Large Risks, Limited Liability and Dynamic Moral Hazard[*]

Bruno Biais[†] Thomas Mariotti[‡]
Jean-Charles Rochet[§] Stéphane Villeneuve[¶]

February 2009

Abstract

We study a continuous-time principal-agent model in which a risk-neutral agent with limited liability must exert unobservable effort to reduce the likelihood of large but relatively infrequent losses. Firm size can be decreased at no cost, and increased subject to adjustment costs. In the optimal contract, investment takes place only if a long enough period of time elapses with no losses occurring. Then, if good performance continues, the agent is payed. As soon as a loss occurs, payments to the agent are suspended, and so is investment if further losses occur. Accumulated bad performance leads to downsizing. We derive explicit formulae for the dynamics of firm size and its asymptotic growth rate, and we provide conditions under which firm size eventually goes to zero, or grows without bounds.

Keywords: Principal-Agent Model, Limited Liability, Continuous Time, Poisson Risk, Downsizing, Investment, Firm Size Dynamics.
JEL Classification: C61, D82, D86, D92.

1 Introduction

Industrial and financial firms are subject to large risks: the former are prone to accidents and the latter are exposed to sharp drops in the value of their assets. Preventing these risks requires managerial effort. Systematic analyses of industrial accidents point to the role of human deficiencies and inadequate levels of care.[1] A striking illustration is offered by the explosion at the BP Texas refinery in March 2005: after investigating the case, the Baker Panel concluded that "BP executive and corporate refining management have not provided effective process safety leadership."[2] Similarly, the large losses incurred by banks and insurance companies during the recent financial crisis were in part due to insufficient risk control. These large risks present a major challenge to firms, investors and citizens. This paper studies the design of incentives to mitigate them.

One way to stimulate the prevention of large risks would be to make managers and firms bear the social costs that they generate. Yet, this is often impossible in practice, because total damages often exceed the wealth of managers and even the net worth of firms, while the former are protected by limited liability and the latter by bankruptcy laws.[3] This curbs managers' incentives to reduce the risk of losses that exceed the value of their assets.[4] Nevertheless, if the risk prevention activities undertaken by managers were observable, it would be straightforward to design compensation schemes that would induce them to take socially optimal levels of risk. To a large extent, however, these activities are unobservable by external parties, which leads to a moral hazard problem.

Besides informational asymmetries, another important aspect of large risks lies in their timing. Large losses are relatively rare events that contrast with day-to-day firm operations and cash-flows.[5] It is therefore natural to study large risk prevention in a dynamic set-up, where the timing of losses differs from that of operations. To do so, we focus on the simplest

---

[1]See for instance Leplat and Rasmussen (1984), Gordon, Flin, Mearns and Fleming (1996) or Hollnagel (2002).

[2]"The Report of the BP U.S. Refineries Independent Safety Review Panel," January 16, 2007. Also, Chemical Safety Board Chairman Carolyn W. Merritt stated that "BP's global management was aware of problems with maintenance, spending, and infrastructure well before March 2005. [...] Unsafe and antiquated equipment designs were left in place, and unacceptable deficiencies in preventative maintenance were tolerated," "CSB Investigation of BP Texas City Refinery Disaster Continues as Organizational Issues are Probed," CSB News Release, October 30, 2006.

[3]For instance, Katzman (1988) report that "In *Ohio v. Kovacs* (U.S.S.C. 83–1020), the U.S. Supreme Court unanimously ruled that an industrial polluter can escape an order to clean up a toxic waste site under the umbrella of federal bankruptcy." Similarly, the social losses created by the recent financial crisis exceeded by far the assets one could withhold from financial executives.

[4]Shavell (1984, 1986) discusses how a party's inability of paying for the full magnitude of harm done dilutes its incentives to reduce risk.

[5]From now on, we shall generically refer to any realization of a large risk as a loss.

possible model: operating cash-flows are constant per unit of time, while losses follow a Poisson process whose intensity depends on the level of risk prevention.

