UNIL | Université de Lausanne
Département de finance

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

"The (ir)resistible rise of agency rents"

Prof. Augustin LANDIER

Toulouse School of Economics

Abstract

The rents agents can extract from principals increase with the magnitude of incentive problems, which the literature usually takes as given. We endogenize it, by allowing agents to choose technologies that are more or less opaque and correspondingly prone to agency problems. In our overlapping generations model, agents compete with their predecessors. We study whether the presence of old-timers earning low rents can keep young managers'rent-seeking in check. With dynamic contracts, long horizons help principals incentivize agents. Hence, old agents are imperfect substitutes for young ones. This mutes down competition between generations, especially if compensation deferral is strong. As a result, young managers can opt for more opaque and complex technologies, and therefore larger rents, than their predecessors. Thus, in equilibrium, complexity and rents rise over time.

Friday, October 3, 2014, 10:30-12:00

Room 126, Extranef building at the University of Lausanne

The (ir)resistible rise of agency rents

Bruno Biais, Toulouse School of Economics (CNRS-CRM & PWRI-IDEI)
Augustin Landier, Toulouse School of Economics[1]

May 2013

Abstract

The rents agents can extract from principals increase with the magnitude of incentive problems, which the literature usually takes as given. We endogenize it, by allowing agents to choose technologies that are more or less opaque and correspondingly prone to agency problems. In our overlapping generations model, agents compete with their predecessors. We study whether the presence of old-timers earning low rents can keep young managers' rent-seeking in check. With dynamic contracts, long horizons help principals incentivize agents. Hence, old agents are imperfect substitutes for young ones. This mutes down competition between generations, especially if compensation deferral is strong. As a result, young managers can opt for more opaque and complex technologies, and therefore larger rents, than their predecessors. Thus, in equilibrium, complexity and rents rise over time.

Keywords: Agency rents, moral hazard, finance sector, dynamic contracts, opacity.

JEL codes: D8, G2, G3.

[1]This research was conducted within the Paul Woolley Research Initiative on Capital Market Dysfunctionalities at IDEI, Toulouse. Many thanks for insightful discussions to participants at seminars in Toulouse University, Lugano University, University of Virginia, NYU Abu Dhabi, City University, the 4^{th} Game Theory Society world meeting in Istanbul, the NBER, the 6^{th} Paul Woolley Conference at the LSE, and the FIRS conference in Dubrovnik, especially Philip Bond, Matthieu Bouvard, Max Bruche, Thomas Mariotti, Roger Myerson, Guillaume Plantin, Larry Samuelson, Lucio Sarno, David Sraer and Jean Tirole. Bruno Biais benefitted from the support of the European Research Council, grant 295484 TAP.

1 Introduction

In the presence of moral hazard, principals must leave rents to agents, to incentivize appropriate actions. The more complex and opaque the task delegated to the agent, the more difficult it is to monitor his actions, the larger his rents. Furthermore, while operating the project, the agent acquires knowledge and network connections, which are key to the success of the project, but are also portable. In this context the agent could leave the firm and use these skills and connections for his own profit. To induce the agent to remain in the firm, the principal must leave him rents. The two problems are compounded if the complexity of the task delegated to the agent exacerbates the consequence that would arise if the agent were to abscond.[2]

Agency problems due to unobservable actions have been developed, notably, by Holmstrom (1979), Grossman and Hart (1983), and Holmstrom and Tirole (1997). The agency problem arising when the agent can leave the firm is related to the limited commitment models of Thomas and Worall (1988) and Kocherlakota (1998). It is also related to the ex–post unobservability problem analyzed, notably, by Townsend (1979), Diamond (1984), and Bolton and Scharfstein (1990).

In these analyses, the severity of the agency problem is taken as given. Yet, in practice, agents can take actions that will affect the magnitude of these problems. For example, they can opt for activities, sectors or products that are extremely complex or opaque. They can even devote resources to increasing such complexity and opacity. Agents can also opt for activities relying primarily on network connections or specific knowledge, which they could carry with them if absconding.[3]

The first innovation of the present paper is to consider a setting in which, when agents are born, they choose the characteristics of the task they propose to conduct for principals. Correspondingly, they choose the magnitude of the incentive problem to which principals will be confronted. Thus, the agents set the potential rents they could obtain if hired.

The agents' desire to capture rents, however, could be kept in check by market forces and com-

[2] A particularly perverse instance of this problem arose in the case of AIG. As was mentioned in the New York Times issue of March 16, 2009: "A.I.G. employees concocted complex... If they leave... they might simply turn around and trade against A.I.G.'s book. Why not? They know how bad it is. They built it."

[3] For example, Hochberg, Ljungqvist and Lu (2007) show empirically the importance of newtworks in venture capital, while Oyer (2008) notes his results are consistent with finance sector managers acquiring skills and developing networks of connections shortly after taking jobs on Wall Street.

petition among managers. If each principal could run an auction with several, otherwise identical, managers, he could select the agent with the smallest incentive problem, and hence the smallest rent. Anticipating this, managers would have to opt for activities with very limited agency problems, in order to be hired. We show, however, the existence of natural forces limiting competition between agents, thus creating the scope for rent capture. And in this context we endogenize the dynamics of agency problems.

We consider an OLG model. At each period a new principal and a new manager are born.[4] Each lives for two periods. At the beginning of his life, the young manager chooses (at a cost) the parameter b specifying the magnitude of the agency problem. The larger b, the larger the incentive rents which must be left to the agent. The young principal meets the young agent, observes his b, and decides whether to hire him or not. When making this choice, the principal bears in mind that she could instead hire the agent born at the previous generation, which would be particularly attractive if that agent had a low b. Thus, if the young and old agents were otherwise perfectly substitutable, competition between generations would deter young agents from choosing higher bs than their predecessors. Consequently, opacity, complexity and rents would remain low. Incentive problems, however, endogenously induce imperfect substitutability between successive generations. The intuition is the following: To reduce rents, principals defer compensation (as in Becker and Stigler, 1974, and Rogerson, 1985). This makes it relatively unattractive for a young principal to hire an old agent. First, to poach the old agent, the young principal would have to pay him the rent he has been promised by his current employer. Second, old agents have a shorter horizon than young ones. This reduces the ability of the young principal to defer these agents' compensation to reduce their rents. Thus, other things equal, it is cheaper to incentivize young agents than old ones. Because old agents are imperfect substitutes for young ones, the latter can afford to choose technologies with greater agency problems than their predecessors, and still be hired. This gives rise to an upward trend in complexity, opacity, incentive problems and rents.

Our analysis also makes a theoretical point which, to be the best of our knowledge, had not been made before: While, at an individual level, long–term contracting and backloading of compensation is beneficial, at an aggregate level, it can generate inefficiencies in equilibrium. This is because,

[4] In Section 6, we show that our analysis extends to the case where, each period, N managers and N investors are born.

when different managers have different horizons, they can have different long–term contracts and therefore be imperfect substitutes, which mutes down competition in the labor market. In that sense, backloading of compensation in individually optimal dynamic contracts, exerts negative externalities on other principals in the labor market.

Because our analysis uncovers that equilibrium can be inefficient, it calls for public policy intervention. We consider the case of a regulator, who can set transparency and simplicity standards, monitor agents, and punish those who don't comply. We focus on proportional monitoring costs. That is, we assume costs are linear in the expected number of agents that are monitored –and thus don't include a fixed component. We show that, even in the absence of fixed monitoring costs, it can be optimal to abstain from monitoring during several periods and then engage in intense monitoring for one period. The intuition is the following: To induce agents to comply with transparency and simplicity standards, it is necessary to choose a large monitoring probability. Small monitoring probabilities have no effect on agents' choices between compliance and non compliance. Therefore monitoring, when it occurs, must be on a large scale. On the other hand, intense monitoring during one period compels all young agents in that period to opt for low bs. This resets agency problems at a low level. While lack of monitoring after that period leads to a decline in standards, this decline is only gradual. It is therefore optimal to wait until standards have declined enough before incurring again the large cost of monitoring.

Empirical implications: While the economic mechanims we analyze can be at play in various settings, our assumptions are particularly relevant for the financial sector. First, innovative financial activities are often complex and difficult to understand for outside investors. For example, managers in high–frequency–trading or private equity understand these activities much better than outside investors do. Second, finance faces much less hard–wired technological constraints than manufacturing. In other words there is more plasticity or flexibility in financial activities. This opens up the scope for creativity. Third, accidents caused by malfunctioning financial innovations are less severe than those caused by innovations in other sectors, e.g., medecine or public transportations. Hence, innovation is less regulated and monitored in finance. Again, this opens up the scope for creativity. Our theoretical analysis shows how finance sector managers can take advantage of these characteristics of the sector to earn rents.

