Wall Street occupations: An equilibrium theory of overpaid jobs

ULF AXELSON and PHILIP BOND*

June 21, 2013

ABSTRACT

Many financial sector jobs naturally feature large amounts of capital per employee, and effort that is hard to monitor. We show how this leads to overpay in equilibrium for financial sector employees, even with full competition in both labor and product markets. The optimal dynamic contract associated with overpaying jobs features up-or-out promotion and long work hours, yet gives more utility to employees than their outside options dictate. We show that moral hazard problems in the financial sector are exacerbated in good economic times, leading to worse investment decisions, even though pay goes up. We also show that employees whose talent would be more valuable elsewhere can be lured into overpaying finance jobs, while the most talented employees 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 University of Minnesota. A previous version of this paper circulated under the title "Investment Banking Careers." We thank Wouter Dessein, Paolo Fulghieri, Andrew Hertzberg, Camelia Kuhnen, Alan Morrison, Paul Oyer, Uday Rajan, David Robinson, Ibola Schindele, and seminar audiences at the American Economic Association meetings, BI Oslo, Boston University, Columbia, Cornell, the European Finance Association meetings, University of Exeter, ESMT Berlin, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of Minneapolis, the Financial Intermediation Research Society meetings, University of Frankfurt (Goethe), University of Glasgow, HEC Paris, University of Houston, IE, IESE, Jackson Hole Finance Group conference, London Business School, London School of Economics, McGill, 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.

Recent years have seen a heated debate about the level and structure of pay in the financial sector, especially in light of the 2008-2009 financial crisis. There is no doubt that financial sector pay is extremely high. Bell and Van Reenen (2013) report an average 2010 compensation level of £1,905,000 for a group of 1,408 senior bankers in the UK. Kaplan and Rauh (2009) show that a significant proportion of incomes from the very top of the US income distribution stem from the financial sector. Moving beyond people at the very top of the income distribution, Oyer (2008) and Philippon and Reshef (2008) provide evidence that the financial sector offers more compensation than other industries, even after controlling for individual characteristics; in particular, Oyer estimates that the lifetime pay premium enjoyed by an MBA in the financial sector relative to other alternatives (such as consulting) is \$1.5-\$5 M, in present value terms.

Much of the public outrage, reflected for example in the recent "Occupy Wall Street" movement, stems from the perception that Wall Street professionals are *overpaid*. This outrage does not seem completely unfounded. We formalize the notion of *overpay* as a situation in which one person enjoys a higher lifetime expected utility than another equally skilled person. As we review below, there is significant empirical evidence that financial sector employees indeed earn more than is justified by their work conditions or skill levels, and hence are indeed overpaid in this sense.

But if financial market employees are overpaid, why does competition among potential bankers fail to drive compensation down? There is no lack of eager smart undergraduates and MBA students who would gladly take an investment banking entry level position with somewhat lower compensation levels than are currently offered. Similarly, if the high compensation is due to tacit collusion (for example, between hedge fund managers or private equity fund managers against their limited partners), why does entry fail to drive down fees in these segments of the market? Indeed, the idea that competition eliminates the possibility of overpay is standard to textbook economic models; related, economists such as Philippon and Reshef (2008) have expressed the view that, going forwards, these standard economic forces will drive down financial sector pay.[1]

In this paper we provide an equilibrium theory based on moral hazard for how overpay can persist in an economy with free entry by firms and competition for jobs among candidates. Our theory features optimal dynamic contracts, to ensure that our results are not driven by arbitrary contracting restrictions; this strikes us as particularly important in a model of the financial sector where age-compensation profiles are often very steep, consistent with dynamic incentives.[2] We

[1] For example, these authors write: "30% to 50% of wage differentials observed in the past 10 years can be attributed to rents, and can be expected to disappear."

[2] That we allow for optimal dynamic contracts is also one of the key distinguishing features of our model relative to the older efficiency wage literature (for example, Shapiro and Stiglitz (1984)), as we explain in detail below.

explain why an industry like finance may be particularly prone to overpay, and how the existence of overpay explains the type of career paths and contracts we see for financial-sector employees, such as up-or-out promotion and long work hours.

The main assumptions of our model are that for many jobs in finance the exact effort of an employee is hard to monitor; and that it is technologically possible for one employee to oversee a large amount of capital. The first of these assumptions leads to a standard moral hazard problem, and by standard arguments the employee may receive large bonuses after success for incentive reasons. These bonuses have to be especially large if there is a lot of capital at stake, as in finance. But this still begs the question of why all workers do not end up with these jobs if they indeed pay more than other jobs and if firms are free to enter and create more of these jobs—in which case the outside option would be exactly the same as the rent given for incentives. The extra step we take relative to the extant contracting literature in order to explain this is to endogenize the return to financial sector activities. In equilibrium, returns are consistent with aggregate activity, and are such that firms just break even. Because the market is unable to sustain returns high enough for firms to break even if all employees are put to oversee high-capital projects, there is a natural limit to entry into the overpaying segment.

We start our analysis by showing how these forces operate to produce overpay in a setting in which employees work for just one period. The heart of our paper then extends this one-period model to a dynamic setting. Analysis of the dynamic setting is important for three reasons.

First, because dynamic contracts ameliorate moral hazard problems, it is possible that dynamic contracts eliminate overpay.[3] We show that dynamic contracts can indeed eliminate overpay for tasks in which each employee oversees only a moderate amount of capital. However, provided that capital-per-employee is sufficiently large, as it plausibly is in the financial sector, overpay persists even after we allow for optimal dynamic contracts.

Second, the dynamic model delivers implications for career structures, which a one-period model is by definition incapable of doing. In particular, a trade-off exists between starting employees in low-moral hazard tasks, where they cannot do too much damage, then assigning them to high-moral hazard tasks only after success; and immediately assigning them to high moral hazard tasks, where steep up-or-out incentives induce them to work very hard. We show that when capital-per-employee is sufficiently large the latter path dominates. As we discuss in detail, we interpret this result as saying that a strict subset of employees enter the financial sector when young, are

[3] In particular, the older efficiency wage literature, to which we are related, attracted substantial criticism for its neglect of optimal—and especially dynamic—contracting, a criticism broadly know as the "bonding critique." Katz (1986) provides a useful review, including a discussion of the bonding critique.

promoted after success but leave the financial sector after failure. Despite the long hours and threat of firing, these employees have strictly higher utilities (i.e., are overpaid) relative to other less lucky employees who miss out on a finance job initially, and are never subsequently able to enter finance.

Third, and somewhat related, the dynamic model delivers implications for the effect of aggregate shocks. In particular, bad aggregate states lead to fewer overpaid jobs, and this has life-long effects on employees who enter the labor force in bad times (see in particular Oyer (2008) for evidence from the financial sector). On the other hand, good aggregate shocks lead to lower success rates and less profitable investments, because moral hazard is procyclical in our model—this is consistent with evidence of overinvestment in booms in the buyout and venture capital sectors, as well as widespread perceptions of careless lending in the run-up to the recent financial crisis. Finally, in our model capital inflows only partially respond to improved profit opportunities (i.e., capital is "slow-moving"). None of these effects arises in the one-period version of our model.

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 term "talent lured," is the observation that jobs like investment banking attract talented employees whose skills might be socially more valuable elsewhere, such as engineers or PhDs. In our model, this type of misallocation follows immediately from the fact that overpaying firms can outbid other employers for employees even if their talent is wasted in investment banking. The second phenomenon, which we term "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 employees, when fired, have higher outside opportunities.

As stated in our opening paragraph, we describe an employee as overpaid if his expected utility exceeds that of another employee with identical skills. It is worth highlighting that under this definition the existence of overpaid employees 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 employees. 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 (2013).

We next discuss why we believe high compensation in finance cannot be fully explained as either a return to skill or as a compensating differential for stressful work conditions. 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 employee characteristics using a fixed effect regression, while Oyer instruments for employee 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 employees different amounts.[4]

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.[5] 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.[6]

Related literature: As noted above, our paper is related to and builds on the older efficiency wage literature, 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. In Sections IV and V we discuss in detail our contribution relative to this literature. In brief, we show that overpay can persist even when there are minimal frictions in both labor and product markets—in particular free entry and/or constant returns to scale at the firm level, and optimal dynamic contracting in the labor market. Indeed, our results on optimal dynamic contracting address the bonding critique of the efficiency wage literature (see footnote 3). Separately, the extensive search literature in labor economics also predicts heterogeneity in wages for homogeneous employees.[7] In common with the efficiency wage literature, this literature largely ignores the possibility of dynamic contracting.

Our analysis is also related to the vast literature on optimal dynamic contracting. The contracting problem for an individual firm in our setting is relatively standard, and several of the contract

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

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

[6] 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.

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

characteristics we derive for high moral hazard tasks have antecedents in the dynamic contracting literature. In particular, contracts exhibit memory, as in Rogerson (1985);[8] backloading of pay, as in Lazear (1981), or more recently, Edmans et al (2012); and an up-or-out flavor, as in Spear and Wang (2006) and Biais et al (2010). As we discuss in more detail in Section IV, the key difference to the extant dynamic contracting literature is that we determine returns (i.e., prices) and outside options endogenously via equilibrium arguments. Moreover, the fact that task-payoffs are determined separately for different tasks means that the different tasks in our model are not isomorphic to the variable project size in papers such as Biais et al (2010). Finally, and as explained below, we impose one-sided commitment.

Finally, 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 employees 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 employees 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, by which we mean the feature that employees can only enter the high-paying sector when young; instead dynamic segregation is an important assumption in Tervio's analysis. 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 specifies the contracting problem. Section III derives the frictionless benchmark, Section IV analyzes the one-period version of our model, while Section V derives the core results of our paper in the dynamic setting. Section VI studies the effects of aggregate shocks. Section VII introduces observable talent

[8] Rogerson's results show that, for any effort profile the principal wants to induce, an agent's consumption reflects the history of outputs. This history dependence is a consequence of risk-aversion, and in particular, how the degree of risk-sharing is set to provide incentives in the most efficient way. In contrast, the agent in our model is risk-neutral, and history dependence instead relates to the optimal sequencing of effort levels and task assignments; Rogerson's analysis is silent on the first point, and his model has just one task.

differences. Section VIII discusses how our model can be used to explain the time series of pay in finance. Section IX concludes.

I Model

We need two key elements in the model: Employees who work for multiple periods, and tasks that vary in their degree of moral hazard problems. There is a countably infinite number of periods, and each period a measure 1 of young employees enter the labor market each period, work for two periods, and then exit. Except for age, employees are identical, and in particular, have the same skill. (Section VII analyzes an extension where skills differ across employees.) Employees are risk neutral, start out penniless, and have limited liability. Employees are employed by firms, who are risk neutral, maximize profits, and have "deep-pockets," so that limited liability constraints never bind for firms. Firms all have access to the same production technology, and are price-takers.

We first describe our model in terms of our leading financial sector example, and then discuss other interpretations below. There are two tasks, labeled H and L, which differ in the amount of the firm's resources they require. Each employee is assigned to one task $i \in \{H, L\}$ per period; but firms are free to operate in both tasks, and to switch employees across tasks in different periods. In both tasks, the employee exerts costly unobservable effort that increases the success probability p of the task he is assigned to (for example, increases the probability of a successful trade), where the private cost of effort $\gamma(p)$ is strictly increasing and strictly convex, with $\gamma(0) = \gamma'(0) = 0$ and $\gamma'(\overline{p}) = \infty$ for some upper bound $\overline{p} \leq 1$ on the success probability. We assume that the effort cost $\gamma(\cdot)$ satisfies Assumption 1 below. Part (i) ensures that a firm's marginal cost of inducing effort is increasing in the effort level. Part (ii) ensures that old employees exert strictly positive effort, even given the agency problem.[9]

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

Task H is a high-stakes task, in that it requires firm resources ("capital") $k_H > 0$. For example, task H might correspond to trading on the firm's own account, or complicated long-short "market-making" trades. If the trade fails, the firm loses k_H, while if the trade succeeds, the firm's profit is $g_H - k_H$, where g_H is the gross payoff from the trade.[10] As we formalize below, we assume that

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

[10] This agency problem captures the problem faced by financial firms of incentivizing their employees to reduce trading risk. However, it does not capture any notion of value-destroying risk-shifting, whereby financial sector workers take actions that are ex ante unprofitable because of the prospect of high rewards after good outcomes. A

the return g_H/k_H is decreasing in the economy-wide resources devoted to H trades; as a trading strategy becomes "crowded" its equilibrium return goes down.

Task L is a low-stakes task, in that it requires few resources (beyond the employee's labor); for simplicity, we assume it requires no resources, $k_L = 0$. Task L can be interpreted as a non-financial sector task or a "lower level" financial sector task such as preparing analyst recommendations. If task L fails, the firm loses $k_L = 0$, while if it succeeds, the firm's payoff is g_L. (Note that an alternative and equivalent specification of task L is that it is a safe investment, in which capital k_L generates gross firm payoffs of $g_L + k_L$ and k_L after success and failure respectively.)

An alternative interpretation is that task L is a safe task in which the amount of capital $k_L > 0$ invested is not lost after failure, and the extra payoff after success is g_L.

As we describe below, competition among firms means that, in equilibrium, the payoffs g_H and g_L must be such that profits net of employee compensation are zero.

The key parameter in our model is k_H, which is the amount the firm loses if the employee fails on task H. Our main results relate to the case in which k_H is large. The case of k_H large arises naturally in the financial sector, as follows.

1. Tasks involving money management or trading are quite scalable in that large amounts of capital can be allocated to a single employee. In fact, as we show in subsection VF, even if it were possible to scale down the task, it would never be in the interest of the firm to do so because the incentive costs to motivate effort do not scale down accordingly. The parameter k_H can then be interpreted as a technological, regulatory, or risk-management upper bound (determined outside of the model) on how much capital one employee can oversee.

   The following back-of-the envelope calculation gives an approximate sense of the amount of capital per employee in the financial industry. The Bureau of Labor Statistics measures total employment in the U.S. financial sector in 2012 at approximately 8 million; while from the Flow of Funds, U.S. households hold, in aggregate, approximately \$55 trillion of financial assets. Hence the average financial sector employee is, in some sense, responsible for approximately \$7 million of financial assets. Moreover, this average obscures a great deal

simple way to add risk-shifting to the model is to add the possibility of abandoning the trade after the worker has exerted effort p, but before the resources have been deployed. Risk-shifting then corresponds to a worker exerting little effort, but then going ahead with the unprofitable trade anyway, instead of abandoning it. However, because the only information the worker has about the trade success probability is p, the addition of this risk-shifting problem has no effect on equilibrium outcomes. The firm will simply pick a contract that induces an amount of effort p such that continuing with the trade is always optimal. (This follows from standard revelation-principle arguments: See Proposition 2 of Myerson (1982).) In contrast, the possibility of equilibrium risk-shifting would arise if the worker received additional information about the trade's success probability *after* making his effort choice p. In Section IX, we make some conjectures about how adding this feature to the model would affect our results.

of cross-sectional variation, so that many of the financial sector employees we are especially interested in when thinking about high compensation levels are responsible for vastly larger quantities.

