i as the marginal product of labor. Workers have a per-period time endowment of 1, 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 $p'(1) = 0$. 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.[11] 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 $\gamma'(p(1)) = \infty$.

For the case of the financial sector, the following specific interpretation of the moral hazard problem is worth spelling out. The success payoff g_i 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_i are lost. By working h hours, the amount of tail risk a worker needs to achieve his target return is such that the probability of tail risk being realized is $1 - p(h)$.

We make the following assumption on the shape of the effort cost function. Part (i) ensures that a firm's marginal cost of inducing effort is increasing in the effort level. Part (ii) ensures that old workers exert strictly positive effort, even given the agency problem.[12]

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

A firm in the economy can operate one or both tasks. To close the model, we need to determine the aggregate supply of the two tasks. For simplicity, we assume there is free entry and perfect

[11] As formulated, the only difference in the degree of moral hazard in the two tasks stems from $k_H > k_L$, which in equilibrium implies $g_H > g_L$. However, we would obtain qualitatively similar results if instead moral hazard varied due to different costs of effort, or different degrees of observability of output.

[12] Moral hazard means that the marginal cost to the firm of inducing effort for an old worker is $\gamma'(p) + p\gamma''(p)$ (see contracting problem below). Part (ii) of Assumption 1 ensures that this quantity approaches 0 as $p \to 0$.

competition. The output prices g_H and g_L are determined in equilibrium by the standard market clearing condition that excess demand must equal zero. (Alternatively, if task i involves trading financial securities, then g_i is inversely related to how many people are following a given trading strategy.) We write y_i for total task i output. We write ζ_i for the inverse demand curve for task i output, i.e., $\zeta_i(y_i)$ is the price such that total demand is y_i.

As will be clear below, the task L moral hazard problem causes no distortion, since when firm profits are zero, there is enough surplus available for the worker to induce him to exert first-best effort. In this sense, task H is the more interesting task, and in order to focus our analysis we make the simplifying assumption that demand for task L output is perfectly elastic, i.e., $\zeta_L \equiv g_L > 0$. (Our results are qualitively unaffected if this assumption is relaxed; details are available on request from the authors.) For task H output, the demand curve slopes strictly down, i.e., ζ_H is strictly decreasing. We also impose the standard Inada condition that $\zeta_H(y_H) \to \infty$ as $y_H \to 0$.

The above specification of demand is partial equilibrium, in the sense that demand comes from outside the model. Because we view the model as relating to a subset of the labor market, this seems appropriate. Nonetheless, one can show that our model is isomorphic to an alternate model in which demand is determined in general equilibrium.[13]

Finally, as a benchmark, consider an economy where worker effort is observable so that there is no moral hazard problem. Effort is at the first best level where the marginal product of labor g_i is equated with the marginal cost of labor $\gamma'(p_i)$. Since there is free entry, equilibrium prices must be such that the surplus from each task is equalized, i.e., $p_H g_H - \gamma(p_H) - k_H = p_L g_L - \gamma(p_L)$. Firms break even and workers earn the surplus. Task H aggregate output is determined by the market clearing condition $\zeta_H(y_H) = g_H$. Critically, and in contrast to the outcome of the moral hazard economy analyzed below, which task a worker is assigned to over his life time is indeterminate and independent of age and success, and all workers earn the same utility.

II Contracts and equilibrium

Taking output prices g_H and g_L as given, firms compete to hire young workers by offering them employment contracts. Consistent with reality, we rule out indentured labor and model workers as

[13] Specifically, consider the following economy: Workers consume only when old, and have utility $c_L + \ln c_H - \gamma(p_1) - \gamma(p_2)$, where p_1 and p_2 are effort levels in period 1 and period 2 respectively. Task L output is the numeraire good (we normalize $g_L = 1$), and g_H is the relative price of task H output. In the production technology, the cost k_H is paid in task L output. Finally, although c_L is allowed to be negative, workers have limited liability in the sense that $c_L + g_H c_H$ must be nonnegative.

having limited commitment, in the sense that they can walk away from the contract after the first period if another firm is willing to hire them at better conditions. In contrast, we assume that firms are able to commit to contract terms.[14] In other words, we assume *one-sided commitment*.[15]

We allow firms to offer arbitrary dynamic contracts, subject only to the constraint of satisfying one-sided commitment. We allow contracts to specify lotteries, although only ones in which the firm (but not necessarily the worker) is indifferent over lottery outcomes, since any lottery in which the firm is not indifferent would be subject to manipulation by the firm.

Any contract offered in equilibrium must satisfy the following *no-poaching* condition, which we formalize below: there is no alternate contract satisfying one-sided commitment that both strictly raises worker utility and gives a firm strictly positive profits. The no-poaching condition replaces the usual condition that firms maximize worker utility subject to breaking even. We work with the no-poaching condition because the equilibrium of our model often features some workers receiving strictly more utility than others, which is inconsistent with the usual utility maximization condition. This equilibrium feature is exactly the "overpay" of our title.

We next characterize the contracting problem in more detail. In Appendix A, we show that in our setting dynamic contracts can be represented in the following sequential way:

- The first period contract consists of a task assignment $i \in \{L, H\}$, a payment $w_S \geq 0$ to the worker after first-period success, and a payment $w_F \geq 0$ after first-period failure.

- The worker enters the second period with wealth w from his first-period payment, which he uses to "buy" a one-period contract $\{i, w_S, w_F\}$ from the firm such that the firm just breaks even and such that the following second-period no-poaching constraint is satisfied: There is no other contract $\{\tilde{\imath}, \tilde{w}_S, \tilde{w}_F\}$ that the worker can buy from another firm with his wealth w such that the firm makes strictly positive profits and the worker is strictly better off.

This representation of contracts turns out to be quite useful for describing the economics of the model in the most transparent way.[16] But we emphasize that although it is analytically useful to

[14] Concretely, in order for a firm to commit to a long-term contract it is sufficient for the firm to be able to commit to severance payments at the end of the first period, where the size of the severance payment is potentially contingent on the first-period outcome.

[15] See, e.g., Phelan (1995), and Krueger and Uhlig (2006).

[16] It might seem as if this representation imposes a stronger condition than one-sided commitment as it allows the worker to take wealth earned in the first period with him if he walks away from the firm. As we show in the appendix, this is not the case. Allowing the worker to take wealth with him is simply the firm's way of commiting to deliver a certain continuation utility to the worker.

think of dynamic contracts in this way, most actual contracts are likely to instead make use of the equivalent device of partially deferring compensation until the end of the second period.

IIA Incentive contracts for old workers

We start by analyzing contracts for old workers, and then use the solution to the old worker problem to analyze the young worker problem. Using the representation of contracts above, we assume that the old worker enters with wealth w, earned in the first period. The wealth is posted with the firm in exchange for an employment contract on which the firm breaks even. Taking the task assignment as given, the second-period no-poaching constraint implies that the firm must give the worker the maximum possible utility.[17] After solving for this maximal utility given a task assignment, we show how the entering wealth determines the task assignment itself.

Conditional on a task assignment, the contracting problem for the old worker is very standard: A one-period contract consists simply of a payment w_S after success and a payment w_F after failure. This gives the worker utility

$$\max_{w_S,w_F} pw_S + (1-p)\, w_F - \gamma(p)\,,$$

such that

$$\gamma'(p) = w_S - w_F,$$

and

$$p\,(g_i - w_S) - (1-p)\, w_F - k_i + w = 0$$

, and the effort level p is determined by the incentive constraint

$$\gamma'(p) = w_S - w_F. \tag{IC-O}$$

As a benchmark, we define $v_{FBi}(g_i) \equiv \max_p pg_i - \gamma(p) - k_i$ as the maximum—"first-best"—one-period total surplus attainable in task i, conditional on the price g_i. Similarly, define the effort level at the first best as p_{FBi}, given by $\gamma'(p_{FBi}) = g_i$.

First, consider assigning the old worker to task L. Since $k_L = 0$, the firm can pay the full revenue after success, g_L, to the worker and still break even. This makes the worker fully internalize the effects of his effort, so he exerts the first-best effort level p_{FBL}, and the first-best surplus level

[17] Formally, this follows from the fact that, holding the task assignment fixed at $i \in \{L, H\}$, firm profits are either strictly concave or strictly decreasing as a function of worker utility. However, this property does not hold once the choice of task is endogenous.

v_{FBL} is attained. Because first-best surplus is attained even when the worker has no wealth, *a fortiori* first-best surplus is also attained when the worker enters the second period with positive wealth. Hence when an old worker with wealth w is assigned to task L, his expected utility is $v_{FBL} + w$; for use throughout the paper, we denote this by $v_L(w)$.

Second, consider assigning the old worker to task H. If the worker has wealth $w \geq k_H$, he can fund the cost k_H in entirety. In this case, exactly the same argument as for task L applies, and the worker's utility is $v_{FBH} + w$. For lower levels of wealth $w \in [0, k_H)$, the worker can fund only part of the cost k_H. Consequently, the firm must pay strictly less than g_H for success, and the worker exerts strictly less effort than the first-best p_{FBH}. Specifically, the firm pays nothing after failure, and the success "bonus" that a firm must pay to induce effort p is $\gamma'(p)$. To get as close as possible to the first-best effort level, the firm raises the bonus as high as possible subject to satisfying its break-even constraint given that the worker has partially funded the cost k_H, i.e.,

$$p(g_H - \gamma'(p)) - (k_H - w) = 0. \tag{1}$$

Define $p(w)$ as the largest solution to (1). The worker's utility level is then $p(w)\gamma'(p(w)) - \gamma(p(w))$. Note that the firm cannot break even under any contract when the agent's wealth is below the critical level $\underline{w}$, defined as the minimal value w such that equation (1) has a solution in p. We denote an old worker's utility when assigned to task H by $v_H(w)$.

Lemma 1 *The function v_H satisfies: (i) $v_H'(w) > 1$ for $w \in (\underline{w}, k_H)$; (ii) $v_H''(w) \leq 0$, with strict inequality for $w \in (\underline{w}, k_H)$; (iii) $v_H'(w) \to \infty$ as $w \to \underline{w}$; (iv) $v_H'(w)$ decreases in the price g_H.*

The functions v_L and v_H give the maximum utility that an old worker with wealth w can be given if assigned to tasks L and H respectively. We now show how task assignment and ultimate utility is determined. If $w < \underline{w}$, there is no other choice but to employ the worker on task L. If $w > \underline{w}$, the no-poaching condition implies that the firm must assign the worker to whatever task gives higher utility. The only tricky case is when $w = \underline{w}$ and $v_L(\underline{w}) < v_H(\underline{w})$. For this case, even if the firm allocates the worker to the lower utility task L, there is no competing firm that can deliver higher utility to the agent and make strictly positive profits. The firm is therefore free to randomize the task allocation. It may indeed be optimal for the firm to allocate the worker to task L for ex ante incentive reasons.

The case of old workers with wealth $\underline{w}$ illustrates how overpay can emerge with *one-period* contracts. Two equivalent workers with wealth $\underline{w}$ could in principle end up with job placements that give them different utilities ($v_L(\underline{w})$ or $v_H(\underline{w})$). This difference is not eliminated in equilibrium, because a firm employing a worker with wealth $\underline{w}$ on task H cannot break even if it pays the worker less, even if a worker currently employed on task L would gladly agree to such a contract. The reason is that such a contract would lead to inefficiently low effort.[18,19] This economic force is also necessary for moral hazard to generate overpay in dynamic contracts, but as we discuss in depth below, is not sufficient.

To summarize, the utility $v(w)$ as a function of entering wealth of the old worker is determined by the following correspondence, which to reiterate is multivalued only at $w = \underline{w}$:

$$v(w) = \begin{cases} \{v_L(w)\} & \text{for } w \in [0, \underline{w}) \\ [v_L(w), \max\{v_L(w), v_H(w)\}] & \text{at } w = \underline{w} \\ \{\max\{v_L(w), v_H(w)\}\} & \text{for } w > \underline{w} \end{cases} . \tag{2}$$

IIB Incentive contracts for young workers

We next consider the young worker problem. A contract for a young worker specifies a first-period task assignment $i \in \{L, H\}$, and first-period payments to the worker of $w_S, w_F \geq 0$ after first-period success and failure. For the case where either payment equals the threshold wealth $\underline{w}$ consistent with employment on task H when old, the firm also has to pick a continuation utility in the set $v(\underline{w})$ described above; we encompass this by having the firm pick continuation utilities $v_S \in v(w_S)$, $v_F \in v(w_F)$ as choice variables in addition to the payments. Except at $w = \underline{w}$, the effort level p for the young worker is given by the incentive constraint

$$\gamma'(p) = v(w_S) - v(w_F). \tag{IC-Y}$$

(At $\underline{w}$ we write this as $\gamma'(p) = v_S - v_F$.) A contract also has to satisfy the no-poaching condition, which we can now write formally as follows: There does not exist an alternate contract $(\tilde{\imath}, \tilde{w}_S, \tilde{w}_F, \tilde{v}_S, \tilde{v}_F)$ and an effort level $\tilde{p}$ determined by (IC-Y) where $\tilde{p}(g_{\tilde{\imath}} - \tilde{w}_S) - (1-\tilde{p})\tilde{w}_F - k_{\tilde{\imath}} > 0$ (firms make positive profits) and $\tilde{p}\tilde{v}_s + (1-\tilde{p})\tilde{v}_F - \gamma(\tilde{p}) > pv_s + (1-p)v_F - \gamma(p)$ (the worker gets strictly higher utility than under the old contract). The one-sided commitment constraint is

[18] This is essentially the same argument as in Shapiro and Stiglitz (1984) and subsequent papers.

[19] Related, it is straightforward to use a *one-period* version of our model to show that efficiency wages can arise even when firms write output-dependent contracts—a question that provoked some debate in the existing literature, as discussed by Moen and Rosen (2006), who develop a model along these lines. (Although workers live many periods in their model, their informational assumptions make dynamic contracts degenerate, and so the model essentially reduces to a one-period model.)

embodied in the definition of the correspondence $v(w)$ together with $w_S, w_F \geq 0$.

Our formulation of the dynamic contracting problem, using the wealth of the worker as a state variable, is closely related to the standard way of writing dynamic contracting problems using the promised continuation utility of the agent as a state variable (see, for example, Spear and Srivastava (1987) and Green (1987)), where the firm's continuation payoff is the cost-minimizing way of delivering this promised utility. The correspondence v defined in (2) is simply the inverse of the usual mapping from promised utilities to firm costs (see Lemma A-1 in the appendix). In other words, v specifies a worker's promised utility as a function of the firm's cost w; and the firm's cost w is in turn equivalent to paying the worker in cash at the end of the first-period, then recontracting.

