Ulf Axelson and Philip Bond

# Wall Street occupations

# Wall Street Occupations

ULF AXELSON and PHILIP BOND*

**ABSTRACT**

Many finance jobs entail the risk of large losses, and hard-to-monitor effort. We analyze the equilibrium consequences of these features in a model with optimal dynamic contracting. We show that finance jobs feature high compensation, up-or-out promotion and long work hours, and are more attractive than other jobs. Moral hazard problems are exacerbated in booms, even though pay increases. Employees whose talent would be more valuable elsewhere can be lured into 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 London School of Economics and SIFR, and Philip Bond is with the University of Washington. A previous version of this paper circulated under the title "Investment Banking Careers." We thank Bruno Biais (the editor), two anonymous referees, Wouter Dessein, Paolo Fulghieri, Andrew Hertzberg, Camelia Kuhnen, Alan Morrison, Christian Opp, Paul Oyer, Uday Rajan, David Robinson, Ibola Schindele, and seminar audiences at the American Economic Association meetings, BI Oslo, Boston University, Carnegie Mellon University, University of Chicago, 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, NBER, Ohio State University, Oxford University, SIFR/SSE, SSE Riga, Temple, UT Austin, Stanford, and Toulouse IDEI for helpful comments and suggestions. Any errors are our own.

In the aftermath of the 2007 financial crisis, there has been much debate about the level and structure of pay in the financial sector. Why are financial sector employees paid so much, and why do they seem to be especially irresponsible in booms, even though their pay is higher than usual? In this paper, we build a parsimonious model that addresses these questions and also explains other distinguishing characteristics of financial sector careers—the extreme work hours, high reliance on bonuses, and up-or-out career structures. We further explain entry and exit patterns for financial sector employees.

There is no question that financial sector pay is extremely high. Bell and Van Reenen (2014) report a 2010 average compensation level of £1,905,000 for a group of 1,408 senior bankers in the UK, and Kaplan and Rauh (2010) show that a significant proportion of incomes from the very top of the U.S. income distribution stem from the financial sector. Moving beyond people at the very top of the income distribution, Oyer (2008) and Philippon and Reshef (2012) provide evidence that the financial sector offers more compensation than other industries, even after controlling for individual characteristics;[1] in particular, Oyer estimates that the lifetime pay premium enjoyed by an MBA graduate in the financial sector relative to other sectors (such as consulting) is \$1.5 to \$5M in present value terms.

One might conjecture that such generous pay is compensation for the notoriously tough and stressful work conditions in many financial sector jobs.[2] Michel (2011) reports that investment banking employees work up to 120 hours per week, while Bertrand, Goldin, and Katz (2010) report average hours worked in investment banking at 73.6 hours per week. In addition, pay is highly variable, job security is low, and firms rely heavily on up-or-out career structures.[3] It is not surprising that jobs in high-risk industries such as mining and oil production have high pay, and maybe financial sector jobs are similar. However, in finance the onerous work conditions are chosen by the employers, and are not unavoidable consequences of the production technology as in, for example, mining. Furthermore, Oyer's estimates of the lifetime pay premium when entering finance explicitly accounts for the risk of exiting the sector.[4] Consequently, instead of explaining high pay, the unusual work conditions pose puzzles of their own.

Oyer also provides evidence for a further characteristic of financial sector jobs, namely, that if an individual fails to land a job in the financial sector upon graduation solely because Wall Street is down, he is unlikely to enter the sector later in life when Wall Street has recovered. Given the attractive compensation of financial sector jobs, this pattern of entry is puzzling, since one might expect people to try very hard to gain entry into the sector even later in their careers. In other words, these labor markets appear to be *dynamically segregated*, in that employees can only enter the high-paying sector when young, so that the state of the economy upon graduation has life-long implications.

In this paper, we build a model that parsimoniously explains these characteristics of financial sector jobs. Our model also predicts that in the time series, high financial sector compensation (especially bonuses) is correlated with more reckless behavior by financial sector employees. This prediction is consistent with many accounts of the run-up to the financial crisis.

Our starting point is to observe that a combination of two characteristics seems especially important in finance. First, for many finance jobs the exact effort of an employee is hard to monitor, leading to a moral hazard problem. Second, one employee may oversee a large amount of capital, making it important that employees take sufficient care.[5] Together, these two features imply that moral hazard problems in financial sector jobs are likely to be large. If employees worked for only one period, this moral hazard problem would induce firms to pay financial sector employees large bonuses. As in the efficiency wage literature (see Shapiro and Stiglitz (1984)), in equilibrium these workers would then enjoy strictly higher utility than employees in other sectors (these statements are formally established in Section IV).

In reality, of course, employees work for many periods. Our key theoretical contribution is to analyze how firms structure dynamic contracts in response to the high rents that may have to be given out on high moral hazard tasks. In a dynamic setting, a natural conjecture is that large moral hazard problems in the financial sector would lead to an equilibrium career structure in which everyone is initially assigned to low moral hazard tasks. Employees who succeed on these tasks earn bonuses—possibly in the form of deferred compensation—which they can subsequently use as performance bonds in higher moral hazard tasks. In this conjectured equilibrium, all employees are on the same career track when they first enter the labor force, they have the same expected compensation, and they work the same amount. But these predictions are inconsistent with the empirical observation that financial sector jobs are notably different from other jobs, even at the start of an employee's career. Hence, explaining why this type of equilibrium does not arise is the key step in our analysis.

We show that when moral hazard problems are sufficiently large, it is more profitable for a firm to assign a young employee to the high moral hazard task immediately, rather than wait until he has succeeded on a low moral hazard task. The reason is that firms must pay large bonuses to experienced employees assigned to the high moral hazard task, and these bonuses incentivize effort earlier in employees' careers, generating serial correlation in effort. The most profitable way for a firm to take advantage of this serial correlation is to assign employees to the high moral hazard task when young. So these employees work very hard, even when young, and face up-or-out career structures. Moreover, their expected compensation is higher than that of other employees assigned to the low moral hazard task throughout their careers, that is, the labor market is dynamically segregated. These predictions are consistent with the observed structure of financial sector jobs.

The fact that some employees earn large rents relative to other equally talented individuals raises the question of why this gap is not eliminated through entry by new firms, expansion of existing firms, or the replacement of highly compensated employees by less compensated employees poached from elsewhere. As in the efficiency wage literature, in our model incentive problems prevent firms from replacing well-paid employees with cheaper ones. However, unlike in the efficiency wage literature, we do not rule out the possibility of new entry or expansion, both of which seem relevant in the financial sector, where one sees both active entry by new hedge funds and very large (and growing) banks. Instead, the equilibrating force in our model is that output prices (returns) in the financial sector fall to the point that further entry and/or expansion is not worthwhile. At this break-even return, the supply of financial sector "output" is infinitely elastic. This feature of our model is important, because it implies that prices are determined solely by the contracting problem and not by the interaction of demand and supply. This in turn implies that returns in finance are relatively insensitive to aggregate shocks. Instead, negative aggregate shocks are absorbed largely by reductions in hiring rather than by firing experienced employees,[6] which in combination with dynamic segregation is why aggregate conditions when a worker enters the labor market have lifelong effects (Oyer (2008)).

The equilibrium response of returns also explains why large bonuses and reckless behavior are observed together in our model. In our framework, both large bonuses and low effort are an equilibrium response to good economic conditions. A positive aggregate shock has the standard positive effect on the return to nonfinance activities. Because the nonfinance sector serves as an outside option for failed finance workers, it becomes harder to incentivize finance workers in good times. Furthermore, since there is no direct positive effect on the return to finance, it does not pay for firms to increase bonuses to the point where finance employees take the same care as they do in busts. The result is that, even though bonuses are increased somewhat, finance employees become more reckless. Put differently, in good economic conditions bankers are confident that they will "land on their feet."This prediction is consistent with countercyclical investment performance in both the buyout (Kaplan and Stein (1993)) and the venture capital markets (Gompers and Lerner (2000)). It is also consistent with claims of irresponsible behavior in financial sector booms.

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 such as investment banking and money management attract talented employees whose skills might be deployed in a socially more valuable way were they to become, for example, engineers or PhDs. In our model, this type of misallocation follows immediately from the fact that the rent earned by financial sector employees implies that financial firms can outbid other employers for employees even if their talent is wasted in finance. The second phenomenon, which we term "talent scorned," is the opposite—jobs that deliver large rents may 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 noted above, our paper is related to and builds on the efficiency wage literature. Relative to this literature, we show that dispersion of agents' utility can arise in equilibrium even when dynamic employment contracts are written optimally, thereby addressing the so-called "bonding critique" of the efficiency wage literature,[7] and that the dynamic contracts used in equilibrium match (at least qualitatively) many features of the employment conditions observed in the financial sector.

Several of the contract characteristics we derive for high moral hazard tasks have antecedents in the partial-equilibrium 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 (2005) and Biais et al. (2010). Relative to this literature, our analysis sheds light on the optimal sequencing of different tasks. In particular, we characterize conditions under which employees are best assigned to a high moral hazard task at the start of their careers. Separately, we also differ from this literature in that we embed the contracting problem in an equilibrium framework, which we use to characterize the distribution of contracts in the economy.

Although our analysis implies that financial sector compensation is negatively correlated with effort in the time series, it is worth reiterating that employment contracts are optimally chosen by shareholders in our paper. In this, our paper is very different from 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 Philippon (2010), Glode, Green, and Lowery (2012), Bolton, Santos, and Scheinkman (2014), or Biais, Rochet, and Woolley (2014). In particular, the last of these papers shares with us an analysis of rents arising from moral hazard problems, and demonstrates how these rents may grow during periods without a crisis (due to Bayesian updating about the true productivity of the sector). In Biais, Rochet, and Woolley (2014), employees differ in unobservable skill, with the consequence that, in equilibrium, employment in the financial sector is larger than in the first-best and the marginal manager shirks. Instead, we abstract from unobservable skill differences and instead focus on the dynamic implications of moral hazard problems, which are absent in Biais, Rochet, and Woolley (2014) because of their assumption that agents work just one period. In a contemporaneous paper, Myerson (2012) also explores the effect of moral hazard on financial sector pay.

Finally, in a very interesting and related recent paper, Tervio (2009) explains high income in a model that builds on talent discovery rather than incentive problems. In his setting, employee rents arise 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; instead dynamic segregation is an 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 are of central importance in the financial sector, and are correspondingly central to our analysis.

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 a dynamic setting. Section VI studies the effects of aggregate shocks. Section VII introduces observable talent differences. Section VIII addresses several remaining issues, including the possible application of our model 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 one of young employees enter the labor market, work for two periods, and then exit. Except for age, employees are identical—in particular, they have the same skill. (Section VII analyzes an extension in which skills differ across employees.) Employees are risk neutral, start out penniless, and have limited liability. Employees are employed by risk-neutral firms that maximize profits and have "deep pockets," so that limited liability constraints never bind for firms. All firms 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, where part (i) ensures that a firm's marginal cost of inducing effort is increasing in the effort level, and 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") $k_H > 0$. Our leading example is that task $H$ corresponds to asset management or trading acitivities, where an employee is responsible for finding good investment opportunities for the asset base $k_H$, as in hedge funds, private equity funds, or mutual funds. It could also represent trading on an investment bank'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. Treating the amount of capital per employee $k_H$ as an exogenous parameter is obviously a simplification; we discuss likely implications of endogenizing the scale in Section VIII. We think of $k_H$ as representing an upper bound on the amount of capital that can be handled by one employee, determined by either regulatory or technological forces. It is easy to show that the firm would always operate at this upper bound as long as the cost of effort and the return on a successful trade are independent of the amount of capital an employee oversees below this maximum amount. However, as we formalize below, we assume that 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. This assumption of aggregate decreasing returns to scale is consistent with empirical evidence from the mutual fund industry (Pastor, Stambaugh, and Taylor (2014)), the private equity industry (Kaplan and Stromberg (2009)), and the venture capital industry (Gompers and Lerner (2000)).