2. Many tasks, such as M&A advising, buyouts, or the conducting of an IPO, affect the entire value of large companies.

Conditional on task assignments, the unobservability of an employee's choice of effort generates a standard moral hazard problem.[11] While a number of studies (see, e.g., the survey of Rebitzer and Taylor (2010)) have suggested that people may work hard even when effort is unobservable because of some type of intrinsic motivation, such motivation is often viewed to be of limited importance in the financial sector (see, e.g., Rajan (2010, chapter 6)).

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 employee 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 the task L payoff is constant, i.e., $g_L > 0$ is a parameter of the model.[12] For task H, write y_H for the economy-wide "supply" of task H, i.e., the expected number of task-H successes in the economy. The equilibrium return g_H/k_H is then determined by $\zeta_H(y_H, k_H)$, where ζ_H is a strictly decreasing function of both arguments; in this sense, ζ_H plays the role of the (inverse) demand function. We also impose

Assumption 2 *(i) For any* k_H, $\lim_{y_H \to 0} \zeta_H(y_H, k_H) = \infty$. *(ii) For any* $y_H > 0$, $\lim_{k_H \to \infty} \zeta_H(y_H, k_H) < 1$.

Part (i) is a standard Inada condition. In terms of our trading interpretation, it says that trading is very profitable if no-one else is trading. Part (ii) says that if the aggregate number of task H successes is bounded away from 0, then the success-return falls below 1 if resources controlled by each employee are large enough. In our trading interpretation, this is just a statement that as total capital deployed to buy an asset grows large, the price paid for the asset eventually exceeds its true value, so that the gross return is eventually less than 1.

Remark: Our specification of $\zeta_H(y_H, k_H)$ is natural in a trading or money management setting where the equilibrium "alpha" from active management should be decreasing in the amount of

[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] Our results are qualitatively unaffected if this assumption is relaxed.

smart money chasing returns. However, it is possible to interpret task H more generally, so that the model may be applied to a range of other occupations, including both non-trading areas of the financial sector, such as M&A advising, and also non-financial occupations, such as train drivers to lawyers. For some of these applications, the natural equilibrium condition is simply $g_H = D(y_H)$, where D is an inverse demand curve that gives the price associated with aggregate output y_H; and there is no direct link with k_H, which measures the cost of failure (presumably high for both M&A-advising and train-driving). Our model nests this specification: simply set $\zeta_H(y_H, k_H) = \frac{1}{k_H} D(y_H)$, where D is a strictly decreasing function with $\lim_{y_H \to 0} D(y_H) = \infty$.[13]

II Contracts and equilibrium

IIA Contracts

We impose minimal contracting restrictions (motivated, in part, by criticisms that previous overpay results were consequences of exogenous restrictions), and allow firms to offer arbitrary dynamic contracts. Firms can commit to contract terms. However, we rule out indentured labor and model employees as having limited commitment, in the sense that they can walk away from the contract after the first period if another firm offers better terms. In other words, we assume *one-sided commitment*.[14,15]

For ease of exposition, we restrict attention to deterministic contracts. We show in an earlier draft of the paper, available upon request, that our results are robust to allowing for contracts that specify lotteries, subject to the constraint that the firm (but not necessarily the employee) is indifferent over lottery outcomes, since any lottery in which the firm is not indifferent would be subject to manipulation by the firm.

Because the firm can commit, we can assume without loss of generality that all compensation payments are deferred to the end of an employee's career. Consequently, a contract is a septuple

[13]When ζ_H is interpreted in this way, demand is determined 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. 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.

[14]See, e.g., Phelan (1995), and Krueger and Uhlig (2006). In our setting, 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]Most of our analysis would be qualitatively unaffected if we instead imposed two-sided commitment, i.e., workers cannot quit an employment contract. The main exceptions are Proposition 4 in Section VI, on procyclical moral hazard, and our discussion of "talent scorned" in Section VII.

$C = (i, i_S, i_F, w_{SS}, w_{SF}, w_{FS}, w_{FF})$ that specifies task assignments when young (i), when old after first-period success (i_S), and when old after first-period failure (i_F); along with compensation payments w_{SS} etc that are contingent on outcomes in both periods.

It is helpful to first compute expected payments and success probabilities for old employees. Given a first-period outcome $X \in \{S, F\}$, an old employee faces a "subcontract" $\boldsymbol{w}_X = (w_{XS}, w_{XF})$ that specifies payments contingent on success or failure in the second period. Given the subcontract $\boldsymbol{w}_X$, the employee chooses second-period effort $p(\boldsymbol{w}_X)$, where

$$p(\boldsymbol{w}_X) \equiv \arg\max_{\tilde{p}} \tilde{p} w_{XS} + (1 - \tilde{p}) w_{XF} - \gamma(\tilde{p}),$$

or equivalently, $p(\boldsymbol{w}_X)$ satisfies the incentive compatibility (IC) constraint

$$\gamma'(p(\boldsymbol{w}_X)) = w_{XS} - w_{XF},$$

implying that a larger success bonus $w_{XS} - w_{XF}$ leads to higher effort. Given effort choice $p(\boldsymbol{w}_X)$, the employee's expected compensation is denoted $E(w|X)$, and is given by

$$E(w|X) \equiv p(\boldsymbol{w}_X) w_{XS} + (1 - p(\boldsymbol{w}_X)) w_{XF}.$$

Hence, the employee's expected future utility after first-period outcome X is $E(w|X) - \gamma(p(\boldsymbol{w}_X))$, and two-period employee utility from a contract C is

$$U(C) = \max_{\tilde{p}} \tilde{p} [E(w|S) - \gamma(p(\boldsymbol{w}_S))] + (1 - \tilde{p}) [E(w|F) - \gamma(p(\boldsymbol{w}_F))] - \gamma(\tilde{p}).$$

Hence first-period employee effort p is determined by the IC constraint

$$\gamma'(p) = [E(w|S) - \gamma(p(\boldsymbol{w}_S))] - [E(w|F) - \gamma(p(\boldsymbol{w}_F))]. \tag{1}$$

The IC constraint (1) illustrates the benefit of dynamic contracts: The utility the employee derives from the second period subcontract after first-period success can be used to motivate work in both periods. Finally, two-period firm profits Π are then given by

$$\Pi(C; g_H) = p [g_i + p(\boldsymbol{w}_S) g_{i_S} - E(w|S) - k_{i_S}] + (1 - p) [p(\boldsymbol{w}_F) g_{i_F} - k_{i_F} - E(w|F)] - k_i,$$

where p satisfies the first-period IC constraint (1).

IIB Equilibrium

As usual, any contract offered in equilibrium must maximize firm profits subject to satisfying an employee participation constraint. An important aspect of our analysis is that the reservation utility that enters the participation constraint is endogenous, and is an equilibrium object.

In full, a (stationary) equilibrium consists of a payoff g_H; a collection of contracts $\mathcal{C}$, together with a probability distribution λ over $\mathcal{C}$; and a reservation utility $\underline{U}$ such that the following conditions hold:

Profit maximization: Each contract $C \in \mathcal{C}$ maximizes firm profits subject to the participation constraint and one-sided commitment constraints, i.e., C solves

$$\max_{\tilde{C}} \Pi\left(\tilde{C}; g_H\right) \text{ s.t. } U\left(\tilde{C}\right) \geq \underline{U}$$

and such that, for $X \in \{S, F\}$, there is no alternative sub-contract $\{\tilde{\imath}_X, \tilde{w}_{XS} \geq 0, \tilde{w}_{XF} \geq 0\}$ that another firm could offer an old employee that both gives strictly positive firm profits and strictly raises employee utility:

$$E\left(\tilde{w}|X\right) - \gamma\left(p\left(\tilde{\boldsymbol{w}}_X\right)\right) > E\left(w|X\right) - \gamma\left(p\left(\boldsymbol{w}_X\right)\right)$$

and

$$p\left(\tilde{\boldsymbol{w}}_X\right) g_{\tilde{\imath}_X} - k_{\tilde{\imath}_X} - E\left(\tilde{w}|X\right) > 0.$$

Zero profits: $\Pi\left(C; g_H\right) = 0$ for each contract $C \in \mathcal{C}$.

Consistency of reservation utility: $\underline{U} = \min_{\tilde{C} \in \mathcal{C}} U\left(\tilde{C}\right)$.

Return consistent with aggregate task H activity: The return $\frac{g_H}{k_H}$ is consistent with equilibrium contracts, i.e., if $y_H\left(C\right)$ is the expected number of successful task H trades generated by contract C, then

$$\zeta_H\left(\sum_{C \in \mathcal{C}} \lambda\left(C\right) y_H\left(C\right), k_H\right) = \frac{g_H}{k_H}.$$

(Note that for some of the other applications discussed, where ζ_H is an inverse demand curve, this condition is simply the requirement that supply equals demand.)

Formally, we define an *overpaying equilibrium* as one in which there is a contract $C \in \mathcal{C}$ such that $U\left(C\right) > \min_{\tilde{C} \in \mathcal{C}} U\left(\tilde{C}\right) = \underline{U}$, that is, some employees earn strictly higher expected life-time utility than otherwise identical employees.

III Frictionless benchmark

Our main results all stem from the moral hazard problem. Before proceeding, we briefly describe the outcomes of a benchmark economy in which effort is fully observable, so that there is no moral hazard. For use throughout the analysis, let $S_i(p; g_i)$ denote the one-period surplus from effort p on task i, given return g_i:

$$S_i(p; g_i) \equiv p g_i - \gamma(p) - k_i.$$

When effort is observable, profit-maximization implies that equilibrium effort maximizes surplus. For use throughout, we denote surplus-maximizing effort given g_i by $p^*(g_i)$, i.e.,

$$p^*(g_i) \equiv \arg\max_{\tilde{p}} S_i(p; g_i).$$

Moreover, because g_L is fixed, we write $p_L^* \equiv p^*(g_L)$.

Profit-maximization also implies that in equilibrium the surplus from each task is equalized, i.e., g_H satisfies $S_H(p^*(g_H); g_H) = S_L(p_L^*; g_L)$.[16] The fraction λ of the population of measure 2 that is assigned to task H in any period is then given by the return consistency condition $\zeta_H(2\lambda p^*(g_H), k_H) = \frac{g_H}{k_H}$. Critically, and in contrast to the outcome of the moral hazard economy analyzed below, which task an employee is assigned to over his life time is indeterminate and independent of age and success, and all employees earn the same utility.

IV A one-period economy

Although our main contribution is the characterization of the dynamic equilibrium, we first analyze a simpler one-period version of our model to fix ideas. This is useful for two reasons: It allows us to introduce some important concepts and quantities that are helpful for deriving the dynamic equilibrium results, and it helps to illustrate some of the key differences relative to the existing literature.

For this section only, assume that there is only one period and (for comparability with the two-period economy) a measure 2 of employees. A contract is now simply a triple (i, w_S, w_F), which specifies the task assignment and payments after success and failure. Moreover, we note immediately that the equilibrium payment after failure must be $w_F = 0$, as follows. If instead $w_F > 0$, a firm can only break-even if $w_S < g_i$. But then the employee's gain from success is less

[16]If task H is very profitable (i.e., the trade-profitability function ζ_H is high), then it is possible that task L is not performed in equilibrium; in this case, surplus is not equalized across tasks.

than the social value g_i, and so the employee's effort is less than the surplus-maximizing level $p^*(g_i)$, and the firm could strictly increase profits (while giving the worker the same utility) by reducing w_F and increasing w_S. Given that $w_F = 0$, the IC is simply $\gamma'(p) = w_S$, and the employee's utility can be written directly in terms of the effort level p; for use throughout, we denote this utility level by $u(p)$,

$$u(p) \equiv p\gamma'(p) - \gamma(p).$$

That is, the employee exerts effort $\gamma(p)$ and is paid a bonus $\gamma'(p)$ after success.

For employees assigned to task L, the firm can set the bonus w_S to the value of success g_L to achieve the surplus-maximizing effort p_L^*. Any employee assigned to task L in equilibrium must receive this surplus-maximizing contract, and receives utility $u(p_L^*)$.

The endogenous reservation utility $\underline{U}$ must be at least $u(p_L^*)$, as follows. Suppose instead that $\underline{U} < u(p_L^*)$ in some equilibrium. So some employees receive a contract that gives them utility $\underline{U}$, and the firm makes zero profits. But this violates the profit-maximization condition, since the contract $(i, w_S, w_F) = (L, g_L - \varepsilon, 0)$ for some $\varepsilon > 0$ gives utility strictly in excess of $\underline{U}$ and delivers strictly positive profits.

An immediate consequence of $\underline{U} \geq u(p_L^*)$ is that no task L employee can be overpaid in equilibrium. We now characterize conditions under which employees on task H are overpaid, which amounts to showing that employees on task H earn strictly more than $u(p_L^*)$, but that some workers are still employed on task L in equilibrium; we then have $\underline{U} = u(p_L^*)$.

Making use of the IC condition $\gamma'(p) = w_S$, we can write the firm's profit maximization problem associated with task H assignment entirely in terms of effort p:

$$\max_p p\left(g_H - \gamma'(p)\right) - k_H \text{ s.t. } u(p) \geq \underline{U}. \tag{2}$$

Firm profits initially increase in the size of the bonus $\gamma'(p)$, since the concomitant increase in effort p justifies the cost; and then eventually decrease, as increasing the bonus affects effort less and less (due to the convexity of γ). For use throughout, denote the effort level that maximizes profits (absent the participation constraint) by $\underline{p}(g_H)$, i.e.

$$\underline{p}(g_H) \equiv \arg\max_{\tilde{p}} \tilde{p}\left(g_H - \gamma'(\tilde{p})\right).$$

The fact that worker utility and firm profits both increase in the incentive bonus initially is a common feature in moral hazard problems and a necessary condition for overpay—if profits uniformly decreased with pay employees would always be held to their reservation utility. To establish overpay, we also need to show that (i) the bonus that maximizes profits above is so high that workers are strictly better off than in task L, and (ii) the return that allows firms to just break even on task H is high enough that, in equilibrium, not all employees can be employed on task H.

It is useful to note that (i) can be restated in terms of effort, since the condition $u(p) > u(p_L^*)$ is equivalent to $p > p_L^*$. In other words, overpay is inseparable from working harder on task H than on task L. We stress, however, that because of the moral hazard problem, an employee who works $p > p_L^*$ receives strictly more utility than $u(p_L^*)$, so overpay is more than a compensating differential.

Denote by $\underline{g}_H$ the return at which the solution to (2) is exactly 0 when $\underline{U}$ is at its lower bound $u(p_L^*)$. The value of $\underline{g}_H$ represents a lower bound on the equilibrium return in the one-period economy: for returns $g_H < \underline{g}_H$, no employee can be assigned to task H, which is inconsistent with equilibrium from Assumption 2(i). It also represents an upper bound on returns consistent with equilibrium overpay, since if $g_H > \underline{g}_H$ all workers would end up with the same contract in task H.