Indeed, the implications of our theory are in line with empirical findings about the financial sector. Our model generates a simultaneaous increase in rents and complexity, which is consistent with the findings of Philippon and Resheff (2008). They find that, while in 1980, managers' earnings were similar in the finance sector and other sectors, in 2006, finance managers earned 70% more than comparable managers in other sectors.[5] Philippon and Resheff (2008) also document that, during this period, the complexity of the tasks and occupations of finance sector managers grew very significantly. Their analysis emphasizes the role played by deregulation in this context. Our theoretical result that market forces may not prevent the rise of rents and opacity, while regulation could stem it, is in line with their findings. While Philippon and Resheff (2008) document the growth in the complexity of the tasks of finance managers, Greenwood and Sharfstein (2012) document an increase in the complexity and opacity of the products and techniques in that sector. They find that as the finance sector was growing, the share of private equity and hedge funds increased relative to the share of standard mutual fund management. This led to an increase in overall complexity and opacity, since hedge funds and private equity funds are more opaque and complex than standard mutual funds. Another manifestation of the increase in complexity and opacity is the growth, documented by Greenwood and Sharfstein (2012), of the shadow banking system. As noted by Greenwood and Sharfstein (2012) shadow banking lengthens the credit intermediation process. This, in turn, increases opacity. Anecdotal evidence also points in the same direction. For example, Greg Smith, an executive director at Goldman Sachs, decided to resign after 12 years in that institution. In the article he wrote on that topic in *The New-York Times* on March 14, 2012, he undescored the rise of complexity, i.e., the increasing tendency to "pitch lucrative and complicated products to clients even if they are not the simplest investments or the ones most directly aligned with the client's goals."[6]

Another implication of our theoretical analysis is that the increase in rents, opacity and complexity

[5] Additional evidence about the finance premium is offered by Goldin and Katz (2008), who find that, in 2005, Harvard graduates working in finance earned 95% than those working in other sectors. This is also consistent with anecdotal evidence, e.g., an article entitled "Bank staff costs take bigger share of pot" (Financial Times, June 5, 2012) stated that, in a survey of 13 global financial institutions in 2011, total bankers' compensation was 81% of the sum of dividends, retained earnings and bankers' pay.

[6] See http://www.nytimes.com/2012/03/14/opinion/why-i-am-leaving-goldman-sachs.html?_r=1&src=me&ref=general

spurred by an initial deregulation shock is not instantaneous. Rather it will be sustained and delayed. Hence our theory implies that sustained increases in rents and opacity will take place at points in time where there is no current changes in exogenous variables.

Yet another implication of our analysis relates the increase in rents to the search for yields. In general, the rise of opacity and rents is limited by the constraint that agents must leave enough return to the principals to convince them to delegate the management of their wealth rather than self–invest it. When the safe return and the return on indexing are low, so is the return on self–investment. This increases the ability of agents to increase opacity to extract rents.

Finally, our analysis implies that experienced managers and junior managers are imperfect substitutes. This should show up in hiring and compensation data. For example, when new slots open up, experienced managers are imperfect substitutes for junior ones. This is consistent with the findings of Oyer (2008) that there is a strong causal effect of initial placement in investment banking on later career. Our theory also predicts that imperfect substitutability, and its consequences, should be stronger when incentive problems are more severe and when compensation is more backloaded.

Literature: Our analysis of the welfare costs induced by agents' opting for socially unproductive rent–seeking is in line with Baumol (1990) and Murphy, Shleifer and Vishny (1991). Both in their analysis and ours, rent–seeking agents impose costs upon the others. But, in Baumol (1990) and Murphy, Shleifer and Vishny (1991) these costs are directly induced by the actions of the rent–seeker, e.g., warfare, litigation or predatory trading. In contrast, in our analysis, the initial choice of the agent (the level of opacity and complexity of the activity) has an indirect impact on the principal, via the agency rent it induces. Furthermore, the equilibrium rise in rents, and the mechanism by which current complexity induces further increases in complexity, is a novel contibution of the present paper.

Our work is also related to Myerson (2012). In both papers overlapping generations of managers are hired by investors, and agency problems (moral–hazard or cash–diversion) lead to compensation deferral. While, both papers show how the dynamic provision of incentives can generate cycles, they focus on different objects. Myerson (2012) shows how long–term contracting designed to cope with moral hazard can generate gradual growth followed by steep recessions. In contrast we show how long–term contracting creates gives rise to a simultaneous increase in rents and complexity and opacity.

Our analysis is related to that of Axelson and Bond (2012) who also analyze an equilibrium model, with overlapping generations, and dynamic contracts designed to mitigate moral hazard. Specific features of our model, which differentiate it from that of Axelson and Bond (2012), are that, in our framework, i) agents initially choose the magnitude of their agency problems, which leads to the rise of rents, and ii) the contract offered by a given principal to his agent has external effects on the actions of other principals and agents, i.e., there are contractual externalities.

Our point that agents in the finance industry choose complex products and techniques to increase the rents they extract from principals, echoes the point made by Carlin (2009) that competing financial institutions design complex products to increase their market power. A major difference is that our analysis hinges on agency problems, which can arise even with large rational investors, while Carlin (2009) focuses on retail investors and abstracts from agency issues.

Our analysis is also related to that of Bolton, Santos and Scheinkman (2013). In their paper also, opportunistic initial occupational decisions lead to rents, opacity and inefficiencies. Also, both papers uncover externalities associated with the development of opaque activities and markets. The economic mechanisms at work in the two papers are different, however. In Bolton et al (2013), when many agents choose to become dealers in the opaque OTC market, this worsens *adverse selection* in the other (transparent) market, which increases the bargaining power, and hence the rents, of the OTC dealers. In contrast, in our analysis, when agents choose opaque investment techniques, this worsens *moral hazard*, and increases rents, for the following generations.

The next section presents our model. Section III analyzes optimal compensation contracts. Section IV characterizes the dynamics of equilibrium rents. Section V discusses policy intervention. Section VI presents extensions and discusses robustness. Section VII concludes.

2 Model

Investors and managers: Consider an overlapping generations model, where, each period, one investor and one manager are born. Both are risk neutral, have limited liability, live for two periods and have discount factor $\beta \in (0, 1)$. Because we consider an overlapping generations model, successive generations of managers coexist in the market at a given point in time. This creates the scope for interactions, and in particular competition, between generations.

Each investor is initially endowed with one unit of investment good. She can invest it in a default technology, which she can operate herself and which returns 1 unit of consumption good per period during two periods. Alternatively, she can delegate the management of her capital to the agent. When entrusted with one unit of capital, the agent can generate return equal to $R > 1$ units of consumption good per period during two periods. For simplicity the choice between self–investment and delegated investment is irreversible.

Managers have zero initial endowment. At the beginning of his life, each young manager must choose among a range of investment techniques indexed by $b \in [b_{\min}, 1)$, with $b_{\min} > 0$.[7] Each technique corresponds to a specific type of skills, knowhow and human capital. The cost of acquiring skills b is equal to cb, with $c \geq 0$. b's are ordered according to the magnitude of the agency problem they involve. Low values of b correspond to simple investment techniques, carried in transparent anonymous markets, for which agency problems are limited. Higher values of b correspond to complex investment techniques, carried in opaque OTC markets, involving networking and repeated bilateral relations. c is the cost of designing complex products and strategies. Importantly, the choice of b is irreversible. The idea is that managers acquire skills, human capital, relations and knowledge of investment techniques at an early stage in their career. Then, they use this informational capital.[8] Finally, for simplicity, we assume R does not depend on b. In Section 6 we show that our qualitative results are robust to letting R increase in b.

Sequence of play: Within each period $t \geq 1$, the timing of actions is the following:

- Stage 1: The young manager chooses $b_t \in [b_{\min}, 1)$.
- Stage 2: The young investor meets the young manager, observes his type (b_t), and decides whether to make him a take–it–or–leave–it offer or reject him. The assumption that investors have all the bargaining power is for simplicity. In Section 6, we present an extension of the

[7] Assuming $b_{\min} > 0$ makes Lemma 3 immediate, but this assumption is not necessary for our analysis.

[8] See Oyer (2008) for empirical evidence on long term effects of initial career paths in the financial sector.

model where the manager have some (but not all of) the bargaining power. It yields the same qualitative insights as the simpler present case.

- Stage 3: If the investor decides to reject the young manager or the latter rejects the offer, then the investor decides whether to self–invest or to approach the old manager (born at $t-1$). In the latter case, the investor meets the old manager, observes his type (b_{t-1}) and current employment status, and can make him a take–it–or–leave–it offer.[9] Also, at this stage, if the previous investor is not currently employing an agent, she can approach a manager and make him a take–it–or–leave–it offer. Similarly, if the previous manager did not get hired, he can approach the young investor and apply for a job.

- Stage 4: Investment takes place.

- Stage 5: Each manager must decide whether to remain in the firm and obtain the wage promised by the manager,[10] or abscond. If the manager absconds, he does not get paid, but obtains $\delta b_t R$, with $\delta \leq 1$. This corresponds to the profits he can earn by working on his own, using the valuable contacts and knowhow obtained while on the job. In that case, the return obtained by the investor is lowered to $R(1-b_t)$.

- Stage 6: Consumption takes place.

For simplicity (and to limit the number of cases in the analysis below), we assume

$$R(1-\delta b_{\min}) > 1,$$

i.e., $b_{\min}$ is sufficiently low that, in a one–period context, self-investment by the principal is dominated by the delegation of investment to an agent with the lowest possible b.