The amount $k_H$ that an employee could potentially lose is key in our model and our main results are for $k_H$ large, a situation we think arises naturally in the financial sector. For example, Kaplan and Rauh (2010) estimate that the average employee in U.S. hedge funds oversaw approximately \$100M of funds in 2004, and the average partner in U.S. private equity firms oversaw approximately \$430M.[10] To address concerns that these large quantities are limited to only a small segment of the financial sector, we next consider the following back-of-the envelope calculation. The Bureau of Labor Statistics measures total employment in the U.S. financial sector in 2012 to be approximately 8 million, while the Flow of Funds indicate that U.S. households hold, in aggregate, approximately \$55 trillion in financial assets. Hence, in some sense, the average financial sector employee is responsible for approximately \$7M of financial assets. Although it is not necessarily the case that all this capital is lost when the employee makes a mistake, as we assume in the model, the potential for value destruction per employee is still very large.

Moving beyond our trading interpretation, there are many other high-stakes tasks in finance, such as M&A advising or conducting an IPO, that affect the entire value of large companies and fit within our framework. Notwithstanding, for the most part we use the language of the asset management/trading interpretation, largely because in this case $k_H$ is related to the empirically observable quantity of funds under management. Likewise, we often refer to $k_H$ as capital.

In contrast to task $H$, task $L$ is a low-stakes task, in that it requires few resources (beyond the employee's labor); for simplicity, we assume that it requires no resources, $k_L = 0$. Task $L$ can be interpreted as a nonfinancial 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, where capital $k_L$ generates gross firm payoffs of $g_L + k_L$ and $k_L$ after success and failure, respectively.)

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.

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, for example, the survey of Rebitzer and Taylor (2010)) suggest 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, for example, 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 thus to focus our analysis we make the simplifying assumption that the task $L$ payoff is constant, that is, $g_L > 0$ is a parameter of the model.[12] For task $H$, we denote by $y_H$ the economy-wide "supply" of task $H$, that is, the expected number of task $H$ successes in the economy.[13] The equilibrium return $g_H/k_H$ is then determined by $\zeta_H(y_H, k_H)$, where $\zeta_H$ is a strictly decreasing function of both arguments; in this sense, $\zeta_H$ plays the role of the (inverse) demand function. We also impose the following assumption.

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) of Assumption 2 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 zero, then the success return falls below one 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 one.

*Remark 1:* 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 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 nontrading areas of the financial sector, such as M&A advising, and nonfinancial occupations, such as train drivers or 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 $y_H$. In this case 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 $\lim_{y_H \to 0} D(y_H) = \infty$.

## II. Contracts and Equilibrium

*A. Contracts*

We impose minimal contracting 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] This assumption rules out contracts in which an employee is punished with zero lifetime pay after failure in the first period, because the employee can always find work on the $L$ task with another firm in the second period.[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 between 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 $C = (i, i_S, i_F, w_{SS}, w_{SF}, w_{FS}, w_{FF})$ 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" $(i_X, \boldsymbol{w}_X)$ where $\boldsymbol{w}_X = (w_{XS}, w_{XF})$ specifies payments contingent on success or failure in the second period. Given $\boldsymbol{w}_X$, 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

$$\gamma'(p(\boldsymbol{w}_X)) = w_{XS} - w_{XF}, \tag{1}$$

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 $E(w|X)$, which 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

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

First-period employee effort $p$ is therefore determined by the IC constraint

$$\gamma'(p) = \left[E(w|S) - \gamma(p(\boldsymbol{w}_S))\right] - \left[E(w|F) - \gamma(p(\boldsymbol{w}_F))\right]. \tag{2}$$

The IC constraint (2) 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 $\Pi$ per worker employed on contract $C$ are given by

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

where $p$ satisfies the first-period IC constraint (2).

*B. Equilibrium*

A (stationary) equilibrium consists of a payoff $g_H$ and a collection of contracts $\mathcal{C}$, together with a probability distribution $\lambda$ with full support on $\mathcal{C}$, such that the following three conditions hold:

1. *Profit maximization:* Each contract in $\mathcal{C}$ maximizes firm profits per employee subject to the "participation" constraint that utility is at least $\min_{C \in \mathcal{C}} U(C)$, and one-sided commitment constraints whereby each contract in $\mathcal{C}$ solves

$$\max_{\tilde{C}} \Pi\left(\tilde{C}; g_H\right) \text{ subject to } U\left(\tilde{C}\right) \geq \min_{C \in \mathcal{C}} U(C)$$

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

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

and

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

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

3. *Return consistent with aggregate task $H$ activity:* The return $\frac{g_H}{k_H}$ is consistent with the equilibrium quantity of $H$ successes. Formally, let $y_H(C)$ be the expected number of task $H$ successes generated by contract $C$.[16] The aggregate number of task $H$ successes is

$$y_H = \sum_{C \in \mathcal{C}} \lambda(C) y_H(C).$$

The return consistency condition is then

$$\zeta_H(y_H, k_H) = \frac{g_H}{k_H}.$$

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

An important feature of our model is that (at least for some parameter values) it gives rise to equilibria in which different employees receive different contracts that deliver different utilities, even though all employees are ex ante identical. Formally, we say that an employee is *overpaid* if he receives a contract $\overline{C} \in \mathcal{C}$ such that $U(\overline{C}) > \min_{C \in \mathcal{C}} U(C)$, that is, his expected lifetime utility strictly exceeds that of otherwise identical employees. Likewise, we say that an equilibrium is *overpaying* if it features overpaid employees.

For use below, we write $\underline{U} \equiv \min_{C \in \mathcal{C}} U(C)$. With a slight abuse of language we often refer to this quantity as an employee's reservation utility.

*Remark 2:* An equivalent way to state the profit-maximization and zero-profit conditions above is as follows: given $g_H$ and a contract $C \in \mathcal{C}$, there is no contract $\tilde{C}$ that gives the employee strictly more utility than $C$ and the firm strictly positive profits. This alternative formulation makes clearer that the participation constraint is driven by what an employee could get in alternative employment, but has the drawback that it cannot be expressed as a straightforward maximization problem.

*Remark 3:* Since our model has constant returns to scale at the firm level, but decreasing returns at the aggregate level, the equilibrium conditions pin down the the aggregate number of employees with each contract $C \in \mathcal{C}$, while firm size is indeterminate.

## 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. Throughout the analysis, let $S_i(p; g_i)$ denote the one-period surplus from effort $p$ on task $i$, given return $g_i$:

$$S_i(p; g_i) \equiv pg_i - \gamma(p) - k_i. \tag{3}$$

When effort is observable, profit maximization implies that equilibrium effort maximizes surplus by equating the marginal cost of effort to $g_i$. We denote this effort level by $p^*(g_i)$, that is,

$$\gamma'(p^*(g_i)) = g_i.$$

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, that is, $g_H$ satisfies $S_H(p^*(g_H); g_H) = S_L(p_L^*; g_L)$.[17] The fraction $\lambda$ of the population of measure two 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 lifetime 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. This helps us illustrate in a transparent way several features that are important in the dynamic model—in particular, how the equilibrium features overpay on task $H$ when the scale $k_H$ becomes sufficiently large, and the determinants of the equilibrium return $g_H$ in an overpaying equilibrium. By its nature, however, the one-period model cannot say anything about either the sequencing of task assignments or career structures, which we analyze in the dynamic version of the model. Moreover, the implications of aggregate shocks are very different in the one-period model relative to the full model.

For this section only, assume that there is only one period and (for comparability with the two-period economy) a measure two 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 $w_F = 0$. 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 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 constraint becomes $\gamma' (p) = w_S$ : the worker chooses effort $p$ to equate his marginal cost of effort, $\gamma' (p)$, to the bonus, $w_S$.

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. We write the utility associated with this contract as $u_L$,

$$u_L \equiv \max_{\tilde{p}} \tilde{p} g_L - \gamma (\tilde{p}) = p_L^* g_L - \gamma (p_L^*) = S_L (p_L^*; g_L) . \tag{4}$$

Moreover, the endogenous reservation utility $\underline{U}$ must be at least $u_L$. Suppose instead that $\underline{U} < u_L$ in some equilibrium. So some employees receive a contract that gives them utility $\underline{U} < u_L$, 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 $\underline{U}$ to the employee and delivers strictly positive profits to the firm.

An immediate consequence of $\underline{U} \geq u_L$ 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 $H$ are paid $w_S > g_L$ after success, but that some workers are still employed on task $L$ in equilibrium; we then have $\underline{U} = u_L$.

The participation constraint is nonbinding for an overpaid employee. Ignoring the participation constraint, a firm's profit-maximization problem for an employee on task $H$ is

$$\max_{w_S, p} p\left(g_H - w_S\right) - k_H \text{ subject to } \gamma'\left(p\right) = w_S. \tag{5}$$

Firm profits initially increase in the size of the bonus $w_S$, since the concomitant increase in effort $p$ justifies the cost. As in the efficiency wage literature, this feature is what makes equilibrium overpay possible: reducing compensation to the point where the employee's participation constraint binds may not be optimal for the firm as it will undermine incentives. For use throughout, denote the bonus that maximizes profits (absent the participation constraint) by $\underline{\Delta}\left(g_H\right)$, that is,

$$\underline{\Delta}\left(g_H\right) \equiv \arg\max_{w_S} \left(\gamma'\right)^{-1}\left(w_S\right)\left(g_H - w_S\right). \tag{6}$$

For any $k_H$, there is a unique return $\underline{g}_H\left(k_H\right)$ at which firms just break even by giving $H$-employees the profit-maximizing success bonus $\underline{\Delta}\left(\underline{g}_H\left(k_H\right)\right)$. This is the equilibrium return under two conditions: 1) the bonus $\underline{\Delta}\left(\underline{g}_H\left(k_H\right)\right)$ is (weakly) larger than $g_L$, the bonus on the $L$ task, so that the participation constraint is satisfied, and 2) the return $\underline{g}_H\left(k_H\right)$ that allows firms to just break even on task $H$ is high enough that, in equilibrium, not all employees can be employed on task $H$.

If 1) holds but not 2), task $H$ returns are high enough to draw everyone into task $H$, and the equilibrium return exceeds $\underline{g}_H\left(k_H\right)$. Competition for employees then raises bonuses above $\underline{\Delta}\left(\underline{g}_H\left(k_H\right)\right)$. If 2) holds but not 1), competition for employees pushes bonuses above $\underline{\Delta}\left(\underline{g}_H\left(k_H\right)\right)$, and the return must rise above $\underline{g}_H\left(k_H\right)$ for firms to break even. In both cases, the economy would not feature overpay, as everyone would earn the same utility.