Our main result in this section is:

Proposition 1 *(A) If (i) $\underline{p}\left(\underline{g}_H\right) > p_L^*$, and (ii) the volume of task H trade that can be sustained at return $\underline{g}_H$ is less than $2\underline{p}\left(\underline{g}_H\right)$, i.e., $\underline{g}_H/k_H > \zeta_H\left(2\underline{p}\left(\underline{g}_H\right), k_H\right)$, then the equilibrium return is $\underline{g}_H$ and task H employees are overpaid relative to task L employees. Otherwise, all employees receive the same utility.*

(B)As k_H becomes large, the equilibrium always features overpay.

Our primary interpretation of Proposition 1 is that it provides an explanation for why financial sector jobs are so highly compensated. In the introduction, we reviewed the evidence that financial sector jobs are overpaid; this is consistent with Proposition 1, where task H is interpreted as a finance task, and task L is interpreted as a non-finance task.

Because the economic forces behind Proposition 1 also operate in the full dynamic model, we walk through the proof here. First, the following lemma formally establishes the natural result that the profit-maximizing effort level $\underline{p}(g_H)$ increases with the marginal product g_H.

Lemma 1 *$\underline{p}(g_H)$ is uniquely defined; is strictly increasing in g_H; and $\underline{p}(g_H) \to \bar{p}$ as $g_H \to \infty$.*

Because the return g_H/k_H must exceed 1 in equilibrium, Lemma 1 implies that as the capital-per-employee k_H increases, profit-maximizing effort $\underline{p}\left(\underline{g}_H\right)$ approaches maximal feasible effort $\bar{p}$.

Hence whenever k_H is large enough, any employee assigned to task H indeed works strictly more than p_L^*, or equivalently, receives utility strictly in excess of $u(p_L^*)$.

Second, to establish overpay, we must still show that some employees are assigned to task L. Since $\underline{p}\left(\underline{g}_H\right)$ approaches $\bar{p}$, it certainly remains bounded away from 0. Consequently, as capital-per-employee grows large, by Assumption 2 the equilibrium return $\zeta_H\left(2\underline{p}\left(\underline{g}_H\right), k_H\right)$ associated with assigning all workers to task H and having them work $\underline{p}\left(\underline{g}_H\right)$ eventually falls below 1. Hence not all employees can be assigned to task H, and some workers must instead be assigned to task L, where they earn strictly less utility.

Figures 1 and 2 illustrate an overpaying equilibrium. Figure 1 shows that, at return $\underline{g}_H$, firms cannot make strictly profitable profits while satisfying the participation constraint, even though some employees are overpaid relative to others. Figure 2 illustrates how the fraction of overpaid workers is determined. In particular, Figure 2 illustrates that the equilibrium return $\underline{g}_H$ of the overpaying equilibrium is determined entirely by the contracting problem, and is independent of the trade-profitability function ζ_H. Economically, and as one can see from the figure, the reason is that at the return $\underline{g}_H$ the "supply" of task H is perfectly elastic, since any division of employees between the two contracts used in equilibrium is consistent with the profit-maximization condition. We make heavy use of this property in our analysis of aggregate shocks in Section VI.

We conclude this section with a discussion of the model ingredients required for equilibrium overpay; and the relation of our analysis to prior literature.

IVA Model ingredients required for equilibrium overpay

A satisfactory theory of overpay should explain why the combination of competition among firms for identical employees, and optimal contracts, does not eliminate overpay in equilibrium.

The *first* and most obvious ingredient required for overpay is that the moral hazard problem must be such that there is some region over which the firm's profit function increases in payments to the employee, or else the firm would always hold the employee down to his participation constraint. Together with an *exogenously* specified reservation utility $\underline{U}$ this relatively standard feature of the profit function can easily lead to a result where the employee's IC constraint binds while the participation constraint is slack, as in much of the partial equilibrium contracting literature. However, by itself this is not an overpay result, since it does not imply that ex ante identical employees receive different utilities in equilibrium.

Consequently, the *second* ingredient needed in a model explaining equilibrium overpay is that the reservation utility is set endogenously by competition among firms for employees, where in our setting free entry drives firm profits down to zero.

The question is then why all employees do not end up with the contract that maximizes utility; in fact, they typically would in our setting if the payoff g_H were set exogenously, since for generic values of g_H, there is a unique profit-maximizing contract. For example, suppose that we specified that $g_H > \underline{g}_H$, which implies that there is a task-H contract that gives strictly positive firm profits and employee utility strictly in excess of $u(p_L^*)$. But free entry would then ensure that the reservation utility goes up to the level at which firms just break even on task H, so that all employees are employed on task H, and earn the same utility. On the other hand, if $g_H < \underline{g}_H$, no-one can be employed on task H, and everyone would earn the utility $u(p_L^*)$ associated with employment in task L. So there is no overpay if either $g_H > \underline{g}_H$ or $g_H < \underline{g}_H$. Hence, the *third* ingredient necessary for overpay is that the return g_H is endogenously set by "supply and demand" for task H and, for a non-empty set of economies, ends up being exactly the return $\underline{g}_H$ that makes profit-maximizing firms just break even on task H when the participation constraint is slack.

IVB Related literature

The relation to the partial-equilibrium contracting literature is discussed above. Here, we consider the relation to the older efficiency wage literature, foremost Shapiro and Stiglitz (1984). Almost all papers in this literature lack optimal contracts, which lead to many to criticize the literature on the grounds that perhaps optimal contracting would eliminate overpay. A more recent incarnation in Acemoglu and Newman (2002) allows for optimal contracts in a static setting. However, they assume a combination of decreasing returns at the firm level, and limited firm entry. The combination of these features limits competition among firms for employees; and relaxing either of these features would eliminate equilibrium overpay. For many parts of the financial sector firms appear relatively easy to scale up, casting doubt on the importance of firm-level decreasing returns in this context; while for other parts of the financial sector, such as asset management, entry by new competitors is very common. In terms of modeling, the main difference relative to Acemoglu and Newman that allows us to derive overpay in equilibrium with more acute competition among firms is that we endogenize the equilibrium price of output (the return g_H).

A further difference between our model and these previous papers is that we allow for continuous (as opposed to binary) effort. This has the primary advantage that projects with more at stake, or larger projects, such as we often observe in the financial sector where it is often easy to scale up tasks, endogenously give higher rents to the agent because effort becomes more important.

V Equilibrium overpay in the dynamic economy

In this section we establish the core result of the paper: If task H stakes k_H are large, in equilibrium some employees are strictly overpaid relative to others; and moreover, overpaid employees have very different career paths relative to other employees. As discussed below, these differences qualitatively match the distinguishing characteristics of highly compensated financial sector jobs.

Note that considering a dynamic setting is important for three reasons, which we flesh out below. First, doing so addresses the "bonding" critique of the old efficiency wage literature, which suggested that optimal dynamic contracts would eliminate efficiency wages (see footnote 3). Second, it delivers implications for career structures, which a one-period model is by definition incapable of doing. Third, and somewhat related, it delivers implications for how aggregate shocks affect an employee's career; bonuses; the profitability and riskiness of investments; and on the response of capital to investment opportunities, where all these implications are very different from those of the one-period economy.

As a preliminary to solving for equilibria in the dynamic setting, observe that a lower bound for the participation constraint in the dynamic economy is given by the utility the employee gets if employed in the L-task for both periods. Analogously to the one-period setting above, such a contract involves giving the employee the full marginal product, which in the dynamic contract means setting $w_{SS} = 2g_L$, $w_{SF} = w_{FS} = g_L$, and $w_{FF} = 0$. This is equivalent to a repeated one-period contract, yields zero profits for the firm, and gives the employee $2u\left(p_L^*\right)$ in lifetime utility. We refer to this contract below as the C^{LL} contract.

We first give an example that shows that—consistent with the bonding critique—overpay may be eliminated by optimal dynamic contracts.

Example: Let $k_H \leq g_L$. Define g_H^* as the equilibrium return in a frictionless world where surplus across the two tasks are equalized, so that:

$$S_H\left(p^*\left(g_H^*\right);g_H^*\right) = S_L\left(p_L^*;g_L\right) = u\left(p_L^*\right). \tag{3}$$

We now show that—given some conditions on trade-profitability ζ_H—there is an equilibrium in the dynamic setting such that g_H^* is the equilibrium return, all employees earn the same utility, and effort is first-best in all periods. As we verify in the appendix, under the same circumstances the one-period economy often features overpay and inefficiently low effort on task H. The result is accomplished by giving the following dynamic contract to some subset $\lambda < 1$ of young employees: When young, the employee is employed on task L. After success, the employee is employed on task

H, where he receives the net profits from both tasks: $(w_{SS}, w_{SF}) = (g_H^* + g_L - k_H, g_L - k_H)$. Note that since $g_L \geq k_H$, these payments satisfy the limited liability constraint. After failure in period 1, the employee is employed on task L, receiving all profits: $(w_{FS}, w_{FF}) = (g_L, 0)$. Since the employee receives all profits net of the invested amount from the tasks he works on during his lifetime, it is as if he owned the firm, and hence he fully internalizes the effect of his effort on total surplus. Since by (3) the maximal surplus in the two tasks is the same, the maximal surplus is $2u(p_L^*)$, and so the employee earns the same utility as under contract C^{LL}.[17] The remaining fraction $1-\lambda$ of employees are assigned contract C^{LL}. As long as the amount of task H trade that can be sustained at g_H^* is lower than the number of task H successes associated with assigning all employees to the contract above, namely $p_L^* p^*(g_H^*)$, there is a $\lambda < 1$ such that the return consistency condition $\zeta_H(\lambda p_L^* p^*(g_H^*), k_H) = \frac{g_H^*}{k_H}$ is satisfied. Furthermore, since the contract generates maximal surplus, it is impossible to find another contract that gives higher profits to the firm without violating the participation constraint.

This example relates to an important insight of contract theory: employees with more wealth are easier to employ, because the wealth can be used as a bond to alleviate moral hazard.[18] In our setting, employing the employee on task L when young creates a payoff g_L after success that can be pledged as a bond on task H when old, which completely solves the moral hazard problem and leads all employees to earn the same utility across tasks, as in the frictionless benchmark of Section III.

We next establish our central result: As k_H grows large, overpay is not eliminated by dynamic contracts, the above example notwithstanding. Furthermore, career paths are very different from in the above example:

Proposition 2 *For all sufficiently large k_H, the unique equilibrium of the economy features:*

1. *Lower returns than in the one-period benchmark: $g_H < \underline{g}_H$.*

2. *Overpay: A strict subset of young employees start on task H, and receive strictly greater expected utility than young employees starting on task L.*

[17] In more detail, the employee's expected utilities after first-period success and failure are, respectively, $u(p^*(g_H^*)) + g_L - k_H$ and $u(p_L^*)$. Since $u(p) - k_H = S_H(p; g_H)$ when $p = p^*(g_H)$, by (3) the expected utility after success reduces to $g_L + u(p_L^*)$. So the employee exerts effort p_L^* in the first-period, and has expected utility $2u(p_L^*)$ across the two periods.

[18] For an early statement of this point, see Jensen and Meckling (1976); or for a more recent statement in a moral hazard problem close to the one in this paper, see Holmström and Tirole (1997). For recent papers that explicitly model the reduction in inefficiency associated with the dynamic accumulation of wealth, see, for example, DeMarzo and Fishman (2007), Biais et al (2007) and Biais et al (2010).

3. *Up-or-out for overpaid employees: Task H employees remain on task H if they succeed, and have higher success rates when old than when young. If they fail they are "demoted" to task L.*

4. *Dynamically segregated labor markets: Task L employees are never "promoted:" They remain in task L when old, and exert the same effort as when young.*

The results in Proposition 2 show that there is a limit to how much can be achieved by using the kind of career path illustrated in the example above, in which employees are assigned to the low moral hazard task before "graduating" to work on the high moral hazard task. Instead, when the capital at stake k_H becomes sufficiently large, it is more efficient to assign some young workers directly to the H task, and these workers will be overpaid relative to their unlucky identical twins who are stuck on the L task throughout their career. As we explain in more detail below, this is because as k_H becomes large, the promise of work on the H task after success—and the large surplus this allows the agent to capture—creates such a strong incentive to work when young that the agent can efficiently be employed on the H task already when young.

We derive the main steps of the proof of Proposition 2 here in order to illustrate the economics of the dynamic equilibrium. The proof is constructive. To proceed, we initially assume that property (1) is satisfied, i.e., $g_H < \underline{g}_H$, and then later confirm that this is indeed the case. An immediate implication of $g_H < \underline{g}_H$ is that a firm cannot break even by hiring an old employee on task H while delivering utility of at least $u\left(p_L^*\right)$. Consequently, the one-sided commitment constraint is

$$E\left(w|X\right) - \gamma\left(p\left(\boldsymbol{w}_X\right)\right) \geq u\left(p_L^*\right), \tag{4}$$

that is, each second-period subcontract must pay the employee at least as much as he would get in a one-period L-task contract.

VA "Up-or-out"

We next show that, if indeed a contract is overpaying so that the participation constraint is slack,[19] the one-sided commitment constraint (4) must bind after failure in the first period, and the employee must be allocated to task L—the failed employee is "out." To see this, note that giving the employee the minimal possible utility $u\left(p_L^*\right)$ after failure has two positive effects on profits: First, it increases incentives to work in period 1. Second, it maximizes profits after failure, subject to

[19] We verify below that the participation constraint is indeed slack for a contract featuring employment on task H when k_H is large.

delivering utility of at least $u(p_L^*)$. (This is so since $g_H < \underline{g}_H$, so that an assignment to H would produce strictly negative profits on the second period task.) Hence, the subcontract after failure is $(i_F = L, \boldsymbol{w}_F = (g_L, 0))$, and firm profits $p(\boldsymbol{w}_F)\, g_L - E(w|F)$ after failure are zero.[20]

We next argue that if the employee is ever to be employed on task H, he must be employed there after success—the successful employee is "up." Suppose this were not the case, so that the employee is assigned to task L both after success and failure, and hence must be assigned to task H when young. But then, the contract effectively becomes a one-period task H contract, which eliminates all dynamic incentives and makes it impossible for the firm to break even. The reason is that given that any subcontract must deliver utility weakly above $u(p_L^*)$, when the employee is assigned to L in the second period he must be given a subcontract of the form $(w_{XS}, w_{XF}) = (g_L + w_{XF}, w_{XF})$ that induces surplus-maximizing effort $p_S = p_L^*$ and pays any additional promised utility in the form of a fixed pay w_{XF}. But then the young employee faces a contract that delivers utility $u(p_L^*) + w_{FF}$ after failure and $u(p_L^*) + w_{SF}$ after success, costing the firm w_{FF} after failure and w_{SF} after success. This is equivalent to a one-period contract on task H.

VB Dynamic incentives

Given the up-or-out characterization, the key contract feature to discuss is why the overpaid employee is initially assigned to task H when stakes k_H are large. The alternative is to assign the employee to task L initially, then "promote" him to task H after success as in the example above. We denote these two alternatives as the HH and LH paths. To illustrate the relative merits of these career paths, it helps to write out the firm optimization problem more explicitly.