In this context, we study the optimal contract between a principal and an agent that provides the latter with appropriate incentives to reduce the risk of losses under dynamic moral hazard. The agent, who can be thought of as an entrepreneur or a manager running a business, is risk-neutral and protected by limited liability. She can exert effort to reduce the instantaneous probability of losses.[6] Effort is costly to the agent and unobservable by other parties. The project run by the agent can expand, through investment, or shrink, through downsizing. While downsizing is unconstrained, we assume that the pace of investment is limited by adjustment costs, in the spirit of Hayashi (1982) or Kydland and Prescott (1982). We also assume constant returns to scale, in that downsizing and investment affect by the same factor the operating profits of the project, the social costs of accidents, and the private benefits that the agent derives from shirking. This assumption implies that the principal's value function is homogeneous in size and enables one to characterize the optimal contract explicitly. However, as discussed in the paper, some of our key qualitative results are robust to relaxing the constant returns to scale assumption.

The optimal contract maximizes the expected value that the principal derives from an incentive feasible risk prevention policy. It relies on two instruments: positive payments to the agent, and project size management through downsizing and investment. While these decisions are functions of the entire past history of the loss process, this complex history dependence can be summarized by two state variables: the size of the project, and the continuation utility of the agent. The former reflects the history of past downsizing and investment decisions, while the latter reflects the prospect of future payments to the agent. The evolution of the agent's continuation utility mirrors the dynamics of losses, and thus serves as a track record of the agent's performance.[7] We characterize the compensation and size management policy arising in the optimal contract.

First consider the compensation policy. To motivate the agent, the optimal contract relies on the promise of payments after good performance and the threat of reductions in her continuation utility after losses. When the track record of the agent is relatively poor, there is a probation phase during which she does not receive any payment. As long as no loss occurs, the size-adjusted continuation utility of the agent increases until it reaches a

---

[6]Unlike in Shapiro and Stiglitz (1984) or Akerlof and Katz (1989), effort merely makes losses less likely, but does not eliminate them altogether. As a result of this, losses do occur on the equilibrium path, and it is no longer optimal to systematically fire the agent following a loss.

[7]That the optimal contract exhibits memory is a standard feature of dynamic moral hazard models, see for instance Rogerson (1985).

threshold at which she receives a constant wage per unit of time and size of the project, such that her size-adjusted continuation utility remains constant. As soon as a loss occurs, the continuation utility of the agent undergoes a sharp reduction and the contract reverts to the probation phase. The magnitude of that reduction in the agent's continuation utility is pinned down by the incentive compatibility constraint. The more severe the moral hazard problem, and the larger the project, the greater the punishment. The induced sensitivity of the agent's continuation utility to the random occurrence of losses is socially costly because the principal's value function is concave in that state variable. Therefore, it is optimal to set the reduction in the agent's continuation utility following a loss to the minimum level consistent with incentive compatibility.

Next consider the dynamics of the size of the project. In the first-best, there is no need for downsizing. Since the project has positive net present value, investment then always takes place at the highest feasible rate in order to maximize the size of the project. In the second-best, however, the size of the project is lower than in the first-best. The intuition is the following. As mentioned above, the agent is partly motivated by the threat of reductions in her continuation utility in case of bad performance. Yet, when the continuation utility of the agent is low, the threat to reduce it further has limited bite, because of limited liability. To cope with this limitation, it can be necessary to lower the agent's temptation to shirk by reducing the scale of operations after losses. Apart from such circumstances, and in particular when no loss occurs, the project is never downsized. In addition to downsizing, moral hazard also affects the size of the project through its impact on investment. Since increases in the size of the project raise the temptation to shirk, investment can take place only when the agent has enough at stake in the project, that is, when her track record has been good enough for her continuation utility to reach a given threshold. While payments when they occur are costly for the principal, investment benefits both parties. As long as investment takes place, the total size of the pie grows, which in turn makes delaying the compensation of the agent less costly. Thus it is efficient to invest before actually compensating the agent. Note that the sequencing of compensation and investment is reversed in the first-best. This is because the agent, who is assumed to be more impatient than the principal, receives then all her compensation at time zero, before any investment actually takes place.