We now claim that if $k_H$ is large enough, both 1) and 2) will hold, so that $\underline{g}_H\left(k_H\right)$ is the equilibrium return and $\underline{\Delta}\left(\underline{g}_H\left(k_H\right)\right)$ the equilibrium bonus. The following intuitive lemma helps establish this claim.

LEMMA 1: *(i)* $\underline{g}_H(k_H)$ *is uniquely defined, is strictly increasing in* $k_H$*, and approaches infinity as* $k_H \to \infty$*. (ii)* $\underline{\Delta}(g_H)$ *is uniquely defined, is strictly increasing in* $g_H$*, and approaches infinity as* $g_H \to \infty$*.*

Part (i) of Lemma 1 simply says that when there is more to lose ($k_H$), in equilibrium there has to be more to gain ($g_H$) for the firm to break even. Part (ii) says that as the marginal productivity of labor $g_H$ increases, the firm finds it optimal to incentivize higher effort with a higher bonus. An implication of the lemma is that as $k_H$ grows large, the bonus $\underline{\Delta}\left(\underline{g}_H(k_H)\right)$ must also grow large, and in particular larger than the bonus on the $L$ task. This establishes that condition 1) above is satisfied as $k_H$ grows large.

That not everyone can be employed on task $H$ when the amount of capital $k_H$ per employee grows large follows immediately from Assumption 2, which says that if too much capital in aggregate is devoted to the $H$ task, the return $\frac{g_H}{k_H}$ is driven below one so that firms cannot break even. Economically, when capital-per-employee is large, the economy does not require many employees to be assigned to task $H$. 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.

Applied to the financial sector, this analysis shows how the combination of moral hazard and scaleability combine to partition the labor market so that some employees work very hard and receive very large bonuses, while other (ex ante identical) employees obtain less desirable jobs. However, the one-period analysis is inherently incapable of shedding light on career structures, or on whether firms respond to aggregate shocks by firing existing employees or reducing hiring. We take up these topics in our analysis of the dynamic model.

Figures 1 and 2 illustrate an overpaying equilibrium. Figure 1 shows that at return $\underline{g}_H$, firms cannot make strictly positive 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 $\zeta_H$. 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.

**[FIGURES 1 AND 2 ABOUT HERE]**

For comparison with the results of Section VI below, we consider here the effect of aggregate shocks, which we model as a shift in the return functions $\zeta_H$ and $g_L$. Focusing on the case of an overpaying equilibrium in the one-period economy, the only effect of aggregate shocks is to the number of employees assigned to the two contracts. The equilibrium contracts themselves are unchanged. In particular, an upwards shift in $\zeta_H(\cdot, k_H)$ and $g_L$, which one can interpret as an economic boom, increases the number of employees assigned to task $H$, but does not affect employee bonuses or effort.

Note that many of the elements in the one-period version of our model are present in existing literature. The fact that firm profits are increasing in employee utility is a standard property of partial-equilibrium analyses of moral hazard problems. Moreover, the implication that employee utilities are not equalized in equilibrium can be found in a range of papers, most notably in Shapiro and Stiglitz (1984), but also in more recent papers such as Acemoglu and Newman (2002) and Biais, Rochet, and Woolley (2014).

However, the one-period version of our model contains a new element: employee utilities are not equalized even though we allow for free entry and/or constant returns to scale at the firm level. In contrast, the aforementioned papers use limited entry and decreasing returns to scale at the firm level to avoid equilibrium utility equalization. Instead, we assume decreasing returns at the aggregate level (that is, $\zeta_H$ is decreasing), and close the model by solving for the equilibrium return.

This new ingredient is important for two reasons. First, we believe it is realistic for much (though certainly not all) of the financial sector: financial sector firms appear relatively easy to scale up, and in some parts of the financial sector, such as asset management, entry by new competitors is very common. Second, as noted above, this feature of the model implies that equilibrium returns are unaffected by aggregate shocks to $\zeta_H$, which drives our results in Section VI below.[18]

## V. Equilibrium Contracts in the Dynamic Economy

In this section we establish the main result of the paper: if task $H$ stakes $k_H$ are large, two very different contracts—that is, career paths—coexist in equilibrium. In one contract, a young employee is immediately assigned to the high moral hazard task $H$. The contract is up-or-out, in the sense that if the employee succeeds, he continues to work in task $H$ and is paid very generously, while if he fails, he is "demoted" to task $L$ and is paid less. In the other equilibrium contract, an employee spends his entire career in task $L$, with no possibility of "promotion" to task $H$. Expected compensation, employee effort, and expected utility are all higher in the first of these contracts. We interpret these two coexisting contracts as financial sector and nonfinancial sector jobs, or alternatively as high- and low-stakes jobs within the financial sector.

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. Analogous 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 $2u_L$ in lifetime utility. We refer to this contract as the $C^{LL}$ contract.

As discussed in the introduction, a natural conjecture would be for the equilibrium of the dynamic model to take a very different form from the one described above, with all young employees starting their careers in the low moral hazard task $L$, and only moving to task $H$ after success. As the following example demonstrates, this conjecture is correct when capital $k_H$ is sufficently small.

EXAMPLE: *Let $k_H \leq g_L$. Define $g_H^*$ as the equilibrium return in the frictionless benchmark of Section III, in which surplus is equalized across tasks, that is, $S_H(p^*(g_H); g_H) = S_L(p_L^*; g_L)$.*

*We now show that, given some conditions on trade-profitability $\zeta_H$, 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 the maximal surplus in the two tasks is the same, the maximal surplus is* $2u_L$, *and so the employee earns the same utility as under contract* $C^{LL}$.[19] *The remaining fraction* $1 - \lambda$ *of employees are assigned contract* $C^{LL}$. *As long as the number of task H trades 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.*

The example above 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.[20] In our setting, assigning the employee to 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, the equilibrium takes a very different form from the above example. Two contracts coexist in equilibrium, with very different compensation, effort, utility, and promotion characteristics.

PROPOSITION 1: *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 and more work for task H employees: a strict subset of young employees start on task H, and receive strictly greater expected utility than young employees starting on task L,*

*but also work harder.*

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: in contrast to task $H$ employees, 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 1 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 task $H$ after success—and the large surplus this allows the employee to capture—creates a very strong incentive to work when young. Profit maximization then leads the young employee to be assigned to task $H$.

The proof of Proposition 1 is constructive, and proceeds in two main steps. In the first step, detailed in Sections V.A and V.B, we solve the partial-equilibrium problem of finding the profit-maximizing contract that assigns an employee to the $H$ task in at least some node. We do not impose the participation constraint in this step: as we will show, the profit-maximizing contract involving some activity on $H$ satisfies this constraint in equilibrium. In the second step, detailed in Section V.C, we use this partial-equilibrium result to solve for the full equilibrium.

We conduct the partial-equilibrium analysis under the assumption that the return $g_H$ satisfies two basic properties (which we then verify are satisfied in equilibrium). First, the return $g_H$ is high enough that there exists some contract with nonnegative profits that assigns an employee to task $H$ at some node. Second, the return satisfies part 1 of Proposition 1, that is, $g_H < \underline{g}_H$. Intuitively, this second property reflects the fact that dynamic contracts give the firm more tools to structure incentives, which makes it possible for them to break even at a lower return. Free entry ensures that the aggregate amount of $H$ activity increases until returns are driven down to this new break-even level.

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 $u_L$. Consequently, the one-sided commitment constraint reduces to

$$\max_p pw_{XS} + (1-p)\, w_{XF} - \gamma(p) \geq u_L \qquad \text{for } X = S, F, \tag{7}$$

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

*A. Up-or-Out*

In this subsection and the next, we solve the partial-equilibrium problem of maximizing profits subject to the one-sided commitment constraint (7) and to assigning the employee to task $H$ in at least some node:

$$\max_C \Pi(C) \text{ subject to (7) and } (i, i_S, i_F) \neq (L, L, L). \tag{8}$$

As noted, we do not impose the participation constraint.

We first show that the one-sided commitment constraint (7) 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 $u_L$ 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 delivering utility of at least $u_L$. (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, we have $i_F = L$, $w_{FS} = g_L$, and $w_{FF} = 0$, which leads to firm profits of zero in the second period after first-period failure.[21]

We next argue that if the employee is ever to be employed on task $H$, he must be employed on task $H$ after success—the successful employee is "up." Suppose this were not the case, so that the employee is assigned to task $L$ after both 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 because any subcontract must deliver utility weakly above $u_L$, when the employee is assigned to $L$ in the second period he must be given payments of the form $(w_{XS}, w_{XF}) = (g_L + w_{XF}, w_{XF})$ that induce surplus-maximizing effort $p_S = p^*(g_L)$, with any additional promised utility paid in the form of fixed pay $w_{XF}$. But then the young employee faces a contract that delivers utility $u_L + w_{FF}$ after failure and $u_L + w_{SF}$ after success, costing the firm $w_{FF}$ after failure and $w_{SF}$ after success. This is equivalent to a one-period contract on task $H$.

*B. Initial Assignment to Task H*

Given the up-or-out characterization, that is, $(i_S, i_F, w_{FS}, w_{FF}) = (H, L, g_L, 0)$, it remains to determine the initial task assignment $i$ and the remaining compensation terms $w_{SS}, w_{SF}$. For notational convenience, we write $\Delta_S = w_{SS} - w_{SF}$. The contract terms $i, w_{SF}$, and $\Delta_S$ solve the following profit-maximization problem, which corresponds to substituting the up-or-out characterization into (8):

$$\max_{i,\Delta_S \geq 0, w_{SF} \geq 0} p(g_i - w_{SF}) - k_i + p(p_S(g_H - \Delta_S) - k_H), \tag{9}$$

where the employee's effort decision $p_S$ in the second period after success and $p$ in the first period are determined by the IC constraints (1) and (2), which simplify to

$$\begin{aligned} \gamma'(p_S) &= \Delta_S \\ \gamma'(p) &= w_{SF} + \max_{\tilde{p}} (\tilde{p}\Delta_S - \gamma(\tilde{p})) - u_L = w_{SF} + p_S\Delta_S - \gamma(p_S) - u_L. \end{aligned}$$

Note that, compared to a one-period problem, $\Delta_S$ has a double effect here: it incentivizes effort in the first period as well as the second. An important consequence is that effort is *serially correlated:* if the firm incentivizes high effort after success, it also inevitably incentivizes high effort in the first period. We next show that when capital $k_H$ is large, the serial correlation of effort makes allocation to task $H$ in the first period profit-maximizing. The reason is that when $k_H$ is large, $g_H$ must be large for firms to break even, so that the difference in firm payoff between failure and success makes high effort on task $H$ very important. Therefore, the bonus for success in the second period must be large, which in turn induces so much effort in the first period that the young worker is most efficiently allocated to task $H$.