The big gain from inverting the usual promised-utilities approach is that it makes the intuition for our main results much easier to give. The cost is that the inversion produces a correspondence so that agent wealth is not always a sufficient statistic for the state. However, in our case the cost is small, because the correspondence is degenerate everywhere except at $\underline{w}$. In fact, whenever contracts specify $w_S, w_F \neq \underline{w}$, it is enough to keep track of only wealth as a state variable. Economically, the simplicity of the correspondence follows from the fact that continuation contracts in our setting are renegotiation proof. Renegotiation proofness is not assumed in our setting, but rather is a consequence of one-sided commitment (see appendix).

We conclude this subsection with a couple of remarks. First, from the definition of v, the minimum utility a firm can threaten a worker with is at least $v_L(0) = v_{FBL} > 0$. Economically, one-sided commitment ensures firms bid up the utility they would give to an old worker with zero wealth to at least this amount.[20]

Second, the definition of v highlights two ways in which higher wealth raises worker utility. One effect is that high wealth reduces inefficiency in the second period, so each extra dollar given to the worker raises his utility by more than a dollar (see Lemma 1), up to the point where the worker has wealth $w = k_H$ and full efficiency is achieved. The second effect is that as wealth crosses the critical level $\underline{w}$, the old worker's employment prospects qualitatively improve, since he can now be assigned to task H as well as task L.

[20] Moreover, when $\underline{w} \leq 0$, an even tighter minimum utility bound may arise.

IIC Equilibrium

Before proceeding with our analysis of the young worker problem, we state equilibrium conditions. An equilibrium specifies a price g_H, and at most two distinct contracts (i, w_S, w_F, v_S, v_F) where, if there are two contracts, a young worker is allocated with probabilities q and $1 - q$ over the two contracts, and q is specified as part of the equilibrium description. The continuation values $\{w_S, w_F, v_S, v_F\}$ determine the worker's task allocation and production when old as described above. Hence, aggregate supply of the two tasks is also determined. A price g_H, the contract set, and the allocation probability q together constitute an equilibrium if the no poaching condition is satisfied and the supply of task H matches demand $\zeta_H(g_H)$.

III Career Paths and Efficiency Contracts

We can now state the core result of the paper, which shows how overpay emerges in our setting and how career paths are intrinsically linked to the degree of overpay in the economy:

Proposition 1 *For all sufficiently large task H stakes k_H we have:*

- *Overpay: A strict subset of young workers start on task H, and receive strictly greater expected utility than young workers starting on task L.*
- *Up-or-out for overpaid workers: Task H workers remain on task H if they succeed, exert more effort and are paid more than when young. If they fail they are "demoted" to task L.*
- *Dynamically segregated labor markets: Task L workers are never "promoted:" they remain in task L when old, and exert the same effort as when young.*

IIIA Dynamic segregation

The existence of overpaid workers is intimately connected to Proposition 1's prediction of "segregated" career tracks: If a worker is not lucky enough to be assigned to the overpaying task H when young, he will never again have the chance to be assigned to it. At first sight, this dynamic segregation result flies in the face of an important insight of contract theory: Workers with more wealth are easier to employ, because the wealth can be used to acquire a stake in the firm and lessen the split between ownership and control that lies at the heart of the moral hazard problem.[21]

[21] See footnote 3.

In our framework, it is old workers who succeeded when young who have wealth. Consequently, it might seem that any old worker who succeeded when young should be assigned to task H, while all young workers, who have no wealth, should be assigned to task L. Under these career paths, *neither* dynamic segregation *nor* equilibrium overpay arises, since all young workers enter the labor force with the same expected utility.

Hence establishing dynamic segregation is the key to explaining why efficiency wages are not eliminated by dynamic contracts. Loosely speaking, dynamic segregation reflects a second insight of contracting theory: the prospect of wealth accumulation in the future can be used to ameliorate moral hazard problems in the present.[22]

Here, we sketch the argument for why dynamic segregation occurs when k_H is large. Consider two potential ways in which demand for task H could be met. First, as in Proposition 1, some young workers can be assigned to task H and remain on task H if successful. Denote this the "HH" career path. Second, all young workers can be assigned to task L, and some successful old workers get promoted to task H. Denote this the "LH" path. (We explain below why the third alternative, "HL", where young workers start on task H and move to task L after success, is never used.)

To sketch the argument, it is easiest to show that the HH career path maximizes firm profits, ignoring the worker's outside option—which is determined by competition from other firms, and formalized by the no-poaching condition. As we explain further below, the no-poaching constraint is non-binding for young task H workers. Moreover, the equilibrium price g_H must be such that firms make zero profits, so no career path that fails to maximize profits is viable.

The maximal profits a firm can generate by employing a young worker using the HH path is given by the following problem:

$$\max_{w_S \geq \underline{w}} p\left(g_H - w_S\right) - k_H \text{ subject to the incentive constraint } \gamma'\left(p\right) = v_H\left(w_S\right) - v_{FBL}.$$

The incentive constraint follows from the fact that when successful, the worker has wealth higher than $\underline{w}$ and so stays on task H, while he has zero wealth after failure and so moves to task L and earns surplus v_{FBL}. Contrast this with the maximal profits a firm can attain by employing an old successful L-worker. Note that this worker has at most wealth $w = g_L$ to reinvest, since firm profits in the young worker problem would be negative otherwise. Hence the firm's profits when

[22] See footnote 4.

employing the old worker on task H are at best given by:

$$\max_{w_S \geq 0} p\left(g_H - w_S\right) - \left(k_H - g_L\right) \text{ subject to the incentive constraint } \gamma'(p) = w_S.$$

The benefit of the LH path is that the reinvested wealth helps the firm cover the cost k_H. The benefit of the HH path is that the agent has stronger incentives to work for a given bonus w. To see this, we can rewrite the incentive constraint for the HH path as

$$\gamma'(p) = \underbrace{w_S}_{\text{Bonus incentive}} + \underbrace{\left(v_H\left(w_S\right) - v_{FBL} - w_S\right)}_{\text{Up-or-out incentive}}$$

Over and above the direct incentive effect from the bonus, the young worker potentially has an extra incentive to work in order to ensure further employment on task H. This is captured by the up-or-out incentive term, which is the utility difference from employment on task H with reinvested wealth w_S relative to consuming the wealth and being employed on task L. (Note that the HL path has neither of these advantages since young workers have no wealth, and assignment to task L after success eliminates up-or-out incentives, since $v_L(w) = v_{FBL} + w$.)

We now show that the up-or-out incentive benefit of the HH path dominates the reinvestment benefit of the LH path when the amount at stake k_H is large. On the one hand, when k_H is large relative to g_L, the benefit of the LH path—being able to reinvest wealth g_L—is relatively unimportant. On the other hand, when k_H is large the equilibrium price g_H is likewise large (in order for firms to break even), which in turn means that $v_H(w_S)$ must be large since when the price is high it is optimal for the firm to give high incentive pay. Hence the up-or-out incentive benefit of the HH path is large when k_H is large. These two forces act in the same direction, and so when k_H is large the HH path is the profit-maximizing one, and dynamic segregation occurs.

The same force that makes up-or-out incentives large when k_H is large also, and directly, implies that the expected utility of a young worker placed on the HH path is high. In other words, such a worker is overpaid relative to his unfortunate contemporaries stuck on task L. The no-poaching condition determines the equilibrium utility of workers starting in task L, but not of workers starting in task H—even if firms paid these workers less, they could still not be profitably poached away. Instead, compensation for young workers starting in task H is determined purely by the need to set incentives so as to maximize profits, and the no-poaching constraint is non-binding.

The dynamic segregation result shows that there is a complementarity between working on task

H when young and when old. Working on task H when young gives workers more wealth because g_H is greater than g_L, and this makes the worker more employable on task H when old. Conversely, having the chance of working on task H when old gives high up-or-out incentives, which makes the worker more employable on task H when young.

Although dynamic segregation always occurs when the stakes k_H are sufficiently high, it is not inevitable. When k_H is small relative to g_L, up-or-out incentives become weaker because the utility difference between tasks becomes smaller. Because the wealth accumulated on the L-task at the same time becomes more significant relative to the task H stakes, promoting people from task L to task H becomes efficient. This captures the point we made at the start of this subsection, namely that there is a force pushing firms to assign only workers with already accumulated wealth to task H. In this case, there is no dynamic segregation, and no equilibrium overpay.

Proposition 2 *Fix $k_H < g_L$. For all sufficiently low levels of demand for task H output, there is an equilibrium in which all workers start on task L, and some are promoted to task H after success. All workers have the same expected life-time utility.*

Propositions 1 and 2 illustrate the trade-off between starting people on low-stakes tasks and letting them work their way up (the LH path), relative to starting some people on a "fast-track" career (the HH path). When tasks are more similar in moral hazard costs, it is efficient to sequence lower moral hazard tasks early in a worker's career and have him work his way up. When tasks differ sufficiently in moral hazard relative to the length of a worker's career, dynamic segregation and overpay emerge as in Proposition 1. For the rest of paper we focus on the case in which k_H is large, and dynamic segregation and overpay arise.

Our dynamic segregation result has the direct implication that random variation in a worker's initial job placement has long-term consequences. Several recent empirical papers strongly support this. Oyer (2008) shows that a missed opportunity to enter investment banking upon MBA graduation due to temporarily lower demand from Wall Street significantly reduces expected life-time income, mainly due to the fact that the worker is very unlikely to enter investment banking later on in his career even if Wall Street recovers. Kahn (2010) shows more generally that graduating college in a recession has a very long-lasting negative impact on salaries and job attainment. Although Proposition 1 is formally about cross-sectional variation in initial conditions, whereas these empirical results relate to time-series variation, in Section IV we introduce aggregate uncertainty and formally derive these and other time-series implications from dynamic segregation.

IIIB Contract characteristics of overpaid jobs

We next discuss characteristics of overpaid jobs relative to normal jobs when dynamic segregation occurs. We discuss three related phenomena: The reliance on up-or-out contracts, the role of promotion, and the inefficient overallocation of menial tasks to overpaid workers early in their careers.

Up-or-out contracts: Proposition 1 states that young workers who start on task H face "up-or-out" promotion prospects. If they fail, they move to task L. If they succeed, they remain on task H, and are promoted in the sense that they now have more responsibility (i.e., are expected to work harder), and receive more pay. In contrast to these workers, workers who spend their entire careers on task L are never promoted. Instead, in *both periods* they receive a bonus of g_L if they succeed, and in both periods exert exactly the same effort.

The "out" half of "up-or-out" is a direct consequence of overpay. Because the no-poaching condition is not binding for an overpaying contract, the payment after failure when young must be set to zero since this enhances effort and increases firm profits. When $w_F = 0$, the worker is "out" in the sense that he is allocated to task L when old.[23]

For the "up" half, observe that because workers reinvest their success payments w_S with the firm, all success payments up to k_H are effectively paid as deferred compensation that is received only if the worker succeeds again when old. For exposition, we focus here on the case of $w_S \leq k_H$ (the general case is handled in the appendix). The worker's effort when young is given by $\gamma'(p) = v_H(w_S) - v_{FBL}$. Writing p_S for the worker's effort when old after he succeeds, $v_H(w_S) = p_S \gamma'(p_S) - \gamma(p_S)$, so that $\gamma'(p_S) = (v_H(w_S) + \gamma(p_S))/p_S$, which is larger than $\gamma'(p)$. Consequently, $p < p_S$, meaning the worker exerts less effort when young than when old (after succeeding). Effectively, the worker is paid a bonus only after two successes, so that when he is young he discounts that bonus by the probability that he fails when young, and by the effort he will have to exert when old; and also by the fact that if he fails, he still receives utility v_{FBL}.

The role of promotion: Both Proposition 1, where dynamic segregation occurs, and Proposition 2, where it does not, have promotion as part of the optimal contract. The promotion result matches the received wisdom that senior employees in organizations such as investment banks and law firms are both especially productive, and compensated especially well.

Our explanation for promotion is new in that it builds solely on the presence of moral hazard.

[23] This is related to the result in Spear and Wang (2005) that a worker should optimally be fired after failure because further employment leads to too high a continuation utility.

Existing theories such as Landers, Rebitzer, and Taylor (1996), Lazear (2004), Levin and Tadelis (2005), and Waldman (1990) all emphasize screening of talented workers into important jobs as the economic rationale for promotion. Although promotion policies are also used for incentive purposes in practice, there has been no good theoretical explanation for why promotion would dominate pure monetary rewards as an incentive device (see Baker, Jensen, and Murphy (1988) for a discussion).[24]

Up-or-out incentive schemes for overpaid workers and the fact that task L workers cannot move to task H (see Proposition 1) together imply that moving "up" to a better job is harder than moving "down" from a good job. This implication fits well with many anecdotal accounts of the labor market, especially in prestigious occupations such as investment banking and management consulting. Hong and Kubik (2003) offer more systematic evidence for security analysts. They show that it is much more common for security analysts to move from a high-paying, more prestigious brokerage firm to a lower-paying, less prestigious one than the other way around.

We also note that our model features a particularly simple explanation for the Peter Principle (Peter and Hull (1969)). The Peter Principle states that workers are promoted to "their level of incompetence", and is based on the empirical observation that workers who perform well get promoted, until they get stuck at a level in which their performance appears worse than before promotion. This is true in our setting in the following sense: workers are promoted only after success, and so the conditional success probability after promotion is necessarily lower than a worker's previous realized success probability.[25]

Dog years: Many anecdotal accounts suggest that overpaid workers often start their careers working extremely long hours on very straightforward and boring tasks. As we show, this characteristic of overpaid jobs emerges naturally as a way for the firm to reduce the surplus it surrenders to workers. Crucially, our model predicts that this surplus extraction occurs only at the start of overpaid workers' careers.

To capture these ideas, we introduce what we call *menial* tasks to workers, over and above the regular task. This menial task could involve gathering data, preparing spreadsheets, copying papers, or fetching lunch 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.

We take the equilibrium of the economy without menial tasks, and then introduce menial tasks

[24] An exception is the theory in Fairburn and Malcomson (2001), in which promotion is preferrable to monetary rewards when managers who make the promotion decision are subject to influence costs.

[25] For alternate but more complicated explanations, see Lazear (2004), Faria (2000), and Fairburn and Malcolmson (2001).

to a null set of firms (this allows us to hold the overall structure of the equilibrium unchanged). 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 very small but positive. A worker can work on both the menial and important tasks: his total hours worked are $\gamma(p) + m$, which must be less than 1, his total time endowment.