We obtain an explicit formula mapping the path of the agent's size-adjusted continuation utility into the size of the project. If one interprets the latter as firm size, this formula exactly spells out how firm size grows, stays constant or declines over time. Relying on asymptotic theory for Markov ergodic processes, we then characterize the long-run growth rate of the

firm. In the first-best, firm size goes to infinity at a constant rate. Our formula for the long-run growth rate of the firm shows how, in the second-best, this trend in firm size is reduced by downsizing and possibly lower investment rates. When the adjustment costs are high, firm size eventually goes to zero. By contrast, when both the adjustments costs and the frequency of losses are low, firm size eventually goes to infinity, although more slowly than in the first-best.

Our paper belongs to the rich and growing literature on dynamic moral hazard that uses recursive techniques to characterize optimal dynamic contracts.[8] Our main contributions are the following. First, we study the case where moral hazard is about large but relatively infrequent risks. As illustrated by recent industrial accidents or by the recent financial crisis, preventing such risks is a major challenge. We show that optimal contracts that mitigate the risk of infrequent but large losses differ markedly from those prevailing when fluctuations in the output process are frequent but infinitesimal. In the latter, as illustrated by the Brownian motion models of DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin and Rochet (2007) or Sannikov (2008), the continuation utility of the agent continuously fluctuates until it reaches zero, an event that is predictable. At this point, the project is liquidated. In contrast, with Poisson risk, the continuation utility of the agent increases smoothly most of the time, but incurs sharp decreases when losses occur. In this context, incentive compatibility together with limited liability imply unpredictable downsizing, unlike in the Brownian case. Second, we analyze the interaction between incentives considerations and firm size dynamics. Our asymptotic analysis offers the first description of the long-run impact of downsizing and investment on firm size under moral hazard.

Building on DeMarzo and Fishman (2007a), DeMarzo, Fishman, He and Wang (2008) analyze investment in a dynamic moral hazard model with quadratic adjustment costs and a constant returns to scale production technology. As in our model, the agent's current continuation utility together with the current capital stock are sufficient statistics for the optimal contract. Our model differs from theirs in that we consider Poisson risk, while they focus on the case of Brownian risk. Moreover, while we are particularly interested in the asymptotic properties of firm size growth, the main emphasis of their paper is on the dynamics of Tobin's q.

In the context of a political economy model, Myerson (2008) contemporaneously offers an

---

[8]See for instance Green (1987), Spear and Srivastava (1987), Thomas and Worrall (1990) or Phelan and Townsend (1991) for seminal contributions along these lines. By focusing on the case where the agent is risk-neutral, with limited liability, our model is in line with the recent papers by Clementi and Hopenhayn (2006) and DeMarzo and Fishman (2007a, 2007b).

analysis of dynamic moral hazard in a Poisson framework. A distinctive feature of our paper is that we analyze the impact of investment on the principal-agent relationship. Moreover, Myerson (2008) considers the case where the principal and the agent have identical discount rates. This case, however, is not conducive to continuous-time analysis, as an optimal contract does not exist. To cope with this difficulty, Myerson (2008) imposes an exogenous upper bound on the continuation utility of the agent. By contrast, we do not impose such a constraint on the set of feasible contracts. Instead, we consider the case where the principal is less impatient than the agent. While this makes the formal analysis more complex, this also restores the existence of an unconstrained optimal contract.[9]

Sannikov (2005) also uses a Poisson payoff structure. A key difference with our analysis lies in the way output is affected by the jumps of the Poisson process. In Sannikov (2005), jumps correspond to positive cash-flow shocks, while in our model they correspond to losses that are less likely to occur if the agent exerts effort.[10] This leads to qualitatively very different results. While downsizing is a key feature of our optimal contract, as it ensures that incentives can still be provided following a long sequence of accidents, it plays no role in Sannikov (2005). Liquidation in his model is still required to provide incentives, but it corresponds to a predictable event: if a sufficiently long period of time elapses during which the agent reports no cash-flow, the firm is liquidated. By contrast, downsizing in our model is unpredictable.