LEMMA 2: *Let* $g_H(k_H)$ *be a sequence of returns indexed by* $k_H$ *such that (a)* $g_H(k_H) < \underline{g}_H(k_H)$*, and (b) there exists an up-or-out contract that gives nonnegative profits. For* $k_H$ *sufficiently large, the solution to the profit-maximization problem (9) is unique, and assigns young workers to task* $H$*.*

The task assignment prediction of Lemma 2 is key to our analysis, and we prove it here (uniqueness is proved in the Appendix). We focus on the case in which there exists an up-or-out contract with $i = L$ that gives nonnegative profits. (For the opposite case in which no up-or-out contract $i = L$ breaks even, the result is immediate.) Consider the contract terms $(L, w_{SF}, \Delta_S)$ that maximize profits within this class, inducing effort $p$ in the first period and $p_S$ in the second period after success. Our goal is to show that when $k_H$ is large, there exists a perturbed contract that sets $i = H$ and generates strictly higher profits. From (9), profits from the original contract are

$$p\left(g_L - w_{SF}\right) + p\left(p_S\left(g_H - \Delta_S\right) - k_H\right).$$

We make two observations about these profits. First, because $g_H < \underline{g}_H$, there is no one-period contract on the $H$ task that breaks even; in particular, the second-period profits $p_S\left(g_H - \Delta_S\right) - k_H$ that stem purely from the $H$ task net of the bonus $\Delta_S$ are negative. Hence, total profits are bounded from above by $g_L$, the revenue generated after first-period success. Second, because (by supposition) total profits are nonnegative, losses on the second-period project are less than $g_L$ :

$$p_S\left(g_H - \Delta_S\right) - k_H \geq -g_L. \tag{10}$$

We now show that when $k_H$ is large, there exists a contract that instead allocates the employee to the $H$ task in the first period, and makes higher profits than $g_L$, completing the proof. In particular, consider the alternate contract $\left(H, \tilde{w}_{SF}, \Delta_S\right)$, which (i) allocates the worker to task $H$ in the first period, (ii) keeps the bonus $\Delta_S$ the same so that second-period effort $p_S$ after success is unchanged, and (iii) sets $\tilde{w}_{SF}$ such that first-period effort is also $p_S$, which from the first-period

IC constraint implies that $\tilde{w}_{SF}$ satisfies[22]

$$\gamma'(p_S) = \tilde{w}_{SF} + p_S \Delta_S - \gamma(p_S) - u_L.$$

Profits from this contract can be decomposed as

$$(1 + p_S)(p_S(g_H - \Delta_S) - k_H) + p_S(\Delta_S - \tilde{w}_{SF}). \tag{11}$$

The first term is the firm's profits from using a one-period $H$ contract with bonus $\Delta_S$ in the first period, and then again in the second period after success. From (10), this term generates losses no larger than $(1 + p_S) g_L$. The second term is the reduction in the cost of incentivizing first-period effort $p_S$ that is afforded by the dynamic contract: in a one-period contract a bonus $\Delta_S$ is required, whereas in the dynamic contract the prospect of future bonuses means that a first-period bonus of $\tilde{w}_{SF}$ is sufficient. By the IC constraints, the reduction in bonus equals

$$\Delta_S - \tilde{w}_{SF} = p_S \Delta_S - \gamma(p_S) - u_L. \tag{12}$$

To complete the proof, we show that $\Delta_S$ increases without bound as $k_H$ grows large.[23] This implies that the bonus savings and hence profits also increase without bound, and in particular grow larger than the upper bound $g_L$ generated by the contract terms $(L, w_{SF}, \Delta_S)$. To show that $\Delta_S$ grows without bound, first note that since $\Delta_S$ maximizes profits in the $(L, w_{SF}, \Delta_S)$ contract, it is easy to see from the profit-maximization problem (9) that $\Delta_S$ must be no smaller than the bonus $\underline{\Delta}(g_H)$ defined by (6) that maximizes one-period profits, since increasing $\Delta_S$ has the added benefit in the dynamic contract of increasing first-period effort. The result then follows from Lemma 1, which shows that $\underline{\Delta}(g_H)$ increases without bound as $g_H$ grows large: as $k_H$ grows large, $g_H$ must grow large for firms to break even. This completes the proof of Lemma 2.

In the above proof of Lemma 2, we use a contract that implements the same success probability on task $H$ in both periods. This is not the profit-maximizing contract—as we now show, the profit-maximizing contract incentivizes more effort and hence a higher success probability on task $H$ after first-period success. If $w_{SF} = 0$, this property is immediate from the IC constraints. Economically, the bonus payment $\Delta_S$ provides more incentives in the second period than in the first, because in the first period the employee discounts the bonus by the cost of second-period effort $p_S$, by the probability that he is unlucky and fails in the first period, and by the fact that even if he fails he still receives a continuation utility of $u_L$. Moreover, this conclusion holds even if $w_{SF} > 0$ :

LEMMA 3: *If $g_H \geq g_L$ and $g_H > k_H$, the solution to the profit-maximization problem (9) has $p_S > p$, that is, effort rises after success.*

The fact that effort increases after success constitutes a further sense in which employees are moved "up" after success: more is expected of them when old.[24]

*C. Completing the Equilibrium Characterization*

So far, we have characterized the solution to the profit-maximization problem (8). We have also shown that when $k_H$ is sufficiently large, and provided that $g_H$ satisifes the two conditions stated, the profit-maximizing contract is $C^{HH}(g_H) \equiv (H, H, L, w_{SS}, w_{SF}, g_L, 0)$, where $(w_{SS}, w_{SF}) = (w_{SF} + \Delta_S, w_{SF})$ solves the profit-maximization problem (9) subject to $i = H$.

We now use this result to characterize the equilibrium. Let $g_H^{eq}$ be such that it is just possible for a firm to break even with contract $C^{HH}(g_H)$,[25] that is,

$$\max_{\Delta_S \geq 0, w_{SF} \geq 0} p(g_H - w_{SF}) - k_H + p(p_S(g_H - \Delta_S) - k_H) = 0. \tag{13}$$

We start by noting that $g_H^{eq}$ is indeed strictly less than $\underline{g}_H$, since otherwise the extra incentives induced by a dynamic contract imply that there exists an up-or-out contract with strictly positive profits, contradicting the definition of $g_H^{eq}$. Concretely, dynamic contracting reduces the rent that needs to be delivered to the employee, and this increases aggregate task $H$ activity and reduces the equilibrium return.

Two things remain to be shown to prove that the equilibrium is as described in Proposition 1. First, we need to show that not enough aggregate $H$ activity can be sustained for everyone to be employed with contract $C^{HH}$, so that some employees need to be employed with the $C^{LL}$ contract that leaves them stuck on the $L$ task for their entire career. Second, we need to show that $C^{HH}$ employees earn strictly higher expected lifetime utility than $C^{LL}$ employees. The last feature follows immediately from the proof of Lemma 2, where we showed that as $k_H$ grows without bound, $g_H$ must grow without bound, so the optimal second-period bonus $\Delta_S$ and hence utility for $C^{HH}$ employees also grow without bound (and in particular become larger than $2u_L$, the expected lifetime utility of $C^{LL}$ employees).

To show that not everyone can be employed with $C^{HH}$ in equilibrium, let $p\left(C^{HH}\right)$ and $p_S\left(C^{HH}\right)$ be the effort levels induced by the $C^{HH}$ contract, so that each young employee given the $C^{HH}$ contract expects $\left(1+p_S\left(C^{HH}\right)\right)p\left(C^{HH}\right)$ task $H$ successes. This quantity is certainly bounded away from zero as $k_H$ grows large, since as just argued, the optimal second-period bonus $\Delta_S$ grows without bound. Moreover, part (ii) of Assumption 2 then implies that if all employees were given the $C^{HH}$ contract as $k_H$ grows large, the return would fall below the break-even level $g_H^{eq}$. Consequently, only a strict subset of young employees are given the $C^{HH}$ contract,[26] with the remaining employees assigned to the $C^{LL}$ contract. With the exception of uniqueness (which is established in the Appendix), this completes the proof of Proposition 1.

Finally, we note that parallel to the one-period benchmark, $g_H^{eq}$ is determined solely by the firm's profit-maximization problem, and is independent of the trade-profitability function $\zeta_H$. We make extensive use of this property in our analysis of aggregate shocks below.

*D. Applications*

Not only does Proposition 1 provide an explanation of high compensation in the financial sector that is neither a compensating differential nor a return to skill, but the equilibrium also has a number of 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 observation that financial sector jobs often entail very long hours. For example, using a sample of University of Chicago MBA alumni, Bertrand, Goldin, and Katz (2010) report that the average hours worked in investment banking is 73.6 hours per week; the next highest figure reported is for consulting, at 60.7 hours per week. At the extreme, as reported by Michel (2011), investment bankers work up to 120 hours per week. We reiterate, however, that Proposition 1 says that 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.

*Serial correlation of hours:* Overpaid employees work longer hours than non-overpaid employees both early in their careers and later in their careers (provided they succeed when young). Again, this is consistent with accounts of the financial sector.

*Heavy use of both performance pay and backloading of pay:* Employees who receive the $C^{HH}$ contract receive larger bonuses than those receiving the $C^{LL}$ contract. Moreover, pay is more backloaded in the $C^{HH}$ contract. First, observe that the firm could pay the employee up to $w_{SF}$ after first-period success, without affecting incentives. Accordingly, we identify $w_{SF}$ with the first-period bonus, and measure the backloading of the pay via the ratio $\frac{w_{SS}-w_{SF}}{w_{SF}}$, that is, the ratio of the second-period bonus to the first-period bonus. For the $C^{LL}$ contract, this ratio is one, that is, there is no backloading. For the $C^{HH}$ contract, this ratio exceeds one.[27] These predictions are consistent with perceptions that the financial sector makes heavy use of both 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. The website WWW.CAREERS-IN-FINANCE.COM/IBSAL.HTM reports investment banking pay and tells a similar story. Bell and Van Reenen (2014) report detailed compensation data for the "code staff" of large banks headquartered in London.[28] For this admittedly senior group of bank employees, 58.3% of total compensation is deferred, while out of the nondeferred portion 64.4% is bonus pay.

*Importance of entering profession soon after graduation:* As noted, Proposition 1 features dynamic segregation: if an employee is not assigned to task $H$ when young, he never enters the high-paying sector in his career. This is consistent with both anecdotal accounts (see, for example, DeChesare (2012)[29]), and Oyer's (2008) econometric finding that Stanford MBAs who fail to get a job in finance upon graduation for reasons unrelated to skill or preferences typically do not manage to make the transition into finance later on, despite the high pay differential. (Oyer uses aggregate economic conditions as an instrument for initial job placement to avoid selection on skill or preferences; we formally study the effect of aggregate shocks in the next section.)

*Up-or-out:* As noted, Proposition 1 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, for example, DeChesare (2012)[30]).

A slightly different way to interpret Proposition 1 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 prestigious brokerage firm to a lower-paying, less prestigious one than the other way around.

## 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. In particular, we predict that good economic conditions lead to both high compensation and low effort, so that compensation and effort are negatively correlated in the time series.

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 (a higher 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 $\zeta_H^G$ is sufficiently close to $\zeta_H^B$ and $g_L^G$ is sufficiently close to $g_L^B$ 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.