We show that the menial task is assigned only for young workers starting on the overpaid task; in all other circumstances, firms prefer workers to work on the more efficient tasks:

Proposition 3 *Suppose k_H is high enough such that there is dynamic segregation. Then, whenever the menial task is sufficiently menial (i.e., ε below some level $\bar{\varepsilon} > 0$), it is assigned* only *to young overpaid workers. Young overpaid workers perform the menial task up to the point where either their time endowment constraint binds, or their utility is reduced to the level of task L workers.*

We want to stress two features of this result. First, 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.

Second, since the menial task is used as an inefficient surplus extraction mechanism, its use is concentrated in overpaid industries. This is our "dog years" result: in overpaid industries, such as investment banking or law, 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.

Our "rent dissipation" explanation for overwork is different from, and arguably substantially simpler than, explanations proposed in the previous literature on inefficiently long hours, such as Holmström (1999), Landers, Rebitzer, and Taylor (1996), or Rebitzer and Taylor (1995), who build on either signalling or screening motives when workers are heterogenous in skill or preferences.

IV The effect of aggregate shocks on career dynamics

We now extend our basic model to allow for aggregate shocks to the economy. This allows us to study the time series implications of our model along three dimensions: Job placement, employment contracts, and firm productivity.

First, we show that entering the labor market in bad economic times has life-long negative effects on job placement, consistent with empirical evidence in Oyer (2008) and Kahn (2010) discussed above.

Second, we show that even if a worker is lucky enough to land an overpaid job in bad economic conditions, the overpaid job is *worse* than it would be in good times. The employment contract pays less not only initially but also later on in the worker's career, even if economic conditions recover. This is consistent with well-established cohort effects as in Baker, Gibbs, and Holmström (1994) and Beaudry and DiNardo (1991). We also show that employment contracts do not insulate the worker from risk beyond his control; there is an element of "pay-for-luck" in the optimal contract.

Last, we show that the productivity on task H is *countercyclical*, consistent with evidence for the latest three decades in the US (see Gali and van Rens (2010)).

We start with a specification of our basic model in which k_H is sufficiently large that young workers who start in task H are overpaid. To keep the analysis as simple as possible, assume the aggregate state of the economy is either "Good" (G) or "Bad" (B), with demand higher in the good state, i.e., $\zeta_H^G(\cdot) \geq \zeta_H^B(\cdot)$ and $g_L^G \geq g_L^B$. We assume throughout that ζ_H^G is sufficiently close to ζ_H^B and g_L^G is sufficiently close to g_L^B so that—as we explain below—the stochastic economy continues to feature overpaid workers.

Throughout, we let all contracts be fully contingent on the aggregate shock realization.

IVA Time series implications: Initial conditions matter

We first extend our dynamic segregation result to a setting with aggregate shocks, to show formally that prevailing labor market conditions at the time when a worker enters the labor force have long-lasting effects on his career. In particular, we show that when demand for task H goes down, firms respond by enacting hiring freezes rather than by firing old workers, so that entering young workers have a lower chance of landing an overpaid job. Furthermore, because of dynamic segregation, they are unable to enter this job later on even if the economy recovers. Instead, it is the next generation of young workers that get these jobs. This hiring pattern (consistent with the evidence in Oyer (2008) and Kahn (2010)) across the business cycle affects the workforce composition of a firm, which in turn affects productivity; we show the net effect is that productivity is countercyclical for the overpaying sector of the economy.

We can make these points by studying the particularly simple case in which the demand shock only affects task H, i.e., $g_L^G = g_L^B$ and $\zeta_H^G(\cdot) > \zeta_H^B(\cdot)$. For this case, prices and hence contracts remain the same regardless of the state of the economy, as we now show. When $g_L^G = g_L^B$, a worker's minimum continuation utility v_{FBL} is independent of the state. When v_{FBL} does not vary, the minimum price g_H where a profit-maximizing firm can break even on a young worker employed on task H is also state independent. Importantly, at this price supply is perfectly elastic: Firms are willing to hire any number of workers into task H at price g_H using the profit maximizing contract, but no workers below this price. Since we assume that workers on task H are overpaid at the profit maximizing contract, firms have no difficulty in attracting workers to task H. Therefore, as long as demand y_H^ω at price g_H does not vary too much over the two states $\omega \in \{G, B\}$, supply responds to demand shocks purely via changes in the number of young workers hired into task H, while prices and contracts remain unaffected.

To be more specific, let λ_t be the number of overpaid young workers hired for task H at date t. From the demand equation, date t output from task H must equal $y_H^{\omega_t}$. Denote by p_1 and p_2 the success probabilities for workers on task H when young and old, respectively: given the conjecture that prices are independent of the state, optimal contracts and hence effort levels are also state-independent. From the supply equation, date t output from task H must equal $p_1\lambda_t + p_1\lambda_{t-1}p_2$, where $p_1\lambda_t$ is the output by the λ_t just-hired young workers and $p_1\lambda_{t-1}p_2$ is the output from the λ_{t-1} old workers who were hired last period and succeeded when young. Consequently, the number of workers hired for task H at date t is

$$\lambda_t = \frac{y_H^{\omega_t}}{p_1} - \lambda_{t-1}p_2. \tag{3}$$

As one would expect, more young workers are assigned to task H in good states, and when fewer workers were hired at the previous date. We verify in the appendix that it is indeed possible to vary the number of workers hired by a sufficient amount to fully absorb the demand shock, as long as demand is not too volatile.[26]

It is easy to see from (3) that if the economy remains in state $\omega \in \{G, B\}$ for a long time, the number of young workers assigned to task H converges to λ^ω, defined by $\lambda^\omega \equiv \frac{y_H^\omega}{p_1(1+p_2)}$, and the age-profile of task H workers converges to p_1 old workers for every young worker. As one would expect, a sustained period in the good state leads to greater hiring of young workers into the overpaid task H jobs, i.e., $\lambda^G > \lambda^B$. Average productivity, on the other hand, is the same in both scenarios.

[26] Formally, this amounts to showing that λ_t remains between 0 (one cannot hire a negative number of new workers), and 1/2 (the total population of young workers).

Proposition 4 *Suppose that after many periods in the good state, the economy suffers an aggregate shock and enters the bad state. Hiring of young workers into task H falls below even λ^B, and young workers who fail to get employment in task H will not get employed in task H later in their career even if the economy recovers. At the same time, average productivity in task H actually increases.*

The proof is almost immediate from (3), and we give it here. In the first period that the economy is in the bad state, the number of young workers hired into task H is

$$\lambda_t = \frac{y_H^B}{p_1} - \lambda^G p_2 < \frac{y_H^B}{p_1} - \lambda^B p_2 = \lambda^B < \lambda^G.$$

The age-profile in task H is now skewed towards old workers. Since old workers work harder than young workers, i.e., $p_2 > p_1$ (see Proposition 1) the average productivity in task H increases when the bad shock hits, implying countercyclical productivity.

The reason task H hiring falls below even λ^B is that in the good state, firms hired many workers into task H, and the optimal contract prescribes that these workers are retained when old even in a downturn, which is at the expense of hiring new young workers. The shortfall in date t hiring translates into an increase in date $t+1$ hiring of the *next* generation of young workers into task H,

$$\lambda_{t+1} = \frac{y_H^{\omega_{t+1}}}{p_1} - \lambda_t p_2 > \lambda^\omega > \lambda_t.$$

In the case that the economy recovers so that the date $t+1$ state is again G, the hiring burst is particularly dramatic, since $\lambda_{t+1} > \lambda^G$. This hiring burst only benefits the date $t+1$ generation of young workers, however; workers who were young in date t and missed out on an overpaid job because of the bad shock are not now hired. Moreover, task H productivity is depressed at date $t+1$, as firms suffer from the lack of a "missing generation" that was not previously hired: the age profile is now unduly tilted towards young workers.

Although we focus primarily on the implications of our model for career dynamics, it is interesting to note that Proposition 4 can also be interpreted in terms of unemployment. To do so, think of task L as corresponding to unemployment, with v_{FBL} the level of utility obtained by unemployed workers. Then Proposition 4 says that if the economy shifts from an extended time in the good state to an extended time in the bad state, unemployment first spikes up even as productivity increases. Subsequently, unemployment *partially* recovers, while productivity drops back to its prior level. Moreover, and consistent with the descriptive evidence of Bewley (1999), wages do not fall when the economy enters bad times.

IVB Time series implications: Procyclical moral hazard

Next, we expand our analysis to the case in which aggregate shocks affect the demand for output from both tasks, i.e., $g_L^G > g_L^B$ and $\zeta_H^G(\cdot) > \zeta_H^B(\cdot)$. The significance of demand shocks for task L output is that they affect v_{FBL}, the minimum continuation level that a worker can be given. This in turn affects the incentives that workers can be given, which has the following two implications, analyzed below. First, contracts are now state contingent, generating time series implications for contract characteristics. Second, the state-contingency of contracts generates further implications for firm productivity and moral hazard over the business cycle.

We make the standard assumption that the state follows a Markov process, with the transition probability of moving from state $\omega \in \{G, B\}$ at date t to state ψ at date $t+1$ denoted by $\mu^{\omega\psi}$. We assume that the state is at least somewhat persistent, in the sense that the state is more likely to be good (respectively bad) tomorrow if it is good (respectively, bad) today, $\mu^{GG} > \mu^{BG}$.

Write $g^\omega = (g_H^\omega, g_L^\omega)$ for the state ω output prices. Write v_{FBL}^ω for v_{FBL} evaluated at output prices g^ω; note that $v_{FBL}^G > v_{FBL}^B$ since $g_L^G > g_L^B$. So when a young worker enters the labor force at date t, the minimum expected continuation utility he can be given is

$$\bar{v}_{FBL}^\omega \equiv \sum_{\psi=G,B} \mu^{\omega\psi} v_{FBL}^\psi.$$

The state-persistence assumption $\mu^{GG} > \mu^{BG}$ implies $\bar{v}_{FBL}^G > \bar{v}_{FBL}^B$, and so workers entering the labor force in good times are harder to incentivize, because the minimum utility they can be threatened with in the case of failure is higher. This is the key economic force driving our results below.

In contracts for young workers starting in task H, firms commit to make success payments of $w^{\omega\psi}$. (Given our focus on the case in which k_H is high and overpaid task H jobs exist, and since there is no failure payment, we omit the subscript S.) It is convenient to keep track of the expected success payment, $\bar{w}^\omega \equiv \sum_{\psi=G,B} \mu^{\omega\psi} w^{\omega\psi}$. The utility an old worker obtains with $w^{\omega\psi}$ depends on *tomorrow's state*, and we capture this dependence by writing v^ψ for the previously-defined function v evaluated using tomorrow's output prices g^ψ. Firms want to maximize a worker's expected utility after success, which means that this utility can be written as a function of $\bar{w}^\omega$ only, i.e.,

$$\bar{v}^\omega(\bar{w}^\omega) \equiv \max_{w^{\omega\psi}} \sum_{\psi=G,B} \mu^{\omega\psi} v^\psi \left(w^{\omega\psi}\right) \text{ s.t. } \sum_{\psi=G,B} \mu^{\omega\psi} w^{\omega\psi} = \bar{w}^\omega. \tag{4}$$

Hence a contract for a young worker is summarized by $\bar{w}^G$ and $\bar{w}^B$, which are the expected payments a firm promises him after success given that *today's* state is G and B respectively.

To determine the equilibrium, we must find the contract terms $\bar{w}^G$ and $\bar{w}^B$ and prices g_H^G, g_H^B. For the case with overpaid workers, this involves solving for the price at which the firm breaks even with the profit maximizing contract:

$$\max_{p^\omega,\, \bar{w}^\omega} p^\omega\left(g_H^\omega - \bar{w}^\omega\right) - k_H = 0 \text{ subject to } \bar{v}^\omega(\bar{w}^\omega) - \bar{v}_{FBL}^\omega = \gamma'(p^\omega). \tag{5}$$

Our main result, stated formally below, is that moral hazard problems in task H endogenously worsen in good times, i.e., are procyclical. The driving force is the incentive compatibility condition of (5), which captures the fact that the higher outside option $\bar{v}_{FBL}^\omega$ in the good state makes it more costly to incentivize workers. To establish procyclical moral hazard, we must show this incentive effect dominates the direct effect that higher demand in good times increases the equilibrium price and hence the available total surplus, which tends to ameliorate the moral hazard problem. However, precisely because workers are overpaid in equilibrium, supply of task H is locally completely elastic, and so the increase in demand has *no* direct impact on prices (exactly as in the previous subsection).[27]

Firms understand that workers are harder to motivate in good times, and adjust contracts to partially offset this effect. However, doing so is expensive, and the equilibrium effect is that even though firms pay more to workers starting in good times, these workers exert less effort.

Proposition 5 *(A) Overpaid young workers work less hard in good times, $p^G \leq p^B$, where the inequality is strict unless all old workers work the socially efficient amount.*

(B) Old workers assigned to task H earn more if they started their careers in a good aggregate state.

Proposition 4 above established one type of cohort effect, namely that entering the labor force in a good aggregate state increases a worker's lifetime utility because it increases his chances of

[27] However, the increase in demand has an indirect effect on equilibrium prices: because workers are more difficult to incentivize, the equilibrium price must rise, as can be seen from the equilibrium profit condition (5). Details are in the proof of Proposition 5.

entering an overpaid job. Part (B) of Proposition 5 establishes a second type of cohort effect: even conditioning on a worker entering an overpaid job, the worker earns more (and has higher lifetime utility) if he enters the labor force in a good aggregate state. Baker, Gibbs, and Holmström (1994) and Beaudry and DiNardo (1991) provide empirical evidence for these type of within-firm cohort effects in wages.

Proposition 4 showed that changes in the composition of the workforce makes task H productivity countercyclical. Proposition 5 establishes a second force in the same direction. Not only is the workforce in a boom tilted towards the less productive young workers, but these workers are even more unproductive because moral hazard is procyclical. In the particular case of the financial sector, this prediction fits well with perceptions that traders and bankers are more careless in financial booms. More generally, there is evidence that aggregate US productivity has been countercyclical since the mid-1980s (see Gali and van Rens (2010)). Indeed, and more speculatively, if one thinks that high-moral hazard tasks account for a larger share of the economy than previously, our model provides an explanation for why aggregate US productivity has shifted from being procyclical prior to the mid-1980s to being countercyclical since.

Finally, we note the following "pay for luck" characteristic of contracts: The worker is strictly better off if the state turns out to be good when he is old ($v^G\left(w^{\omega G}\right) > v^B\left(w^{\omega B}\right)$), even though he has no control over the state.[28] This follows simply from the fact that the worker's marginal productivity is higher in the good state since the price is higher in the good state; hence, it is cheaper to deliver utility to workers in the good state. The standard argument in the incentive literature is that optimal contracts should insure risk averse agents against risks that they have no control over. Although our setting has risk neutral agents, so that there is no direct benefit from insurance, our result points to a cost of insurance that is often ignored in the incentive literature: If marginal product is higher in some states than in others, insuring the agent against risk makes him provide too little effort in high productivity states and too much effort in low productivity states. The same economic force towards pay for luck operates in, for example, DeMarzo *et al* (forthcoming).