Our paper is also related to the literature on accident law. Shavell (1986, 2000) argues that the desirability of liability insurance depends on the ability of insurers to monitor the firm's prevention effort, and to link insurance premia to the observed level of care. If insurers cannot observe the firm's level of care, making full liability insurance mandatory results in no care at all being taken.[11] In our dynamic analysis, the optimal contract ties the firm's allowed activity level to its performance record: following a series of losses, the firm can be forbidden to engage at full scale in its risky activity. These instruments provide the manager of the firm with dynamic incentives to exert the appropriate risk prevention effort, although the latter is not observed by the principal.

The paper is organized as follows. In Section 2, we present the model. In Section 3, we formulate the incentive compatibility and limited liability constraints. In Section 4, we

---

[9]Poisson processes have also proved useful in the theory of repeated games with imperfect monitoring, see for instance Abreu, Milgrom and Pearce (1991), Kalesnik (2005) and Sannikov and Skrzypacz (2007). Our focus differs from theirs in that we consider a full commitment contracting environment, in which we explicitly characterize the optimal incentive compatible contract.

[10]Thus jumps in our model are bad news in the sense of Abreu, Milgrom and Pearce (1991).

[11]See Jost (1996) and Polborn (1998) for important extensions and qualifications of this argument.

characterize the optimal contract under maximal risk prevention. Based on this analysis, Section 5 studies the dynamics of firm size. Section 6 discusses the robustness of our results. In Section 7, we derive some empirical implications of our theoretical analysis. Section 8 concludes. Sketches of proofs are provided in the appendix. Complete proofs are available in the supplement to this paper (Biais, Mariotti, Rochet and Villeneuve (2009)).

2   The Model

There are two players, a principal and an agent. The agent can run a potentially profitable project for which she has unique necessary skills.[12] However, this project entails costs, and the agent has limited liability and no initial cash. By contrast, the principal has unlimited liability and is able to cover the costs. One can think of the agent as an entrepreneur or a manager running a business, and of the principal as a financier, an insurance company, or society at large.

Time is continuous and the project can be operated over an infinite horizon. The two players are risk-neutral. The principal discounts the future at rate $r > 0$ and the agent at rate $\rho > r$, which makes her more impatient than the principal. This introduces a wedge between the valuation of future transfers by the principal and the agent, and rules out indefinitely postponing payments to the agent. Without loss of generality, we normalize to 0 the set-up cost of the project.

At any time t, the size X_t of the project can be scaled up or down. There are no constraints on downsizing: any fraction of the assets between 0 and 1 can be instantaneously liquidated. For simplicity, we normalize the maximal possible initial size of the project to 1, and assume that the liquidation value of the assets is 0. The project can also be expanded, at unit cost $c \geq 0$. The rate at which such investments can take place is constrained, however. This reflects for instance that new plants cannot be built instantaneously, or that the inflow of new skilled workers is constrained by search and training. Consistent with this, we shall assume that the instantaneous growth rate g_t of the project is at most equal to $\gamma \in (0, r)$. This is in line with the macroeconomic literature emphasizing the delays and costs associated with investment, such as time-to-build constraints (Kydland and Prescott (1982)) or convex adjustment costs (Hayashi (1982)). Our formulation corresponds to a simple version of the adjustment cost model in which there are no adjustment costs up to an instantaneous size adjustment $X_t \gamma dt$, and infinite adjustment costs beyond this point.