Discussion of the agency problem: One motivation for the agency problem we consider is the following. While managing the investment on behalf of the principal, the agent acquires specific knowledge about this technique and develops a network of contacts in this business. These techniques

[9] The assumption that agents' types are observed by investors only when they meet them is also made for simplicity. As will be explained in Section 4, it simplifies the structure of beliefs and thus facilitates the analysis of the game.

[10] For simplicity also we assume the investor can commit to the contract offered.

and contacts are useful to make the investment profitable when undertaken for the principal. But they can also be useful for the agent when absconding, and working on his own.

This stylized agency problem, where the agent can abscond or not, in line with the limited commitment models of Thomas and Worall (1988) and Kocherlakota (1998). In these models the agent cannot commit not to leave the principal and earn his autarky payoff. The equivalent of this payoff in our model is $b_t R$. The agency problem we consider is also in line with the cash–diversion model of Townsend (1979), Diamond (1984), Bolton and Scharfstein (1990), and, more recently, DeMarzo and Fishman (2007) and Biais, Mariotti, Plantin and Rochet (2007).

Instead of the simple agency problem we consider, we could consider a variant of the Holmstrom and Tirole (1997) moral hazard model. At the beginning of each period, the agent can exert unobservable effort or not. The effort can involve spending time and energy on risk control and monitoring, or screening good investment projects. Its cost is $B_t R$. Equivalently, $B_t R$ can be interpreted, as in Holmstrom and Tirole (1997), as the private benefit from shirking. When the agent exerts effort, the project generates cash flow R for sure. When the agent fails to exert effort, cash flows can be R with probability $1 - \Delta$, or 0 with probability Δ. If the project fails (and generates 0 cash flow) at the first period, then it is bound to fail for sure at the second period. Relabeling $\delta b_t = \frac{B_t}{\Delta}$, this setting is exactly equivalent to the agency problem described above. As will be made clear in the next sections, the incentive compatibility conditions and the equilibrium actions are exactly the same in the two environments.

Finally, two comments are in order to complete the discussion of the agency problem: First, when managers abscond, they are busy working on their own, to get $b_t R$, thus they are not available to be hired by another investor. Second, while we assume $\delta \leq 1$, all our major qualitative results are upheld if $\delta = 1$, thus, our analysis holds both if absconding is inefficient and if it is ex–post efficient.

First best: To conclude this section, we describe the first best benchmark, i.e., the situation where the benevolent social planner can observe and decide what investors and managers do. Since utilities are linear, there is a unique Pareto optimum regarding real decisions, and the points on the Pareto frontier differ only in terms of purely redistributive transfers between investors and managers. Since $R > 1$, the optimum is reached when investors delegate the management of their capital. Since

$c \geq 0$ and R is indedendant of b, the socially efficient level of b is 0. The corresponding utilitarian welfare for each generation is

$$(1+\beta)R. \tag{1}$$

3 Moral hazard and managerial compensation

In this section we analyze the compensation contract offered at time t by an investor hiring a young manager with given type b_t. We first analyze the contract offered by the investor hiring the manager for two periods. Next, we show that this dominates hiring the agent for one period only, and then another agent for the second period (whatever the type b of the agent at the next period).

The compensation contract offered at time t by the investor is a pair of wages, w_t^t and w_{t+1}^t, paid to the manager if he remains in the firm at the end of periods t and $t+1$. In the context we consider, the optimal contract is such that the manager takes the efficient action and does not abscond. The corresponding incentive compatibility conditions are

$$w_{t+1}^t \geq \delta b_t R \tag{2}$$

and

$$w_t^t + \beta w_{t+1}^t \geq \delta b_t R. \tag{3}$$

The optimal contract maximizes the investor's net return subject to the incentive compatibility and limited liability constraints. It is spelled out in our first lemma.

Lemma 1: *At time t, for a given choice of b_t, and when the manager is hired for two periods, (3) binds and there is a continuum of optimal contracts, indexed by $\lambda_t \in [1, \frac{1}{\beta})$,*

$$\{w_t^t, w_{t+1}^t\} = \{(1-\beta\lambda_t)\delta b_t R, \lambda_t \delta b_t R\}.$$

Remark 1: Complexity and Rents. Lemma 1 implies that the present value of the fund manager's earnings is

$$w_t^t + \beta w_{t+1}^t = \delta b_t R, \tag{4}$$

while the present value of the investor's returns is

$$R(1+\beta) - \delta b_t R. \tag{5}$$

Thus, the investor's net return is decreasing in b_t, while the manager's earning is increasing. As the complexity and opacity of the investment technology increase, the investor must give the manager a larger and larger fraction of the total revenue of the investment.

Remark 2: Backloading and inefficiency. The larger λ_t, the greater the fraction of the compensation that is deferred to the second period. For example, when $\lambda_t = \frac{1}{\beta}$, managerial compensation is entirely deferred to $t+1$. Both the principal and the agent are indifferent between all values of $\lambda_t \in [1, \frac{1}{\beta})$. In the next section, however, we show that the contractual choices of one generation have external effects on the next ones. Thus, different choices of λ_t lead to different inefficiencies.

Remark 3: Compensation and seniority. Lemma 1 implies that $w_t^t < w_{t+1}^t$, since

$$\frac{1}{1+\beta} < 1 \leq \lambda_t.$$

Thus, for a given generation, compensation rises with seniority, in line with stylized facts. In the next section, however, we show that the structure of the equilibrium we analyze can imply that, when comparing generations, senior managers from the previous generations earn less than junior managers from the current generation.

Remark 4: Equivalence with an unobservable effort model ï¿½ la Holmstrom and Tirole (1997). What if the agent could not abscond, but had to exert unobservable effort to raise the probability of success, as explained in the discussion of the agency problem in the previous section? In that case, the incentive compatibility condition at time $t+1$ would be:

$$w_{t+1}^t \geq (1-\Delta) w_{t+1}^t + B_t R,$$

that is

$$w_{t+1}^t \geq \frac{B_t}{\Delta} R = \delta b_t R, \tag{6}$$

as in (2). Similarly, the incentive compatibility condition at time t would be:

$$w_t^t + \beta w_{t+1}^t \geq (1-\Delta)(w_t^t + \beta w_{t+1}^t) + B_t R,$$

that is

$$w_t^t + \beta w_{t+1}^t \geq \frac{B_t}{\Delta} R = \delta b_t R, \tag{7}$$

as in (3).

Lemma 1 characterizes the optimal contracts when the manager is hired for two periods. What if, instead, the investor hires the manager for one period only, and then hires another manager at the second period. In this case, the incentive compatibility condition for the first period is

$$w_t^t \geq \delta b_t R.$$

At the second period, the investor could hire the young manager, with type b_{t+1}. This would require matching the wage offered by the young investor at time $t+1$ and also providing the young manager with the incentives not to abscond. Thus the investor would have to pay that agent at least

$$\delta b_{t+1} R.$$

Overall, the present value of the returns obtained by the investor would be no larger than

$$R(1+\beta) - \delta(b_t + \beta b_{t+1})R. \tag{8}$$

(5) is larger than (8), whatever teh value of $b_{t+1} \in [b_{\min}, 1)$. This reflects that, with repeated moral hazard, long term contracts help the principal reduce agency rents. In line with previous analyses of dynamic contracting (Becker and Stigler, 1974, and Rogerson, 1985), the future promised rents (here w_{t+1}) is also useful to incentivize current effort (and thus helps limiting w_t).[11] This cannot be achieved when the principal hires a sequence of agents compensated with short–term contracts. Hence, it is cheaper for the investor to hire the manager for two periods. We state this result in the following lemma:

Lemma 2: *Hiring the manager for two periods dominates hiring him for one period and then hiring another manager.*

[11] See also Biais, Mariotti, Plantin and Rochet (2007) or DeMarzo and Fishman (2007).

4 The dynamics of equilibrium rents

We now turn to the dynamics of the b's chosen by managers before they are matched with investors.

Given an initial b_0, an equilibrium is a sequence $\mathcal{E} = \{b_t^*, w_t^{t^*}, w_{t+1}^{t^*}\}_{t\geq 1}$, satisfying the following conditions:

- **Optimization:** At each time t, the young manager chooses b_t^*, to maximize his gains and the investor makes an optimal hiring decision.
- **Rational expectations:** Investors and managers at time t have rational expectations about equilibrium dynamics at all times $\tau > t$, and when each market participant expects the others to follow the equilibrium, he/she himself/herself finds it optimal to also play according to $\mathcal{E}$.
- **Labor market clearing:** At each generation, the young manager is hired by the young investor.

From Lemma 1, the young manager's gains, if he opts for b_t and is employed, are

$$(\delta R - c)b_t. \tag{9}$$

The young manager chooses b_t to maximize (9), subject to the constraint that the investor is willing to hire him. Equilibrium requires that, when the agent anticipates all the others played as specified by $\mathcal{E}$, he will choose $b_t = b_t^*$.

If $\delta R < c$, then the agent opts for $b_t = 0$. To raise the possibility that the agent would choose $b_t > 0$, we assume hereafter the cost c is not too large, i.e.,

$$\delta R > c.$$

To characterize the optimal decision of the investor meeting a young manager with type b_t, we need to compare her payoff if she opts for the equilibrium action (hiring the young manager at Stage 2) to what she gets if she takes an out–of–equilibrium action.