[28] The formal proof is in the appendix.

V Distortions in the allocation of talent

We argued in the introduction that the available evidence suggests that the high compensation of financial sector workers is not a skill premium. Accordingly, in our basic model we have abstracted from skill differences by assuming that workers are ex ante identical. However, our model can be extended to produce interesting implications for the matching of heterogeneously-skilled workers to different jobs. In particular, our model makes precise two forces that affect how talent is matched to jobs. First, talent may be "lured," in the sense that, for example, people who "should" (for maximization of total output) be doctors or scientists become investment bankers instead. Second, talent may be "scorned," in the sense that the most able people do not necessarily get the best jobs.

We introduce differences in talent by assuming that only a null set of workers have higher skills, while the remaining "ordinary" 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 have a cost $c_i\gamma(p)$ of achieving success p in task i, where $c_i < 1$ for both task $i = L, H$. One would expect these talented workers to be more generously rewarded than other workers; and maximization of total output would dictate that they be given more responsibility (in the sense of working harder) at all stages of their careers. As we show below, however, this does not necessarily happen.

As in much of the preceding analysis, we focus here on the case in which k_H is sufficiently high that overpaid task H jobs emerge in equilibrium.

To understand how talent is lured in our model, consider a worker who is more skilled at both tasks, but is especially skilled at task L, i.e., $c_L < c_H < 1$. Provided c_L is sufficiently below c_H, such a worker would be best allocated to task L (for maximization of total output). However, any firm employing young workers at overpaid terms in task H can profitably "lure" this worker. For example, the worker may increase task L output by \$100,000 but task H output by just \$10,000. But if the utility premium offered by the overpaid task H jobs is \$200,000, firms can lure him to take such a job, and task L firms cannot compete. The key driving force for this effect is that the moral hazard problem stops utilities from being equated across jobs in equilibrium. This talent-lured force in our model is very much in line with popular impressions of investment banks hiring away talented scientists from research careers.

Note, however, that a distinct "talent scorned" force operates in the opposite direction: at the same time as the talented worker is more valuable, he is also harder to motivate on tasks where up-or-out incentives are used, in the following sense. If the more talented worker fails, his continuation utility is higher than an ordinary worker's, because one-sided commitment leads firms to compete for his talents. This better outside option after failure makes the more talented worker harder to incentivize when young. (Note that this is the same force as operates in the aggregate shocks analysis of Section IV above.) Colloquially, he is "difficult," or "hard-to-manage." Holding task L talent fixed, the talent scorned force dominates whenever the worker's talent advantage in task H is sufficiently small, i.e., c_H close enough to 1. In this case, and perhaps surprisingly, the most talented worker in the economy does not get the best job, even though he would prefer to.[29]

As the worker's task H talent advantage grows, however, the talent lured force becomes the dominant one. Of course, if the task H advantage is very large, social efficiency would dictate that the worker *should* be assigned to task H, and there is no longer a sense in which talent is lured away from its most productive use. But numerical simulations (available upon request) show that, given task L talent c_L, there is an interval of task H talents c_H such that workers are employed in task H even though they would increase output more if employed in task L. In this case, talent is truly lured.

VI Equilibrium existence and secondary labor markets

At the heart of our analysis is the result that, in equilibrium, old workers need wealth above some (endogenous) critical value, $\underline{w}$, in order to be assigned to task H. As we have discussed, this is the driving force behind both dynamic segregation and the emergence of overpaid workers. However, the critical wealth level $\underline{w}$ also gives rise to a fundamental difficulty in establishing equilibrium existence, as we next explain. It is worth noting that this issue did not arise in the older efficiency wage literature precisely because it did not analyze dynamic contracts with deferred pay.

The difficulty that arises from the critical wealth level $\underline{w}$ is that the minimum continuation utility that a worker can be threatened with after failure, namely $\min v(0)$, is not continuous as a function of the output price g_H—see next paragraph. Because $\min v(0)$ directly affects the incentives that a young worker can be given, and thus how hard he works, this means that the correspondence from prices to possible equilibrium production levels may fail to be upper hemi-continuous (UHC). This

[29] Ohlendorf and Schmitz (2011) study a similar repeated moral hazard problem in which they also show that more talented workers may sometimes be avoided by employers. In their model, the firm avoids more talented workers as a commitment device to avoid renegotiation after failure; in contrast, our result stems from competition from other firms.

greatly complicates showing that the excess demand correspondence is UHC, which is the key step in most proofs of equilibrium existence. Other papers have confronted broadly related problems in establishing existence in economies with agency problems; see, for example, Acemoglu and Simsek (2010), and the papers cited therein.

In more detail, the continuation utility $\min v(0)$ is discontinuous precisely in the neighborhood of the price g_H such that the minimum wealth $\underline{w}$ needed for an old worker to be assigned to task H is zero. On the one hand, if g_H is very slightly lower, then $\underline{w} > 0$ and so penniless old workers are always assigned to task L; hence $v(0) = v_L(0)$. On the other hand, if g_H is very slightly higher, then $\underline{w} < 0$, meaning even penniless old workers can be assigned to task H. In this case, $v(0) = \max\{v_H(0), v_L(0)\}$. If—as is quite possible—task H pays workers more utility, i.e., $v_H(0) > v_L(0)$, it follows that $v(0)$ is discontinuous in the price g_H.

To resolve this problem, note that the equilibrium conditions stated in Section II are more stringent than necessary when $\underline{w} = 0$ and $v_L(0) < v_H(0)$. The reason is that the no-poaching condition assumes that a poaching firm can offer a contract that incentivizes a worker by assigning him to task L with certainty after failure. However, if a secondary labor market exists, such a threat may be impossible: since $\underline{w} = 0$, firms are happy to assign a penniless old worker to either task L or task H, and provided both types of jobs are offered in the secondary labor market, a worker's minimum utility strictly exceeds $v_L(0)$.

Accordingly, when $\underline{w} = 0$ we augment our definition of an equilibrium with a pair of parameters $\mu_1,\ \mu_2 \in [0,1]$ (one for each contract) which determine the conditions of the secondary labor market. A contract $j \in \{1,2\}$ is feasible only if the utility v_x offered after outcome x exceeds $\left(1-\mu_j\right) v_L(0) + \mu_j v_H(0)$. The parameter μ_j is the probability that a penniless old worker who originally received contract j is assigned to task H in the secondary labor market.

Note that when $\mu_1 = \mu_2 = 0$, the equilibrium conditions coincide with those in Section II. Consequently, contracts satisfying the conditions stated in Section II do indeed constitute an equilibrium. Moreover, when $\underline{w} \neq 0$ the conditions above coincide completely with those in Section II. Note that all results in the paper relate to the case $\underline{w} \neq 0$.

By entertaining all possible secondary labor market conditions $\mu_1, \mu_2 \in [0,1]$, we ensure that the excess demand correspondence is UHC, and hence has a fixed point. The fixed point pins down the equilibrium secondary labor market conditions.

Proposition 6 *An equilibrium exists.*

VII Conclusion

In this paper we develop a parsimonious dynamic equilibrium model in which some workers are overpaid relative to other workers, even when firms employ fully optimal dynamic contracts. We further show how this same model matches a variety of empirical observations about both cross-sectional variation of job characteristics, and time-series variation of labor force conditions. All of these predictions hinge crucially on solving for the optimal dynamic contract. For example, our model predicts that overpaid jobs rely heavily on up-or-out promotion, and demand long hours for entry-level workers, often on surprisingly mundane tasks. They are most commonly entered when young, implying that cross-sectional variation in workers' initial employment conditions have long lasting effects. In the time-series, our model predicts that workers who enter the labor force in bad economic times are less likely to get an overpaid job; that even if they do, the overpaid job is worse; and that they work harder, implying countercyclical productivity. We have reviewed the empirical support for these results in the text above.

For tractability, we analyzed the simplest possible model with both multiple tasks and long-lived workers, both of which are essential for the subject of the paper. However, we believe the main insights of our analysis would remain in settings with more than two tasks and/or workers who live more than two periods.

Throughout, we have conducted our analysis under the realistic assumption that indentured labor is impossible, and a worker can quit an employment contract whenever he wants (one-sided commitment). As we observed, this assumption also implies that all equilibrium contracts are renegotiation proof. Nonetheless, it is worth noting that most of our analysis would be qualitively unaffected if instead workers could not quit an employment contract. The main exceptions are Proposition 5, on procyclical moral hazard, and our discussion of "talent scorned."

One obviously counterfactual prediction of our analysis is that young workers who are overpaid and fail receive literally nothing after failure. This is a direct consequence of our assumption of risk-neutrality. If instead workers are risk-averse, firms would generally pay strictly positive payments after failure. Establishing overpay in a model with risk-averse agents could potentially be difficult, however: One might conjecture that firms could punish risk-averse workers very heavily for failure, by making consumption after failure very low (but still strictly positive), thereby eliminating equilibrium overpay since all workers' utilities would be equalized.[30] However, this conjecture is

[30] This is related to a point made in Carmichael (1985).

not correct in our model. One-sided commitment prevents a worker's continuation utility from ever falling very low, since otherwise competing firms would poach him away using a new contract. Hence we conjecture that generalizing our model to a wider class of preferences would lead to strictly positive pay after failure, even for overpaid workers, while still preserving the central prediction of equilibrium overpay. We plan to explore this avenue in future research.

We have completely abstracted from unobservable skill differences in our model. Clearly, if perceptions of an individual's skill increase by enough mid-career, then this individual may be promoted and escape dynamic segregation. We do not mean to suggest that unobservable skill differences are unimportant; our focus on the single friction of moral hazard is meant to isolate an economic force leading to dynamic segregation among sufficiently identical individuals.

We conclude with a brief discussion of economic efficiency. As we noted in the introduction, we use "overpaid" to refer to a situation in which high pay is neither a return to skill nor a compensating differential. In our model, shareholders willingly consent to overpay workers in this sense. A natural question to ask is then whether the decentralized equilibrium of our model satisfies standard notions of economic efficiency. Unfortunately, the one-sided commitment assumption makes this hard to answer in a satisfactory way. A social planner could trivially improve incentives, and hence potentially improve overall welfare also, if he is allowed to deliver lower continuation utilities than allowed by one-sided commitment. However, arguably a more appropriate social planning problem to examine is one in which the social planner is constrained by competition from firms for old workers. But once this constraint is introduced, it seems natural to consider competition for young workers also—but this is simply the decentralized equilibrium we have analyzed.

References

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

Acemoglu, Daron, and Alp Simsek, 2010, Moral Hazard and Efficiency in General Equilibrium with Anonymous Trading, *Working Paper*.

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.

Baker, G., Gibbs, M., and Bengt Holmström, B., 1994, The Wage Policy of a Firm, *The Quarterly Journal of Economics* 109, 921–955.

Baker, George P., Michael C. Jensen, Kevin J. Murphy, 1988, Compensation and Incentives: Practice vs. Theory, *The Journal of Finance* 43, 593–616.

Bebchuk, Lucian, and Jesse Fried, 2004, Pay Without Performance: The Unfulfilled Promise of Executive Compensation, *Harvard University Press.*

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

Beaudry, Paul, and John DiNardo, 1991, The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence From Micro Data, *The Journal of Political Economy* 99 (4), 665–668.

Bewley, Truman F., 2002, Why Wages Don't Fall during a Recession, *Harvard University Press.*

Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet, 2007, Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications, *The Review of Economic Studies* 74, 345–390.

Bolton, P., T. Santos, and J. A. Scheinkman, 2011, Cream Skimming in Financial Markets, Working Paper, Columbia University.

Border, Kim C., 1989, Fixed Point Theorems with Applications to Economics and Game Theory, *Cambridge University Press.*

Bulow, Jeremy I., and Lawrence H. Summers, 1986, A Theory of Dual Labor Markets with Applications to Industrial Policy, Discrimination, and Keynesian Unemployment, *Journal of Labor Economics* 4, 376–414.

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

DeMarzo, Peter, and Michael J. Fishman, 2007, Agency and Optimal Investment Dynamics, *Review of Financial Studies* 20, 151–188.

DeMarzo, Peter, Michael J. Fishman, Zhiguo He and Neng Wang, Forthcoming, Dynamic Agency and the q Theory of Investment, *Journal of Finance.*

Fairburn, James A., and James M. Malcomson, 2001, Performance, Promotion, and the Peter Principle, *Review of Economic Studies* 68: 45–66.

Faria, Joao Ricardo, 2000, An Economic Analysis of the Peter and Dilbert Principles, Manuscript, Sydney: Univ. Technology.

Green, E. J., 1987, Lending and Smoothing of Uninsurable Income, in E. C. Prescott and N. Wallace (eds.), *Contractual Arrangements for Intertemporal Trade,* 3–25. Minneapolis: University of Minnesota Press.

Holmström, Bengt, 1999, Managerial Incentive Problems: A Dynamic Perspective, *The Review of Economic Studies* 66, 169–182.

Hong, Harrison and Jeffrey D. Kubik, 2003, Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts, *The Journal of Finance* 58, 313–351.

Jensen, Michael C., and Meckling, William H., 1976, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics* 3, 305–360.

Kahn, Lisa B., 2010, The Long-term Labor Market Consequences of Graduating from College in a Bad Economy, *Labour Economics* 17, 303–316.

Katz, Lawrence F., 1986, Efficiency Wage Theories: A Partial Evaluation, *NBER Macroeconomics Annual* 1, 235–276.

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.

Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor, 1996, Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms, *The American Economic Review* 86, 329–348.

Lazear, Edward P., 1981, Agency, Earnings Profiles, Productivity, and Hours Restrictions, *The American Economic Review* 71, 606–620.

Lazear, Edward P., 2004, The Peter Principle: A Theory of Decline, *The Journal of Political Economy* 112:1, S141–S163.

Moen, Espen R., and Åsa Rosén, 2006, Equilibrium Incentive Contracts and Efficiency Wages, *Journal of European Economic* Association 4, pages 1165–1192.

Mortensen, Dale T., 2003, Wage Dispersion: Why are Similar Workers Paid Differently? *MIT Press.*

Murphy, K. M., A. Shleifer and R. W. Vishny, 1991, The Allocation of Talent: Implications for Growth, *Quarterly Journal of Economics* 106, 503–530.

Ohlendorf, Susanne, and Patrick W. Schmitz, Forthcoming, Repeated Moral Hazard and Contracts with Memory: The Case of Risk-neutrality, *International Economic Review*

Oyer, Paul, 2008, The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Life-time Income, *Journal of Finance* 63 2601–2628.