---

[12]Empirically, this assumption is particularly relevant in the case of small businesses, where the entrepreneur-manager is often indispensable for operating the firm efficiently (Sraer and Thesmar (2007)).

Operating profits per unit of time are equal to $X_t\mu$, where $\mu > 0$ is a constant representing day-to-day size-adjusted operating profits. While such profits are constant, the project is subject to the risk of large losses. In the case of a manufacturing firm, such losses can be generated by a severe accident. In the case of a financial firm, they can result from a sudden and sharp decrease in the value of the assets that the firm invested in. The occurrence of these losses is modeled as a point process $N = \{N_t\}_{t\geq 0}$, where for each $t \geq 0$, N_t is the number of losses up to and including time t. Denote by $(T_k)_{k\geq 1}$ the successive random times at which these losses occur. A loss generates costs that are borne by the principal rather than by the agent. For example, an oil spill imposes huge damages on the environment and on the inhabitants of the affected region, but has limited direct impact on the manager of the oil company. Or, in the case of financial firms, the losses incurred by many banks in 2007 and 2008 exceeded what they could cope with, and governments and taxpayers had to bear the costs. To capture this in our model, we assume that the agent has limited liability and cannot be held responsible for these losses in excess of her current wealth, so that it is the principal who has to incurs the costs. As for operating profits, we assume that losses increase linearly with the size of the project. Thus, if there is a loss at time t, the corresponding cost is X_tC, where $C > 0$ is the size-adjusted cost. Overall, the net output flow generated by the project during the infinitesimal time interval $(t, t + dt]$ is $X_t(\mu dt - C dN_t)$.

By exerting effort, the agent affects the probability with which losses occur: a higher effort reduces the probability $\Lambda_t dt$ that a loss occurs during $(t, t + dt]$. For simplicity, we consider only two levels of effort, corresponding to $\Lambda_t = \lambda > 0$ and $\Lambda_t = \lambda + \Delta\lambda$, with $\Delta\lambda > 0$. To model the cost of effort, we adopt the same convention as Holmström and Tirole (1997): if the agent shirks at time t, that is if $\Lambda_t = \lambda + \Delta\lambda$, she obtains a private benefit X_tB; by contrast, if the agent exerts effort at time t, that is if $\Lambda_t = \lambda$, she obtains no private benefit. This formulation is similar to one in which the agent incurs a constant cost per unit of time and per unit of size of the project when exerting effort, and no cost when shirking.

Remark It is natural to assume that operating profits and losses are increasing in the size of the project. It is also natural to assume that the opportunity cost of risk prevention is increasing in the size of the project: it takes more time, effort and energy to check compliance and monitor safety processes in two plants than in a single plant, or for a large trading room with many traders than for a small one. Note however that we demand more than monotonicity, since we assume that operating profits, losses and private benefits are linear in the size of the project. This constant returns to scale assumption is made

for tractability. As shown in Section 4, it implies that the value function solution to the Hamilton–Jacobi–Bellman equation (23) is homogeneous of degree one, which considerably simplifies the characterization of the optimal contract. Yet, even without this assumption, some of the qualitative features of our analysis are upheld, as discussed in Section 6.

We assume throughout the paper that

$$\frac{\mu - \lambda C}{r} > c \tag{1}$$

and that

$$\Delta \lambda C > B. \tag{2}$$

The left-hand side of (1) is the present value of the net expected cash-flow generated by one unit of capacity over an infinite horizon when the agent always exerts effort. The right-hand side of (1) is the cost of an additional unit of capacity. Condition (1) implies that the project has positive net present value and that investment is desirable when the agent always exerts effort. The left-hand side of (2) is the size-adjusted expected social cost of increased risk when the agent shirks. The right-hand side of (2) is the size-adjusted private benefit from shirking. Condition (2) implies that in the absence of moral hazard, it is socially optimal to require the agent to always exert effort. The first-best policy can therefore be characterized as follows: first, the project is initiated at its maximal capacity of 1, and then grows at the maximal feasible rate γ with no downsizing ever taking place; second, a maximal risk prevention policy is implemented in which the agent always exerts effort.