### *A. 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 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 $H$, that is, $g_L^G = g_L^B$ and $\zeta_H^G(\cdot, k_H) > \zeta_H^B(\cdot, k_H)$. This assumption makes the analysis 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 1, the return $g_H^{eq}$ is independent of the trade-profitability function $\zeta_H$, and is determined instead by the condition that the firm has zero profits under the profit-maximizing contract (see (13), and the discussion in Section IV). Because $g_L^G = g_L^B$, the employee's minimum continuation utility $(u_L)$ is independent of the state, and so the firm's profit-maximization problem is the same as in the case without aggregate shocks. Consequently, $g_H^{eq}$ 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 $g_H^{eq}$, 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 $\zeta_H^\omega$—"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 $\lambda_t$ be the number of overpaid young employees hired for task $H$ at date $t$. Denote by $y_H^\omega$ the task $H$ supply that can be sustained at the equilibrium return $g_H^{eq}$ in state $\omega$, that is, $y_H^\omega$ solves $g_H^{eq}/k_H = \zeta_H^\omega(y_H^\omega, k_H)$. 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. Date $t$ output from task $H$ equals $p\lambda_t + p\lambda_{t-1}p_S$, where $p\lambda_t$ is the output by the $\lambda_t$ just-hired young employees and $p\lambda_{t-1}p_S$ is the output from the $\lambda_{t-1}$ 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

$$\lambda_t = \frac{y_H^{\omega_t}}{p} - \lambda_{t-1}p_S. \tag{14}$$

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.[31]

It is easy to see from (14) 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 $\lambda^\omega$, 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, that is, $\lambda^G > \lambda^B$. Average success rates, on the other hand, are the same in both scenarios.

PROPOSITION 2: *Suppose that after many periods in the good state, the economy suffers the aggregate shock and enters the bad state. For $k_H$ sufficiently large and an aggregate shock sufficiently small, hiring of young employees into task $H$ falls below even $\lambda^B$, 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 $H$ actually increases.*

The proof is almost immediate from (14), 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, that is, $p_S > p$ (see Proposition 1), the average success rate in task $H$ increases when the bad shock hits.

### *A.1. Implication: Initial Conditions Matter*

The most immediate implication of Proposition 2 is that the aggregate conditions when someone first enters the labor force have lifelong consequences. This complements the result of the previous section that idiosyncratic shocks—in the form of an individual's outcome in the labor market "lottery"–have lifelong effects, and 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 10 (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.

### *A.2. Implication: Financial Sector Firms Respond to Bad Times by Hiring Less Rather Than Increasing Firing*

The reason task $H$ hiring falls below $\lambda^B$ 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 comes at the expense of hiring new young employees. According to the U.S. 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 show that total separations also fell in 2008: our model predicts no change in firing, while one might naively expect that separations would increase.

### *A.3. Implication: Investments Undertaken in Bad Times are More Profitable and Have Higher Success Rates*

Proposition 2 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 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 of such countercyclical investment performance in both the buyout (Kaplan and Stein (1993)) and the venture capital (Gompers and Lerner (2000)) markets.

*A.4. Remaining Observations*

Above, we note 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 U.S. productivity has been countercyclical since the mid-1980s (see Gali and van Rens (2014)). 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 U.S. 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 2 can also be interpreted in terms of unemployment. To do so, think of task $L$ as corresponding to unemployment, with $u_L$ the level of utility obtained by the unemployed. Then Proposition 2 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.

*B. Time-Series Implications: Procyclical Moral Hazard*

Next, we expand our analysis to the case in which aggregate shocks affect both tasks, that is, $g_L^G > g_L^B$ and $\zeta_H^G(\cdot, k_H) > \zeta_H^B(\cdot, k_H)$. The significance of shocks for task $L$ output is that they affect $u_L$, 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 as employees receive more generous bonuses. Moreover, we show that capital does not fully respond to improvements in investment opportunities, that is, 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 $\psi$ 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 (bad) tomorrow if it is good (bad) today, $\mu^{GG} > \mu^{BG}$.

Denote by $g_H^\omega$ the task $H$ state $\omega$ return. Define $u_L^\omega$ analogously to $u_L$ in (4), with $p^*(g_L^\omega)$

replacing $p_L^*$. 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. Put simply, employees expect to "land on their feet"even if they fail. This is the key economic force driving our results below.[32]

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 $g_H^\omega$ 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:

$$\max_{p^\omega, w_{SS}^{\omega\psi}, w_{SF}^{\omega\psi}} p^\omega \left( g_H^\omega + \sum_{\psi=G,B} \mu^{\omega\psi} \left[ p\left(\mathbf{w}_S^{\omega\psi}\right) g_H^\psi - E\left[w^{\omega\psi}|S\right] - k_H \right] \right) - k_H \tag{15}$$

subject to the first-period IC constraint

$$\gamma'\left(p^\omega\right) = \sum_{\psi=G,B} \mu^{\omega\psi} \left[ E\left[w^{\omega\psi}|S\right] - \gamma\left(p\left(\mathbf{w}_S^{\omega\psi}\right)\right) \right] - \bar{u}_L^\omega. \tag{16}$$

Our main result, stated formally below, is that moral hazard problems in task $H$ endogenously worsen in good times, that is, are procyclical. The driving force is the IC constraint (16), 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 that this incentive effect dominates the direct effect that, for any fixed level of task $H$ activity $y_H$, returns are higher in good times, that is, $\zeta_H^G\left(y_H, k_H\right) > \zeta_H^B\left(y_H, k_H\right)$, 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 $\zeta_H$ has *no* direct impact on equilibrium returns (exactly as in the previous subsection).[33]

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 3: *For $k_H$ sufficiently large and aggregate shocks sufficiently small:*

1. *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, $E\left[w_{SS}^{G\psi}|\omega = G\right] > E\left[w_{SS}^{B\psi}|\omega = B\right]$.*

2. *Equilibrium returns are strictly higher in good times, $g_H^G > g_H^B$.*

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

Part (1) 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 2, the mechanism is different. Whereas the previous result reflects 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 (1) also establishes that booms generate higher (promised) bonuses—but the rise in bonuses is insufficient to offset of effect of improved outside options.[34] Consequently, our model predicts that compensation and failure rates are positively correlated in the time series

Part (2) 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, that is, $\zeta_H^G(y_H, k_H) > \zeta_H^B(y_H, k_H)$. 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 $\zeta_H$ pulls in so much new capital that the equilibrium return remains at $\underline{g}_H$.[35] Consequently, in the one-period economy bonuses and success probabilities are unaffected by the state of the world.

Similar to other dynamic contracting papers (see, for example, DeMarzo et al. (2012)), our model predicts "pay for luck" (Part (3)). 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.[36] Hence, in a dynamic setting such as ours, Holmstrom's (1979) well-known informativeness principle, which holds that compensation should only be made contingent on variables that depend on an agent's effort, does not hold. A number of empirical papers document 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.[37] 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 argue in the introduction that the available evidence suggests that high compensation in the financial sector is not a pure skill premium. Accordingly, in our basic model we abstract 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 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 tasks $i = L, H$. 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 $L$, that is, $c_L < c_H < 1$. 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 the talented employee is more valuable, he is 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 the one operating in the aggregate shocks analysis of Section VI above.) Put simply, 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, that is, $c_H$ close enough to one. 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.[38]

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. Discussion

*A. 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 (2012), Kaplan and Rauh (2010), and Bell and Van Reenen (2014)). 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 has been a 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 Custodio, Ferreira, and Matos (2013). In our model, we can associate the importance of general skills with $u_L$, the utility an employee gets on the "normal" task $L$. An upward trend in the outside option $u_L$ 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, which has increased over the past three decades: see, for example, Kaplan and Rauh (2010) for evidence of increases in assets under management per employee within hedge funds, private equity funds, and venture capital funds.

To the extent 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 (2012), 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 in 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 (2012) 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 employees 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 (that is, limits on $k_H$) would reduce overpay, but would also reduce economy-wide surplus, since this would force firms to do something they could freely do on their own (as we argued above, firms would never want to scale down $k_H$). 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 both reduce overpay and increase economy-wide surplus. Third, and contrary to the claims of many commentators, implicit government guarantees may actually 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 from the two tasks is equalized, and overpay is eliminated.

*B. Leaving to Start a Hedge Fund*

The up-or-out terminology that we use to describe optimal contracts for overpaid employees suggests that successful employees should gain promotion to senior positions with their employer (for example, becoming Vice President,Director, Principal,or Managing Director). An alternative and not uncommon path for successful financial sector employees is to leave their firm, take their accumulated bonuses, and combine them with outside capital to, for example, start their own hedge fund. Anecdotal accounts suggest that many hedge funds are started in this way (see, for example, journalistic accounts by Fishman (2004)[39] and Makan (2012)[40]). At first blush, this type of career path might seem at odds with the optimal dynamic contract in our setting as it cuts the link between the employee and the initial firm after the first period. However, we now show that our optimal dynamic contract can be reinterpreted as a sequence of optimal one-period contracts in which the worker uses accumulated wealth from his previous employment to coinvest with new financiers, and in which firms (or, in the hedge fund example, new investors) break even period-by-period.

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 thus be the same (zero) 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 then the employee left the firm.

Armed with the bonus $W_S$, the employee can 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). Note that this gives the employee exactly the same state-contingent payoffs as in the optimal dynamic contract. These financing terms are sufficient to attract investors. It is straightforward to show that $w_{SF} = 0$ when $W_S < k_H$;[41] it then follows from the 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.[42]

*C. Weaker Scale Economies*

Above we assume a strong form of scale economies at the employee level, in the sense that an employee's cost of effort is independent of the amount of capital he manages up to a maximum of $k_H$. As discussed, we think this assumption makes sense for the financial sector. Nonetheless, it is worth considering how our results would change if we relaxed this assumption. Biais et al. (2010) and DeMarzo et al. (2012) analyze partial-equilibrium contracting models with just one type of project, but where the cost of effort is proportional to project size. They find that the optimal contract entails an increase in project size after success, for reasons that parallel Lemma 3, which shows that effort increases after success. Based on these results in existing literature, we conjecture that if we adopted the same assumption about effort costs, we would obtain a further form of promotion in our model: overpaid employees who succeed when young would be "promoted" in the sense of managing more capital when old. At the same time, our result that otherwise-identical employees are given different contracts would remain, because under the proportional-effort cost assumption, agency costs are proportional to capital managed, and so it is best to adopt the maximal feasible capital for old employees, so as to free up employees to work on task $L$. Moreover, we conjecture that this result is robust to having effort costs that grow slightly faster than proportionately in capital, as considered in an extension discussed in Biais et al. (2010).

## IX. Conclusion

In this paper, we analyze a model in which financial sector employees capture rents (are "overpaid") as a natural consequence of the large-scale nature of financial activities and the difficulty of directly observing the actions of employees. The equilibrium dynamic contract features punishing hours and up-or-out promotion structures consistent with observed work conditions in the financial sector. 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. In particular, we show that financial employees act less responsibly exactly during boom times when their pay is the highest. Finally, an extension to observable skill differences delivers implications for when talent is "lured" away from socially more useful occupations, as well as when high-skilled individuals are "scorned" and do not receive the most attractive 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 enough by mid-career, 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 VIII, this is consistent with our model.

In our model, the employee's only choice is the amount of effort to exert in order to increase the probability of success of the trade. To deepen the analysis of the effect of moral hazard on risk-taking, an interesting extension might be one in which the employee learns something relevant about the probability of success after effort is sunk, and then has the possibility of aborting unpromising trades. This extension would capture a particular form of risk-shifting: after low effort, the employee is faced with a choice between a safe investment (trade abandonment), and a less profitable but higher variance investment.[43] A possible response to this risk-shifting problem would be for firms 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 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 full development of a richer model of this sort for future research.