Peter, Lawrence J., and Raymond Hull,. 1969, The Peter Principle: Why Things Always Go Wrong, New York: Morrow

Phelan, Christopher, 1995, Repeated Moral Hazard and One-sided Commitment, *Journal of Economic Theory*, 66, 488–506.

Philippon T., 2010, Engineers vs. Financiers: Should the Financial Sector be Taxed or Subsidized, *American Economic Journal: Macro* 2, 158–182.

Philippon T., and Ariell Reshef, 2008, Wages and Human Capital in the U.S. Financial Industry: 1909-2006, *Working Paper.*

Rebitzer, James B. and Lowell J. Taylor, 1991, A Model of Dual Labor Markets When Product Demand is Uncertain, *The Quarterly Journal of Economics* 106, 1373–1383.

Shapiro, Carl, and Joseph E. Stiglitz, 1984, Equilibrium Unemployment as a Worker Discipline Device, *The American Economic Review* 74, 433–444.

Spear, Stephen E. and Sanjay Srivastava, 1987, On Repeated Moral Hazard with Discounting, *The Review of Economic Studies* 54, 599–617.

Spear, Stephen E. and Cheng Wang, 2005, When to Fire a CEO: Optimal Termination in Dynamic Contracts, *Journal of Economic Theory* 120, 239–256.

Terviö, Marko, 2009, Superstars and Mediocrities: Market Failure in the Discovery of Talent, *Review of Economic Studies* 76, 829–850.

Waldman, Michael, 1990, Up-or-Out Contracts: A Signaling Perspective, *Journal of Labor Economics* 8, 230–250.

Appendix

A Analysis of the contracting problem for young workers

In full generality, a dynamic contract constitutes an assignment to a task in each period, where the second-period task assignment is contingent on whether the first task succeeded or not, and an end-of-life set of state-contingent non-negative payments.[31] Formally, a contract is a septuple $(i, i_S, i_F, w_{SS}, w_{SF}, w_{FS}, w_{FF})$, where $i \in \{L, H\}$ is the initial task assignment, i_S and i_F are the second-period task assignments after first-period success and failure, and w_{SS} etc. are the payments contingent on success/failure in the two periods. As discussed in the main text, conditional on an outcome $x \in \{S, F\}$, the contract may specify a lottery over a set of "continuation contracts"

[31] Given commitment by firms and equal discount rates, there is no loss to postponing all payments until the end.

of the type (i_x, w_{xS}, w_{xF}), subject to the no-manipulability restriction that the firm is indifferent between all lottery outcomes.

It is a standard result that a dynamic contracting problem can be written recursively in terms of the firm committing to deliver outcome-contingent promised utilities to the worker, where the firm's continuation payoff is then determined by the cost-minimizing way of delivering this promised utility. Write $W(z)$ for a firm's minimum cost of providing a continuation utility of z to the worker.

In these terms, a contract for young workers specifies a first-period task assignment i and continuation utilities v_S and v_F after first-period success and failure respectively. The one-sided commitment restriction is that a contract can specify a continuation utility v_x after first period outcome $x \in \{S, F\}$ only if $W(z) \geq 0$ for all $z > v_x$: otherwise, another firm could poach a worker in the second period by offering z, and make strictly positive profits (since $W(z) < 0$). The equilibrium no-poaching condition is that there exists no alternative $\tilde{\imath}$, $\tilde{v}_S$, $\tilde{v}_F$ such that the worker exerts first-period effort p and $\tilde{p}$ under the two contracts, the alternative contract strictly raises his utility, $\tilde{p}\tilde{v}_s + (1-\tilde{p})\tilde{v}_F - \gamma(\tilde{p}) > pv_s + (1-p)v_F - \gamma(p)$, and produces strictly positive profits, $\tilde{p}(g_{\tilde{\imath}} - W(\tilde{v}_S)) - (1-\tilde{p})W(\tilde{v}_F) - k_{\tilde{\imath}}$.

Write $\hat{v}$ for the *inverse* of the cost function W. It is convenient to define $\hat{v}$ so that it contains the one-sided commitment constraint:

$$\hat{v}(w) \equiv \{z : W(z) = w \text{ and } W(\tilde{v}) \geq 0 \text{ for all } \tilde{v} > z\}. \tag{A-1}$$

The inverse $\hat{v}$ is potentially a non-degenerate correspondence. So under this formulation, a contract specifies both firm costs w_S and w_F and continuation utilities $v_S \in \hat{v}(w_S)$, $v_F \in \hat{v}(w_F)$, as well as the first-period task assignment i. The equilibrium no-poaching condition is that there is no alternative $\tilde{\imath}$, $\tilde{w}_S$, $\tilde{w}_F$, $\tilde{v}_S \in \hat{v}(\tilde{w}_S)$, $\tilde{v}_F \in \hat{v}(\tilde{w}_F)$ such that $\tilde{p}\tilde{v}_s + (1-\tilde{p})\tilde{v}_F - \gamma(\tilde{p}) > pv_s + (1-p)v_F - \gamma(p)$ and $\tilde{p}(g_{\tilde{\imath}} - \tilde{w}_S) - (1-\tilde{p})\tilde{w}_F - k_{\tilde{\imath}}$. Given the definition of $\hat{v}$, the one-sided commitment restriction is simply that $w_S, w_F \geq 0$. Economically, $\hat{v}(w)$ is the (set of) utility that an old worker can "buy" with wealth w.

This is exactly the problem stated in the main text, except that it involves $\hat{v}$ rather than v defined by (2). So to complete the derivation of the contracting problem stated in the main text, we establish:

Lemma A-1 *The correspondence $\hat{v}$ equals the correspondence v defined by (2).*

Proof of Lemma A-1: Certainly $v(w) \subset \hat{v}(w)$, since if a firm can deliver utility $\tilde{v} \in v(w)$ at a cost strictly below w, it can provide utility strictly in excess of $\tilde{v}$ at a cost w, contradicting the definition of v.

The remainder of the proof establishes $\hat{v}(w) \subset v(w)$. Suppose to the contrary that there exists $\tilde{v} \in \hat{v}(w)$ such that $\tilde{v} \notin v(w)$. So $\tilde{v} < \min v(w)$,[32] since by the definition of v_H and v_L, a firm cannot deliver utility $\tilde{v} > \max\{v_H(w), v_L(w)\}$ at a cost w.

Observe that $W(v_{FBL} - \varepsilon) < 0$ for $\varepsilon > 0$ small enough, since a firm can make strictly positive profits by assigning an old worker to task L and paying 0 after failure and just less than g_L after success. Since $\tilde{v} \in \hat{v}(w)$, this implies $\tilde{v} \geq v_{FBL}$. So $v_L^{-1}(\tilde{v})$ is well-defined and single-valued, and $W(\tilde{v}) \leq v_L^{-1}(\tilde{v})$, since continuation utility $\tilde{v}$ can certainly be provided at cost $v_L^{-1}(\tilde{v})$.

Since $\tilde{v} \in \hat{v}(w)$, it follows that $w = W(\tilde{v}) \leq v_L^{-1}(\tilde{v})$. Since v_L is strictly increasing, we obtain $v_L(0) \leq v_L(w) \leq \tilde{v}$. Combined with the earlier observation that $\tilde{v} < \min v(w)$, together with the shape of v, it follows that there exists $\tilde{w} < w$ such that $\tilde{v} \in v(\tilde{w})$. But then $W(\tilde{v}) \leq \tilde{w} < w$, giving a contradiction. **End Proof.**

B Proofs of results stated in main text

Proof of Lemma 1

Differentiation implies that, for $w \in (\underline{w}, k_H)$, $v_H'(w) = p'(w)\, p(w)\, \gamma''(p(w))$, where from (1), $p'(w)\left[g_H - \gamma'(p(w)) - p(w)\gamma''(p(w))\right] = -1$. Hence

$$v_H'(w) = \left(1 - \frac{g_H - \gamma'(p(w))}{p(w)\,\gamma''(p(w))}\right)^{-1} = \left(1 - \frac{k_H - w}{p(w)^2\,\gamma''(p(w))}\right)^{-1}, \tag{B-1}$$

where the second equality follows from (1). As either w or g_H increases, $p(w)$ increases, and hence $v_H'(w)$ decreases, establishing concavity and that $v_H'(w)$ is decreasing in g_H. As $w \to k_H$, $\gamma'(p(w)) \to g_H$, establishing $v_H'(w) > 1$ for $w \in (\underline{w}, k_H)$.

Note that $p(\underline{w})$ must maximize firm profits, and so is given implicitly by the first order condition $(g_H - \gamma'(p(\underline{w}))) - p(\underline{w})\,\gamma''(p(\underline{w})) = 0$. As $w \to \underline{w}$, $p(w) \to p(\underline{w})$ and so $g_H - \gamma'(p(w)) - p(w)\,\gamma''(p(w)) \to 0$, establishing $v_H'(w) \to \infty$ as $w \to \underline{w}$. **End Proof.**

Proof of Proposition 1

We prove the result via a series of Lemmas. The key results for dynamic segregation are Lemma B-2, which says that if a worker is initially assigned to task H he remains there after success,

[32] Recall $v(w)$ is potentially a non-degenerate set, since v is a correspondence.

and Lemma B-7, which says that the minimum wealth needed for assignment to task H when old eventually exceeds the maximum wealth a worker can accumulate in task L.

Lemma B-1 *If $v_H(\underline{w}) \geq v_L(\underline{w})$, the only case in which $w_S = \underline{w}$ is if the young worker is initially assigned to task L and $w_S = \underline{w} = g_L$.*

Proof of Lemma B-1: Write i for the young worker's task assignment. From Lemma 1, $v'_H(w) \to \infty$ as $w \to \underline{w}$. Hence if $w_S = \underline{w}$, and $p(g_i - w_S) > 0$ (which, from the zero-profit condition, is the case for all feasible contracts except when $i = L$ and $w_S = g_L$), there exists an alternative contract in which w_S is slightly increased, and both worker utility and firm profits are strictly increased, violating the no-poaching condition. **End Proof.**

Lemma B-2 *If an old worker is sometimes assigned to task L after success, he must have been assigned to task L when young. Equivalently, if a young worker is initially assigned to task H, he remains there with probability* 1 *if he succeeds.*

Proof of Lemma B-2: Suppose contrary to the claimed result that a worker who is sometimes assigned to task L after success is initially assigned to task H. For the worker in question, let w_S and w_F be first-period success and failure payments. Let p be the worker's effort when young. By the hypothesis that the old worker is sometimes assigned to task L after success, $v_H(w_S) \leq v_L(w_S)$ if $w_S > \underline{w}$. From Lemma B-1 it follows that $v_H(w_S) \leq v_L(w_S)$ if $w_S \geq \underline{w}$.

We first show that $w_F = 0$. Suppose to the contrary that this is not the case, and $w_F > 0$. We must have $w_F < w_S$ for the firm to break even. Since $v_H(w_S) \leq v_L(w_S)$ if $w_S \geq \underline{w}$, and since $v'_H(w) \geq 1 = v'_L(w)$ we must have $v(w) = v_L(w)$ for all $w \leq w_S$. From the firm's break-even condition, $g_H - w_S > 0$. Consider a perturbation in which w_S is slightly raised by dw_S while w_F is changed by $dw_F = -\frac{p}{1-p} dw_S$. This perturbation leads the worker's first-period effort to strictly increase by $dp > 0$. Consequently, the firm's profits are strictly increased by $dp(g_H - w_S + w_F) - p dw_S - (1-p) dw_F > 0$. The worker's utility is at least weakly increased. So there exists a further perturbation that strictly increases both worker utility and firm profits, implying the original contract is Pareto dominated, contradicting the equilibrium condition. Hence, we must have $w_F = 0$.

The above arguments imply that either $\underline{w} > 0$ or $v_H(0) < v_L(0)$; if instead $\underline{w} \leq 0$ and $v_H(0) \geq v_L(0)$, Lemma 1 implies $v_H(w_S) > v_L(w_S)$, a contradiction to the above.

The above arguments also imply that the young worker's expected utility is $v_L(0)+p(v_L(w_S)-v_L(0))-\gamma(p)$, which equals $v_L(0)+\max_{\tilde{p}}\tilde{p}w_S-\gamma(\tilde{p})$. The cost to the firm of providing incentives to the young worker is hence exactly the same as providing incentives to an old worker. Since the firm makes zero profits, it follows that $\underline{w}\leq 0$ and $\max_{\tilde{p}}\tilde{p}w_S-\gamma(\tilde{p})=v_H(0)$. Hence the young worker's utility is strictly smaller than $2v_L(0)$. But this violates the no-poaching condition, since it is possible to produce strictly positive profits while delivering utility arbitrarily close to $2v_L(0)$ to a young worker by assigned him in both periods to task L. The contradiction completes the proof. **End Proof.**

Lemma B-3 *As $k_H\to\infty$, the price $g_H\to\infty$; the payment given after success to an old worker assigned to task H grows without bound; the effort p exerted by the worker approaches $p(1)$; and the continuation utility of the old worker grows without bound.*

Proof of Lemma B-3: The fact that $g_H\to\infty$ as $k_H\to\infty$ is implied by the zero-profit condition for firms: if any young worker is assigned to task H, then the result is immediate; if instead only old workers are assigned to task H, then the maximum wealth of any such worker is g_L, and the result is again immediate.

An old worker assigned to task H exerts effort at least $p(\underline{w})$ and has continuation utility at least $v(\underline{w})$ (see main text). The effort level $p(\underline{w})$ solves $g_H=\gamma'(p)+p\gamma''(p)$. By (i) and (ii) of Assumption 1, $\gamma'(p)+p\gamma''(p)$ increases from 0 to ∞ as p increases from 0 to $p(1)$. Hence $p(\underline{w})\to p(1)$ as $g_H\to\infty$. The bonus required to induce this effort is at least $\gamma'(p(\underline{w}))$, and so grows without bound. Finally, the utility $v(\underline{w})$ grows without bound, since for any p_0 utility $v(\underline{w})$ is bounded below by $p_0\gamma'(p(\underline{w}))-\gamma(p_0)$, which grows without bound. **End Proof.**

Lemma B-4 *Suppose that $\underline{w}$ remains both strictly positive and bounded above as $k_H\to\infty$. Then there exists a young worker contract that delivers strictly positive profits and worker utility strictly in excess of $v_H(\underline{w})$.*

Proof of Lemma B-4: By definition, $p(\underline{w})(g_H-\gamma'(p(\underline{w})))-k_H+\underline{w}=0$. Consider assigning a young worker to task H with $w_F=0$, and w_S defined by $\gamma'(p(\underline{w}))=v_H(w_S)-v_L(0)$. Observe that $w_S>\underline{w}$ since $v_H(\underline{w})<\gamma'(p(\underline{w}))$. This contract induces effort of at least $p(\underline{w})$, since $v(w_S)\geq v_H(w_S)$, and, since $\underline{w}>0$, $v(0)=v_L(0)$. Hence the contract gives firm profits of at least $p(\underline{w})(g_H-w_S)-k_H$, which by the definition of $p(\underline{w})$ equals $p(\underline{w})(\gamma'(p(\underline{w}))-w_S)-\underline{w}$, which in turn equals $p(\underline{w})(v_H(w_S)-v_L(0)-w_S)-\underline{w}$. From Lemma 1, $v_H(w_S)-w_S\geq v_H(\underline{w})-\underline{w}$.