The first out–of–equilibrium option for the investor is self–investment at Stage 3 of period t. She does not choose that option if her gains from hiring the young manager

$$R(1+\beta) - \delta b_t R. \tag{10}$$

are at least as large as the gains from self–investment $(1 + \beta)$. This is is the case if

$$b_t \leq b_{\max} \equiv \frac{R-1}{R}\frac{1+\beta}{\delta}. \quad (11)$$

The second out–of–equilibrium option for the investor is to "poach" an old employed agent at Stage 3 of period t and then hire the generation t manager, unemployed at t and available at $t + 1$. She would have to give the former at least as much as his current wage, which she rationally expects to be $w_t^{t-1*} = \delta\lambda_{t-1}^* b_{t-1}^* R$, and would have to incentivize the latter. To do so, she would have to pay him at least $\delta b_t R$ at $t + 1$. On the other hand, the old manager at $t + 1$ cannot ask for more than that, since he could not obtain more from the time $t + 1$ investor. Hence, overall, if she were to opt for that deviation, the time t investor would expect to get

$$R(1+\beta) - \delta(\lambda_{t-1}^* b_{t-1}^* + \beta b_t)R. \quad (12)$$

The third out–of–equilibrium option for the investor is to poach an old employed manager at t and then a young one at $t+1$. When evaluating the gains from this out–of–equilibrium move, the investor rationally expects both the previous and the subsequent generation to follow $\mathcal{E}$. Indeed, the extensive form of the game is such that the next generation's managers and investors make their decisions before observing if their predecessor deviated. Therefore they make their decisions based on the belief that their predecessor sticked to $\mathcal{E}$, and find it optimal to do the same themselves. Thus, when deviating, the generation t investor expects to pay $\delta\lambda_{t-1}^* b_{t-1}^* R$ to the old manager she hires at time t, and $\delta b_{t+1}^* R$ to the young manager she hires at time $t + 1$. Consequently, the deviating investor expects to earn

$$R(1+\beta) - \delta(\lambda_{t-1}^* b_{t-1}^* + \beta b_{t+1}^*)R. \quad (13)$$

Equating (10), evaluated at b_t^*, to the maximum of (12) and (13), as long as the no–self–investment constraint (11) does not bind, the equilibrium choice of the generation t young manager is

$$b_t^* = \lambda_{t-1} b_{t-1}^* + \beta \min[b_t^*, b_{t+1}^*]. \quad (14)$$

(14) and $\lambda_{t-1} \geq 1$ directly imply our next lemma.

Lemma 3: *At time t, if (11) does not bind, then*

$$b_t^* > b_{t-1}^*.$$

Lemma 3 and (14) imply that, as long as the no–self–investment constraint (11) does not bind,

$$b_t^* = \frac{\lambda_{t-1}}{1-\beta} b_{t-1}^*. \tag{15}$$

Thus, we obtain our first proposition.

Proposition 1: *Set the initial manager's type to $b_0 \geq b_{\min}$. There exists a continuum of equilibria, indexed by $\lambda \in [1, 1/\beta]$, such that, for all $t > 0$, b_t^* is strictly increasing in t, until it reaches $b_{\max}$ at time $T_{\max}$ and*

$$b_t^* = \min[(\frac{\lambda}{1-\beta})^t b_0, b_{\max}]. \tag{16}$$

The rise of rents: For simplicity, the proposition only presents the equilibria in which λ_t is constant through time. There also exists equilibria in which λ_t varies as time goes by, but they yield the same qualitative insights as those described in Proposition 1. Interestingly, even if λ_t is constant through time, b_t is non–stationary. Indeed, (16) implies that, until $T_{\max} - 1$, b_t grows at rate

$$\rho_t = \frac{b_t^* - b_{t-1}^*}{b_{t-1}^*} = \frac{\lambda + \beta - 1}{1-\beta} > 0. \tag{17}$$

Thus opacity, complexity and rents rise in equilibrium. This is because investors can offer long–term contracts to young managers, not to old ones. This makes the former attractive relative to the latter, other things equal. Hence, at time t, the young manager can afford to choose $b_t > b_{t-1}$ and yet be employed.

Backloading exerts contractual externalities: Note that ρ_t is increasing in λ. While, at each time t, the investor is indifferent between all $\lambda_t \in [1, \frac{1}{\beta})$, his choice affects competition between successive generations of managers. The greater λ_t, the more compensation is backloaded, the lower the competition between generations, the stronger the growth of rents, ρ_t. The lowest possible equilibrium growth rate of rents, arises for $\lambda = 1$, where

$$\rho_t = \frac{\beta}{1-\beta}. \tag{18}$$

Note that, even for $\lambda_t = 1$, $\rho_t > 0$.

Patience and rents: The more patient the investor, the more he suffers from the cost of incentivizing a new agent next period, the less competition between generations there is, the larger is b_t^*.

Equivalence with an unobservable effort model ï¿½ la Holmstrom and Tirole (1997): As explained above, in Remark 4, the incentive compatibility conditions in our simple setting where the agent can abscond with $\delta b_t R$ are the same as in a simple two–period version of Holmstrom and Tirole (1997). Similarly, the gains the principal can obtain on the equilibrium path or in the deviations involving poaching, (12) and (13), are the same in the two settings. Hence, equilibrium actions are also the same in the two settings. Thus, all our analysis and results, and in particular Proposition 1, also obtain in the simple two–period version of Holmstrom and Tirole (1997).

Welfare: The utilitarian welfare for generation t,

$$R(1+\beta) - cb_t.$$

is lower than welfare in the first best, (1). This is because managers incur wasteful costs (cb_t) to come up with complex and opaque technologies, that increase rents but don't create value for society.

Note however that, when b_T reaches $b_{\max}$, utilitarian welfare becomes

$$R(1+\beta) - c\frac{R-1}{R}\frac{1+\beta}{\delta} = (1+\beta)(R - \frac{c}{\delta R}(R-1)),$$

which is greater than welfare under self–investment under our assumption that $R > 1$ and $\delta R > c$. Thus, while managers choose to incur wasteful costs, these don't wipe out all the efficiency gains from investment delegation.

5 Policy

The analysis above shows that the equilibrium choice of b is inefficient. This begs the question whether policy intervention can improve efficiency.

5.1 Monitoring managers

To analyze this issue we assume there is a benevolent regulator, maximizing utilitarian welfare. To model regulatory intervention we introduce an additional stage in the game. Each period t, at stage 0 the regulator announces a cap $\underline{b}_t$, and an inspection probability α_t, to which we assume he can commit. Then, at stage 1, the young agent chooses b_t, and, with probability α_t, is inspected by the regulator. In case of inspection, if $b_t > \underline{b}_t$ the agent is prevented from working. Otherwise, he enters the market and meets the investor, and the game unfolds. In particular, compensation is determined as in Lemma 1. For simplicity, we hereafter focus on the equilibrium where $\lambda_t = 1, \forall t$. Finally, we assume the monitoring technology is linear, i.e., there exists a constant $\gamma > 0$ such that the cost of monitoring with probability α_t during one period is $\alpha_t \gamma$. In this section we compare the performance of three policies:

- Laissez faire, in which the regulator never intervenes.
- Permanent monitoring, in which the regulator sets a constant cap $\underline{b}$, and monitors with constant probability α to ensure that agents always comply with the regulatory cap.
- Periodic monitoring, which operates as follows: during $T-1$ periods (with T finite and strictly larger than 1), laissez faire prevails, i.e., $\alpha = 0$. Then, at the T^{th} period, the regulator intervenes, sets the maximum $\underline{b}$ for this period, and monitors the agent with probability α_T, sufficiently large to ensure compliance. Then a new cycle starts.

5.1.1 Permanent monitoring

We analyze for which values of the permanent monitoring policy $(\underline{b}, \alpha)$ it is an equilibrium for agents to always comply. The expected gain of the agent if he complies is

$$\delta \underline{b} R.$$

When the agent chooses $\hat{b} > \underline{b}$ and is not inspected he is matched with an investor. The latter is willing to hire him at t, rather than poaching an old agent (with type $\underline{b}$) at t and then hiring the generation $t+1$ agent (also with type $\underline{b}$) if

$$R(1+\beta) - \delta \hat{b} R \geq R(1+\beta) - (1+\beta)\delta \underline{b} R.$$

Binding the constraint yields

$$\hat{b} = (1+\beta)\underline{b}. \tag{19}$$

The agent prefers to comply if his gain under compliance exceeds is expected gain under non compliance. Binding this constraint we have

$$\underline{b} = (1-\alpha)\hat{b}. \tag{20}$$

Substituting (19) in (20) yields the following lemma:

Lemma 4: *An optimal permanent monitoring policy* $(\underline{b}, \alpha)$, *is such that*

$$\alpha = \frac{\beta}{1+\beta} \text{ and } \underline{b} = b_{\min}.$$

Consistent with intuition, the lemma states that the agent complies if the probability of inspection (α) is sufficiently large. Furthermore, it is optimal to set $\underline{b} = b_{\min}$ because the cost of monitoring is independent of the cap $\underline{b}$. Thus, if the regulator opts for a permanent monitoring policy, it is optimal to minimize the cap.