Given the up-or-out feature, an equilibrium contract must have $(i_S, i_F, w_{FS}, w_{FF}) = (H, L, g_L, 0)$, which as noted above leads to zero profits after failure. The remaining contract terms solve

$$\max_{i, w_{SS}, w_{SF} \geq 0} p\left[g_i + p(\boldsymbol{w}_S)\, g_H - E(w|S) - k_H\right] - k_i, \tag{5}$$

subject to the first-period IC and participation constraints,

$$\begin{aligned} \gamma'(p) &= E(w|S) - \gamma(p(\boldsymbol{w}_S)) - u(p_L^*), \\ p\left(E(w|S) - \gamma(p(\boldsymbol{w}_S))\right) + (1-p)\, u(p_L^*) - \gamma(p) &\geq 2u(p_L^*), \end{aligned}$$

and the one-sided commitment constraint (4).

[20] Our argument here is somewhat informal; we give the full details in the proof in the appendix.

Similar to the one-period problem in Section IV, by substituting in the IC constraint and making use of the notation S_H for total surplus, we can rewrite this problem purely in terms of first-period effort p and effort-after-success $p_S = p(\boldsymbol{w}_S)$:

$$\max_{i,p,p_S \geq 0} p\left[g_i - \gamma'(p) + S_H(p_S; g_H) - u(p_L^*)\right] - k_i, \tag{6}$$

subject to

$$\begin{aligned} \gamma'(p) &\geq u(p_S) - u(p_L^*) && (7) \\ u(p) &\geq u(p_L^*). && (8) \end{aligned}$$

Similar to the one-period setting, one benefit of this formulation is that it relates overpay—here, whether (8) binds—to whether the employee's first-period effort p exceeds the task L level p_L^*. In particular, provided that some employees receive the C^{LL} contract, an employee is overpaid if and only if $p > p_L^*$.

Formulation (6) also illustrates the power of dynamic incentives. In a one-period setting, the firm must promise a bonus $\gamma'(p)$ in order to incentivize effort p. However, in the dynamic setting, the bonus can be postponed[21] and paid partly in the form of the incremental surplus $S_H(p_S; g_H) - u(p_L^*)$ the employee creates on the H task after success relative to the surplus $u(p_L^*)$ created after failure. Constraint (7) shows the link between first and second period effort; this constraint is binding when $w_{SF} = 0$ so that both first and second period incentives are provided purely by a bonus w_{SS} after two successes. The constraint becomes non-binding when second-period effort p_S reaches the surplus-maximizing level $p^*(g_H)$; if first-period incentives beyond this level are needed they are paid as an additional fixed bonus after first-period success.

Whenever constraint (7) binds, it is immediate that $p_S > p$, so that an experienced (and previously successful) employee has a higher success rate than an employee at the start of his career.[22] In the appendix, we show that—as stated in Proposition 2—this conclusion remains valid even when constraint (7) is slack. Note that this positive correlation between success rates and experience occurs even without any human capital accumulation, and instead is purely due to the power of dynamic incentives.

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

[22] Manove (1997) contains a related result in a setting where, by assumption, only simple wage contracts are possible.

VC Overpaid employees initially assigned to task H

The LH path has the advantage that the profits g_L produced on the low moral hazard task can be used to alleviate employee limited liability and hence agency problems on the high moral task—the employee "works his way up" to higher responsibility. The advantage of instead assigning the young employee to task H is that if the surplus differential between the high and the low task becomes big, the large first-period incentives created by the up-or-out contract are most efficiently used to motivate work on the H task. We now show that the second effect dominates as k_H becomes large.

If g_H were exogenous, this property would follow naturally whenever g_H is sufficiently large, for the simple reason that task H is then very profitable. However, free entry would then allow all employees to be assigned initially to task H, so there would be no sense of equilibrium overpay in this case. Conversely, if the return g_H were set as in the example above so that surplus is equalized across the two tasks (as it would be in the frictionless economy), up-or-out incentives would be useless, and there would be no way for a firm in our economy to break even on the HH path. Key to our result is to show that, as k_H grows large, the equilibrium return in our setting with agency frictions grows high enough to create high surplus on task H, and therefore high up-or-out incentives, but not so high that all employees are drawn into task H.

To establish the result, we define the return G^{LH} such that firms can just break even on the LH path. We then show that, for k_H large, at the return G^{LH} firms can make strictly positive profits from the HH path. Competition among firms then drives the equilibrium return below G^{LH}, so that overpaid employees must be assigned to the HH path.

Formally, G^{LH} is the return g_H such that the solution to problem (6) subject to (7) (but *not* to (8)), and to the additional constraint that the initial task assignment is $i = L$, is exactly 0. First, note that G^{LH} has a simple and intuitive characterization:

Lemma 2

$$g_L + \max_p \left[p \left(G^{LH} - \gamma' (p) \right) - k_H \right] = 0. \tag{9}$$

The second term in (9) is the maximal profits a firm can get by employing the employee in the second period on task H with a success bonus of $\gamma' (p)$. The lemma states that the profits g_L earned after first period success must just cover the losses on such a contract, so the firm breaks even across the two periods.

As a next step in showing that the HH path dominates, the following lemma shows that the surplus on task H must grow large as k_H grows large if the return is such that the LH contract breaks even:

Lemma 3 *$S_H\left(\underline{p}\left(G^{LH}\right);G^{LH}\right) \to \infty$ as $k_H \to \infty$.*

The economics behind Lemma 3 is as follows. Certainly G^{LH} must grow large as k_H grows large. Hence (by Lemma 1) the effort level $\underline{p}\left(G^{LH}\right)$ must approach the maximal feasible effort $\bar{p}$. For this to happen, the bonus $w_{SS} - w_{SF}$ given to the employee for second-period success must grow very large. But the firm can only afford to give this large bonus if the total surplus $S_H\left(\underline{p}\left(G^{LH}\right);G^{LH}\right)$ grows large, since it has to recoup the amount $k_H - g_L$ (the investment amount less the bond posted from the first period).

To complete the argument, note that by Lemma 2,

$$\max_p p\left(G^{LH} - \gamma'(p) + \frac{g_L}{p}\right) - k_H = 0.$$

In other words, at G^{LH} the firm can almost break even on a one-period contract on task H, in the sense that if the cost of incentive provision were reduced by just $\frac{g_L}{p}$ then the firm would break even. By Lemma 3, when k_H is large the dynamic incentives associated with "promoting" an employee to task H after success reduce the cost of incentive provision by a large amount, and in particular, more than $\frac{g_L}{p}$; formally, for k_H large, certainly

$$\max_p p\left(G^{LH} - \gamma'(p) + S_H\left(\underline{p}\left(G^{LH}\right);G^{LH}\right) - u\left(p_L^*\right)\right) - k_H > 0.$$

Define G^{HH} analogously to G^{LH}, i.e., with the constraint that the initial task assignment is $i = H$. We have just established:

Lemma 4 *For all k_H sufficiently large, $G^{HH} < G^{HL}$.*

This completes the construction of the equilibrium of Proposition 2. The equilibrium return is G^{HH}. Since (trivially) $G^{LH} \leq \underline{g}_H$, we know $G^{HH} < \underline{g}_H$ for k_H large, as we have been assuming. Let C^{HH} be the contract associated with G^{HH}; this is the contract received by overpaid employees in equilibrium. From a comparison of problems (2) and (6), the return comparison $G^{HH} < \underline{g}_H$ implies that

$$S_H\left(p_S;G^{HH}\right) > u\left(p_L^*\right) = S_L\left(p_L^*;g_L\right), \tag{10}$$

so that, in contrast to the example above, surplus is not equalized across the two tasks; instead, at the equilibrium contract, the second-period surplus from task H strictly exceeds maximal surplus in task L.

To establish the overpay result of Proposition 2, the two main conditions[23] to check are as in the one-period case, namely that (i) the contract C^{HH} indeed delivers strictly higher utility than the C^{LL} contract, and (ii) not all employees are given the C^{HH} contract, so that C^{HH}-employees are indeed overpaid.

For (i), recall from above that overpay (i.e., $U\left(C^{HH}\right) > U\left(C^{LL}\right)$) is equivalent to the effort condition $p > p_L^*$. This indeed holds for large k_H: Given (10), certainly p exceeds the profit-maximizing effort $\underline{p}\left(G^{HH}\right)$. Since in equilibrium $G^{HH}/k_H \geq 1$ for firms to break even, and $\underline{p}\left(G^{HH}\right) \to \bar{p}$ as G^{HH} grows large (Lemma 1), it follows that $p > p_L^*$ for large k_H.

For (ii), the argument is exactly as in the one-period economy: task H effort-per-employee in the C^{HH} contract is bounded below by $\underline{p}\left(G^{HH}\right)$, so if all young employees were given the C^{HH} contract the equilibrium return is bounded above by $\zeta_H\left(\underline{p}\left(G^{HH}\right), k_H\right)$, which falls below 1 when the capital-per-employee k_H is large enough. Hence not all employees can get contract C^{HH}, and given Lemma 4 the only alternative is contract C^{LL}.

VD Applications

As with Proposition 1, an important interpretation of Proposition 2 is that it provides an explanation of why financial sector jobs are so highly compensated. In the introduction, we reviewed the evidence that financial sector jobs are overpaid; this is consistent with Proposition 2, where task H is interpreted as a finance task, and task L is interpreted as a non-finance task.

In addition, the equilibrium of Proposition 2 has several other features that match, at least qualitatively, the characteristics of financial sector jobs.

Long hours: Employees who receive the C^{HH} contract work harder than employees who receive the C^{LL} contract. This is consistent with the commonplace observation that financial sector jobs entail very long hours. For example, using a sample of University of Chicago MBA alumni, Bertrand et al (2009, WP) report that the average hours worked in investment banking is 73.6 hours/week; the next highest figure reported is for consulting, at 60.7 hours/week. We reiterate, however, that Proposition 2 says the pay received by financial sector employees is more than a compensating differential for these long hours. Hence MBA students who land an investment job have effectively won a lottery, which is consistent with casual empiricism.

Heavy use of both performance-pay and backloading of pay: Employees who receive the C^{HH} receive larger bonuses than those receiving the C^{LL} contract. Moreover, pay is also more back-loaded in the C^{HH} contract, as follows. First, observe that the firm could pay the employee up

[23] The proof in the appendix takes care of some other details.

to w_{SF} after first-period success, without affecting incentives. Accordingly, we will identify w_{SF} with the first-period bonus, and measure the backloading of the pay via the ratio $\frac{w_{SS}-w_{SF}}{w_{SF}}$, i.e., the ratio of the second-period bonus to the first-period bonus. For the C^{LL} contract, this ratio is 1, i.e., no backloading. For the C^{HH}, this ratio exceeds 1.[24]

These predictions are consistent with perceptions that the financial sector makes heavy use both of performance pay and backloaded pay. Using a sample of Stanford MBA alumni, for investment banking Oyer (2008) documents a very steep slope in the relation between annual compensation and years since graduation.[25] The website WWW.CAREERS-IN-FINANCE.COM/IBSAL.HTM reports investment banking pay, and tells a similar story. Bell and Van Reenen (2013) report detailed compensation data for the "code staff" of large banks headquartered in London.[26] For this admittedly senior group of bank employees, 58.3% of total compensation is deferred; while out of the non-deferred portion, 64.4% is bonus pay.

Importance of entering profession soon after graduation: As noted, Proposition 2 features dynamic segregation: If an employee is not assigned to task H when young, he never is. This is consistent both with anecdotal accounts (see, e.g., DeChesare (2012)), and with Oyer's econometric finding that the probability that a Stanford alumuni works in the financial sector is heavily influenced by whether the alumni's first job after graduation was in the financial sector. Moreover, to mitigate endogeneity concerns, Oyer shows this finding is robust to instrumenting the initial job placement by aggregate economic conditions. (In the following section, we formally add aggregate shocks to our basic model.)

Up-or-out: As noted, Proposition 2 predicts that the overpaying contract C^{HH} has an up-or-out feature: employees who fail in the first period are assigned to task L in the second period. This is consistent with anecdotal accounts of bankers moving to a "normal company" as an "exit option" (see, e.g., DeChesare (2012)).

A slightly different way to interpret Proposition 2 is to map task H to a high-stakes financial sector job, and task L to a lower-stakes financial sector job. This is consistent with anecdotal accounts of people exiting investment banking to enter other lower-paid parts of the financial sector, but not the reverse. It is also consistent with Hong and Kubik's (2003) study of security analysts. They show that it is much more common for security analysts to move from a high-paying, more

[24] To see this, note that if $w_{SF} = 0$ then certainly $\frac{w_{SS}-w_{SF}}{w_{SF}} > 1$. If instead $w_{SF} > 0$, then we know $w_{SS} - w_{SF} = g_H$. But in this case, $w_{SF} < g_H$, since if instead $w_{SF} \geq g_H$ the firm has strictly negative profits. Hence $\frac{w_{SS}-w_{SF}}{w_{SF}} > 1$.

[25] Note that this slope may be affected by survivorship bias.

[26] "Code staff" include senior management and anyone whose professional activities could have a material impact on a firm's risk profile.

prestigious brokerage firms to a lower-paying, less prestigious one than the other way around.

VE Leaving to start a hedge fund

In our discussion of "up-or-out" immediately above, we focused on the "out" half, in which financial sector employees who perform poorly early in their career either exit the financial sector, or else move to less good jobs within the financial sector. We now consider the "up" half in more detail.

One path for financial sector employees who perform well is to remain with the same employer, and gain promotion to senior positions (e.g., "Vice President," "Director," "Principal," "Managing Director"). As we discussed above, our analysis predicts that this promotion is accompanied by larger bonuses. This path is consistent with the representation we have used for dynamic contracts, in which the employee remains with the same firm. (Since the equilibrium subcontract after failure, i.e., $(i_F, w_{FS}, w_{FF}) = (L, g_L, 0)$, delivers zero profits, it can be delivered either in the same firm, or in a different firm.)

An alternative path for successful financial sector employees is to take their accumulated bonuses, and combine them with outside capital to start a hedge fund. Anecdotal accounts suggest that many hedge funds are started in this way (see, e.g., journalistic accounts by Fishman (2004) and Makan (2012)). Such a path is consistent with our analysis, as we next discuss.

In detail, consider the overpaying C^{HH} contract characterized above. After an employee succeeds in the first period, he anticipates payments $\boldsymbol{w}_S = (w_{SS}, w_{SF})$ which depend on whether he succeeds or fails in the second period. These payments induce him to exert effort $p(\boldsymbol{w}_S)$. Consequently, the firm's expected revenue is $p(\boldsymbol{w}_S) g_H - k_H$, and its expected compensation bill is $E(w|S)$. The firm's overall profits would hence be the same if it paid the employee a bonus

$$W_S = E(w|S) - (p(\boldsymbol{w}_S) g_H - k_H)$$

for first period success, and the employee left the firm.