## Appendix: Proofs

*Proof of Lemma 1:* Part (i) is immediate. For part (ii), define $\underline{p}(g_H) = \arg\max_p p(g_H - \gamma'(p))$, that is, $\underline{p}(g_H)$ solves $g_H = \gamma'(p) + p\gamma''(p)$. By a change of variables in the firm's maximization problem, one can see that $\underline{\Delta}(g_H) = \gamma'\left(\underline{p}(g_H)\right)$. By Assumption 1, the expression $\gamma'(p) + p\gamma''(p)$ is strictly increasing in $p$, and ranges from from zero to $\infty$ as $p$ ranges from zero to maximal effort $\bar{p}$. Hence, $\underline{p}(g_H)$ is well defined and strictly increasing in $g_H$, and $\underline{p}(g_H) \to \bar{p}$ as $g_H \to \infty$. We then have

$$\lim_{g_H \to \infty} \underline{\Delta}(g_H) = \lim_{g_H \to \infty} \gamma'\left(\underline{p}(g_H)\right) = \lim_{p \to \bar{p}} \gamma'(p) = \infty.$$

QED

*Verification that the one-period economy features overpay in the example in Section V:* Here, we show that there are conditions under which the one-period economy features overpay under the conditions of the Example in Section V, namely, $k_H \leq g_L$ and $\zeta_H(p^*(g_L)\,p^*(g_H^*), k_H) < \frac{g_H^*}{k_H}$. Suppose that the cost function $\gamma$ has the property $\gamma'(p) > p^2\gamma''(p)$ for some $p$.[44] Given this, choose $\check{p}$ and $k_H$ such that $\gamma'(\check{p}) > \check{p}^2\gamma''(\check{p}) = k_H$, and choose $g_L \in [k_H, \gamma'(\check{p}))$.

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 receive utility strictly above $u_L$. So provided $\zeta_H(\check{p}, k_H)$ is sufficiently small, the one-period economy does indeed feature overpay.

*Proof of Lemma 2:* The main text establishes that for $k_H$ large enough, any solution to (9) features $i = H$. Here we establish uniqueness of the remaining contract terms $w_{SF}$ and $\Delta_S$. Let

$(H, w_{SF}, \Delta_S)$ be a solution to (9), and let $\Pi$ be associated profits. By supposition, $\Pi \geq 0$. The first-order conditions for the solution to the profit-maximization problem (9) are

$$-pp_S + p\frac{\partial p_S}{\partial \Delta_S}(g_H - \Delta_S) + \frac{\partial p}{\partial \Delta_S}(g_i - w_{SF} + p_S(g_H - \Delta_S) - k_H) \leq 0 \tag{A.1}$$

$$-p + \frac{\partial p}{\partial w_{SF}}(g_i - w_{SF} + p_S(g_H - \Delta_S) - k_H) \leq 0, \tag{A.2}$$

where $\frac{\partial p_S}{\partial \Delta_S} = \frac{1}{\gamma''(p_S)}$ and $\frac{\partial p}{\partial \Delta_S} = p_S \frac{\partial p}{\partial w_{SF}} = p_S \frac{1}{\gamma''(p)}$, and where the inequalities hold with equality if $\Delta_S > 0$ and $w_{SF} > 0$, respectively. First, note that $i = H$, $g_H > k_H$, and (A.1) and (A.2) imply that $\Delta_S = w_{SF} = 0$ cannot be a solution. Second, note that these same conditions imply that $\Delta_S = g_H$ if $w_{SF} > 0$.

It follows that $\Delta_S > 0$. It then follows from (A.1) and (A.2) that if $w_{SF} = 0$, then $\Delta_S \leq g_H$. Combined with the second-period IC constraint, it follows that $\Delta_S = \gamma'(p_S) \leq g_H$, regardless of the value of $w_{SF}$.

Substituting the fact that maximized profits are $\Pi$ into (A.1), and dividing by $pp_S$, we obtain

$$\frac{1}{p_S \gamma''(p_S)}\left(g_H - \gamma'(p_S)\right) + \frac{1}{p\gamma''(p)}\frac{\Pi + k_i}{p} = 1. \tag{A.3}$$

Suppose that, contrary to the claimed result, the solution is not unique. Let $(w_{SF}, \Delta_S)$ and $\left(\tilde{w}_{SF}, \tilde{\Delta}_S\right)$ be two distinct solutions. Let $(p, p_S)$ and $(\tilde{p}, \tilde{p}_S)$ be the effort levels induced by these two contracts, and note that $(p, p_S) \neq (\tilde{p}, \tilde{p}_S)$. By Assumption 1, $p\gamma''(p)$ is increasing in $p$. Moreover, from above, $\gamma'(p_S) \leq g_H$. Hence, the two terms on the left-hand side of (A.3) are weakly decreasing in $p_S$ and $p$, respectively. Consequently, without loss of generality $\tilde{p}_S > p_S$ and $\tilde{p} < p$. Hence, $\tilde{\Delta}_S > \Delta_S$ and $\tilde{w}_{SF} < w_{SF}$. But this gives a contradiction since, as noted above, $w_{SF} > 0$ implies $\Delta_S = g_H$, and $\tilde{\Delta}_S \leq g_H$. QED

*Proof of Lemma 3:* The main text establishes the result if $w_{SF} = 0$. Here, suppose instead that $w_{SF} > 0$. The first-order conditions (A.1) and (A.2) from the proof of Lemma 2 imply $\Delta_S = g_H$. Substitution into (A.2) then implies that

$$-p + \frac{1}{\gamma''(p)}(g_i - w_{SF} - k_H) = 0. \tag{A.4}$$

Define $u(p) \equiv p\gamma'(p) - \gamma(p)$, and observe that $u'(p) = p\gamma''(p)$. Hence, the left-hand side of (A.4) has the same sign as $-u'(p) + g_i - w_{SF} - k_H$.

To complete the proof, suppose that contrary to the claimed result $p \geq p_S$. By convexity of $u$ (Assumption 1) and $u(0) = 0$, we know $u'(p) \geq u'(p_S) > p_S u'(p_S) \geq u(p_S) = \gamma'(p) - w_{SF} + u_L$, where the last equality follows from the IC constraints. Since $\gamma'(p) \geq \gamma'(p_S) = g_H \geq g_i$, this implies that the left-hand side of (A.4) is strictly negative, giving a contradiction. QED

*Proof of Proposition 1:* Most of the proof is in the main text. Here, we fill in some formal details (Claim A), and establish equilibrium uniqueness (Claim B).

*Claim A: Let $g_H(k_H)$ be a sequence of returns indexed by $k_H$ such that (a) $g_H(k_H) < \underline{g}_H(k_H)$, and (b) maximized profits in problem* (8) *are nonnegative. For $k_H$ sufficiently large, $C^{HH}$ solves the problem*

$$\max_C \Pi(C; g_H(k_H)) \text{ subject to (7) and } U(C) \geq 2u_L. \tag{A.5}$$

In words, the contract $C^{HH}$ maximizes profits even without the restriction $(i, i_S, i_F, w_{FS}, w_{FF}) = (H, H, L, g_L, 0)$, and delivers utility of at least $2u_L$.

The main step is to establish the following subclaim.

*Subclaim: For $k_H$ sufficiently large, $C^{HH}$ is the unique solution to the maximization problem (8).*

The Subclaim establishes Claim A as follows. From the Subclaim, no contract satisfying (7) and the contract restriction that $(i, i_S, i_F) \neq (L, L, L)$ gives profits higher than $C^{HH}$. Maximal profits from a contract with $(i, i_S, i_F) = (L, L, L)$ and $U(C) \geq 2u_L$ are zero. Finally, for $k_H$ large enough, the main text establishes $U\left(C^{HH}\right) > 2u_L$. Hence, $C^{HH}$ solves problem (A.5).

*Proof of Subclaim:* 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 (7), that gives weakly higher profits than $C^{HH}$. By supposition, one can choose $C$ such that it gives nonnegative profits. 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. 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 $g_H < \underline{g}_H$ and $C$ satisfies (7), the contract $C$ must generate strictly negative profits after failure, that is, $p(\mathbf{w}_F) g_{i_F} - E(w|F) - k_{i_F} < 0$, and hence must generate strictly positive profits after success, that is, $g_i + p(\mathbf{w}_S) g_{i_S} - E(w|S) - k_{i_S} > 0$. Hence, the new contract $\tilde{C}$ strictly raises profits after failure, and because $C$ satisfies (7), 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 higher than those from $C$.

The main text establishes that no contract satisfying (7) and $(i, i_S, i_F) = (H, L, L)$ gives non-negative profits. Hence, $\tilde{C}$ has the up-or-out feature $(i_S, i_F, w_{FS}, w_{FF}) = (H, L, g_L, 0)$. So Lemma 2 applies, and implies that any contract with $(i, i_S, i_F, w_{FS}, w_{FF}) = (L, H, L, g_L, 0)$ gives strictly lower profits. Thus, the only possibility is that the contract $\tilde{C}$ features $(i, i_S, i_F, w_{FS}, w_{FF}) = (H, H, L, g_L, 0)$. But by Lemma 2, $C^{HH}$ is the unique maximizer of profits within this class, giving a contradiction and completing the proof of the Subclaim.

*Claim B: The equilibrium is unique.* To establish uniqueness, we first show that for $k_H$ large there is no equilibrium with $g_H < g_H^{eq}$. Suppose to the contrary that an equilibrium exists. By Assumption 2, the equilibrium entails a contract $C$ with $(i, i_S, i_F) \neq (L, L, L)$, and this contract generates nonnegative profits. Hence, the conditions of the Subclaim above are satisfied, and so $C^{HH}$ is the most profitable contract with $(i, i_S, i_F) \neq (L, L, L)$. But since $g_H < g_H^{eq}$, $C^{HH}$ has strictly negative profits, and so, a fortiori, the contract $C$ also has strictly negative profits, a contradiction.

Second, we show that for $k_H$ large there is no equilibrium with $g_H > g_H^{eq}$. Suppose to the contrary that there exists such an equilibrium. The contract $C^{HH}$ delivers strictly positive profits. Hence, the reservation utility $\underline{U}$ must exceed $U\left(C^{HH}\right)$. Moreover, the contract $C^{HH}$ must use a second-period bonus $\Delta_S = w_{SS} - w_{SF}$ of at least $\underline{\Delta}(g_H)$, which by Lemma 1 grows arbitrarily large as $k_H \to \infty$. Hence, $U\left(C^{HH}\right) \to \infty$ as $k_H \to \infty$. The utility delivered by a contract that assigns the employee to task $L$ with certainty is bounded above by $2u_L$. Thus, 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 $g_H \to \infty$, the expected task $H$ output of any contract that assigns the employee to task $H$ in some node remains bounded away from zero: this is true if $(i_S, i_F) = (L, H)$, $(H, L)$, $(H, H)$, 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 zero, so that $\zeta_H(y_H, k_H)$ remains bounded above.