The steady state utilitarian welfare of each generation under a stationary inspection policy incentivizing permanent compliance is

$$W(1) \equiv R(1+\beta) - \frac{\beta}{1+\beta}\gamma - cb_{\min}. \tag{21}$$

On the other hand, the steady state utilitarian welfare under laissez faire is

$$W(\infty) \equiv R(1+\beta) - cb_{\max}. \tag{22}$$

Comparing $W(1)$ and $W(\infty)$ we obtain the following proposition.

Proposition 2: *Welfare is larger under a stationary policy incentivizing permanent compliance than under laissez faire if*

$$\gamma < \frac{1+\beta}{\beta}c(b_{\max} - b_{\min}). \tag{23}$$

The proposition states that welfare is greater under permanent monitoring than under laissez faire if the inspection cost, γ, is relatively low compared to the cost of opting for complex investment techniques, c, and if $b_{\max}$ is large while $b_{\min}$ is low.

5.1.2 Periodic monitoring

Now turn to the analysis of periodic monitoring with finite periodicity $T > 1$. To evaluate the efficiency of a periodic monitoring policy we consider the average welfare over T generations, which we denote by $W(T)$. Our next proposition states that for intermediary values of the monitoring costs γ, periodic monitoring dominates laissez–faire and permanent monitoring.

Proposition 3:
There exist thresholds $0 < \gamma_1 < \gamma_2 < \gamma_3 < \gamma_4$ *and* $B > 0$ *such that:*

- *If* $\gamma < \gamma_1$, *constant monitoring dominates laissez–faire and periodic monitoring.*
- *If* $\gamma > \gamma_4$, *laissez-faire dominates constant and periodic monitoring.*
- *If* $\gamma_2 < \gamma < \gamma_3$ *and* $b_{\min} < B$, *periodic interventions, at intervals of two or more periods, are better than laissez–faire and constant monitoring.*

The economic intuition behind Proposition 3 combines two insights already analyzed in the paper. First, new generations cannot deviate too much from elder generations in their choice of b. So there is some form of disciplining effect from elder generations onto new generations. To save on monitoring costs, the regulator might rely for some periods on this disciplining effect. Second, monitoring is only effective if a critical mass of the population is inspected. Similarly to fixed costs, this makes infinitesimal levels of monitoring suboptimal. Putting these two insights together provides the intuition for periodic interventions: $-b_t$ can be thought of as a form of technological capital that depreciates endogenously as successive generations choose higher and higher levels of b. Since effective monitoring is non–infinitesimal, it can be optimal to intervene periodically to reset b to a low level, then let the technological capital depreciate, and intervene again only when b reaches a threshold. This is not unlike the dynamics arising for optimal investment with fixed costs.

5.2 Normative and predictive implications

Our modelling of monitoring by a regulator predicts that it can be optimal to have periods of intense monitoring where the regulator performs a "crack-down" that resets rents at a low level, followed by several periods of lighter monitoring. This implies that monitoring expenses can be cyclical and should be allowed to experience spikes: it is optimal to grant from time to time regulators with abnormally high levels of funding, so that they can hire human capital to perform exceptional amounts of monitoring. In other words, the usual flow cost of supervision is optimally smaller than what is spent in special times. The governance and mandate of regulators should thus allow them to expandes their budget temporarily when a "reset" is needed.

Our result on monitoring cycles also has an implication for empirical research: In our model, regulatory changes are slowly offset by agents who bring back opacity and rents to their initial level. This means that the short-term and long-term effects of regulations should be expected to be different. This potentially matters for empirical studies using regulatory shocks as natural experiments: the short-term impact of a reform should be expected to be higher than its long-term impact. In our model, the speed at which the offsetting of regulation takes place is endogenous. Our model predicts a progressive build-up in rents and complexity following deregulation episodes, in line with empirical results from Philippon and Resheff (2008).

5.3 Restricting compensation deferral

In the equilibrium we have described, the choice of compensation deferral by investors at time t exercises an externality on the welfare of future investors. Indeed, investors born at time t are indifferent among all levels of deferral $\lambda_t \in [1, \frac{1}{\beta})$. However, the welfare all future investors is negatively impacted by a choice of $\lambda_t > 1$, as this increases the rents extracted by managers in future periods. This is also bad for aggregate welfare as choosing high level of b is inefficient. The intuition is that, at time $t+1$, a higher λ_t makes it more expensive for new investors to hire an old-timer and thus allows the time $t+1$ managers to extract more rents, by choosing a higher b_{t+1}. It follows that a regulation that imposes $\lambda_t = 1$ for all t improves utilitarian welfare. We believe our model sheds light on a rarely discussed side-effect of compensation-deferral, namely the fact that it potentially reduces competition between managers by making it more costly to hire already employed managers. Several researchers have advocated clawbacks or vesting of compensation to curb managerial short-termism (see e.g. Edmans et al. (2012)). Our model suggests that one should carefully distinguish between deferred payoffs that are not contingent on the manager still working in the same firm at a future date, and payoffs that vest only if the manager still works at the same fim in the future: the latter type of compensation deferral can potentially harm competition on the market for managers.

6 Extensions

Coming back to the model without regulatory intervention, we now show that our qualitative results hold after relaxing several of our assumptions.

6.1 When R increases with b

In this subsection, we show that our qualitative insights are robust to assuming that delegated investment generates $R(b)$, an increasing function of b. For simplicity, and without affecting the qualitative results below, we normalize c to 0 and δ to 1. The surplus earned by the investor when hiring an agent with type b is:

$$\sigma(b) = (1 + \beta - b)R(b).$$

We assume R is twice differentiable, increasing and concave with respect to b. So

$$\sigma'(b) = (1 + \beta - b)R'(b) - R(b) \text{ and } \sigma''(b) = (1 + \beta - b)R''(b) - 2R'(b).$$

We also, assume, in the same spirit as the Inada condition, that $R'(1) = 0$ and $(1+\beta-b_{\min})R'(b_{\min}) \geq R(b_{\min})$. Thus, σ is concave and there exists a point $b^* \in [b_{\min}, 1)$, such that σ increases before b^* and then decreases.

For investors, the optimal level of b is b^*. The agent will never choose b lower than b^*, since the investor would prefer him to choose b^*, which would lead to a Pareto improvement. Thus, the relevant set of values of b that can arise in equilibrium is $[b^*, 1)$. Note that σ is decreasing over this interval. In this context, redefining $b_{\max}$ as

$$b_{\max} = Sup_{b\in[b_{\min},1)}\{b, \sigma(b) \geq 1 + \beta\},$$

we obtain our next proposition:

Proposition 4:

In equilibrium, b_t increases until $b_{\max}$.

Thus, even if R is increasing in b, we have that b's increase above what's optimal for investors. In fact, that R increases with b should amplifies the growth in bs if anything. After all, when R increases with b, large bs are not pure waste. But, as shown above, in equilibrium bs increase beyond b^*, to increase the rents of the agents, at the expense of the net returns earned by the principal.

6.2 When the agent has some bargaining power

In the above analysis we assumed the principal made take–it–or–leave–it offers to the agent. Suppose instead that, when the young time t agent and investor are matched, with probability ϕ it is the former who makes a take–it–or–leave–it offer, while with probability $1-\phi$ it is the latter. Otherwise, the sequence of play is exactly as described above, in Section 2. As shown below, when ϕ is not too large, the outcome of the game is exactly the same with this alternative extensive form.

Suppose the young principal and the young agent at time t expect the outcome of the game at all other times to be as stated in Proposition 1. What is the rent, W, the young agent can request from the young principal when making his offer? W is limited by the constraint that the principal should prefer to hire the agent at t, rather than poaching an old, time $t-1$, agent. The principal expects that, if she were to do so, she would have to pay to the old agent $\lambda_{t-1}b^*_{t-1}$. And then, at time $t+1$, she would have to hire either the old unemployed time t agent, of the new, time $t+1$, agent. Thus, the take–it–or–leave it offer made by the time t agent must be such that

$$W \leq \lambda_{t-1}b^*_{t-1} + \beta \min[b^*_t, b^*_{t+1}].$$

Binding this constraint, W is exactly equal to the rent in Proposition 1. In this context, the young agent chooses b_t to solve

$$\max_b \phi WR + (1-\phi)\delta b_t R - b_t c,$$

s.t. the employability constraint

$$b_t \leq \lambda_{t-1}b^*_{t-1} + \beta \min[b_t, b^*_{t+1}]. \tag{24}$$

If

$$(1-\phi)\delta R \geq c, \tag{25}$$

the employability constraint is binding and it is optimal for the time t agent to choose the same b_t and request the same compensation as in Proposition 1. Thus, as stated above, the outcome of the game is identical to that prevailing when the principal always makes a take–it–or–leave it offer.

The intuition is that, under (25), the ability to set b_t gives the agent the same bargaining power as the ability to make take–it–or–leave–it offers. In both cases, it is competition from the other generations that limits the bargaining power of the agent.

6.3 Intra–generational competition

In the model analyzed above, for simplicity, there is only one agent per generation. Our qualitative results are upheld, however, when there are several agents, as long as intra-generational competition is not perfect. To illustrate this point, consider the following extension of our basic model.