Armed with the bonus W_S, the employee can then start his own investment fund. If $W_S \geq k_H$, he can do so without raising outside financing. If instead $W_S < k_H$, he requires additional financing of $k_H - W_S$. To raise this financing, he promises to pay investors an amount $g_H - w_{SS}$ contingent on success (and nothing after failure). These financing terms are sufficient to attract investors, as follows. It is straightforward to show that $w_{SF} = 0$ when $W_S < k_H$;[27] it then follows from the

[27] In particular, the inequality $W_S < k_H$ is equivalent to $p(\boldsymbol{w}_S) w_{SS} + (1-p) w_{SF} < p(\boldsymbol{w}_S) g_H$, which implies $w_{SS} < g_H$.

definition of W_S that

$$p(\boldsymbol{w}_S)(g_H - w_{SS}) = k_H - W_S,$$

so that investors receive the required rate of return in expectation.

VF Should firms respond to overpay by scaling down task H?

Our analysis shows that when the amount k_H at stake is large, some employees are overpaid in equilibrium. As we discussed earlier, the financial sector is a place where high k_H naturally arises, since it is relatively easy to place a large quantity of resources under the supervision of a single person. In other words, it is straightforward to increase the scale of tasks in the financial sector, at least up to some point (effectively k_H in our analysis) where decreasing returns become severe.

However, and by the same token, it is also possible to scale *down* tasks in the financial sector. Given that overpay arises precisely when k_H is large, at first sight this might seem to be an attractive option: By scaling down the resources of each financial sector employee, perhaps firms could eliminate overpay.

To examine this possibility, suppose that task H can be scaled down by any factor $\delta \in [0,1]$, so that it yields $\delta(g_H - k_H)$ after success but $-\delta k_H$ after failure. However, it is easy to see that no firm would ever want to scale down, as follows. In any contract with a task assignment H in some node, the derivative of profits with respect to δ holding everything else constant is $pg_H - k_H$, where p is effort at the node. This expression must be non-negative in any equilibrium contract: If instead it is negative, the firm is better off assigning the employee to L in that node without changing payments, which leads to no change in employee utility but a change in profits of $pg_L - (pg_H - k_H) > 0$. But given $pg_H - k_H \geq 0$, choosing the maximal possible scale $\delta = 1$ is optimal.

The scalability of financial sector tasks is what makes it possible for a relatively small fraction of the workforce to oversee assets worth several multiples GDP. This same scalability makes k_H large for the financial sector, generating equilibrium overpay. Nonetheless, financial sector firms would not benefit from scaling down the resources of each employee, since by doing so, they fail to exploit the available economies of scale.

In Section VIII below we discuss how changes in the feasible scale of assets per employee can potentially explain the large rise in pay in the financial sector over the last three decades.

VI The effect of aggregate shocks on career dynamics

We now extend our basic model to allow for aggregate shocks. This allows us to study the time series implications of our model along several dimensions, including: the effects of initial conditions on an employee's career; bonuses; profitability and riskiness of investments; and the response of capital to investment opportunities.

We start with a specification of our basic model in which k_H is sufficiently large so that young employees who start in task H are overpaid. To keep the analysis as simple as possible, assume the aggregate state is either "Good" (G) or "Bad" (B), where the good state supports more aggregate activity (number of trades) y_i in task i for a given success payoff: $\zeta_H^G(\cdot, k_H) \geq \zeta_H^B(\cdot, k_H)$ 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 employees.

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

VIA 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 an employee enters the labor force have long-lasting effects on his career. In particular, we show that when the economy enters the bad state, firms respond on the hiring rather than the firing margin, so that entering young employees 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 employees that get these jobs. This hiring pattern is consistent with Oyer's (2008) evidence for the financial sector, and more broadly, with Kahn's (2010) finding for college graduates in general.

In this subsection we assume the shock only affects task H, i.e., $g_L^G = g_L^B$ and $\zeta_H^G(\cdot, k_H) > \zeta_H^B(\cdot, k_H)$. This assumption makes the analysis very straightforward, because it implies that both the equilibrium return g_H and contracts are independent of the aggregate state, as we now show. The key is to recall that in the overpay equilibrium characterized by Proposition 2, the return G^{HH} is independent of the trade-profitability function ζ_H, and instead is determined by the condition that the firm has zero profits under the profit-maximizing contract (see in particular the discussion in Section IV). Because $g_L^G = g_L^B$, the employee's minimum continuation utility $u(p_L^*)$ is independent of the state, and so the firm's profit maximization problem is the same as in the case without aggregate shocks. Consequently, G^{HH} is again the return at which a firm can just break even assigning employees to task H, and remains the equilibrium return, independent of the state; and hence C^{HH} remains the contract received by overpaid employees. In essence, the task H "supply" curve is perfectly elastic at the return G^{HH}, because it is consistent with any division of young employees across contracts C^{LL} and C^{HH}. So as long as the trade-profitability function ζ_H^ω—"demand"—does not vary too much across states, shocks are absorbed purely via changes in the number of young employees given contract C^{HH}. To be more specific, let λ_t be the number of overpaid young employees hired for task H at date t. Write y_H^ω for the task H supply that can be sustained at the equilibrium return G^{HH} in state ω, i.e., y_H^ω solves $G^{HH}/k_H = \zeta_H^\omega\left(y_H^\omega, k_H\right)$. Denote by p and p_S the success probabilities for employees on task H when young and old, respectively: Given the conjecture that returns 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 $p\lambda_t + p\lambda_{t-1}p_S$, where $p\lambda_t$ is the output by the λ_t just-hired young employees and $p\lambda_{t-1}p_S$ is the output from the λ_{t-1} old employees who were hired last period and succeeded when young. Consequently, the number of young employees hired for task H at date t is

$$\lambda_t = \frac{y_H^{\omega_t}}{p} - \lambda_{t-1}p_S. \tag{11}$$

As one would expect, more young employees are assigned to task H in good states, and when fewer employees were hired at the previous date. We verify in the appendix that it is indeed possible to vary the number of employees hired by a sufficient amount to fully absorb the aggregate shock, with no effect on the equilibrium return g_H, as long as the shock is not too large.[28]

It is easy to see from (11) that if the economy remains in state $\omega \in \{G, B\}$ for a long time, the number of young employees assigned to task H converges to λ^ω, defined by $\lambda^\omega \equiv \frac{y_H^\omega}{p(1+p_S)}$, and the age-profile of task H employees converges to p old employees for every young employee. As one would expect, a sustained period in the good state leads to greater hiring of young employees into the overpaid task H jobs, i.e., $\lambda^G > \lambda^B$. Average success rates, on the other hand, are the same in both scenarios.

Proposition 3 *Suppose that after many periods in the good state, the economy suffers an aggregate shock and enters the bad state. Hiring of young employees into task H falls below even λ^B, and young employees 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, the average success rate in task H*

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

actually increases.

The proof is almost immediate from (11), and we give it here. In the first period that the economy is in the bad state, the number of young employees hired into task H is

$$\lambda_t = \frac{y_H^B}{p} - \lambda^G p_S < \frac{y_H^B}{p} - \lambda^B p_S = \lambda^B < \lambda^G.$$

The age-profile in task H is now skewed towards experienced employees. Since experienced employees have higher success rates, i.e., $p_S > p$ (see Proposition 2) the average success rate in task H increases when the bad shock hits.

Implication: Initial conditions matter.
The most immediate implication of Proposition 3 is that the conditions when someone first enters the labor force have lifelong consequences. This is consistent with Oyer's (2008) finding that Stanford MBAs who graduate when the stock market is performing well are much more likely to be working in the financial sector ten (and more) years later. Given that, as documented by Oyer, expected compensation in the financial sector is so high, one might expect MBAs who graduated during depressed financial markets to switch into the financial sector at some point subsequent to graduation. Our model, in which we identify task H with a high-paying financial sector job, gives an explanation for why this does not happen.

Implication: Financial sector firms respond to bad times by hiring less.
The reason task H hiring falls below even λ^B is that in the good state, firms hired many employees into task H, and the optimal contract prescribes that these employees are retained when old even in a downturn, which is at the expense of hiring new young employees. According to the US Bureau of Labor Statistics, and as predicted by our model, hiring by the financial sector fell in 2008. While firing is harder to empirically identify, the same data shows that total separations also fell in 2008: our model predicts no change in firing, while one might naively expect that separations would increase.

Implication: Investments undertaken in bad times are more profitable and have higher success rates.
Proposition 3 predicts that average success rates in task H are higher in bad times. Applied to the financial sector, this prediction says that investments have lower success rates in good times, when the financial sector has a higher proportion of less experienced employees. Conversely, investments will appear to grow more prudent in bad times, even though (by definition) attitudes towards

risk are unchanged in our model. Related, the expected profitability of investments (gross of compensation to managers) is countercyclical. This finding is consistent with anecdotal evidence about poor investments made during the internet and biotech bubbles by venture capital firms, as well as some of the most successful deals being initiated during busts. Academic studies have also found evidence that suggests such countercyclical investment performance in both the buyout (Kaplan and Stein (1993)) and the venture capital markets (Gompers and Lerner (2000)).

Remaining observations:
Above, we noted the prediction of countercyclical success rates on investments. More generally, this prediction can be interpreted as countercyclical productivity in some segments of the economy. 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.

Although we focus primarily on the implications of our model for career dynamics, it is interesting to note that Proposition 3 can also be interpreted in terms of unemployment. To do so, think of task L as corresponding to unemployment, with $u\left(p_L^*\right)$ the level of utility obtained by the unemployed. Then Proposition 3 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.

VIB Time series implications: Procyclical moral hazard

Next, we expand our analysis to the case in which aggregate shocks affect both tasks, i.e., $g_L^G > g_L^B$ and $\zeta_H^G\left(\cdot, k_H\right) > \zeta_H^B\left(\cdot, k_H\right)$. The significance of shocks for task L output is that they affect $u\left(p_L^*\right)$, the minimum continuation utility that an employee can be given. This in turn affects incentives. We show that in good times investments fail more frequently, even at the same time as employees receive more generous bonuses. Moreover, we show that capital does not fully respond to improvements in investment opportunities, i.e., is "slow-moving."

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_H^ω for the task H state ω return. Write u_L^ω for $u\left(p_L^*\right)$ evaluated at g_L^ω; note that $u_L^G > u_L^B$ since $g_L^G > g_L^B$. So when a young employee enters the labor force at date t, the minimum expected continuation utility he can be given is

$$\bar{u}_L^\omega \equiv \sum_{\psi=G,B} \mu^{\omega\psi} u_L^\psi.$$

The state-persistence assumption $\mu^{GG} > \mu^{BG}$ implies $\bar{u}_L^G > \bar{u}_L^B$, and so employees entering the labor force in good times are harder to incentivize, because the minimum utility they can be threatened with is higher. Colloquially, employees expect to "land on their feet" even if they fail. This is the key economic force driving our results below.[29]

In contracts for young employees starting in task H, firms commit to make success payments of $w_{SS}^{\omega\psi}$ and $w_{SF}^{\omega\psi}$. (Given our focus on the case in which k_H is high and overpaid task H jobs exist, we know the payments after failure are $\left(w_{FS}^{\omega\psi}, w_{FF}^{\omega\psi}\right) = \left(g_L^\psi, 0\right)$.) So to determine the equilibrium, we must find the contract terms $\left(w_{SS}^{\omega G}, w_{SS}^{\omega B}, w_{SF}^{\omega G}, w_{SF}^{\omega B}\right)$ and return g_H^ω for each of today's state realizations $\omega = G, B$. For the case with overpaid employees, this involves solving for the return at which the firm breaks even with the profit maximizing contract:

$$\max_{p^\omega, w_{SS}^{\omega\psi}, w_{SF}^{\omega\psi}} p^\omega \left(g_H^\omega + \sum_{\psi=G,B} \mu^{\omega\psi} \left[p\left(\mathbf{w}_S^{\omega\psi}\right) g_H^\psi - E\left[w^{\omega\psi}|S\right] - k_H \right] \right) - k_H \tag{12}$$

subject to the first-period IC

$$\gamma'\left(p^\omega\right) = \sum_{\psi=G,B} \mu^{\omega\psi} \left[E\left[w^{\omega\psi}|S\right] - \gamma\left(p\left(\mathbf{w}_S^{\omega\psi}\right)\right) \right] - \bar{u}_L^\omega. \tag{13}$$

As in the non-aggregate shock case, by substituting the IC into the profit expression, the profit-maximization problem can be written entirely in terms of the effort choices, $p^\omega, p_S^{\omega G}, p_S^{\omega B}$:

$$\max_{p^\omega, p_S^{\omega G}, p_S^{\omega B}} p^\omega \left(g_H^\omega - \gamma'\left(p^\omega\right) + \sum_{\psi=G,B} \mu^{\omega\psi} S_H\left(p_S^{\omega\psi}; g_H^\psi\right) - \bar{u}_L^\omega \right) - k_H \tag{14}$$

[29] Acemoglu and Newman (2002) note the existence of a similar effect of outside options, and use this observation to consider cross-country differences in corporate structure. In contrast to their stationary model, we examine how outside options fluctuate over time in response to aggregate shocks.

subject to

$$\gamma'(p^{\omega}) \geq \sum_{\psi=G,B} \mu^{\omega\psi} u\left(p_S^{\omega\psi}\right) - \bar{u}_L^{\omega}. \tag{15}$$

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 IC (13), which captures the fact that the higher outside option $\bar{u}_L^{\omega}$ in the good state makes it more costly to incentivize employees. To establish procyclical moral hazard, we must show this incentive effect dominates the direct effect that, for any fixed level of task H activity y_H, returns are higher in good times, i.e., $\zeta_H^G(y_H, k_H) > \zeta_H^B(y_H, k_H)$, which tends to ameliorate the moral hazard problem. However, precisely because employees are overpaid in equilibrium, the supply of task H activity is completely elastic (see the one-period benchmark model for a discussion of this point), so that the trade-profitability function ζ_H has *no* direct impact on equilibrium returns (exactly as in the previous subsection).[30]

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

Proposition 4 *(A) Overpaid young employees work less hard in good times, $p^G \leq p^B$ (where the inequality is strict unless all old employees work the socially efficient amount) but receive strictly higher bonuses, $E\left[w_{SS}^{G\psi}|\omega = G\right] > E\left[w_{SS}^{B\psi}|\omega = B\right]$.*

(B) Pay for luck: regardless of today's state ω, the success bonus $w_{SS}^{\omega\psi} - w_{SF}^{\omega\psi}$ is strictly higher when next period's state is good ($\psi = G$).

(C) Equilibrium returns are strictly higher in good times, $g_H^G > g_H^B$.

Part (A) is our formal result that failure rates are higher in good times and lower in bad times. Although the implication is the same as Proposition 3, the mechanism is different. Whereas the previous result reflected a change in the ratio of experienced to inexperienced employees, this new result reflects a decrease in incentives of overpaid employees. In the particular case of the financial sector, this prediction fits well with perceptions that traders and bankers are more careless in financial booms. Part (A) also establishes that booms generate higher (promised) bonuses—but the rise in bonuses is insufficient to offset of effect of improved outside options.[31]

[30] However, the increase in g_L has an indirect effect on equilibrium returns: Because workers are more difficult to incentivize, the equilibrium return g_H must rise, as can be seen from the equilibrium profit condition (12), and formally established in Proposition 4.