Consequently, for $k_H$ sufficiently large the only possible equilibrium return is $g_H^{eq}$. By the Subclaim above, $C^{HH}$ is the unique contract satisfying (7) and $(i, i_S, i_F) \neq (L, L, L)$ that breaks even at $g_H^{eq}$. Hence, $C^{HH}$ and $C^{LL}$ are the only possible equilibrium contracts. The number of employees receiving each contract is uniquely determined by the condition that $g_H^{eq}$ is consistent with aggregate task $H$ activity. Hence, the equilibrium is unique. QED

*Proof of Proposition 2:* Most of the details are in the main text. Here, we verify that returns and hence contracts are state-independent and that $\lambda_t$ 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 $\lambda_t$ remains between zero (one cannot hire a negative number of new employees) and one (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 $\lambda_t$ remains in the interval $\left[\underline{\lambda}, \bar{\lambda}\right]$.[45] Consider what happens as the shock size shrinks, that is, $\zeta_H^G$ and $\zeta_H^B$ approach some common $\bar{\zeta}_H$. Let $\bar{y}_H$ be the output level associated with $\bar{\zeta}_H$ and the payoff $g_H$, that is, $\bar{\zeta}_H(\bar{y}_H, k_H) = \frac{g_H}{k_H}$. Then $y_H^B$ and $y_H^G$ both approach $\bar{y}_H$ and $\underline{\lambda}$ and $\bar{\lambda}$ both approach $\frac{\bar{y}_H}{p(1+p_S)}$. Thus, 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 $\lambda_t$ converges, simply note that iteration of the hiring equation (14) gives

$$\lambda_t = (-p_S)^t \lambda_0 + \frac{1}{p} \sum_{s=0}^{t-1} (-p_S)^s y_H^{\omega_{t-s}}, \tag{A.6}$$

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 task $H$ converges to $\lambda^\omega$. QED

*Proof of Proposition 3:* By substituting the IC constraint (16) and recalling the definition of surplus (3), the profit maximization (15) can be rewritten as

$$\max_{p^{\omega},w_{SS}^{\omega\psi},w_{SF}^{\omega\psi}} p^{\omega}\left(g_{H}^{\omega}-\gamma'\left(p^{\omega}\right)+\sum_{\psi=G,B}\mu^{\omega\psi}S\left(p\left(\mathbf{w}_{S}^{\omega\psi}\right);g_{H}^{\psi}\right)-\bar{u}_{L}^{\omega}\right)-k_{H}.$$

As in the proof of Lemma 3, define

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

Substituting this definition and the second-period IC constraint into IC constraint (16) yields

$$\gamma'\left(p^{\omega}\right)=\sum_{\psi=G,B}\mu^{\omega\psi}\left[w_{SF}^{\omega\psi}+u\left(p\left(\mathbf{w}_{S}^{\omega\psi}\right)\right)\right]-\bar{u}_{L}^{\omega}.$$

Hence, (given the employee limited-liability constraint) the maximization problem (15) subject to (16) is equivalent to

$$\max_{p^{\omega},p_{S}^{\omega G},p_{S}^{\omega B}} p^{\omega}\left(g_{H}^{\omega}-\gamma'\left(p^{\omega}\right)+\sum_{\psi=G,B}\mu^{\omega\psi}S_{H}\left(p_{S}^{\omega\psi};g_{H}^{\psi}\right)-\bar{u}_{L}^{\omega}\right)-k_{H} \tag{A.7}$$

subject to

$$\gamma'\left(p^{\omega}\right)\geq\sum_{\psi=G,B}\mu^{\omega\psi}u\left(p_{S}^{\omega\psi}\right)-\bar{u}_{L}^{\omega}. \tag{A.8}$$

Given this equivalency, we write contracts entirely in terms of their associated effort levels, $\left(p^{\omega},p_{S}^{\omega\psi}\right)$.

*Step 1: We show*

$$\frac{k_{H}}{p^{\omega}u'\left(p^{\omega}\right)}-1+\frac{S_{H}'\left(p_{S}^{\omega\psi};g_{H}^{\psi}\right)}{u'\left(p_{S}^{\omega\psi}\right)}=0. \tag{A.9}$$

Suppose first that $p_{S}^{\omega\psi}\neq p^{*}\left(g_{H}^{\psi}\right)$ for some $\psi=G,B$. Then (A.8) must hold with equality in state $\omega$. Differentiating the profit expression (A.7) 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

$$\frac{\partial p^{\omega}}{\partial p_{S}^{\omega\psi}}\left(\frac{k_{H}}{p^{\omega}}-p^{\omega}\gamma''\left(p^{\omega}\right)\right)+p^{\omega}\mu^{\omega\psi}S_{H}'\left(p_{S}^{\omega\psi};g_{H}^{\psi}\right)=0.$$

From (A.8) 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.9).

If instead $p_S^{\omega\psi} = p^*\left(g_H^\psi\right)$ for both $\psi = G, B$, then the Lagrange multiplier on constraint (A.8) is zero, and differentiation of (A.7) with respect to $p^\omega$ 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.9) since $S_H'\left(p^*\left(g_H^\psi\right); g_H^\psi\right) = 0$.

*Step 2: $g_H^G > g_H^B$, that is, part (2) 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.9),

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

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 (A.8) 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 (A.8) 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 (A.8) 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

$S_H\left(\cdot; g_H^{\psi}\right) - u(\cdot)$ 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 (A.8) 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 $\omega, \psi$.*

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.9), 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 (A.8) 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.9) 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 $\psi = B, G$,

$$p_S^{B\psi} \geq p_S^{G\psi}, \tag{A.10}$$

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

$$p_S^{\omega G} \geq p_S^{\omega B}, \tag{A.11}$$

which again using convexity of $u$ implies

$$S_H'\left(p_S^{\omega G}; g_H^G\right) \geq S_H'\left(p_S^{\omega B}; g_H^B\right). \tag{A.12}$$

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 (A.8) and using the fact that it holds at equality for $\omega = G$,

$$\sum_{\psi=G,B} \mu^{G\psi}\left(S_H\left(p_S^{G\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right) < \sum_{\psi=G,B} \mu^{B\psi}\left(S_H\left(p_S^{B\psi}; g_H^{\psi}\right) - u\left(p_S^{G\psi}\right)\right). \tag{A.13}$$

The left-hand side 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.11) and (A.12) 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.13) 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.10) and completing the proof.

*Step 4: Pay for luck, that is, part (3).*

Given $g_H^G > g_H^B$, it follows by the same argument as used repeatedly above that (A.9) implies $p_S^{\omega G} > p_S^{\omega B}$. The pay for luck implication is then immediate from the second-period IC constraint.

*Step 5: Completing part (1) by establishing higher bonuses.*

Given $p^G \leq p^B$, it follows by the same argument as used repeatedly above that (A.9) 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 the second-period IC constraint. If instead $p^G = p^B$, then given $\bar{u}_L^G > \bar{u}_L^B$, $p_S^{G\psi} \geq p_S^{B\psi}$, and $p_S^{\omega G} > p_S^{\omega B}$ (from Step 4), the first-period IC constraint (16) implies the result. QED

# Notes

[1]More generally, a large empirical literature argues that different jobs pay otherwise identical employees different amounts; see, for example, Krueger and Summers (1988) and Abowd, Kramarz, and Margolis (1999).

[2]In common with high pay levels, these conditions have also attracted considerable public interest. See, for example, The Cult of Overwork, James Surowiecki, *The New Yorker*, January 27, 2014.

[3]For evidence on bonuses, see, for example, Bell and Van Reenen (2014). Consistent with up-or-out career structures, Hong and Kubik (2003) show that it is much more common for security analysts to move from a high-paying, more prestigious brokerage firms to a lower-paying, less prestigious one than the other way around.

[4]Further, although more anecdotal, evidence against high pay being a compensating differential for onerous work conditions is that people who obtain a job with an investment bank typically act as if they have won a lottery, a reaction that our model explains but that is inconsistent with the compensating differential explanation. (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.) Moreover, Philippon and Reshef (2012) control for hours worked and still find excess pay in the financial sector.

[5]See Section I for estimates of capital per employee in the financial sector.

[6]This prediction is consistent with data from the U.S. Bureau of Labor Statistics; see the discussion in Section VI.

[7]In particular, the efficiency wage literature attracted substantial criticism for its neglect of optimal—and especially dynamic—contracting, a criticism broadly known as the "bonding critique." Katz (1986) provides a useful review, including a discussion of the bonding critique.

[8]Rogerson's (1985) 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, of 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.

[9]Given moral hazard, the firm's marginal cost of inducing effort for an old worker is $\frac{\partial}{\partial p}\left(p\gamma'\left(p\right)\right)$. Part (ii) of Assumption 1 ensures that this quantity approaches zero as $p \to 0$.

[10]In particular, Kaplan and Rauh (2010) report total money under management in 177 private equity firms in 2004 as \$461.2B (Table 7). Given six partners per firm (p. 1029), this equates to approximately \$430M per partner.

[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.

[13]Formally, $y_H = \mu p$, where $\mu$ is the measure of employees (young and old) on task $H$ in a given period, and $p$ is their average success probability.

[14]See, for example, 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, that is, employees cannot quit an employment contract. The main exceptions are Proposition 3 in Section VI, on procyclical moral hazard, and our discussion of "talent scorned" in Section VII.

[16]That is, $y_H(C) = p1_{i=H} + pp_S 1_{i_S=H} + (1-p) p_S 1_{i_F=H}$, where 1 is the indicator function and $p, p_S$, and $p_F$ are success probabilities in contract $C$ when young, when old after success, and when old after failure, respectively.

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

[18]In contrast, in Biais, Rochet, and Woolley (2014) an increase in the value of output would increase an employee's equilibrium compensation. In Acemoglu and Newman (2002), an increase in output value has no effect on equilibrium compensation, but only because in their model effort choice is binary, and low effort leads to zero output. Consequently, all equilibrium employment contracts induce the higher effort level, regardless of the value of output. Shapiro and Stiglitz (1984) make the same assumption. In contrast, in our setting effort is a continuous variable.

[19]More specifically, the employee's expected utilities after first-period success and failure are, respectively, $p^*(g_H^*) g_H^* - \gamma(p^*(g_H^*)) - k_H + g_L$ and $u_L$. Since $p^*(g_H^*) g_H^* - \gamma(p^*(g_H^*)) - k_H$ is the surplus on task $H$ when effort is $p^*(g_H^*)$, and since surplus across the two tasks at this effort level are equalized, expected utility after success reduces to $g_L + u_L$. So the employee exerts effort $p_L^*$ in the first period, and has expected utility $2u_L$ across the two periods.

[20]For an early statement of this point, see Jensen and Meckling (1976); for a more recent statement in a moral hazard problem close to the one in this paper, see Holmstrom 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).

[21]Our argument here is somewhat informal. For full details, see Claim A in the proof of Proposition 1 in the Appendix.

[22]Note that the second-period IC constraint $\gamma'(p_S) = \Delta_S$ implies $\tilde{w}_{SF} > 0$.

[23]Note that the right-hand side of (12) can be written as $\max_{\tilde{p}} \tilde{p}\Delta_S - \gamma(\tilde{p}) - u_L$.

[24]At first sight, the prediction that senior employees work harder than junior employees may seem inconsistent with anecdotal accounts of the financial sector that emphasize especially long hours by junior employees. However, the prediction relates to effort on a high-productivity task. An earlier draft of this paper included an extension of the model in which employees can also expend effort on a "menial" task that is both low in value and easily monitored by the firm. In this extension, senior employees do not perform the menial task, and junior employees work more hours in total, even though they work less on the high-productivity task.