The sequence of play is almost the same as in the basic model, but, at each generation N managers and N agents are born. Then, at Stage 1 of period t, each young manager chooses his b. We focus on symmetric equilibria where all generation t young managers choose the same b_t^*. At Stage 2 of period t, each young investor is matched with one young manager. She observes his type and can make him a take it or leave it offer. If they don't strike a deal, then at Stage 3, as in the basic model, the young investor can approach an old employed agent. Alternatively, in contrast with the basic model, the investor can approach a young manager who has just been employed. In that case, she observes the type b of this manager and can make him a take it or leave it offer. From that point on, the game unfolds as in the basic model.

In equilibrium, all agents choose the same b, and investors rationally anticipate that. Hence, there is no point for the young investor to contact another young manager at Stage 3. She would get exactly the same gain as with the agent she had been initially matched with. Thus, as in the basic model, all generation t managers choose their b to leave the investor indifferent between hiring them or their predecessors, while intra–generational competition does not impose any additional constraint on managers. Hence the equilibrium is exactly the same as in Proposition 1.

7 Conclusion

We develop an overlapping generation model where the intensity of moral hazard that prevails in principal-agent relationships is endogenous and varies over time. Older workers are imperfect substitutes for new workers, due to their shorter horizons and due to their backloaded compensation. This allows younger generations of workers to pick technologies that increase the severity of moral hazard vis-ï¿½-vis previous generations and thus allows them to extract larger rents. This model of endogenous rents is particularly well suited to understand the dynamics of rents in the finance industry, where complexity and opacity are endogenous dimensions that ultimately determine the size of rents. Innovations in the financial industry might be partly driven by the desire of agents to use technologies that lend themselves to rent extraction due to the severity moral hazard problems. Our model implies that optimal regulation might take the form of periodic “crack-downs”: Regulators reset rents periodically at a low level through intense monitoring and then let them drift upwards, as the industry progressively bypasses the regulatory constraints with ad-hoc rent-enhancing technologies. Our theoretical analysis also implies that excess back–loading of compensation, while being privately efficient, might negatively affect the level of competition among managers, leading to inefficiently high rents in the future.

References

Baumol, W.J., 1990, “Entrepreneurship: Productive, unproductive and destructive.” *The Journal of Political Economy*, 5:893–921.

Biais, B., T. Mariotti, G. Plantin and J-C. Rochet, 2007, “"Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications", *Review of Economic Studies,* vol. 74: 345-390.

Becker, G. and G. Stigler, 1974, “Law enforcement, malfeasance, and compensation of enforcers,” *Journal of Legal Studies,* 3:1-18.

Bolton, P., and D. Scharfstein, 1990, “A theory of predation based on agency problems in financial contracting,” *American Economic Review*, 80, 94–106.

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

Bond, P., and U. Axelson, 2012, “Wall Street occupations: An equilibrium theory of overpaid jobs”, LSE working paper.

Carlin, B., 2009, “Strategic Price Complexity in Retail Financial Markets,” *Journal of Financial Economics*, 91, 278–287.

DeMarzo, P.M., and M.J. Fishman, 2007, “Optimal Long-Term Financial Contracting,” *Review of Financial Studies*, 20, 2079–2128.

Diamond, D., 1984, “Financial intermediation and delegated monitoring,” *Review of Economic Studies*, 51, 393–414.

Edmans, A., X. Gabaix, T. Sadzik and Y. Sannikov, 2012, Dynamic CEO compensation, *The Journal of Finance,* 67*,* 1603-1647.

Greenwwod, R. and D. Scharfstein, 2012, “The Growth of Modern Finance,” Working paper, Harvard Business School.

Grossman, S., and O. Hart, 1983, “An analysis of the principal–agent problem,” *Econometrica*, 51, 7–45.

Hochberg, Y., A. Ljungqvist, and Y. Lu, 2010, “Networking as a barrier to entry and the competitive supply of venture capital”, *Journal of Finance*, 65, 829-859.

Holmstrom, B., 1979, "Moral Hazard and Observability", *The Bell Journal of Economics*, 10, 74–91.

Holmstrom, B., and J. Tirole, 1997, "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112, 663–691.

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

Myerson, R., 2012, "A model of moral–hazard credit cycles," Working paper, University of Chicago.

New York Times, 2009, http://www.nytimes.com/2009/03/17/business/17sorkin.html?_r=1)

Oyer, P., 2008, "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *Journal of Finance*, 63, 2601–2628.

Rogerson, W., 1985, "Repeated Moral Hazard," *Econometrica*, 58, 69-76.

Philippon, T. and A. Resheff, 2008, "Wages and Human Capital in the U.S. Financial Industry: 1909-2006," Working Paper, New York University.

Townsend, R., 1979, "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, 22, 265–293.

Proofs

Proof of Lemma 1:

The program of the investor is

$$\max_{w_t^t, w_{t+1}^t} R(1+\beta) - w_t^t - \beta w_{t+1}^t, \text{s.t.}, \; w_{t+1}^t \geq \delta b_t R \text{ and } w_t^t + \beta w_{t+1}^t \geq \delta b_t R.$$

The Lagrangian is

$$\mathcal{L} = R(1+\beta) - w_t^t - \beta w_{t+1}^t + \mu_t(w_t^t + \beta w_{t+1}^t - \delta b_t R) + \mu_{t+1}(w_{t+1}^t - \delta b_t R),$$

where μ_t and μ_{t+1} are the multipliers of the time t and $t+1$ incentive constraints, respectively. The first order condition with respect to w_t is:

$$-1 + \mu_t = 0.$$

Hence the incentive compatibility constraint at time $t+1$ binds, i.e.,

$$w_t^t + \beta w_{t+1}^t = \delta b_t R,$$

and the set of optimal values for w_t^t is the interval $[0, (1-\beta)\delta b_t R]$.

QED

Proof of Lemma 3:

Since $\lambda_{t-1} \geq 1$ and $b_t^* \geq b_{\min} > 0$, (14) implies $b_t > b_{t-1}$.

QED

Proof of Proposition 1:

To complete the analysis of the equilibrium, we must rule out a last possible deviation by managers. If the generation t manager chooses $b_t = b_t^*$, he gets hired and works for two periods. It could be tempting for the time t manager to deviate to $b_t = \hat{b} > b_t^*$, aware that he would not be hired at time t, but hoping he would earn large rents at $t+1$.

To evaluate his gains under this deviation, the agent must form beliefs about the reaction of the next generation. Our assumption that managers and investors make choices at stages 1 and 2 facilitates this analysis. Since they make their decision before having the opportunity to observe previous actions, they stick to the equilibrium even after deviations.

When deviating to $\hat{b}$ at period t, the manager must ensure he will be hired at period $t+1$. At that time, when making the equilibrium choice and hiring the young manager, the young investor obtains

$$R(1+\beta) - \delta b_{t+1}^{*} R.$$

If, instead, the period $t+1$ investor hires the previous deviator, and, at the following period, the (now old) time $t+1$ manager, he obtains

$$R(1+\beta) - \delta \hat{b} R - \beta \delta b_{t+1}^{*} R.$$

To be employed at time $t+1$, the deviator must choose $\hat{b}$ such that

$$\hat{b} \leq (1-\beta) b_{t+1}^{*}. \tag{26}$$

The equilibrium choice of b_t^* dominates the deviation to $\hat{b}$ if

$$\delta b_t^* R \geq \beta \delta \hat{b} R. \tag{27}$$

Binding (26) and substituting it into (27), the condition under which the agent is better off taking the equilibrium action rather than deviating to $\hat{b}$ is

$$b_t^* \geq \beta(1-\beta) b_{t+1}^{*}.$$

Substituting b_{t+1}^{*} from (16), the condition is

$$b_t^* \geq \beta(1-\beta)\lambda \frac{b_t^*}{1-\beta}.$$

That is

$$\frac{1}{\beta} \geq \lambda,$$

which holds since $\lambda \in [0, 1/\beta]$.

QED

Proof of Proposition 3:

To prove proposition 3, we first characterize the equilibrium dynamics of b_{kT+t}, over a cyclical intervention regime where the regulator resets b to $\underline{b}$ every T periods. We then study the condition under which the agent complies after T periods. Finally we compare $W(1)$ and $W(\infty)$ to $W(T)$, for T finite and larger than 2. A series of technical lemmas are needed to complete the proof.