From Lemma B-3, $v_H(\underline{w}) \to \infty$ as $k_H \to \infty$. Since $\underline{w}$ is bounded above, it follows that profits from the contract described grow arbitrarily large, and in particular, are strictly positive for all k_H large enough. Finally, worker utility is at least $v_L(0) + \max_{\tilde{p}} \tilde{p}\gamma'(p(\underline{w})) - \gamma(\tilde{p})$, which equals $v_L(0) + v_H(\underline{w})$. **End Proof.**

Lemma B-5 *For k_H sufficiently large, a successful old worker is assigned to task H with probability 0 or 1.*

Proof of Lemma B-5: Suppose to the contrary that a successful old worker is assigned to task H with probability strictly between 0 and 1. This is possible only if $w_S = \underline{w}$ and $v_H(\underline{w}) \geq v_L(\underline{w})$, and so by Lemma B-1, only if he is assigned to task L when young, and $\underline{w} = g_L$. Hence, when k_H is large, such a contract can only arise if $\underline{w}$ remains bounded as $k_H \to \infty$. But then Lemma B-4 implies that the contract violates the no poaching condition, completing the proof. **End Proof.**

Lemma B-6 *If a young worker's expected lifetime utility grows without bound as $k_H \to \infty$, the young worker's expected lifetime output in task H must be bounded away from 0.*

Proof of Lemma B-6: If a worker is always assigned to task L when old, by Lemma B-2 he must also be assigned to task L when young. In this case, the worker's utility is bounded above. So the only way for a young worker's utility to grow without bound is for him to be assigned to task H when old, at least after he succeeds when young. From Lemma B-3, the only way for such a young worker's expected lifetime output in task H to approach 0 is for his probability of being assigned to task H when old to approach 0, while still being strictly positive. Also from Lemma B-3, this means that the only way for the worker's utility to grow without bound while still having lifetime output in task H approach zero is for the success payment when young to be exactly $\underline{w}$, and for the successful worker to be assigned to task H with a probability approaching 0, while remaining strictly positive. By Lemma B-5, this is impossible. **End Proof.**

Lemma B-7 *As $k_H \to \infty$, the minimum wealth $\underline{w}$ needed for an old worker to be assigned to task H grows without bound.*

Proof of Lemma B-7: Suppose to the contrary that $\underline{w}$ is bounded above as $k_H \to \infty$. On the one hand, if $\underline{w}$ remains strictly positive then Lemmas B-3 and B-4 imply that the utility of all young workers must grow without bound (or else the no-poaching condition is violated). But from Lemma B-6, this means that total output in task H is bounded away from 0, which since (by

Lemma B-3) $g_H \to \infty$ violates the equilibrium condition that supply equals demand, completing the proof. On the other hand, if $\underline{w} = 0$ when k_H is large, then all workers who succeed when young are assigned to task H, and hence (by B-3) the utility of all workers grows without bound as $k_H \to \infty$. The proof is then completed in the same way as in the first case. **End Proof.**

Completing the proof:

From Lemma B-2, at least some old workers must be assigned to task H, for otherwise there is no task H output in the economy, and supply cannot equal demand. From Lemmas B-3 and B-5, the expected lifetime utility of these workers grows without bound as $k_H \to \infty$. From Lemma B-6, it follows that for all k_H sufficiently large, at least some young workers are initially assigned to task L, since otherwise there is too much task H output for demand to equal supply. By the firm's break-even condition, a worker starting in task L has wealth of at most g_L entering the second period. So from Lemma B-7, for k_H sufficiently large any young worker initially assigned to task L is assigned there when old also, i.e., dynamic segregation. The utility of such a worker is bounded above by $2v_{FBL}$, and so the young workers who start in task H receive strictly more utility, i.e., are overpaid. Since workers who start in task H are overpaid, they must have zero wealth after they fail, since otherwise a firm could perturb the contract by reducing wealth after failure, thereby increasing effort; this perturbed contract could then be used to strictly increase firm profits by poaching a young worker who starts in task L. Zero wealth is associated with assignment to task L, from Lemma B-7. So workers who start in task H move to task L after failure, but (by Lemma B-2) remain in task H after success. Finally, the effort and pay implications follow from Lemma B-8 below. **End Proof.**

Lemma B-8 *Suppose a young worker starts on task H; remains on task H after success; receives a continuation utility v_{FBL} after failure; and receives strictly more expected utility than some other young workers (i.e., is overpaid). Then the worker exerts strictly more effort when old after he succeeds than when young, and moreover, receives more pay.*

Proof of Lemma B-8: There are two cases to consider. The first case, in which $w_S \leq k_H$, is handled in the main text. Here, we deal with the second case in which $w_S > k_H$, and so the worker's effort after success is p_{FBH}. Let p denote the worker's effort when young. For any effort level $\tilde{p}$, let $S(\tilde{p}) = \tilde{p}g_H - \gamma(\tilde{p}) - k_H$ be total one-period surplus (i.e., the sum of firm profits and worker utility) associated with effort $\tilde{p}$. Because $w_S \geq k_H$, $v_H(w_S) - w_S = S(p_{FBH})$. Hence firm profits from employing the young worker can be written as $S(p) + \gamma(p) + p(S(p_{FBH}) - v_H(w_S))$. Denote by $U(p)$ the one-period utility for a worker from being induced to work p by receiving a bonus $\gamma'(p)$ after success, $U(p) \equiv p\gamma'(p) - \gamma(p)$. Substituting in for $U(\cdot)$ and $\gamma'(p) = v_H(w_S) - v_{FBL}$, firm profits equal $S(p) - U(p) + p(S(p_{FBH}) - v_{FBL})$. Since the worker is overpaid, the derivative of profits with respect to p, namely $S'(p) - U'(p) + S(p_{FBH}) - v_{FBL}$, must be weakly positive. To complete the proof, suppose that, contrary to the claimed result, $p \geq p_{FBH}$. By (i) of Assumption 1, U is convex in p. So $U'(p) \geq U'(p_{FBH})$. Combined with $S'(p) \leq 0$, this implies

$$0 \leq -U'(p_{FBH}) + S(p_{FBH}) - v_{FBL}. \tag{B-2}$$

Finally, note that $S(p_{FBH}) = p_{FBH}g_H - \gamma(p_{FBH}) - k_H \leq U(p_{FBH})$; and $U(0) = 0$ together with the convexity of U in p implies $U(p_{FBH}) \leq p_{FBH}U'(p_{FBH}) < U'(p_{FBH})$. Hence the righthand side of (B-2) is strictly negative, giving a contradiction and completing the proof that the worker exerts more effort.

Finally, the pay implication is obtained as follows. Since the worker exerts first-best effort p_{FBH} when old after first-period success, he must receive a bonus of at least g_H after second-period success. Since the worker's first period effort is strictly below p_{FBH}, and his payment after first-period failure is 0, his first-period payment must be strictly less than g_H. **End Proof.**

Proof of Proposition 2

The proof is constructive. Define the candidate equilibrium price g_H^* of task H output by $v_{FBH}(g_H^*) = v_{FBL}$, where recall that $v_{FBH}(g_H) = \max_p pg_H - \gamma(p) - k_H$. Write p_{FBH}^* for the maximizing value of p, i.e., p_{FBH}, evaluated at g_H^*. We show that when demand is low enough such that $\zeta_H\left(\frac{1}{2}p_{FBL}p_{FBH}^*\right) \leq g_H^*$, there is an equilibrium with price g_H^*, in which all workers start in task L, are paid $w_F = 0$ and $w_S = g_L$, and a fraction $\mu \in [0,1]$ of successful workers are assigned to task H when old (where μ is defined by $\zeta_H\left(\frac{1}{2}\mu p_{FBL}p_{FBH}^*\right) = g_H^*$.

This is an equilibrium as follows. By the definition of g_H^*, $v_H(w) = v_L(w)$ for all $w \geq k_H$, and so the stated assignments of old workers are optimal. Moreover, note that $v(w) = v_{FBL} + w$.

Since $g_L > k_H$, any successful old worker can be assigned to task H while exerting first-best effort p_{FBH}^*. So the goods market clears. Firms make zero profits from young workers. There is no alternate contract that would produce higher profits from assigning a young worker to task L. Finally, because $v(w) = v_{FBL} + w$, a firm would lose money by assigning a young worker to task H: dynamic incentives are nonexistent here (i.e., $v'(w) \equiv 1$), and young workers have no wealth.

This completes the proof.

Proof of Proposition 3

Fix k_H sufficiently large that the equilibrium of the benchmark economy is of the type described in Proposition 1, and such that $\underline{w} > 0$ (see Lemma B-7). In particular, young workers are either initially assigned to task L and remain there with probability 1, or else are initially assigned to task H using a contract that pays $w_F = 0$ after failure (in which case they move to task L) or w_S after success. Firms make zero profits from this contract. For use below, we establish the following interim lemma, which implies that any other contract for a young worker starting in task H would generate strictly negative profits:

Lemma B-9 *Let (w_S, w_F) be a contract given to a young worker who is assigned to task H such that firm profits are weakly positive; and the derivative of firm profits with respect to w_S is weakly negative. Then profits are strictly decreasing in w_S for all higher values of w_S.*

Proof of Lemma B-9: Firm profits are $p(g_H - w_S) - (1-p)w_F - k_H$. Since $w_F \geq 0$ and profits are weakly positive, $g_H - w_S + w_F > 0$. By Lemma B-2 , the worker is assigned to task H after success, and so his continuation utility after success is $v_H(w_S)$. So a small increase dw_S in w_S affects effort p according to $dp\gamma''(p) = dw_S v_H'(w_S)$ (from differentiation of (IC-Y)). Consequently, the increase dw_S affects profits by $\frac{1}{\gamma''(p)}(v_H'(w_S)(g_H - w_S + w_F) - \gamma''(p)p)\,dw_S$. We know p strictly increases in w_S, v_H is concave (by Lemma 1), and $\gamma''(p)p$ increases in p by (i) of Assumption 1. Hence the expression $v_H'(w_S)(g_H - w_S + w_F) - \gamma''(p)p$ is strictly decreasing in w_S, establishing the result. **End Proof.**

We now consider the contract a firm would give to a worker when the menial task is a possibility. We study the relaxed problem in which the old worker's time constraint is disregarded. We show the menial task is never assigned to old workers in the solution to the relaxed problem. Consequently, the solution to the relaxed problem coincides with the solution to the full problem.

We assume for now that $v_H(\underline{w}) \geq v_L(\underline{w})$. As we explain below, the opposite case $v_H(\underline{w}) < v_L(\underline{w})$ is considerably easier. Given this assumption, in the equilibrium under consideration, $v(w) = v_{FBL} + w$ for $w \in [0, \underline{w})$, $v(w) = v_H(w)$ for $w > \underline{w}$, and $v(\underline{w}) = [v_{FBL} + \underline{w}, v_H(\underline{w})]$.

Consider an old worker entering with wealth w. When the menial task is introduced, the new $v_i(\cdot)$ mappings (for $i = L, H$) are given by

$$v_i^*(w) \equiv \max_{m \geq 0} v_i(w + m\varepsilon) - m.$$

This follows since a firm can just break even on an old worker that puts up wealth w, and spends time m on the menial task when old, by giving him a contract that delivers utility $v_i(w + m\varepsilon)$ by employment on task i, whilst keeping the profits $m\varepsilon$ produced on the menial task. This results in net utility $v_i(w+m\varepsilon)-m$ to the agent, and the no poaching condition for old workers requires this utility to be maximized.

Analogous to the mapping v, define $v^*(w)$ as the maximum promised utility a firm can deliver to a worker entering with wealth w, i.e., $v^*(w) \equiv \max_{i \in L,H} v_i^*(w)$. From this maximization problem, it is straightforward to show that the menial task is used in the second period only if w is both below $\hat{w}$ defined by $v_H'(\hat{w}) = \frac{1}{\varepsilon}$ and above $\check{w}$ defined by $\frac{v_H(\hat{w})-(v_{FBL}+\check{w})}{\hat{w}-\check{w}} = \frac{1}{\varepsilon}$. Consequently, for $w \notin [\check{w}, \hat{w}]$, the possibility of the menial task makes no difference to continuation utilities, i.e., $v^*(w) = v(w)$. For use below, note that both $\check{w}$ and $\hat{w}$ approach $\underline{w}$ as $\varepsilon \to 0$.

Case: Young workers assigned to task H

As noted, for the non-menial task case there is a unique contract that gives non-negative profits. Write w_S for this contract (recall $w_F = 0$). Consequently, for all $\alpha > 0$ sufficiently small, there exists some $\delta(\alpha) > 0$ such that losses of at least α are produced by any contract $(\tilde{w}_S, \tilde{w}_F)$ with $\tilde{w}_S \notin (w_S - \delta(\alpha), w_S + \delta(\alpha))$ and/or $\tilde{w}_F \notin [0, \delta(\alpha))$. Moreover, $\delta(\alpha) \to 0$ as $\alpha \to 0$. From Lemma B-1, $w_S > \underline{w}$. Fix α sufficiently small such that $w_S > \underline{w} + 2\delta(\alpha)$ and $\underline{w} > 2\delta(\alpha)$.

Next, consider how the contract changes when menial tasks are possible. Given α, choose $\varepsilon \in (0, \alpha)$ small enough such that $\check{w} > \delta(\alpha)$ and $\hat{w} < \underline{w} + \delta(\alpha)$.

Since the direct profits from a young worker performing the menial task are bounded above by ε, and $\varepsilon < \alpha$, it follows that any equilibrium contract (w_S^*, w_F^*) with menial tasks must have $w_S^* \in (w_S - \delta(\alpha), w_S + \delta(\alpha))$ and $w_F^* \in [0, \delta(\alpha))$. Hence $w_S^* > \hat{w}$ and $w_F^* < \check{w}$, implying that the menial task is never assigned to old workers.