From now on, we focus on the case where there is asymmetric information. Specifically, we assume that, unlike profits and losses, the agent's effort decisions are not observable by the principal. This leads to a moral hazard problem, whose key parameters are B and $\Delta \lambda$. The larger the size-adjusted private benefit B is, the more attractive it is for the agent to shirk. The lower $\Delta \lambda$ is, the more difficult it is to detect shirking. The contract between the principal and the agent is designed and agreed upon at time 0. The agent reacts to this contract by choosing an effort process $\Lambda = \{\Lambda_t\}_{t \geq 0}$. We assume that the players can fully commit to a long-term contract.

Remark Throughout the paper, we thus abstract from imperfect commitment problems and focus on a single source of market imperfection: moral hazard in risk prevention. This assumption is standard in the dynamic moral hazard literature, see for instance Rogerson (1985), Spear and Srivastava (1987) or Phelan and Townsend (1991). More precisely, our analysis is in line with Clementi and Hopenhayn (2006), DeMarzo and Sannikov (2006), Biais,

Mariotti, Plantin and Rochet (2007), DeMarzo and Fishman (2007a, 2007b) or Sannikov (2008), where limited liability reduces the ability to punish the agent. This compels the principal to substitute to such punishments actions, such as downsizing or liquidation, that are ex-post inefficient.[13] When the principal is more patient than the agent and there is no investment, as in DeMarzo and Fishman (2006) and Biais, Mariotti, Plantin and Rochet (2007), this leads to the result that the firm eventually ceases to exist. By contrast, in the present model, this negative trend can be outweighed by investment.

A contract specifies downsizing, investment and liquidation decisions, as well as payments to the agent, as functions of the history of past losses. The size process $X = \{X_t\}_{t \geq 0}$ is positive, with initial condition $X_0 \leq 1$. The size of the project can be decomposed as:

$$X_t = X_0 + X_t^d + X_t^i \tag{3}$$

for all $t \geq 0$, where $X^d = \{X_t^d\}_{t \geq 0}$, the cumulative downsizing process, is decreasing, and $X^i = \{X_t^i\}_{t \geq 0}$, the cumulative investment process, is increasing. Our assumptions imply that X^i is absolutely continuous with respect to time, that is:

$$X_t^i = \int_0^t X_s g_s \, ds, \tag{4}$$

where the instantaneous growth rate of the project satisfies

$$0 \leq g_t \leq \gamma \tag{5}$$

for all $t \geq 0$. The positive and increasing process describing the cumulative transfers to the agent is denoted by $L = \{L_t\}_{t \geq 0}$. The time at which liquidation occurs, if ever, is denoted by τ. We allow τ to be infinite and we let $X_t = 0$ and $L_t = L_\tau$ for all $t > \tau$.

At any time t prior to liquidation, the sequence of events during the infinitesimal time interval $[t, t + dt]$ can heuristically be described as follows:

1. The size X_t of the project is determined, that is, there is downsizing, or investment, or the size remains constant.

2. The agent takes her effort decision Λ_t.

3. With probability $\Lambda_t dt$, there is a loss, in which case $dN_t = 1$.

4. The agent receives a positive transfer dL_t.

---

[13]For a discussion of renegotiation in this context, see Quadrini (2004), DeMarzo and Sannikov (2006, Section IV.B) or DeMarzo and Fishman (2007a, Appendix B2, 2007b, Section 2.9).