Lemma 5: Consider a policy that resets b to $\underline{b}$ with periodicity T . The dynamics of b_t is described by:

$$b_{kT+n} = \min(\frac{b_{kT+n-1}}{1-\beta}, b_{\max}) = \min\left(\frac{\underline{b}}{(1-\beta)^n}, b_{\max}\right), \forall n \in \{1, ...T-2\}, \text{if } T > 2,$$

while

$$b_{(k+1)T-1} = \min(b_{(k+1)T-2} + \beta\underline{b}, b_{\max}) = \min\left((\frac{1}{(1-\beta)^{T-2}} + \beta)\underline{b}, b_{\max}\right).$$

Proof of Lemma 5:

Assume for now that $b_{\max}$ is not reached during the cycle. First consider the case $T > 2$. At time $t = kT + n$, where $n \in \{1, ...T-2\}$, young managers choose the highest possible b_t at which they are still employed. Assume this does not bind $b_{\max}$. Following the same reasoning as in the previous analyses, binding the employability constraint, yields

$$R(1+\beta) - \delta b_t R = R(1+\beta) - \delta b_{t-1} R - \beta\delta \min[b_t, b_{t+1}]R,$$

which simplifies to

$$R(1+\beta) - \delta b_t R = R(1+\beta) - \delta b_{t-1} R - \beta\delta b_t R.$$

Thus

$$b_t = \frac{b_{t-1}}{1-\beta}.$$

This implies

$$b_{kT+n} = \frac{b_{kT+n-1}}{1-\beta}, \forall n \in \{1, ...T-2\}.$$

For the last generation before regulatory intervention, $t = (k+1)T - 1$, things are slightly different since the young manager anticipates that $b_{t+1} = \underline{b}$. Hence, the manager in that generation chooses b_t such that

$$R(1+\beta) - \delta b_t R = R(1+\beta) - \delta b_{t-1} R - \beta\delta\underline{b}R.$$

That is

$$b_t = b_{t-1} + \beta\underline{b}.$$

Since $t - 1 = kT + T - 2$, $b_{t-1} = \frac{\underline{b}}{(1-\beta)^{T-2}}$. Thus we have

$$b_t = (\frac{1}{(1-\beta)^{T-2}} + \beta)\underline{b}. \tag{28}$$

Second, consider the case $T = 2$. At $t = 1$, we have $b_1 = \underline{b}$, at $t = 2$, we have $b_2 > \underline{b}$, ... b_2 binds the employability constraint:

$$R(1+\beta) - \delta b_2 R = R(1+\beta) - \delta\underline{b}R - \beta\delta\underline{b}R.$$

Hence

$$b_2 = (1+\beta)\underline{b}, \tag{29}$$

which is consistent with (28) evaluated at $T = 2$.

In the lemma, we allow for the possibility that $b_{\max}$ is reached during the cycle. This means that $b_{kT+n} \leq b_{\max}$ is also part of the employability constraints that can bind, hence the min in the formulas.

QED

Now, turn to the monitoring probability ensuring compliance. At the time of intervention, $t = kT$, the regulator sets the desired level of complexity below the previous level, i.e., $\underline{b} < b_{t-1}$. If the generation t agent complies, he gets

$$\delta\underline{b}R.$$

If he deviates, he chooses the maximum level of complexity at which he is still employed, i.e., $\hat{b}$ such that

$$R(1+\beta) - \delta\hat{b}R = R(1+\beta) - \delta b_{t-1}R - \beta\delta\min[\hat{b}, b^*_{t+1}]R. \tag{30}$$

The left–hand–side is the profit of the investor if she hires at time t the manager who deviated to $\hat{b}$. The right–hand–side is her expected profit if she poaches an old manager at time t, and then, at time $t+1$, hires either the (unemployed) generation t manager (and pays him $\delta\hat{b}R$) or the young manager (whom she expects to follow the candidate equilibrium strategy and choose b^*_{t+1}). This simplifies to

$$\hat{b} = b_{t-1} + \beta \min[\hat{b}, b^*_{t+1}]. \tag{31}$$

Given rational expectations about $\hat{b}$, the regulator will choose the lowest possible level of monitoring ensuring compliance, so that (20) holds, i.e. : $\underline{b} = (1-\alpha)(b_{t-1} + \beta \min[\underline{b}, b^*_{t+1}])$

Lemma 6: for a given $\underline{b}$, longer cycles imply weakly higher monitoring expense at the time of reset, i.e. $\alpha(T) \leq \alpha(T+1)$.

Proof of Lemma 6:

We know that $\underline{b} = (1-\alpha(T))(b_{kT-1}(T) + \beta \min[\underline{b}, b_{kT+1}(T)]) = (1-\alpha(T+1))(b_{kT-1}(T+1) + \beta \min[\underline{b}, b_{k(T+1)+1}(T+1)])$, where $b_{kT+i}(T)$ is the equilibrium path on a T-cycle of length T and reset at $\underline{b}$ as given by Lemma 5.

From Lemma 5, we also know that $b_{kT-1}(T) \leq b_{kT-1}(T+1)$ and $b_{kT+1}(T) \leq b_{k(T+1)+1}(T+1)$, from which we conclude that $\alpha(T) \leq \alpha(T+1)$.

QED

This allows us to now characterize further policies that cannot be optimal in the set of cyclical policies. This is useful later in establishing proposition 3.

Lemma 7: A periodic monitoring policy of period T that is optimal among cyclical policies (including laissez-faire) must be such that $b_{(k+1)T-1} < b_{\max}$.

Proof of Lemma 7:

Consider a periodic monitoring policy such that $b_{T-1} = b_{\max}$ and $b_T = \underline{b}$. We are going to show that it is dominated by either laissez faire or by the periodic monitoring policy with period $T-1$.

Either $W(T) \leq W(\infty)$, and the lemma is true, or $W(T) > W(\infty)$. Consider the latter case. We want to establish that $W(T-1) > W(T)$. We can explicit welfare over the cycle of length T with reset at $\underline{b}$ as follows: $W(T) = R(1+\beta) - cb_{average}(T) - \frac{\gamma}{T}\alpha(T)$ where $b_{average}(T)$ is the average of b over the cycle. The $(T-1)-$cycle with starting point $\underline{b}$ has an identical dynamics of b as the cycle of length T

along the first $(T-2)$ periods , a weakly lower b at period $T-1$ (a consequence of lemma 5) and a smaller cost of monitoring ($\alpha(T-1) \leqslant \alpha(T)$ from Lemma 6). Thus, we have $b_{average}(T-1) \leqslant b_{average}(T)$, such that: $W(T) \leqslant \frac{T-1}{T}W(T-1) + \frac{1}{T}W(\infty)$. This implies $W(T-1) > W(T)$ because we know that $W(T) > W(\infty)$.

QED

To finally estimate welfare over cycles of various lengths, we need to compute the monitoring costs associated to them. This is what the following lemma does:

Lemma 8: To ensure compliance with periodic monitoring, under an optimal periodic monitoring cycle of period T , the regulator sets $\underline{b} = b_{\min}$ (i.e. the regulator sets $\underline{b}$ at its minimum level) and sets the monitoring probability to:

$$\alpha_T = 1 - \frac{1}{\frac{1}{(1-\beta)^{T-2}} + \beta\frac{2-\beta}{1-\beta}} \text{ if } T > 2 \text{ and } 1 - \frac{1}{(1+\beta)^2} \textit{ if } T = 2.$$

Proof of Lemma 8:

From lemma 7, we know that we can consider that $b_{\max}$ is not reached during the cycle. First consider the case $T > 2$. In this case, the dynamic choice equation for b for the generation that is being monitored writes as $\hat{b} = b_{t-1} + \beta \min[\hat{b}, \frac{\underline{b}}{1-\beta}]$.

If $\min[\hat{b}, \frac{\underline{b}}{1-\beta}] = \hat{b}$, that is $\hat{b} < \frac{\underline{b}}{1-\beta}$, then

$$\hat{b} = \frac{b_{t-1}}{1-\beta}$$

which is in contradiction with $\hat{b} < \frac{\underline{b}}{1-\beta}$ since $\underline{b} < b_{t-1}$. Hence $\min[\hat{b}, \frac{\underline{b}}{1-\beta}] = \frac{\underline{b}}{1-\beta}$, and $\hat{b}$ is set by

$$\hat{b} = b_{t-1} + \beta\frac{\underline{b}}{1-\beta}.$$

Thus (20) writes as

$$\underline{b} = (1-\alpha)[b_{t-1} + \beta\frac{\underline{b}}{1-\beta}].$$

That is

$$\alpha = 1 - \frac{1}{\frac{b_{t-1}}{\underline{b}} + \frac{\beta}{1-\beta}} = 1 - \frac{1}{\frac{b_{kT-1}}{\underline{b}} + \frac{\beta}{1-\beta}}.$$

Note that the right–hand–side is between 0 and 1. Substituting b_{kT-1} from Lemma 5, this simplifies to

$$\alpha = 1 - \frac{1}{\frac{1}{(1-\beta)^{T-2}} + \beta\frac{2-\beta}{1-\beta}}.$$

Second, consider the case $T = 2$. Subsituting (29), we have

$$\hat{b} = (1+\beta)\underline{b} + \beta \min[\hat{b}, (1+\beta)\underline{b}].$$

If $\hat{b} \leq (1+\beta)\underline{b}$, then this equation is equivalent to

$$\hat{b} = \frac{1+\beta}{1-\beta}\underline{b},$$

which is a contradiction. Hence, $\hat{b} > (1+\beta)\underline{b}$ and the employability constraint yields

$$\hat{b} = (1+\beta)^2\underline{b}.$$

Thus (20) yields

$$\alpha = 1 - \frac{1}{(1+\beta)^2}.$$

Thus, whether $T = 2$ or $T > 2$, the monitoring cost is independent of the chosen reset level $\underline{b}$, so, choosing $\underline{b} = b_{\min}$ guarantees the highest level of welfare.