[25]Note that the solution to (9) subject to $i = H$ is strictly increasing in $g_H$, is negative for $g_H$ sufficiently small, and is positive for $g_H$ sufficiently large.

[26]Formally, a fraction $\lambda$ of young employees are given the $C^{HH}$ contract, where $\lambda$ solves

$$\zeta\left(\lambda\left(1+p_S\left(C^{HH}\right)\right)p\left(C^{HH}\right),k_H\right)=\frac{g_H}{k_H}.$$

As $k_H$ grows large, the solution $\lambda$ to this equation is guaranteed to be in the interval $(0,1)$.

[27]To see this, note that if $w_{SF}=0$, there is no first-period bonus, so pay is maximally backloaded. 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, $\frac{w_{SS}-w_{SF}}{w_{SF}}>1$.

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

[29]DeChesare, Brian, 2012, From analyst monkey to king of the jungle MD: The investment banking hierarchy, `http://www.mergersandinquisitions.com/investment-banking-hierarchy/`

[30]Ibid.

[31]Formally, this amounts to showing that $\lambda_t$ remains between zero (one cannot hire a negative number of new workers) and one (the total population of young workers).

[32]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.

[33]However, the increase in $g_L$ has an indirect effect on equilibrium returns: because employees are more difficult to incentivize, the equilibrium return $g_H$ must rise, as can be seen from the equilibrium profit condition (15) and is formally established in Proposition 3.

[34]Related, Proposition 2 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 job. Part (1) of Proposition 3 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 Holmstrom (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.

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

[36]A distinct explanation for the pay-for-luck result is given by Oyer (2004), who shows that it may be optimal to index worker wages to firm profits over which the worker has no control when firm profits are correlated with worker outside options. In contrast to our result, which builds on the benefits of providing the right incentives to work, Oyer's result is driven purely by the participation constraint of the worker.

[37]Since employees 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.

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

and 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.

[39]Fishman, Steve, 2004, Get richest quickest, *New York Magazine*, November 22.

[40]Makan, Ajay, 2012, Skin in the game is crucial, but how much? *Financial Times*, November 18.

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

[42]Moreover, the employee's best use of his accumulated bonuses is indeed to start an investment fund in this way. This follows because the optimal contract $C^{HH}$ is renegotiation-proof. An earlier draft of this paper contains a formal proof of this property (which emerges as a result rather than an assumption of our model).

[43]At a formal level, this is closely related to Gromb and Martimort's (2007) analysis of experts. Moreover, note that it is straightforward to show that, without additional information about trade success probabilities, the possibility of trade abandonment by itself has no effect on equilibrium contracts and 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).)

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

[45]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

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

## REFERENCES


Abowd, John M., Francis Kramarz, and David N. Margolis, 1999, High wage workers and high wage firms, *Econometrica* 67, 251–333.

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

Baker, George, Gibbs, Michael, and Bengt Holmstrom, 1994, The Wage Policy of a Firm, *Quarterly Journal of Economics* 109, 921–955.

Beaudry, Paul, and John DiNardo, 1991, The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data, *Journal of Political Economy* 99, 665–668.

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

Bell, Brian, and John Van Reenen, 2014, Bankers and their bonuses, *Economic Journal* 124, F1–F21.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz, 2010, Dynamics of the gender gap for young professionals in the financial and corporate sectors, *American Economic Journal: Applied Economics* 2, 228–255.

Bewley, Truman F., 1999, *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, *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.

Biais, Bruno, Jean-Charles Rochet, and Paul Woolley, 2014, The dynamics of innovation and risk, forthcoming, *Review of Financial Studies.*

Bolton, Patrick, Tano Santos, and Jose Scheinkman, 2014, Cream skimming in financial markets, Working paper, Columbia University.

Custodio, Claudia, Miguel A. Ferreira, and Pedro Matos, 2013, Generalists versus specialists: Lifetime work experience and CEO pay, *Journal of Financial Economics* 108, 471–492.

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, 2295–2340.

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

Gali, Jordi, and Thijs van Rens, 2014, The vanishing procyclicality of labor productivity, Working paper, CREI.

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.

Glode, Vincent, Richard C. Green, and Richard Lowery, 2012, Financial expertise as an arms race, *Journal of Finance* 67, 1723–1759.

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

Holmstrom, Bengt, 1979, Moral hazard and observability, *Bell Journal of Economics* 10, 74–91.

Holmstrom, Bengt, and Jean Tirole, 1997, Financial intermediation, loanable funds, and the real sector, *Quarterly Journal of Economics* 112, 663–691.

Hong, Harrison, and Jeffrey D. Kubik, 2003, Analyzing the analysts: Career concerns and biased earnings forecasts, *Journal of Finance* 58, 313–351.

Jensen, Michael C., and William H. Meckling, 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 Joshua Rauh, 2010, Wall Street and Main Street: What contributes to the rise in the highest incomes? *Review of Financial Studies* 23, 1004–1050.

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 Per Stromberg, 2009, Leveraged buyouts and private equity, *Journal of Economic Perspectives* 23, 121–146.

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, *American Economic Review* 71, 606–620.

Michel, Alexandra, 2011, Transcending socialization: A nine-year ethnography of the body's role in organizational control and knowledge workers' transformation, *Administrative Science Quarterly* 156, 325–368.

Murphy, Kevin M., Andrei Shleifer and Robert 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.

Myerson, Roger B., 2012, A model of moral-hazard credit cycles, *Journal of Political Economy* 120, 847–878.

Ohlendorf, Susanne, and Patrick W. Schmitz, 2012, Repeated moral hazard and contracts with memory: The case of risk-neutrality, *International Economic Review* 53, 433–452.

Oyer, Paul, 2004, Why do firms use incentives that have no incentive effects? *Journal of Finance* 59, 1619–1649.

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

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2014, Scale and skill in active management, Working paper, University of Chicago.

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

Philippon Thomas, 2010, Engineers vs. financiers: Should the financial sector be taxed or subsidized? *American Economic Journal: Macro* 2, 158–182.

Philippon Thomas, and Ariell Reshef, 2012, Wages and human capital in the U.S. financial industry: 1909-2006, *Quarterly Journal of Economics* 127, 1551–1609.

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

Rebitzer, James B., and Lowell J. Taylor, 2010, Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets, in Orley Ashenfelter and David Card, eds.: *Handbook of Labor Economics 4A (North Holland).*

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, *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.

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

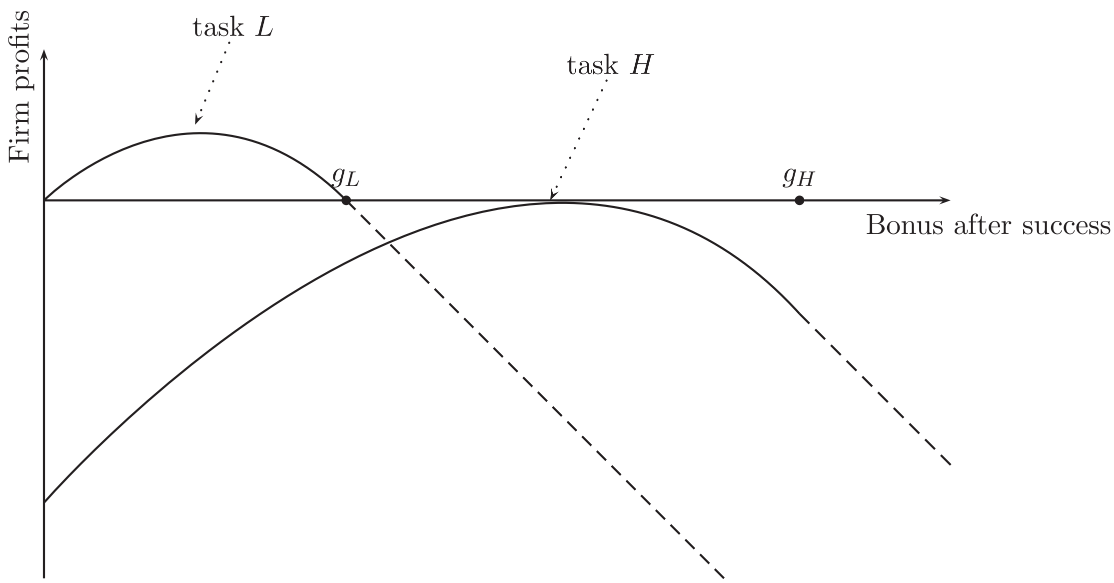


**Figure 1. One-period firm profits.** The graph displays firm profits from using a one-period contract to incentivize the employee in tasks $L$ and $H$, as a function of the success bonus. The graph is drawn for $g_H = \underline{g}_H$, so that maximal profits in task $H$ are exactly zero. For bonuses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

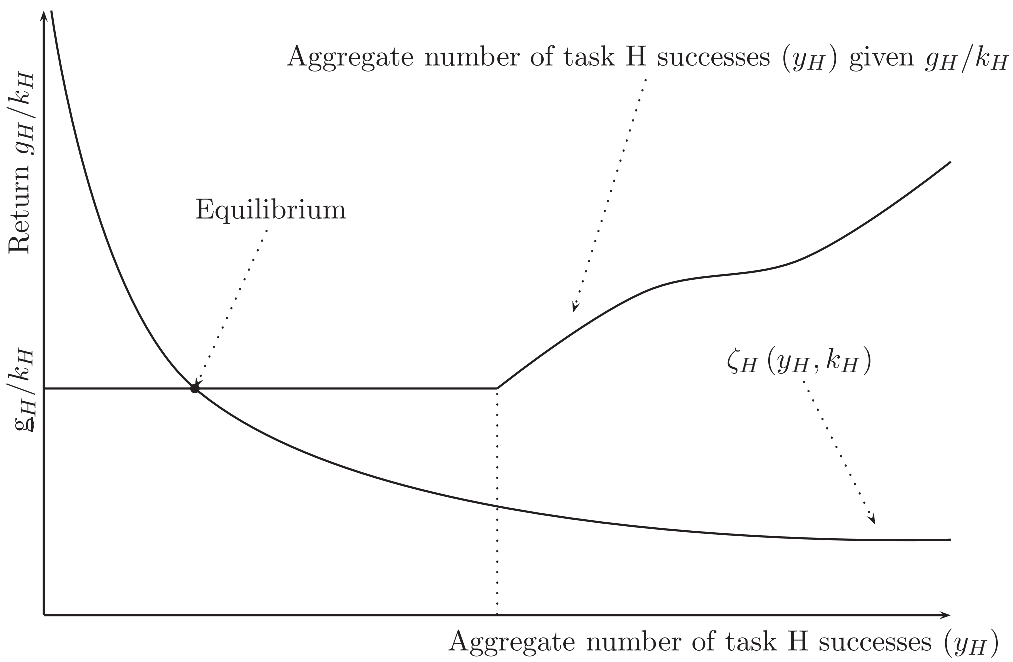


**Figure 2. An overpaying equilibrium for the one-period economy.** The graph shows an equilibrium with overpay. At $\underline{g}_H$, firms make exactly zero profits in task $H$ using the profit-maximizing contract, and this contract delivers strictly higher employee utility than the task $L$ contract (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.