5. The project is either liquidated or continued.

According to this timing, the downsizing and effort decisions are taken before knowing the current realization of the loss process. Formally, the processes X and Λ are $\mathcal{F}^N$–*predictable*, where $\mathcal{F}^N = \{\mathcal{F}^N_t\}_{t \geq 0}$ is the filtration generated by N. By contrast, payment and liquidation decisions at any time are taken after observing whether or not there was a loss at this time. Hence L is $\mathcal{F}^N$–*adapted* and τ is an $\mathcal{F}^N$–*stopping time*.[14] An effort process Λ generates a unique probability distribution $\mathbf{P}^\Lambda$ over the paths of the process N. Denote by $\mathbf{E}^\Lambda$ the corresponding expectation operator.

Given a contract $\Gamma = (X, L, \tau)$ and an effort process Λ, the expected discounted utility of the agent is

$$\mathbf{E}^\Lambda\left[\int_0^\tau e^{-\rho t}(dL_t + 1_{\{\Lambda_t = \lambda + \Delta\lambda\}}X_t B\, dt)\right], \tag{6}$$

while the expected discounted profit of the principal is[15]

$$\mathbf{E}\left[\int_0^\tau e^{-rt}\{X_t[(\mu - g_t c)dt - C dN_t] - dL_t\}\right]. \tag{7}$$

An effort process Λ is incentive compatible with respect to a contract Γ if it maximizes the agent's expected utility (6) given Γ. The problem of the principal is to find a contract Γ and an incentive compatible effort process Λ that maximize its expected discounted profit (7), subject to delivering to the agent a required expected discounted utility level. It is without loss of generality to focus on contracts Γ such that the present value of the payments to the agent is finite, that is:

$$\mathbf{E}^\Lambda\left[\int_0^\tau e^{-\rho t}dL_t\right] < \infty. \tag{8}$$

Indeed, by inspection of (7), if the present value of the payments to the agent were infinite, the fact that $\rho > r$ would imply infinitely negative expected discounted profits for the principal. The latter would be better off proposing no contract altogether.

3 Incentive Compatibility and Limited Liability

To characterize incentive compatibility, we rely on martingale techniques similar to those introduced by Sannikov (2008). When taking her effort decision at a time t, the agent

---

[14]See for instance Dellacherie and Meyer (1978, Chapter IV, Definitions 12, 49 and 61) for definitions of these concepts.

[15]All integrals are of the Lebesgue–Stieltjes kind. For each s and t, we write $\int_s^t$ for $\int_{[s,t]}$ and $\int_s^{t^-}$ for $\int_{[s,t)}$.

considers how it will affect her continuation utility, defined as:

$$W_t(\Gamma, \Lambda) = \mathbf{E}^\Lambda\left[\int_t^\tau e^{-\rho(s-t)}(dL_s + 1_{\{\Lambda_s = \lambda + \Delta\lambda\}} X_s B ds)\,|\,\mathcal{F}_t^N\right] 1_{\{t < \tau\}}. \tag{9}$$

Denote by $W(\Gamma, \Lambda) = \{W_t(\Gamma, \Lambda)\}_{t \geq 0}$ the agent's continuation utility process. Note that, by construction, $W(\Gamma, \Lambda)$ is $\mathcal{F}^N$–adapted. In particular, $W_t(\Gamma, \Lambda)$ reflects whether or not there was a loss at time t. To characterize how the agent's continuation utility evolves over time, it is useful to consider her lifetime expected utility, evaluated conditionally upon the information available at time t, that is:[16]

$$U_t(\Gamma, \Lambda) = \mathbf{E}^\Lambda\left[\int_0^\tau e^{-\rho s}(dL_s + 1_{\{\Lambda_s = \lambda + \Delta\lambda\}} X_s B ds)\,|\,\mathcal{F}_t^N\right]$$

$$= \int_0^{t \wedge \tau^-} e^{-\rho s}(dL_s + 1_{\{\Lambda_s = \lambda + \Delta\lambda\}} X_s B ds) + e^{-\rho t} W_t(\Gamma, \Lambda). \tag{10}$$

Since $U_t(\Gamma, \Lambda)$ is the expectation of a given random variable conditional on $\mathcal{F}_t^N$, the process $U(\Gamma, \Lambda) = \{U_t(\Gamma, \Lambda)\}_{t \geq 0}$ is an $\mathcal{F}^N$–martingale under the probability measure $\mathbf{P}^\Lambda$. Its last element is $U_\tau(\Gamma, \Lambda)$, which is integrable by (8).