QED

We can now compute welfare over a cycle that does not reach $b_{\max}$, which yields the following lemma:

Lemma 9: Average welfare under an optimal periodic monitoring cycle of period T strictly larger than 2, is:

$$W(T) = R(1+\beta) - \frac{1}{T}\{cb_{\min}[(1-\beta)\frac{(\frac{1}{1-\beta})^{T-1} - 1}{\beta} + (\frac{1}{1-\beta})^{T-2} + \beta] + (1 - \frac{1}{(\frac{1}{1-\beta})^{T-2} + \beta\frac{2-\beta}{1-\beta}})\gamma\}. \quad (32)$$

For $T = 2$:

$$R(1+\beta) - \frac{1+(1+\beta)}{2}cb_{\min} - \frac{1}{2}(1 - \frac{1}{(1+\beta)^2})\gamma. \quad (33)$$

Proof of lemma 9:

Averaging welfare across generations, for any finite T strictly larger than 2, social welfare under periodic monitoring is :

$$W(T) = R(1+\beta) - \frac{1}{T}\{cb_{\min}[1+\frac{1}{1-\beta}+...+\frac{1}{(1-\beta)^{T-2}}+(\frac{1}{(1-\beta)^{T-2}}+\beta)]+(1-\frac{1}{\frac{1}{(1-\beta)^{T-2}}+\beta\frac{2-\beta}{1-\beta}})\gamma\}. \tag{34}$$

For $T=2$, the formula computes welfare over two periods where b_t oscillates between the two values $b_{\min}$ and $(1+\beta)b_{\min}$.

QED

We now compare welfare under periodic monitoring to its counterpart under permanent monitoring and under laissez faire. The next proposition states a preliminary result.

Lemma 10: Assume $b_{\max} \geq (1+\beta)b_{\min}$.Welfare under monitoring with $T=2$ is larger than welfare under laissez faire and than welfare under permanent monitoring iff:

$$\frac{(1+\beta)^2}{\beta}cb_{\min} < \gamma < c\frac{b_{\max} - \frac{1+(1+\beta)}{2}b_{\min}}{\frac{1}{2}(1-\frac{1}{(1+\beta)^2})}.$$

Proof of Lemma 10:

(33) is greater than (22), iff

$$\frac{1+(1+\beta)}{2}cb_{\min} + \frac{1}{2}(1-\frac{1}{(1+\beta)^2})\gamma < cb_{\max}. \tag{35}$$

That is

$$\gamma < c\frac{b_{\max} - \frac{1+(1+\beta)}{2}b_{\min}}{\frac{1}{2}(1-\frac{1}{(1+\beta)^2})}.$$

Note that the right–hand–side of is positive, since $b_{\max} \geq (1+\beta)b_{\min}$.

(33) is greater than (21), iff

$$\frac{\beta}{2}cb_{\min} < [\frac{\beta}{1+\beta} - \frac{1}{2}(1-\frac{1}{(1+\beta)^2})]\gamma.$$

$$\frac{\beta}{2}cb_{\min} < \frac{1}{(1+\beta)^2}[\beta(1+\beta) - \frac{1}{2}((1+\beta)^2-1)]\gamma.$$

$$\frac{\beta}{2}cb_{\min} < \frac{1}{(1+\beta)^2}[\beta(1+\beta) - \frac{1}{2}(2\beta+\beta^2)]\gamma.$$

$$\frac{\beta}{2}cb_{\min} < \frac{\beta}{(1+\beta)^2}[(1+\beta) - (1+\frac{\beta}{2})]\gamma.$$

$$\frac{(1+\beta)^2}{\beta}cb_{\min} < \gamma.$$

QED

We can now complete the proof of proposition 5:

Call N the smallest integer such that $n > N \Rightarrow \frac{b_{\min}}{(1-\beta)^n} > b_{\max}$. Thus, we know that for any value of γ, a cycle of length longer than N induces lower welfare than a cycle of length N. Thus, looking for the optimal policy is equivalent to solving the following program:

$$\max_{T\in\{1,2,\ldots,N,+\infty\}} W(T,\gamma)$$

where we note $W(1,\gamma)$ the welfare of permanent monitoring and $W(+\infty,\gamma)$ the welfare of laissez-faire.

$$W(T,\gamma) = \begin{cases} R(1+\beta) - \frac{\beta}{1+\beta}\gamma - cb_{\min} & \text{for} \quad T=1 \\ R(1+\beta) - \frac{1+(1+\beta)}{2}cb_{\min} - \frac{1}{2}(1-\frac{1}{(1+\beta)^2})\gamma & \text{for} \quad T=2 \\ R(1+\beta) - \frac{1}{T}\{cb_{\min}[(1-\beta)\frac{(\frac{1}{1-\beta})^{T-1}-1}{\beta} + (\frac{1}{1-\beta})^{T-2}+\beta] & \\ \quad +(1-\frac{1}{(\frac{1}{1-\beta})^{T-2}+\beta\frac{2-\beta}{1-\beta}})\gamma\} & \text{for} \quad T>2 \\ R(1+\beta) - cb_{\max} & \quad T=+\infty \end{cases}$$

In particular, for $\gamma = 0$:

$$W(T,0) = \begin{cases} R(1+\beta) - cb_{\min} & \text{for} \quad T=1 \\ R(1+\beta) - \frac{1+(1+\beta)}{2}cb_{\min} & \text{for} \quad T=2 \\ R(1+\beta) - \frac{1}{T}\{cb_{\min}[(1-\beta)\frac{(\frac{1}{1-\beta})^{T-1}-1}{\beta} + (\frac{1}{1-\beta})^{T-2}+\beta]\} & \text{for} \quad T>2 \\ R(1+\beta) - cb_{\max} & \quad T=+\infty \end{cases}$$

It follows that $\forall T \in \{2,...,N,+\infty\} : W(1,0) > W(T,0)$. This is economically intuitive: if there is no monitoring cost, always monitoring is optimal. Thus, by continuity of W, this inequality holds in a neighborhood of $\gamma = 0$.I.e., $\exists\gamma_1 > 0; \gamma < \gamma_1 \Rightarrow W(1,\gamma) = \max_{T\in\{1,2,\ldots,N,+\infty\}} W(T,\gamma)$.

Similarly, we see that

$$\lim_{\gamma\to+\infty} W(T,\gamma) = \begin{cases} -\infty & \text{for} \quad T\in\{1,2,...,N\} \\ R(1+\beta) - cb_{\max} & \text{for} \quad T=+\infty \end{cases}$$

It follows that $\forall T \in \{1,...,N\} : \lim_{\gamma\to+\infty} W(+\infty,\gamma) > \lim_{\gamma\to+\infty} W(T,\gamma)$. Thus, by continuity, this inequality holds in a neighborhood of $\gamma = +\infty$. I.e., $\exists\gamma_4 > 0; \gamma > \gamma_4 \Rightarrow W(+\infty,\gamma) = \max_{T\in\{1,2,\ldots,N,+\infty\}} W(T,\gamma)$.

Now, let's prove the existence of (γ_1, γ_2). One just needs to notice that:

$$\gamma_2 = \frac{(1+\beta)^2}{\beta} c b_{\min} < \gamma_3 = c \frac{b_{\max} - \frac{1+(1+\beta)}{2} b_{\min}}{\frac{1}{2}(1 - \frac{1}{(1+\beta)^2})}$$

is equivalent to:

$$b_{\min} < \frac{b_{\max}}{\frac{(1+\beta)^2}{2\beta} + 1 - \frac{1}{2\beta} + \frac{\beta}{2}}$$

$$b_{\min} < \frac{b_{\max}}{(2+\beta)} = B.$$

Note that $b_{\min} < B$ implies $b_{\max} \geq (1+\beta) b_{\min}$ which validates the use Lemma 10.

From lemma 10, we can finally conclude that for $\gamma_1 < \gamma < \gamma_2$:

$$Arg \max_{T \in \{1,2,\ldots,N,+\infty\}} W_T(\gamma) \notin \{1, +\infty\}.$$

QED

Proof of Proposition 4:

Suppose the players expect b's to increase in equilibrium. Then, the employability constraint faced by the new agent at time t is

$$\begin{aligned} \sigma(b_t) &\geq R(b_{t-1})(1-b_{t-1}) + \beta(R(b_t)(1-b_t)) \\ &= R(b_{t-1})(1+\beta-b_{t-1}) - \beta R(b_{t-1}) + \beta(R(b_t)(1-b_t)) \\ &= \sigma(b_{t-1}) + \beta(R(b_t) - R(b_{t-1})) - \beta R(b_t) b_t. \end{aligned}$$

Binding the employability constraint we have

$$\sigma(b_t) - \sigma(b_{t-1}) = \beta(R(b_t) - R(b_{t-1})) - \beta R(b_t) b_t.$$

We now prove by contradiction that this implies $b_t \geq b_{t-1}$. Assume the opposite: $b_t < b_{t-1}$. Then, since we are in the decreasing part of σ, $\sigma(b_t) - \sigma(b_{t-1}) > 0$. Thus, the binding employability constraint implies

$$\beta(R(b_t) - R(b_{t-1})) - \beta R(b_t) b_t > 0.$$

But, since $b_t < b_{t-1}$ and since R is increasing, the left–hand–side of this inequality is negative. Hence a contradiction.

QED