[31] Related, Proposition 3 above established one type of cohort effect, namely that entering the labor force in a good aggregate state increases an employee's lifetime utility because it increases his chances of entering an overpaid

Part (C) is our slow-moving capital result. By definition, for any given level of trading activity, in good times task H investments are more profitable than in bad times, i.e., $\zeta_H^G(y_H, k_H) > \zeta_H^B(y_H, k_H)$. Other things equal, this pulls more capital into task H. However, because moral hazard is countercyclical, the inflow of capital is tempered by the increased cost of incentivizing employees. It is worth contrasting this result with the effect of aggregate shocks in the one-period economy (when it exhibits overpay): There, an increase in trade-profitability ζ_H pulls in so much new capital that the equilibrium return remains at $\underline{g}_H$.[32] Consequently, in the one-period economy bonuses and success probabilities are likewise unaffected by state.

Part (B) is our "pay for luck" result: The employee is strictly better off if the state turns out to be good when he is old, even though he has no control over the state. This follows simply from the fact that the employee's marginal productivity is higher in the good state since the return is higher in the good state; hence, it is cheaper to deliver utility to employees in the good state. Hence, in a dynamic setting such as ours, Holmström's (1979) well-known informativeness principle, which states that compensation should only be made contingent on variables that depend on an agent's effort, does not hold. A number of empirical papers have documented that pay for luck is a pervasive phenomenon, and have interpreted this as evidence of inefficient contracting—a conclusion that our analysis casts some doubt on.[33,34] In the specific context of the financial sector, this result says that optimal contracts *should not* be fully indexed for aggregate market returns, as is often argued.

VII Distortions in the allocation of talent

We argued in the introduction that the available evidence suggests that high compensation in the financial sector is not a skill premium. Accordingly, in our basic model we have abstracted from skill differences by assuming that employees are ex ante identical. However, our model can be extended to produce interesting implications for the matching of heterogeneously-skilled employees 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 bankers instead. Second, talent may

job. Part (A) of Proposition 4 establishes a second type of cohort effect: even conditioning on an employee entering an overpaid job, the employee earns more if he enters the labor force in a good aggregate state. Although we are not aware of any direct evidence for financial sector jobs, Baker, Gibbs, and Holmström (1994) and Beaudry and DiNardo (1991) provide empirical evidence for these type of within-firm cohort effects in wages for other sectors of the economy.

[32] In the one-period case, this is true regardless of whether g_L changes across states.

[33] Since workers in our model are risk-neutral, pay for luck has no direct utility cost. However, since pay for luck is strictly optimal, we conjecture that it would remain optimal even after some degree of risk-aversion is introduced.

[34] The same economic force towards pay for luck operates in, for example, DeMarzo *et al* (2012).

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 employees have higher skills, while the remaining "ordinary" employees are homogeneous as before. This assumption ensures that the basic structure of the equilibrium remains unchanged. Specifically, suppose that a null set of employees 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 employees to be more generously rewarded than other employees; 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. We show, however, that this is not necessarily the case.

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 an employee 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 an employee would be best allocated to task L (for maximization of total output). However, any firm employing young employees at overpaid terms in task H can profitably "lure" this employee. For example, the employee 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 employee 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 employee fails, his continuation utility is higher than an ordinary employee's, because one-sided commitment leads firms to compete for his talents. This better outside option after failure makes the more talented employee harder to incentivize when young. (Note that this is the same force as operates in the aggregate shocks analysis of Section VI above.) Colloquially, he is "difficult," or "hard-to-manage." Holding task L talent fixed, the talent scorned force dominates whenever the employee'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 employee in the economy does not get the best job, even though he would prefer to.[35]

[35] A contemporaneous paper of Ohlendorf and Schmitz (2012) studies a similar repeated moral hazard problem, and

As the employee'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, surplus-maximization would dictate that the employee *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 employees are employed in task H even though they would increase output more if employed in task L. In this case, talent is truly lured.

VIII Increasing compensation in the financial sector

It is well-documented that pay has increased remarkably in the financial sector over the last three decades, and in particular relative to pay in the rest of the economy (see Philippon and Reshef (2008), Kaplan and Rauh (2010), and Bell and Van Reenen (2013)). As Philippon and Reshef show, much of the rise from the mid 1980s to 2007 cannot be attributed to an increase in human capital or hours worked in the sector—in other words, the last three decades have seen a steady rise in *overpay* in the financial sector. We believe our model of overpay can shed some light on this trend in pay.

First, Section VI predicts that pay is higher when expectations about the future level of g_L are higher; this increases employees' outside options after failure, and necessitates an increase in incentives. To the extent that the last three decades is a time period when, in general, the future has looked good, this predicts an increase in pay—and especially bonus pay—over time.

Second, and related, there is a perception that general skills have increased in importance over the same time period; see, for example, both references and evidence (from the CEO market) in Custódio *et al* (forthcoming). Because such a trend would increase employees' outside options after failure, it again would generate an increase in pay over time, for the same reasons as above.

Third, our model ties overpay to the amount at stake k_H that each employee is responsible for, and there is evidence that this amount has increased over the last three decades. As documented in Kaplan and Rauh (2010), capital employed per worker in the top 50 US securities firms increased by a factor of 9 in real terms between 1987 and 2004. They also document that the last three decades have seen a large increase in assets under management per employee within both hedge funds, private equity funds, and venture capital funds.

similarly shows that employers may avoid more talented employees. In their model, the firm avoids more talented employees as a commitment device to avoid renegotiation after failure; in contrast, our result stems from competition from other firms.

To the extent to which an increase in k_H is responsible for increasing pay, this begs the question of why k_H has increased. One possibility is technological advance. Another possibility, which is consistent with the evidence of Philippon and Reshef, is that it is the result of deregulation. In particular, these authors document a strong relation between deregulation events and the rise in pay, where at least three of the four deregulation events they analyze have increased the potential scale of projects in the financial sector (softening of bank branch regulation, repeal of the Glass Steagall act, and an easing of restrictions on mergers between insurance companies and banks).

If our model-based explanations for the increase of overpay are correct, this has implications both for predictions of overpay going forward and for the current policy debate on how to curb financial market pay. In line with standard economic models, Philippon and Reshef interpret overpay as a rent that is incompatible with competition, and hence unsustainable in equilibrium. In contrast, our paper shows that overpay is consistent with competition, and so may survive going forwards. Note, however, that our paper also predicts overpaid workers are not immune from recessions, and so is consistent with declines in bonuses during the recent "great recession."

In terms of policy, our analysis has a number of implications. First, limits on capital-per-employee would reduce overpay, but would also reduce economy-wide surplus, since from subsection VF this forces firms to do something they could freely do on their own. Second, expanding the scope and/or enforcement of employment bans within the financial sector following failure would potentially relax the one-sided commitment constraint, and hence serve both to reduce overpay and increase economy-wide surplus. Third, and contrary to the claims of many commentators, implicit government guarantees may actually serve to reduce rather than increase overpay. To see this, consider the extreme case in which the government guarantees the full capital at risk per employee, k_H. In this case, firms can afford to reward successful employees with the full success payoff g_H, which results in surplus-maximizing effort $p^*(g_H)$ on task H as well as task L. In equilibrium, the return g_H falls until surplus form the two tasks is equalized, i.e., (3), and overpay is eliminated.

IX Conclusion

Many financial sector employees are extremely highly compensated. Public opinion generally views these employees as overpaid; and empirical research is largely consistent with this view, in the sense that it finds that high compensation is hard to attribute to skill differences. However, a basic prediction of standard economic models is that overpay is inconsistent with competitive labor markets.

In this paper, we present a model in which some employees are overpaid relative to others. The key ingredients are moral hazard in effort, and projects where failure has severe consequences. Both features arise naturally in the financial sector, where intrinsic motivation is likely to be low, and many employees are responsible for large financial positions. Relative to the existing partial equilibrium contracting literature, our paper explains why otherwise identical employees receive different contracts, instead of all receiving the (generically unique) profit-maximizing contract; and relative to the efficiency wage literature, our paper explains how overpay can survive both optimal dynamic contracts and full competition among firms.

In addition to the overpay prediction, we also explain a number of other features of financial sector jobs, such as punishing hours and up-or-out promotion structures. Moreover, by allowing for aggregate shocks, we obtain implications for the effects of initial conditions on an employee's career; bonuses; profitability and riskiness of investments; and the response of capital to investment opportunities. Finally, an extension to observable skill differences delivers implications for when talent is "lured" away occupations, and when high-skilled individuals are "scorned" and do not receive the best jobs.

For tractability, we analyze the simplest possible model with both multiple tasks and long-lived employees, 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 employees who live more than two periods.

We have completely abstracted from unobservable skill differences in our model. We do not mean to suggest that unobservable skill differences are unimportant; our focus on the single friction of moral hazard is to isolate an economic force leading to dynamic segregation among sufficiently similar individuals. Clearly, if perceptions of an individual's skill increase by enough mid-career, then this individual may be promoted and escape dynamic segregation. Indeed, casual empiricism suggests that investment bankers who are unusually successful are sometimes poached by higher-paying firms. On the other hand, for deal-making firms such as hedge funds and private equity funds, a first-order concern for investors is the amount of "skin in the game," or personal wealth reinvested in the firm, that deal-makers have; as explicitly discussed in Section V, this is consistent with our model.

In the paper, we have assumed that the only information relevant for predicting the success probability of a trade is employee effort, represented by p. As noted, this implies that, in equilibrium, trades would never be aborted. To deepen the analysis of the effect of moral hazard on risk taking, an interesting extension might be one in which the employee can learn something relevant about the probability of success before the trade, but after the search effort has been sunk.[36] It might then potentially be valuable to structure contracts such that the employee has an incentive to abandon trades that look unpromising, which can be done by giving the employee some positive pay if the trade is abandoned. In fact, much of the critique of banker contracts in the wake of the financial crisis is that the high level of bonuses relative to fixed pay induce excessive risk taking. However, our analysis makes clear that fixed-pay contracts would dampen search effort, since they make lazy employees better off. Hence, an optimal contract would trade off the agency cost of excessive risk taking (pursuing unpromising risky trades) against the agency cost of underprovision of effort. Somewhat speculatively, it seems likely that when effort provision is very important, as in our high stakes tasks, a higher level of excess risk taking is tolerated in the optimal contract. Furthermore, building on our results on procyclical moral hazard, it also seems plausible that excess risk taking will be procyclical; because the effort problem is worse in good times, a firm might be willing to accept more excess risk taking to alleviate the effort problem. We leave a full development of a richer model of this sort for future research.

One obviously counterfactual prediction of our analysis is that young employees who are overpaid and fail receive literally nothing after failure. This is a direct consequence of our assumption of risk-neutrality. If instead employees 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 employees very heavily for failure, by making consumption after failure very low (but still strictly positive), thereby eliminating equilibrium overpay since all employees' utilities would be equalized.[37] However, this conjecture is not correct in our model. One-sided commitment prevents an employee'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 employees, while still preserving the central prediction of equilibrium overpay. We plan to explore this avenue in future research.

References

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

[36]At a formal level, this is closely related to Gromb and Martimort's (2007) analysis of experts.
[37]This is related to a point made in Carmichael (1985).

Acemoglu, Daron, and Andrew F. Newman, 2002, The labor market and corporate structure, *European Economic Review*, 46, 1733 – 1756.

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.

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.

Bell, Brian, and John Van Reenen, 2013, Bankers and their Bonuses, *Working Paper.*

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.

Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet, and Stephane Villeneuve, 2010, Large Risks, Limited Liability, and Dynamic Moral Hazard, *Econometrica* 78, 73–118.

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

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.

Custódio, Cláudia, Miguel A. Ferreira and Pedro Matos, forthcoming. Generalists versus Specialists: Lifetime Work Experience and CEO Pay. *Journal of Financial Economics.*

DeChesare, Brian, 2012, From Analyst Monkey to King of the Jungle MD: The Investment Banking Hierarchy, `http://www.mergersandinquisitions.com/investment-banking-hierarchy/`

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, 2012, Dynamic Agency and the q Theory of Investment, *Journal of Finance*, 67:6, 2295-2340.

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov, 2012, Dynamic CEO Compensation, *Journal of Finance* 67: 1603–1647.

Fishman, Steve, 2004, Get Richest Quickest, *New York Magazine*, November 22.

Gompers, Paul, and Josh Lerner, 2000, Money chasing deals?: The impact of fund inflows on the valuation of private equity investments, *Journal of Financial Economics* 55: 281–325.

Gromb, Denis and David Martimort, 2007, Collusion and the organization of delegated expertise, *Journal of Economic Theory*, 137:1, 271–299.

Holmström, Bengt, 1979, Moral Hazard and Observability, *Bell Journal of Economics* 10:1, 74-91.

Holmström, Bengt, and Jean Tirole, 1997, Financial Intermediation, Loanable Funds, and the Real Sector, *Quarterly Journal of Economics*, 112:3, 663-691.

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.

Kaplan, Steven N., and Jeremy Stein, 1993, The evolution of buyout pricing and financial structure in the 1980s, *Quarterly Journal of Economics* 108, 313–358.

Kaplan, Steven N., and Joshua Rauh, 2010, Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes? *Review of Financial Studies* 23, 1004-1050.

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.

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

Makan, Ajay, 2012, Skin in the game is crucial, but how much? *Financial Times*, Novermber 18.

Manove, Michael, 1997, Job responsibility, pay and promotion, *Economic Journal*, 107, 85-103.

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.

Myerson, Roger B., 1982, Optimal Coordination Mechanisms in Generalized Principle-Agent Problems, *Journal of Mathematical Economics* 10, 67-81.

Ohlendorf, Susanne, and Patrick W. Schmitz, 2012, Repeated Moral Hazard and Contracts with Memory: The Case of Risk-neutrality, *International Economic Review*, 53:2, 433–452.

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

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

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.*

Rajan, Raghuram G., 2010, Fault Lines: How Hidden Fractures Still Threaten the World Economy, *Princeton.*

Rebitzer, James B. and Lowell J. Taylor, 2010, Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches To Agency and Labor Markets, in the *Handbook of Labor Economics* 4A, North Holland, edited by Orley Ashenfelter and David Card.

Rogerson, William P., 1985, Repeated Moral Hazard, *Econometrica* 53, 69-76.

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

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.

Appendix

Proof of Lemma 1: The effort level $\underline{p}(g_H)$ solves $g_H = \gamma'(p) + p\gamma''(p)$. By Assumption 1 , the expression $\gamma'(p) + p\gamma''(p)$ is strictly increasing in p, and ranges from from 0 to ∞ as p ranges from

0 to maximal effort $\bar{p}$. Hence $\underline{p}(g_H)$ is well-defined, is strictly increasing in g_H, and $\underline{p}(g_H) \to \bar{p}$ as $g_H \to \infty$. **QED**

Verification that the one-period economy features overpay in the example in SectionV: Suppose that the cost function γ has the property $\gamma'(p) > p^2\gamma''(p)$ for some p;[38] and let k_H be such that the solution, $\check{p}$ say, to $p^2\gamma''(p) = k_H$ satisfies $\gamma'(\check{p}) > k_H$; and let $g_L \in [k_H, \gamma'(\check{p}))$.