Finally, it is optimal to have the *young* worker do the menial task until either his time constraint binds, or his utility is reduced to the utility of workers assigned to task L.

Case: Workers starting in sector L

For the non-menial task case, the equilibrium contract for workers starting on task L is simply $w_S = g_L$ and $w_F = 0$, and the worker's utility is $2v_{FBL}$. When menial tasks are possible, the contract must still deliver utility of at least $2v_{FBL}$ to the worker. By an exactly parallel argument to the task H case, it follows that for all $\varepsilon > 0$ sufficiently small, an equilibrium menial task contract is close to the equilibrium contract without menial tasks, and that no menial task is assigned to old workers. In particular, an equilibrium menial task contract has $w_S, w_F < \check{w}$, and the worker remains in task L when old.

Finally, since an equilibrium menial task contract must deliver utility at least $2v_{FBL}$, and the worker remains in task L, and the menial task is socially inefficient, it follows that the equilibrium menial task contract must remain $w_S = g_L$ and $w_F = 0$, and no menial task is assigned to the young worker.

Finally, consider the case in which $v_H(\underline{w}) < v_L(\underline{w})$. In this case, $v(w)$ is a monotonically increasing function. For ε sufficiently small, the menial task is never used. The result is then very straightforward. **End Proof.**

Analysis for subsection IVA

To verify the conjecture that prices and hence contracts are state-independent, we need to show that it is possible to vary the number of workers hired by a sufficient amount to fully absorb the demand shock. Formally, this amounts to showing that λ_t remains between 0 (one cannot hire a negative number of new workers), and 1/2 (the total population of young workers). Define $\underline{\lambda} \equiv \frac{y_H^B - p_2 y_H^G}{p_1\left(1-p_2^2\right)}$ and $\bar{\lambda} \equiv \frac{y_H^G - p_2 y_H^B}{p_1\left(1-p_2^2\right)}$. It is straightforward to establish that λ_t remains in the interval $\left[\underline{\lambda}, \bar{\lambda}\right]$.[33] Consider what happens as the shock size shrinks, i.e., ζ_H^G and ζ_H^B approach some common value $\bar{\zeta}_H$. Let $\bar{y}_H$ be the output level associated with demand $\bar{\zeta}_H$ and the price g_H, i.e., $\bar{\zeta}_H(\bar{y}_H) = g_H$. Then y_H^B and y_H^G both approach $\bar{y}_H$ and $\underline{\lambda}$ and $\bar{\lambda}$ both approach $\frac{\bar{y}_H}{p_1(1+p_2)}$. Hence provided the shocks are sufficiently small, there is indeed enough flexibility to absorb the shocks via hiring decisions, verifying the conjecture that prices are independent of the state.

[33] If $\lambda_{t-1} \in \left[\underline{\lambda}, \bar{\lambda}\right]$, then

$$\lambda_t \geq \frac{y_H^B}{p_1} - \bar{\lambda} p_2 = \frac{y_H^B\left(1-p_2^2\right) - \left(y_H^G - p_2 y_H^B\right) p_2}{p_1\left(1-p_2^2\right)} = \frac{y_H^B - p_2 y_H^G}{p_1\left(1-p_2^2\right)} = \underline{\lambda}$$

and

$$\lambda_t \leq \frac{y_H^G}{p_1} - \underline{\lambda} p_2 = \frac{y_H^G\left(1-p_2^2\right) - \left(y_H^B - p_2 y_H^G\right) p_2}{p_1\left(1-p_2^2\right)} = \frac{y_H^G - p_2 y_H^B}{p_1\left(1-p_2^2\right)} = \bar{\lambda}.$$

To confirm that λ_t converges, simply note that iteration of the hiring equation (3) gives

$$\lambda_t = (-p_2)^t \lambda_0 + \frac{1}{p_1} \sum_{s=0}^{t-1} (-p_2)^s y_H^{\omega_{t-s}}, \tag{B-3}$$

which determines date t hiring as a function of the history of shock realizations. Hence if the economy remains in state $\omega \in \{G, B\}$ for a long time, the number of young workers assigned to task H converges to λ^ω.

Proof of Proposition 5

Proof of Part (A): We first show that $\bar{v}_{FBL}^G > \bar{v}_{FBL}^B$ implies that the equilibrium price of task H output must be higher in good times, $g_H^G > g_H^B$, as follows. Suppose to the contrary that $g_H^G \leq g_H^B$. Note that because $v^\psi\left(w^{\omega\psi}\right)$ is increasing in g_H^ψ (from Lemma 1), state persistence implies that $\bar{v}^G(\cdot) \leq \bar{v}^B(\cdot)$ if $g_H^G \leq g_H^B$. From the incentive compatibility condition, it is then more expensive to induce a level of effort p^G in the good state, and hence impossible to satisfy (5) in both states unless $g_H^G > g_H^B$.

Next, suppose that, contrary to the claimed result in the proposition, there is an equilibrium in which either $p^G \geq p^B$ and old workers sometimes depart from the socially efficient effort level; or in which $p^G > p^B$.

The supposition $p^G \geq p^B$ and the zero-profit conditions for the two states imply $g_H^G - \bar{w}^G \leq g_H^B - \bar{w}^B$, and hence $0 < g_H^G - g_H^B \leq \bar{w}^G - \bar{w}^B$. Similarly, the supposition $p^G \geq p^B$ and the profit-maximization conditions for the two states imply (given Assumption 1) $\bar{v}^{G\prime}\left(\bar{w}^G\right)\left(g_H^G - \bar{w}^G\right) \geq \bar{v}^{B\prime}\left(\bar{w}^B\right)\left(g_H^B - \bar{w}^B\right)$ and hence $\bar{v}^{G\prime}\left(\bar{w}^G\right) \geq \bar{v}^{B\prime}\left(\bar{w}^B\right)$. Note that this inequality is strict if $p^G > p^B$.

To obtain a contradiction, we show that $\bar{w}^G > \bar{w}^B$ implies $\bar{v}^{G\prime}\left(\bar{w}^G\right) \leq \bar{v}^{B\prime}\left(\bar{w}^B\right)$, with strict inequality if old workers sometimes depart from the socially efficient effort level. Maximization of worker utility implies that the expected payment $\bar{w}^\omega$ is distributed across the two states so that $\bar{v}^{\omega\prime}\left(\bar{w}^\omega\right) = v^{G\prime}\left(w^{\omega G}\right) = v^{B\prime}\left(w^{\omega B}\right)$. Lemma 1 and $g_H^G > g_H^B$ imply that $w^{\omega B} \geq w^{\omega G}$, i.e., the worker receives some insurance against the realization of tomorrow's state. Observe that $\bar{w}^G = \mu^{GG} w^{GG} + \mu^{GB} w^{GB}$ can be rewritten as

$$\begin{aligned}
\bar{w}^G &= \mu^{BG} w^{GG} + \mu^{BB} w^{GB} + \left(\mu^{GG} - \mu^{BG}\right) w^{GG} - \left(\mu^{BB} - \mu^{GB}\right) w^{GB} \\
&= \mu^{BG} w^{GG} + \mu^{BB} w^{GB} + \left(\mu^{GG} - \mu^{BG}\right)\left(w^{GG} - w^{GB}\right).
\end{aligned}$$

Since $\mu^{GG} > \mu^{BG}$ and $w^{GB} \geq w^{GG}$, the final term is weakly negative. Hence $\bar{w}^G > \bar{w}^B$ implies that at least one of $w^{GG} > w^{BG}$ and $w^{GB} > w^{BB}$ must hold. By concavity of v (see Lemma 1), either of these inequalities implies

$$\bar{v}^{G\prime}\left(\bar{w}^G\right) = v^{G\prime}\left(w^{GG}\right) = v^{B\prime}\left(w^{GB}\right) \leq v^{G\prime}\left(w^{BG}\right) = v^{B\prime}\left(w^{BB}\right) = \bar{v}^{B\prime}\left(\bar{w}^B\right),$$

where the inequality is strict unless $w^{\omega\psi} \geq k_H$ for all ω, ψ. If old workers sometimes depart from the socially efficient effort level, we know that $w^{\omega\psi} < k_H$ for at least some ω, ψ, and so $\bar{v}^{G\prime}\left(\bar{w}^G\right) < \bar{v}^{B\prime}\left(\bar{w}^B\right)$. This establishes the required contradiction and completes the proof of part (A).

Proof of Part (B): From Part (A), $p_G \leq p_B$. We first deal with the case of $p_G < p_B$. The zero-profit conditions for the two states imply $g_H^G - \bar{w}^G > g_H^B - \bar{w}^B$. The profit-maximization conditions for the two states imply (given Assumption 1) $\bar{v}^{G\prime}\left(\bar{w}^G\right)\left(g_H^G - \bar{w}^G\right) < \bar{v}^{B\prime}\left(\bar{w}^B\right)\left(g_H^B - \bar{w}^B\right)$ and hence $\bar{v}^{G\prime}\left(\bar{w}^G\right) < \bar{v}^{B\prime}\left(\bar{w}^B\right)$. Since firms pay workers in the most efficient way, $\bar{v}^{\omega\prime}\left(\bar{w}^\omega\right) = v^{G\prime}\left(w^{\omega G}\right) = v^{B\prime}\left(w^{\omega B}\right)$, and so $v^{G\prime}\left(w^{GG}\right) < v^{G\prime}\left(w^{BG}\right)$ and $v^{B\prime}\left(w^{GB}\right) < v^{B\prime}\left(w^{BB}\right)$. By concavity of v (see Lemma 1), $w^{GG} > w^{BG}$ and $w^{GB} > w^{BB}$.

Finally, consider the case $p_G = p_B$. The zero-profit conditions for the two states imply $g_H^G - \bar{w}^G = g_H^B - \bar{w}^B$, and so, since $g_H^G > g_H^B$, $\bar{w}^G > \bar{w}^B$. From Part (A), old workers always work the socially efficient amount. Consequently, there is indeterminacy in exactly how the expected payments $\bar{w}^\omega$ are delivered across tomorrow's future states. However, a natural way to deliver these payments is to pay the same amount in both tomorrow's states, which gives the result, and completes the proof of part (B).

Proof of "pay for luck," subsection IVB

We need to show that $v^G\left(w^{\omega G}\right) > v^B\left(w^{\omega B}\right)$. Denote by $p_2^{\omega\psi}$ the effort on task H in the second period for $\psi \in \{G, B\}$. As in the proof of Proposition 5, we know $v^{G\prime}\left(w^{\omega G}\right) = v^{B\prime}\left(w^{\omega B}\right)$. There are two cases. First, it can be the case that $p_2^{\omega\psi}$ is at the first best level $p_{FBH}^{\omega\psi}$ for both states, so that $v^\psi\left(w^{\omega\psi}\right) = v_{FBH}^\psi + w^{\omega\psi}$. Since $g_H^G > g_H^B$, we have $v_{FBH}^G > v_{FBH}^B$. If $w^{\omega G} = w^{\omega B} = 0$ the result follows. If $w^{\omega\psi} > 0$ for some state, any contract in which the resource constraint $\sum_{\psi=G,B} \mu^{\omega\psi} w^{\omega\psi} = \bar{w}^\omega$ is satisfied is equivalent, so without loss of generality we can set $w^{\omega G} = w^{\omega B}$ and the result follows.

The other case is when $p_2^{\omega\psi}$ is below the first best level for both states. From (B-1) in the proof of Lemma 1 and (i) of Assumption 1, the conditions $v^{G\prime}\left(w^{\omega G}\right) = v^{B\prime}\left(w^{\omega B}\right)$ and $g_H^G > g_H^B$ imply $p_2^{\omega G} > p_2^{\omega B}$. Since $v^{\psi}\left(w^{\omega\psi}\right) = p_2^{\omega\psi}\gamma'\left(p_2^{\omega\psi}\right) - \gamma\left(p_2^{\omega\psi}\right)$ when $p_2^{\omega\psi} < p_{FBH}^{\omega\psi}$, the result follows (again using (i) of Assumption 1).

C Proof of Proposition 6 (equilibrium existence)

Throughout, we routinely write $\underline{w}(g)$, $v_L(w;g)$, $v_H(w;g)$, $v(w;g)$ to emphasize the dependence of the previously defined quantities $\underline{w}$ etc on prices $g = (g_L, g_H)$. Define $\underline{v}(g) = \min v(0;g)$ and $\bar{v}(g) = \max v(0;g)$.

As discussed in the main text, a contract for young workers is a quintuple (i, w_S, w_F, v_S, v_F). We write C for a representative contract. Write $\pi(C;g)$ for the firm's profits from contract C given prices g, and $u(C)$ for a young worker's expected utility from contract C.

Given secondary labor market conditions $\mu \in [0,1]$, we add the constraint

$$v_S, v_F \geq (1-\mu)\underline{v}(g) + \mu\bar{v}(g). \tag{C-1}$$

(Note that whenever $\underline{w} \neq 0$ this constraint is already implied by $v_x \in v(w_x)$ for $x \in \{S, F\}$.) We also relax the no-poaching condition so that it applies to contracts satisfying this extra constraint.

Formally, for given secondary labor market conditions μ, write $\mathcal{C}(g;\mu)$ for the set of feasible contracts, i.e., (i, w_S, w_F, v_S, v_F) satisfying $w_x \geq 0$ and $v_x \in v(w_x)$ for $x \in \{S, F\}$, along with the secondary labor market constraint (C-1). Write $\mathcal{E}(g;\mu)$ for the subset of feasible contracts that satisfy the no-poaching condition,

$$\mathcal{E}(g;\mu) \equiv \left\{C \in \mathcal{C}(g;\mu) : \pi(C;g) \geq 0, \text{ and } \nexists \tilde{C} \in \mathcal{C}(g;\mu) \text{ with } \pi\left(\tilde{C};g\right) > 0, u\left(\tilde{C}\right) > u(C)\right\}.$$

Then define

$$\mathcal{E}(g) \equiv \bigcup_{\mu\in[0,1]} \mathcal{E}(g;\mu).$$

The set $\mathcal{E}(g)$ is the set of possible equilibrium contracts. The basic outline of the proof of equilibrium existence is then as follows. First, we conjecture a level of task H output y_H. For the goods market to clear, the price must be $\zeta_H(g_H)$. (Recall we assume g_L is fixed, i.e., demand for task L output is perfectly elastic.)[34] The price in turn implies a set of possible equilibrium contracts, $\mathcal{E}(g)$. The equilibrium contracts determine task H output. If output coincides with our

[34] The proof of existence easily extends to the case in which demand for task L output is less than perfectly elastic.

initial conjecture, we have found an equilibrium. Formally, we define a correspondence mapping task H output to task H output, and use Kakutani's fixed point theorem to prove a fixed-point exists. The key step is Lemma C-3, which establishes upper hemi-continuity.