On the one hand, for the one-period economy, straightforward combination of the zero-profit and profit-maximization conditions implies that in any overpaying equilibrium, the overpaid employees exert effort $\check{p}$. Since $\gamma'(\check{p}) > g_L$, these employees indeed receive utility strictly above $u(p_L^*)$. So provided $\zeta_H(\check{p}, k_H)$ is sufficiently small, the one-period economy indeed features overpay.

Proof of Proposition 2: (The proof below uses Lemma 4, which is proved in the main text using Lemmas 2 and 3; these latter two results are proved below.)

We establish that the following four statements hold for k_H large enough. To understand the first two statements, recall that, at $g_H = G^{HH}$, the contract C^{HH} defined in the main text maximizes profits subject to the "up-or-out" contract restriction embedded in (6) that $(i_S, i_F, w_{FS}, w_{FF}) = (H, L, g_L, 0)$, and to the restriction that the initial task assignment is $i = H$; and these maximized profits are zero. We establish:

(1) The contract C^{HH} is unique, i.e., when $g_H = G^{HH}$ it is the unique maximizer of (6) subject to (7) and $i = H$. (2) The contract C^{HH} maximizes profits even without the restriction to "up-or-out" contracts with initial task assignment $i = H$; and moreover, satisfies the participation constraint. That is, C^{HH} solves

$$\max_C \Pi\left(C; G^{HH}\right) \text{ s.t. (4) and } U(C) \geq 2u\left(p_L^*\right). \tag{A-1}$$

(3) The equilibrium is unique. (4) For overpaid employees, the success probability is greater after first-period success than when young.

Step 1: At $g_H = G^{HH}$, C^{HH} is the unique maximizer of (6) subject to (7) and $i = H$: The solution to this problem has either (a) $p_S \in [\underline{p}(g_H), p^*(g_H))$ and p determined by (7) at equality, with profits $p[g_H + S_H(p_S; g_H) - u(p_S)] - k_H$; or else (b) $p_S = p^*(g_H)$ and profits $p[g_H - \gamma'(p) + S_H(p^*(g_H); g_H) - u(p_L^*)] - k_H$. For case (a), the derivative of profits with re-

[38]For example, $\gamma(p) = \frac{p^2}{1-p}$ has this property; and satisfies Assumption 1, along with the other conditions required of γ.

spect to p_S equals (using $\frac{\partial p}{\partial p_S} = \frac{u'(p_S)}{\gamma''(p)}$)

$$\frac{u'(p_S)}{\gamma''(p)}\left[g_H + S_H(p_S; g_H) - u(p_S)\right] + p\left(S'_H(p_S; g_H) - u'(p_S)\right),$$

which has the same sign as

$$\frac{1}{p\gamma''(p)}\left[g_H + S_H(p_S; g_H) - u(p_S)\right] + \frac{S'_H(p_S; g_H)}{u'(p_S)} - 1. \tag{A-2}$$

For case (b), the derivative of profits with respect to p is

$$g_H - \gamma'(p) + S_H(p^*(g_H); g_H) - u(p_L^*) - p\gamma''(p),$$

which by (7) is less than

$$g_H + S_H(p^*(g_H); g_H) - u(p^*(g_H)) - p\gamma''(p). \tag{A-3}$$

Suppose first that there is a solution with $p_S \in [\underline{p}(g_H), p^*(g_H))$. So expression (A-2) equals 0. It then follows that expression (A-2) is strictly negative for all higher p_S, and expression (A-3) is strictly negative: This follows from the facts that S_H is strictly concave; u is strictly convex (Assumption 1); p is increasing in p_S, and $p\gamma''(p)$ is strictly increasing in p (Assumption 1 again); and $S_H(p_S; g_H) - u(p_S)$ is strictly decreasing over $p_S \geq \underline{p}(g_H)$. Hence the solution is unique in this case.

Second, suppose there is no solution with $p_S \in [\underline{p}(g_H), p^*(g_H))$. Hence all solutions are in case (b), and from (A-3), it is immediate that profits are strictly concave in p, establishing uniqueness.

Step 2: C^{HH} solves problem (A-1): The main step is to establish:

Claim: For k_H large enough that Lemma 4 holds, at return $g_H = G^{HH}$ the contract C^{HH} is the unique contract that gives non-negative profits while satisfying (4) and the contract restriction that $(i, i_S, i_F) \neq (L, L, L)$ (i.e., the employee is assigned to task H at some node).

The Claim establishes Step 2, as follows. From the Claim, no contract satisfying (4) and the contract restriction that $(i, i_S, i_F) \neq (L, L, L)$ gives weakly higher profits than C^{HH}. Maximal profits from a contract with $(i, i_S, i_F) = (L, L, L)$ and $U(C) \geq 2u(p_L^*)$ are 0. Finally, for k_H large enough, the main text establishes $U\left(C^{HH}\right) > 2u(p_L^*)$. Hence C^{HH} solves problem (A-1).

Proof of Claim: Suppose to the contrary that there exists a contract $C \neq C^{HH}$, with $(i, i_S, i_F) \neq (L, L, L)$, and satisfying (4), that gives non-negative profits. This implies that there must exist a contract $\tilde{C}$ with the "out" feature $(i_F, w_{FS}, w_{FF}) = (L, g_L, 0)$ that satisfies the same criteria, as follows. If C already has this property, we are done. Otherwise, let $\tilde{C}$ be the contract obtained by adding the "out" feature $(i_F, w_{FS}, w_{FF}) = (L, g_L, 0)$ to C, while leaving all other components of C unchanged. Because $G^{HH} < \underline{g}_H$ and C satisfies (4), the contract C must generate strictly negative profits after failure, i.e., $p(\mathbf{w}_F) g_{i_F} - E(w|F) - k_{i_F} < 0$, and hence must generate strictly positive profits after success, i.e., $g_i + p(\mathbf{w}_S) g_{i_S} - E(w|S) - k_{i_S} > 0$. Hence the new contract $\tilde{C}$ strictly raises profits after failure, and because C satisfies (4), it also weakly increases the employee's first-period effort p. Consequently, total expected two-period profits from the perturbed contract $\tilde{C}$ are strictly positive.

The main text establishes that no contract satisfying (4) and $(i, i_S, i_F) = (H, L, L)$ gives non-negative profits. Hence $\tilde{C}$ has the up-or-out feature $(i, i_F, w_{FS}, w_{FF}) = (H, L, g_L, 0)$. By Lemma 4, $G^{HH} < G^{HL}$, and so no contract satisfying (4) with $(i, i_S, i_F, w_{FS}, w_{FF}) = (L, H, L, g_L, 0)$ gives non-negative profits. Hence the only possibility is that the contract C has $(i, i_S) = (H, H)$. By the argument above, if $(i_F, w_{FS}, w_{FF}) \neq (L, g_L, 0)$ in contract C, then there exists a perturbed contract $\tilde{C}$ with $(i, i_S, i_F, w_{FS}, w_{FF}) = (H, H, L, g_L, 0)$ that gives strictly positive profits, contradicting the definition of G^{HH}. Hence contract C features $(i, i_S, i_F, w_{FS}, w_{FF}) = (H, H, L, g_L, 0)$. But by Step 1, C^{HH} is the unique such contract that gives non-negative profits, giving a contradiction and completing the proof both of the claim and Step 2.

Step 3: Equilibrium uniqueness: First, we show that for k_H large there is no equilibrium with $g_H < G^{HH}$. Step 2 above establishes that for k_H large and $g_H < G^{HH}$ there is no contract that assigns task H at any node, solves Problem (A-1), and gives non-negative profits. Consequently , it is impossible to satisfy the required equilibrium condition that the return g_H is consistent with aggregate task H activity.

Second, we show that for k_H large there is no equilibrium with $g_H > G^{HH}$. Suppose on the contrary that there exists such an equilibrium. The contract C^{HH} delivers strictly positive profits. Hence the reservation utility $\underline{U}$ must exceed $U(C^{HH})$. Moreover, the contract C^{HH} must induce effort $p(\boldsymbol{w}_S) \geq \underline{p}(G^{HH})$. As $k_H \to \infty$, $G^{HH} \to \infty$ also, so $\underline{p}(G^{HH})$ approaches the maximal feasible effort $\bar{p}$, and the bonus payment $w_{SS} - w_{SF} = \gamma'(\underline{p}(G^{HH}))$ that is needed to induce this effort level grows arbitrarily large. Hence $U(C^{HH}) \to \infty$ as $k_H \to \infty$. The utility delivered by a contract that assigns the employee to task L with certainty is bounded above by $2u(p_L^*)$. Hence for all k_H large enough, all employees must receive a contract that assigns them to task H in at least one node. It is straightforward to show that as $k_H \to \infty$ and hence $G^{HH} \to \infty$ and $\underline{p}(G^{HH}) \to \infty$, the expected task H output of any contract that assigns the employee to task H

in some node remains bounded away from 0: this is true if $(i_S, i_F) = (L, H)$; or (H, L); or (H, H); or (L, L), with $i = H$. But then for k_H sufficiently large it is impossible to satisfy the required equilibrium condition that the return g_H is consistent with aggregate task H activity, since we have shown that task H output y_H remains bounded away from 0, so that $\zeta_H(y_H, k_H)$ remains bounded above.

Consequently, for k_H sufficiently large the only possible equilibrium return is G^{HH}. By the Claim in Step 2 above, C^{HH} and C^{LL} are the only possible equilibrium contracts. The number of employees receiving each contract is uniquely determined by the condition that G^{HH} is consistent with aggregate task H activity. Hence the equilibrium is unique.

Step 4: For overpaid employees, the success probability after first-period success exceeds the success probability when young, i.e., $p_S > p$: The effort levels p and p_S induced by the C^{HH} contract solve Problem (6). As noted in the main text, the result is immediate if constraint (7) binds. Here, we show that the result also holds if (7) is slack. Differentiation of (6) with respect to first-period effort p, and making use of $p\gamma''(p) = u'(p)$, profit-maximization implies that p satisfies

$$g_H - \gamma'(p) + S_H(p_S; g_H) - u(p_L^*) - u'(p) = 0. \tag{A-4}$$

Note that, since (7) is slack, $p_S = p^*(g_H)$. Suppose that, contrary to the claimed result, $p \geq p_S$. Hence $g_H - \gamma'(p) \leq 0$. To complete the proof, we obtain a contradiction to (A-4) by establishing

$$S_H(p_S; g_H) < u'(p). \tag{A-5}$$

To establish (A-5), note that $p \geq p_S$ implies

$$u'(p) \geq u'(p_S) > p_S u'(p_S) \geq u(p_S),$$

where the third inequality follows from the convexity of u (by Assumption 1) and $u(0) = 0$. Moreover, $p_S = p^*(g_H)$ implies $u(p_S) - S_H(p_S; g_H) = k_H$. This establishes (A-5) and completes the proof. **QED**

Proof of Lemma 2: At $g_H = G^{LH}$, the solution to (6) subject to (7) (but not to (8)), and to the additional constraint that the initial task assignment is $i = L$, is given by

$$\max_{p_S} g_L - (u(p_S) - u(p_L^*)) + \left(p_S G^{LH} - \gamma(p_S) - k_H\right) - u(p_L^*) = 0.$$

After cancelling terms, this gives the result. **QED**

Proof of Lemma 3: From Lemma 2, the zero-profit contract associated with G^{LH} sets $p_S = \underline{p}\left(G^{LH}\right)$ and p to satisfy (7) with equality. From (6),

$$g_L - \gamma'(p) + S_H\left(\underline{p}\left(G^{LH}\right); G^{LH}\right) - u\left(p_L^*\right) = 0. \tag{A-6}$$

As $k_H \to \infty$, certainly $G^{LH} \to \infty$, and hence $\underline{p}\left(G^{LH}\right) \to \bar{p}$ and $\gamma'\left(\underline{p}\left(G^{LH}\right)\right) \to \infty$. Since for any $\tilde{p}$ and $\check{p}$, $u\left(\check{p}\right) \geq \tilde{p}\gamma'\left(\check{p}\right) - \gamma\left(\tilde{p}\right)$, it follows that $u\left(\underline{p}\left(G^{LH}\right)\right) \to \infty$ as $k_H \to \infty$, and hence from (7), $\gamma'(p) \to \infty$ also. By (A-6), this implies the result. **QED**

Proof of Proposition 3: Most of the details are in the main text. Here, we verify that returns and hence contracts are state-independent; and that λ_t converges.

To show that returns are state-independent, we need to show that it is possible to vary the number of employees hired by a sufficient amount to fully absorb the aggregate shock. Formally, this amounts to showing that λ_t remains between 0 (one cannot hire a negative number of new employees), and 1 (the total population of young employees). Define $\underline{\lambda} \equiv \frac{y_H^B - p_S y_H^G}{p\left(1-p_S^2\right)}$ and $\bar{\lambda} \equiv \frac{y_H^G - p_S y_H^B}{p\left(1-p_S^2\right)}$. It is straightforward to establish that λ_t remains in the interval $\left[\underline{\lambda}, \bar{\lambda}\right]$.[39] Consider what happens as the shock size shrinks, i.e., ζ_H^G and ζ_H^B approach some common $\bar{\zeta}_H$. Let $\bar{y}_H$ be the output level associated with $\bar{\zeta}_H$ and the payoff g_H, i.e., $\bar{\zeta}_H\left(\bar{y}_H, k_H\right) = \frac{g_H}{k_H}$. Then y_H^B and y_H^G both approach $\bar{y}_H$ and $\underline{\lambda}$ and $\bar{\lambda}$ both approach $\frac{\bar{y}_H}{p\left(1+p_S\right)}$. Hence provided the shocks are sufficiently small, there is indeed enough flexibility to absorb the shocks via hiring decisions, verifying the conjecture that returns are independent of the state.

To confirm that λ_t converges, simply note that iteration of the hiring equation (11) gives

$$\lambda_t = \left(-p_S\right)^t \lambda_0 + \frac{1}{p}\sum_{s=0}^{t-1}\left(-p_S\right)^s y_H^{\omega_{t-s}}, \tag{A-7}$$

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 employees assigned to

[39] 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

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

task H converges to λ^{ω}. **QED**

Proof of Proposition 4: Let g_H^{ω} denote the equilibrium returns and $\left(p^{\omega}, p_S^{\omega\psi}\right)$ denote the equilibrium contracts (here we are using the simple representation from problem (14)).