We give the details of this argument below. First, however, we expand on the main text's description of the form the equilibrium takes.

The equilibrium potentially entails randomization of several different initial contracts. That is, when young workers initially enter the labor force, they are randomly assigned to one of several different contracts. For concreteness, note that since the correspondence we construct maps to one-dimensional output sets, we know that there exists an equilibrium with just two contracts (formally, this is Carathéodory's theorem). Let q and $1-q$ be the probabilities of being assigned to contracts 1 and 2. For each contract $m = 1, 2$, there exists μ^m such that $C^m \in \mathcal{E}(g; \mu^m)$.

Fix $m \in \{1, 2\}$, and consider the workers given contract C^m. If $v_F^m, v_S^m \geq \bar{v}(g)$ then these continuation utilities are delivered by retention within the firm: by construction, it is impossible for a firm to make strictly positive profits while delivering utility strictly above $\bar{v}(g)$ to an old worker. Next, consider the case of $v_x^m \in [\underline{v}(g), \bar{v}(g))$ for at least one of $x \in \{S, F\}$. These continuation utilities are delivered via a lottery over assignment to task L with continuation utility $\underline{v}(g)$, and assignment to task H with continuation utility $\bar{v}(g)$. Because the cost of delivering continuation levels below $\bar{v}(g)$ is zero, we assume these continuation levels are delivered outside the original firm; in other words, there is a secondary labor market for old workers who started under contract C^m, and are owed a continuation utility that has no cost to deliver it. Any worker initially assigned contract C^m is free to enter this secondary labor market; hence by construction, there is no contract $\tilde{C}$ that can simultaneously deliver strictly positive profits and strictly improve a worker's utility over C^m. Finally, note that the secondary labor markets for workers starting on the two contracts are separate.

The proof of the existence of a fixed point follows:

Lemma C-1 *Let $\{g^n\}$ be a sequence of prices such that $g^n \to g$ and $\underline{v}(g^n)$ converges. Then* $\lim \underline{v}(g^n) \in [\underline{v}(g), \bar{v}(g)]$.

Proof of Lemma C-1: First, we show $\lim \underline{v}(g^n) \geq \underline{v}(g)$. If $\underline{w}(g) \neq 0$, then $v(0; \tilde{g})$ is a continuous function of prices $\tilde{g}$ in the neighborhood of g, and the result is immediate. Consider instead the case $\underline{w}(g) = 0$, in which case $\underline{v}(g) = v_L(0; g)$, and suppose to the contrary that $\lim \underline{v}(g^n) < \underline{v}(g) = v_L(0; g)$. By the continuity of $v_L(\cdot; \tilde{g})$ in $\tilde{g}$, for all n sufficiently large,

$\underline{v}(g^n) < v_L(0; g^n)$. But this contradicts the definition of $\underline{v}(g^n)$.

Second, we show $\lim \underline{v}(g^n) \leq \bar{v}(g)$. Suppose to the contrary that $\lim \underline{v}(g^n) > \bar{v}(g) = \max\{v_L(0, g), v_H(0, g)\}$. By the continuity of $v_L(\cdot; \tilde{g})$ in $\tilde{g}$, for all n sufficiently large, $\underline{v}(g^n) > v_L(0; g^n)$. Hence for all n sufficiently large, $\underline{w}(g_n) \leq 0$ and $\underline{v}(g^n) = v_H(0; g^n)$. Hence $\underline{w}(g) = 0$ and $v_H(0; g) = \lim \underline{v}(g^n)$, which contradicts $\lim \underline{v}(g^n) > v_H(0, g)$ and completes the proof. **End Proof.**

Lemma C-2 *Let $\{g^n\}$ be a sequence of prices such that $g^n \to g$, $\underline{v}(g^n)$ and $\bar{v}(g^n)$ converge, and $\underline{v}(g^n) < \bar{v}(g^n)$. Then $\lim \underline{v}(g^n) = \underline{v}(g)$ and $\lim \bar{v}(g^n) = \bar{v}(g)$.*

Proof of Lemma C-2: Since $\underline{v}(g^n) < \bar{v}(g^n)$, we know $\underline{w}(g_n) = 0$, $\underline{v}(g^n) = v_L(0; g^n)$ and $\bar{v}(g^n) = v_H(0; g^n)$. Then $\underline{w}(g) = 0$, $v_L(0; g) = \lim v_L(0; g^n)$ and $v_H(0; g) = \lim v_H(0; g^n)$. Also, since $v_L(0; g^n) < v_H(0; g^n)$, we know $v_L(0; g) \leq v_H(0; g)$. Hence $\underline{v}(g) = v_L(0; g)$ and $\bar{v}(g) = v_H(0; g)$, implying the result. **End Proof.**

Lemma C-3 *The correspondence $\mathcal{E}$ is non-empty, compact valued, and upper hemi-continuous.*

Proof of Lemma C-3:

For any g, the set $\mathcal{E}(g)$ is non-empty since there always exists a contract that delivers non-negative profits by assigning the worker to task L, and because $\mathcal{E}(g)$ certainly contains the contract that maximizes worker utility subject to non-negative firm profits. Moreover, for any given g, the set $\mathcal{E}(g)$ is bounded.

Consider a sequence $g^n \to g$ with $C^n \in \mathcal{E}(g^n)$ such that $\{C^n\}$, $\{\underline{v}(g^n)\}$ and $\{\bar{v}(g^n)\}$ are all convergent. Let C be the limit of $\{C^n\}$. Below, we establish $C \in \mathcal{E}(g)$. This has two implications:

First, applied to the special case in which g^n is simply constant at g, this establishes that $\mathcal{E}(g)$ is closed-valued, and hence compact-valued.

Second, by Proposition 11.11 in Border (1989), and given the Bolzano-Weierstrass theorem, it implies that $\mathcal{E}(\cdot)$ is an upper hemi-continuous correspondence.

We now show $C \in \mathcal{E}(g)$. Note that $\pi(C; g) \geq 0$. The proof is by contradiction: Suppose that $C \notin \mathcal{E}(g)$.

First, we consider the case in which for all n large enough, $\underline{v}(g^n) = \bar{v}(g^n)$. From Lemma C-1, let $\mu \in [0, 1]$ be such that $\lim \underline{v}(g^n) = (1-\mu)\underline{v}(g) + \mu\bar{v}(g)$. In particular, we know $C \notin \mathcal{E}(g; \mu)$. Also, we know C specifies continuation utilities $v_S, v_F \geq \lim \underline{v}(g^n)$, and so belongs to $\mathcal{C}(g; \mu)$. So there exists $\tilde{C} \in \mathcal{C}(g; \mu)$ such that $u\left(\tilde{C}\right) > u(C)$, $\pi\left(\tilde{C}; g\right) > 0$ and has $v_F, v_S > (1-\mu)\underline{v}(g) + \mu\bar{v}(g) =$

$\lim \underline{v}(g^n)$. So for n sufficiently large, $\tilde{C} \in \mathcal{C}(g^n; 0)$, $u\left(\tilde{C}\right) > u(C)$ and $\pi\left(\tilde{C}; g^n\right) > 0$. But since $\underline{v}(g^n) = \bar{v}(g^n)$ for all n large enough, the set $\mathcal{C}(g^n; \tilde{\mu})$ is independent of $\tilde{\mu}$, and hence for all $\tilde{\mu} \in [0,1]$, $C^n \notin \mathcal{E}(g^n; \tilde{\mu})$, implying $C^n \notin \mathcal{E}(g^n)$, a contradiction.

Second, we consider the alternate case in which there is no subsequence such that for all n large enough, $\underline{v}(g^n) = \bar{v}(g^n)$. This implies that there is a subsequence such that $\underline{v}(g^n) < \bar{v}(g^n)$. From Lemma C-2, $\lim \underline{v}(g^n) = \underline{v}(g)$ and $\lim \bar{v}(g^n) = \bar{v}(g)$.

There are two subcases.

In the first and easier subcase, the limit contract C has $v_F, v_S \geq \bar{v}(g)$, and so $C \in \mathcal{C}(g; \mu = 1)$. Since $C \notin \mathcal{E}(g; \mu = 1)$, there exists a contract $\tilde{C}$ with $\tilde{v}_F, \tilde{v}_S > \bar{v}(g)$ such that $u\left(\tilde{C}\right) > u(C)$ and $\pi\left(\tilde{C}; g\right) > 0$. So for all n large enough, $\tilde{v}_S, \tilde{v}_F > \bar{v}(g^n)$, $u\left(\tilde{C}\right) > u(C^n)$ and $\pi\left(\tilde{C}; g^n\right) > 0$. But then for all n large enough, for all $\tilde{\mu} \in [0,1]$, $C^n \notin \mathcal{E}(g^n; \tilde{\mu})$, and hence $C^n \notin \mathcal{E}(g^n)$, a contradiction.

In the second subcase, the limit contract C has $\min\{v_F, v_S\} \in [\underline{v}(g), \bar{v}(g))$. Let μ be such that $\min\{v_F, v_S\} = (1-\mu)\underline{v}(g) + \mu\bar{v}(g)$. So $C \in \mathcal{C}(g; \mu)$. Since $C \notin \mathcal{E}(g; \mu)$, there exists $\tilde{C}$ such with $\tilde{v}_F, \tilde{v}_S > (1-\mu)\underline{v}(g) + \mu\bar{v}(g)$, $u\left(\tilde{C}\right) > u(C)$ and $\pi\left(\tilde{C}; g\right) > 0$. So there exists $\varepsilon > 0$ such that $\tilde{v}_S, \tilde{v}_F > (1-\mu-\varepsilon)\underline{v}(g) + (\mu+\varepsilon)\bar{v}(g)$. For all n large enough, for all $\tilde{\mu} \in [0, \mu+\varepsilon]$, $\tilde{v}_S, \tilde{v}_F > (1-\tilde{\mu})\underline{v}(g^n) + \tilde{\mu}\bar{v}(g^n)$. So for all n large enough, for all $\tilde{\mu} \in [0, \mu+\varepsilon]$, $C^n \notin \mathcal{E}(g^n; \tilde{\mu})$. Moreover, for all n large enough, C^n has $\min\{v_F^n, v_S^n\} < (1-\tilde{\mu})\underline{v}(g^n) + \tilde{\mu}\bar{v}(g^n)$ for all $\tilde{\mu} \in [\mu+\varepsilon, 1]$, and so for all $\tilde{\mu} \in [\mu+\varepsilon, 1]$, $C^n \notin \mathcal{C}(g^n; \tilde{\mu})$ and hence $C^n \notin \mathcal{E}(g^n; \tilde{\mu})$. But then for all n large enough, for all $\tilde{\mu} \in [0,1]$, $C^n \notin \mathcal{E}(g^n; \tilde{\mu})$, and hence $C^n \notin \mathcal{E}(g^n)$, a contradiction. **End Proof.**

Lemma C-4 *Let $\alpha : E \rightarrow\rightarrow F, \beta : F \rightarrow\rightarrow G$ be upper hemi-continuous, α closed-valued, and $\beta(y)$ bounded for all $y \in F$. Then $\beta \circ \alpha : E \rightarrow\rightarrow G$ is upper hemi-continuous and compact-valued.*

Proof of Lemma C-4: Upper hemi-continuity is standard (see Proposition 11.23 of Border). We show that $\beta \circ \alpha$ is compact-valued. Given that $\beta(y)$ is bounded for all $y \in F$, it suffices to show that $\beta \circ \alpha$ is closed-valued. Fix $x \in E$, and consider any convergent sequence $\{z^n\} \subset \beta \circ \alpha(x)$, with limit z. For each n, there exists $y^n \in \alpha(x)$ such that $z^n \in \beta(y^n)$. By Bolzano-Weierstrass, y^n has a convergent subsequence. By upper hemi-continuity of β, $z \in \beta(\lim y^n)$. By closed-valuedness of $\alpha(x)$, $\lim y^n \in \alpha(x)$. Hence $z \in \beta \circ \alpha(x)$, completing the proof. **End Proof.**

Lemma C-5 *For any continuation utility v, let $\mathcal{Y}^c(v)$ be the set of expected task H outputs that are associated with the cost-minimizing way of delivering v. Then $\mathcal{Y}^c$ is compact-valued and upper hemi-continuous.*

Proof of Lemma C-5: The proof is standard, and omitted.

Proof of Proposition 6:

We write y_H for total output of task H. Even if all workers work in task H, and always succeed, total output is still just 1, and so we know $y_H \in [0,1]$. To establish existence, we construct a correspondence that maps the set of possible task H output levels, $[0,1]$, into itself, and then apply Kakutani's fixed point theorem. We first define a correspondence on $(0,1]$, and then extend it to cover $[0,1]$.

For any $y_H \in (0,1]$, the associated output price is $g_H = \zeta_H(y_H)$. (Recall we assume g_L is fixed.) Given g_H, define $\mathcal{Y}(g_H)$ as the set of per-period expected task H outputs associated with giving young workers contracts $C \in \mathcal{E}(g)$, i.e.,

$$\mathcal{Y}(g_H) = \bigcup_{C=(i,w_S,w_F,v_S,v_F)\in\mathcal{E}(g)} \left\{ \begin{array}{l} \frac{1}{2}\left(p\mathbf{1}_{(i=H)} + py_S + (1-p)\,y_F\right) \text{ such that} \\ \gamma'(p) = v_S - v_F,\ y_S \in \mathcal{Y}^c(v_S) \text{ and } y_F \in \mathcal{Y}^c(v_F) \end{array} \right\}.$$

It follows straightforwardly from Lemma C-4 that $\mathcal{Y}$ is upper hemi-continuous and compact-valued. It is also non-empty because $\mathcal{E}$ is. Define $\bar{\mathcal{Y}}(g_H)$ as the convex hull of $\mathcal{Y}(g_H)$. The correspondence $\bar{\mathcal{Y}}$ is compact and convex valued, and by Proposition 11.29 of Border, it is upper hemi-continuous.

Consequently, $\bar{\mathcal{Y}}(\zeta_H(y_H))$ defines a correspondence from $(0,1]$ into $[0,1]$. Note that as $y_H \to 0$ the price $g_H(y_H) \to \infty$, so the set $\mathcal{Y}(g)$ converges to $\{(0, p(1))\}$, where recall that $p(1)$ is the maximal attainable success probability. So defining $\bar{\mathcal{Y}}(\zeta_H(0))$ as $\{(0, p(1))\}$ ensures upper hemi-continuity of the correspondence $\bar{\mathcal{Y}}$.

By Kakutani's fixed point theorem, $\bar{\mathcal{Y}}$ has a fixed point, y_H^* say. Let the associated price be $g_H(y_H^*)$. **End Proof.**