Step 1: We show

$$\frac{k_H}{p^{\omega} u'(p^{\omega})} - 1 + \frac{S_H'\left(p_S^{\omega\psi}; g_H^{\psi}\right)}{u'\left(p_S^{\omega\psi}\right)} = 0. \tag{A-8}$$

Suppose first that $p_S^{\omega\psi} \neq p^*\left(g_H^{\psi}\right)$ for some $\psi = G, B$. Then (15) must hold with equality in state ω. Differentiating the profit expression (14) with respect to $p_S^{\omega\psi}$, and then substituting in the zero-profit and profit-maximization conditions that must hold at equilibrium values, gives

$$\frac{\partial p^{\omega}}{\partial p_S^{\omega\psi}} \left(\frac{k_H}{p^{\omega}} - p^{\omega}\gamma''(p^{\omega})\right) + p^{\omega}\mu^{\omega\psi} S_H'\left(p_S^{\omega\psi}; g_H^{\psi}\right) = 0.$$

From (15) we have $\gamma''(p^{\omega}) \frac{\partial p^{\omega}}{\partial p_S^{\omega\psi}} = \mu^{\omega\psi} u'\left(p_S^{\omega\psi}\right)$. Hence

$$\frac{\mu^{\omega\psi} u'\left(p_S^{\omega\psi}\right)}{\gamma''(p^{\omega})} \left(\frac{k_H}{p^{\omega}} - p^{\omega}\gamma''(p^{\omega})\right) + p^{\omega}\mu^{\omega\psi} S_H'\left(p_S^{\omega\psi}; g_H^{\psi}\right) = 0.$$

Rearranging, and using $u'(p) = p\gamma''(p)$, delivers (A-8).

If instead $p_S^{\omega\psi} = p^*\left(g_H^{\psi}\right)$ for both $\psi = G, B$, then the Lagrange multiplier on constraint (15) is 0, and differentiation of (14) with respect to p^{ω} combined with the zero-profit and profit-maximization conditions yields $\frac{k_H}{p^{\omega}} - p^{\omega}\gamma''(p^{\omega}) = 0$, which rearranges to (A-8) since $S_H'\left(p^*\left(g_H^{\psi}\right); g_H^{\psi}\right) = 0$.

Step 2: $g_H^G > g_H^B$, i.e., Part (C) holds.

Suppose to the contrary that $g_H^G \leq g_H^B$. We first show that Step 1 implies that $p_S^{\omega G} \leq p_S^{\omega B}$ and $S_H\left(p_S^{\omega G}; g_H^G\right) \leq S_H\left(p_S^{\omega B}; g_H^B\right)$. From (A-8),

$$\frac{S_H'\left(p_S^{\omega G}; g_H^G\right)}{u'\left(p_S^{\omega G}\right)} = \frac{S_H'\left(p_S^{\omega B}; g_H^B\right)}{u'\left(p_S^{\omega B}\right)}.$$

From Assumption 1, u is convex, so $\frac{S_H'(\tilde{p}; g_H)}{u'(\tilde{p})}$ is decreasing in $\tilde{p}$ for all $\tilde{p}$ such that $\gamma'(\tilde{p}) \leq g_H$, which is the relevant range here. So $g_H^G \leq g_H^B$ implies $p_S^{\omega G} \leq p_S^{\omega B}$, which in turn implies $S_H\left(p_S^{\omega G}; g_H^G\right) \leq S_H\left(p_S^{\omega B}; g_H^B\right)$.

Next, consider the contract $\left(p^G, p_S^{GG}, p_S^{GB}\right)$, which delivers zero profits in state $\omega = G$. If this contract satisfies constraint (15) for $\omega = B$, then $-\bar{u}_L^G < -\bar{u}_L^B$, $g_H^G \leq g_H^B$, $S_H\left(p_S^{GG}; g_H^G\right) \leq S_H\left(p_S^{GB}; g_H^B\right)$, and $\mu^{GB} \leq \mu^{BB}$ together imply that the contract delivers strictly positive profits when used in $\omega = B$, contradicting the equilibrium condition. If instead the contract violates (15) for $\omega = B$, then consider the profits from using the contract $\left(\tilde{p}, p_S^{GG}, p_S^{GB}\right)$ in $\omega = B$, where $\tilde{p}$ is chosen to set (15) to equality. Note that $\tilde{p} > p^G$. The profits from this contract are

$$\tilde{p}\left(g_H^B + \sum_{\psi=G,B} \mu^{B\psi}\left(S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right)\right) - k_H.$$

Using $g_H^G \leq g_H^B$, $p_S^{GG} \leq p_S^{GB}$ and $\mu^{GB} \leq \mu^{BB}$, along with the fact that $p_S^{\omega\psi} \geq \underline{p}\left(g_H^{\psi}\right)$ so that $S_H\left(\cdot; g_H^{\psi}\right) - u\left(\cdot\right)$ is decreasing, this expression is greater than

$$\tilde{p}\left(g_H^G + \sum_{\psi=G,B} \mu^{G\psi}\left(S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right)\right) - k_H,$$

which by (15) exceeds

$$\tilde{p}\left(g_H^G - \gamma'\left(p^G\right) + \sum_{\psi=G,B} \mu^{G\psi} S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - \bar{u}_L^G\right) - k_H,$$

which by zero-profits equals $\tilde{p}\frac{k_H}{p^G} - k_H > 0$, again contradicting the equilibrium condition.

Step 3: $p^G \leq p^B$, with the inequality strict unless $p_S^{\omega\psi} = p^*\left(g_H^{\psi}\right)$ for all ω, ψ.

If $p_S^{G\psi} = p^*\left(g_H^{\psi}\right)$ for $\psi = G, B$, the implication $p^G \leq p^B$ is immediate from (A-8), and is strict provided $p_S^{B\psi} \neq p^*\left(g_H^{\psi}\right)$ for $\psi = G, B$. The remainder of the proof deals with the case in which $p_S^{G\psi} \neq p^*\left(g_H^{\psi}\right)$ for $\psi = G, B$. Note that this implies that (15) holds with equality for $\omega = G$.

Suppose to the contrary that $p^G \geq p^B$. So by Assumption 1, $p^B u'\left(p^B\right) \leq p^G u'\left(p^G\right)$, and (A-8) implies

$$\frac{S_H'\left(p_S^{BB}; g_H^B\right)}{u'\left(p_S^{BB}\right)} = \frac{S_H'\left(p_S^{BG}; g_H^G\right)}{u'\left(p_S^{BG}\right)} \leq \frac{S_H'\left(p_S^{GB}; g_H^B\right)}{u'\left(p_S^{GB}\right)} = \frac{S_H'\left(p_S^{GG}; g_H^G\right)}{u'\left(p_S^{GG}\right)}.$$

The same argument as used in Step 2 implies that, for $\psi = B, G$,

$$p_S^{B\psi} \geq p_S^{G\psi}; \tag{A-9}$$

and given that $g_H^G \geq g_H^B$, also implies that for $\omega = G, B$,

$$p_S^{\omega G} \geq p_S^{\omega B}, \tag{A-10}$$

which again using convexity of u implies

$$S_H'\left(p_S^{\omega G}; g_H^G\right) \geq S_H'\left(p_S^{\omega B}; g_H^B\right). \tag{A-11}$$

Separately, the zero-profit condition and $p^G \geq p^B$ together imply

$$g_H^G - \gamma'\left(p^G\right) + \sum_{\psi=G,B} \mu^{G\psi} S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - \bar{u}_L^G \leq g_H^B - \gamma'\left(p^B\right) + \sum_{\psi=G,B} \mu^{B\psi} S_H\left(p_S^{B\psi}; g_H^{\psi}\right) - \bar{u}_L^B.$$

Since $g_H^G > g_H^B$, it follows that

$$-\gamma'\left(p^G\right) + \sum_{\psi=G,B} \mu^{G\psi} S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - \bar{u}_L^G < -\gamma'\left(p^B\right) + \sum_{\psi=G,B} \mu^{B\psi} S_H\left(p_S^{B\psi}; g_H^{\psi}\right) - \bar{u}_L^B.$$

Substituting in (15), and using the fact that it holds at equality for $\omega = G$,

$$\sum_{\psi=G,B} \mu^{G\psi}\left(S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right) < \sum_{\psi=G,B} \mu^{B\psi}\left(S_H\left(p_S^{B\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right). \tag{A-12}$$

The LHS of this inequality can be written as

$$\begin{aligned}&\sum_{\psi=G,B} \mu^{B\psi}\left(S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right)\\ &+\left(\mu^{GG} - \mu^{BG}\right)\left(\left(S_H\left(p_S^{GG}; g_H^G\right) - u\left(p_S^{GG}\right)\right) - \left(S_H\left(p_S^{GB}; g_H^B\right) - u\left(p_S^{GB}\right)\right)\right).\end{aligned}$$

Note that, for any $\tilde{p}$ and g_H, $S_H\left(\tilde{p}; g_H\right) - u\left(\tilde{p}\right) = \tilde{p}S_H'\left(\tilde{p}\right) - k_H$. Hence (A-10) and (A-11) imply

$$S_H\left(p_S^{GG}; g_H^G\right) - u\left(p_S^{GG}\right) \geq S_H\left(p_S^{GB}; g_H^B\right) - u\left(p_S^{GB}\right),$$

so that (A-12) implies that for at least one of $\psi = G, B$,

$$S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right) < S_H\left(p_S^{B\psi}; g_H^{\psi}\right) - u\left(p_S^{B\psi}\right).$$

Hence $p_S^{G\psi} > p_S^{B\psi}$ for at least one of $\psi = G, B$, contradicting (A-9) and completing the proof.

Step 4: Pay for luck, i.e., Part (B).

Given $g_H^G > g_H^B$, it follows by the same argument as used repeatedly above that (A-8) implies $p_S^{\omega G} > p_S^{\omega B}$. The pay for luck implication is then immediate from the second-period IC.

Step 5: Completing Part (A) by establishing higher bonuses.

Given $p^G \leq p^B$, it follows by the same argument as used repeatedly above that (A-8) implies $p_S^{G\psi} \geq p_S^{B\psi}$, with the inequality strict if $p^G < p^B$. So if $p^G < p^B$, the result is then immediate from second-period IC. If instead $p^G = p^B$, then given $\bar{u}_L^G > \bar{u}_L^B$, $p_S^{G\psi} \geq p_S^{B\psi}$, $p_S^{\omega G} > p_S^{\omega B}$ (from Step 4), the first-period IC (13) implies the result. **QED**

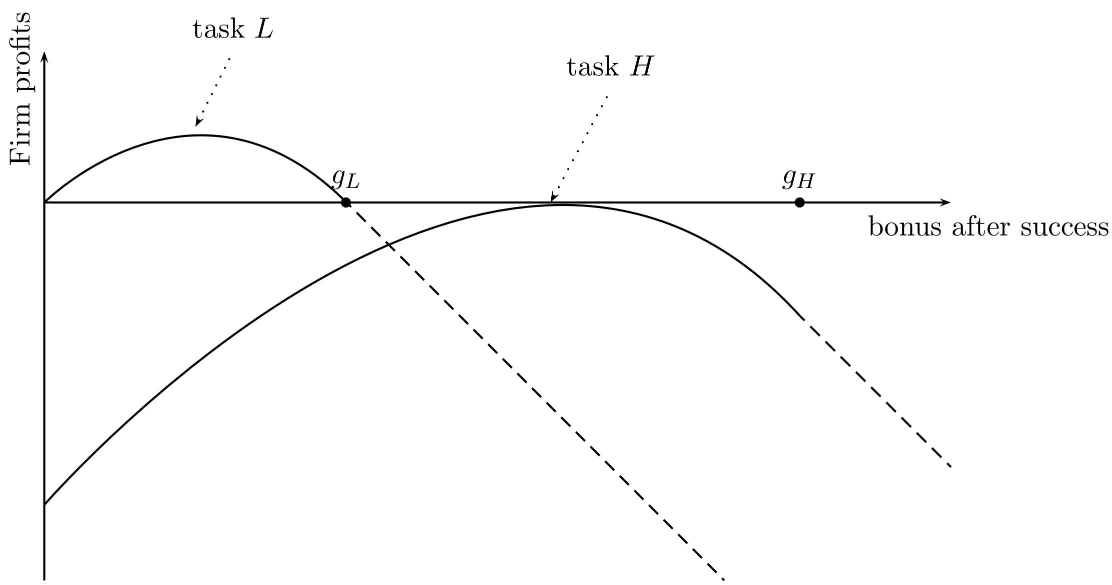

Figure 1: The graph displays firm profits from using a one-period contract to incentivize the employee in each of tasks L and H, as a function of the success bonus. The graph is drawn for $g_H = \underline{g}_H$, so that maximal profits in task H are exactly 0. For bonuses below g_i in task i, the employee receives nothing after failure. The dashed lines have slope -1, and reflect that the fact that once the bonus reaches g_i, profit-maximization is achieved by paying the worker after failure (while maintaining the bonus g_i).

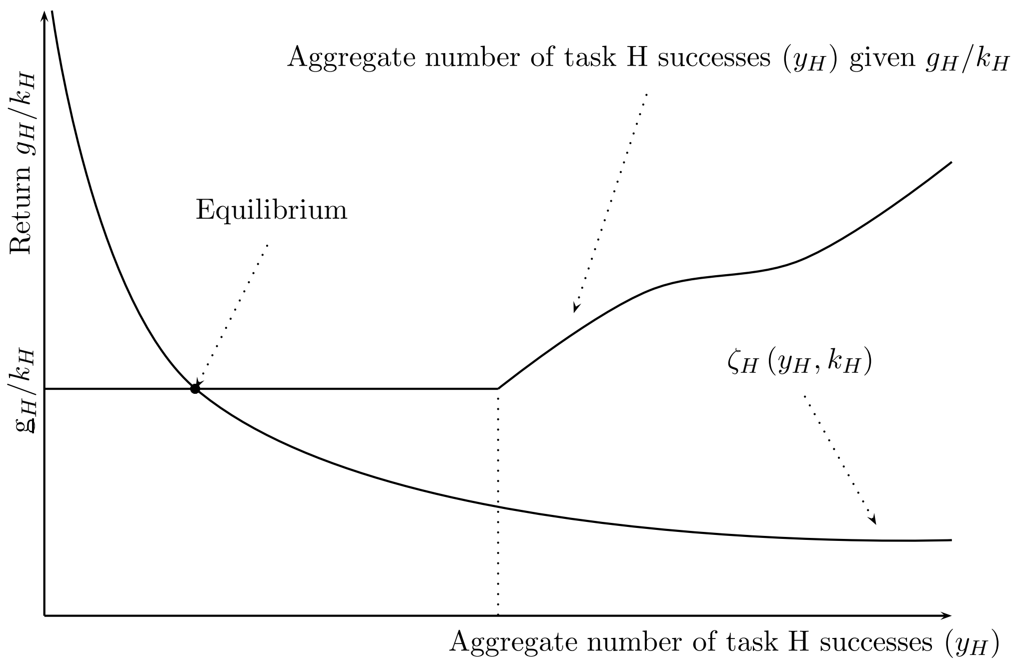

Figure 2: The graph shows an equilibrium with overpay. At $\underline{g}_H$, firms make exactly zero profits in task H using the profit-maximizing contract, and this contract delivers strictly higher employee utility than the task L contract (see Figure 1). The horizontal line corresponds to allocating different fractions of employees to these two contracts. The graph is drawn for the case in which if everyone is allocated to the H-contract, it is impossible to sustain the return $\underline{g}_H/k_H$. Consequently, in equilibrium a strict subset of employees receive each of the two contracts, and overpay exists.