Relying on this martingale property, we now offer an alternative representation of $U(\Gamma, \Lambda)$. Consider the process $M^\Lambda = \{M_t^\Lambda\}_{t \geq 0}$ defined by

$$M_t^\Lambda = N_t - \int_0^t \Lambda_s\, ds \tag{11}$$

for all $t \geq 0$. Equation (11) is best understood when Λ is a constant process. In that case, M_t^Λ is simply the number of losses up to and including time t, minus its expectation. More generally, a basic result from the theory of point processes is that M^Λ is an $\mathcal{F}^N$–martingale under $\mathbf{P}^\Lambda$. Changes in the effort process Λ induce changes in the distribution of losses, which essentially amount to Girsanov transformations of the process N. The martingale representation theorem for point processes then implies the following lemma.[17]

Lemma 1 *The martingale $U(\Gamma, \Lambda)$ satisfies*

$$U_t(\Gamma, \Lambda) = U_0(\Gamma, \Lambda) - \int_0^{t \wedge \tau} e^{-\rho s} H_s(\Gamma, \Lambda)\, dM_s^\Lambda \tag{12}$$

for all $t \geq 0$, $\mathbf{P}^\Lambda$–almost surely, for some $\mathcal{F}^N$–predictable process $H(\Gamma, \Lambda) = \{H_t(\Gamma, \Lambda)\}_{t \geq 0}$.

Along with (11), (12) implies that the lifetime expected utility of the agent evolves in

---

[16]For each x and y, we denote by $x \wedge y$ the minimum of x and y, and by $x \vee y$ the maximum of x and y.

[17]See for instance Brémaud (1981, Chapter III, Theorems T9 and T17, and Chapter VI, Theorems T2 and T3) for the relevant results.

response to the jumps of the process N. At any time t, the change in $U_t(\Gamma, \Lambda)$ is equal to the product between a $\mathcal{F}^N$–predictable function of the past, namely $e^{-\rho t} H_t(\Gamma, \Lambda)$, and a term $-dM_t^\Lambda$ reflecting the events occurring at time t. This term is in turn equal to the difference between the instantaneous probability $\Lambda_t dt$ of a loss, and the instantaneous change dN_t in the total number of losses, which is equal to 0 or 1. Equations (10) and (12) imply that the continuation utility of the agent evolves as:

$$dW_t(\Gamma, \Lambda) = [\rho W_t(\Gamma, \Lambda) - 1_{\{\Lambda_t = \lambda + \Delta \lambda\}} X_t B] dt + H_t(\Gamma, \Lambda)(\Lambda_t dt - dN_t) - dL_t \qquad (13)$$

for all $t \in [0, \tau)$. Equation (13) states that, net of private benefits and wages, the expected instantaneous change in the continuation utility of the agent is equal to her discount rate ρ, while $H(\Gamma, \Lambda)$ is the sensitivity to losses of this utility. Building on this analysis, and denoting $b = \frac{B}{\Delta \lambda}$, we obtain the following result, in line with Sannikov (2008, Proposition 2).

Proposition 1 *A necessary and sufficient condition for the effort process Λ to be incentive compatible given the contract $\Gamma = (X, L, \tau)$ is that*

$$\Lambda_t = \lambda \ \ \text{if and only if} \ \ H_t(\Gamma, \Lambda) \geq X_t b \qquad (14)$$

for all $t \in [0, \tau)$, $\mathbf{P}^\Lambda$–almost surely.

It follows from (13) that, if there is a loss at some time $t \in [0, \tau)$, the agent's continuation utility must be instantaneously reduced by an amount $H_t(\Gamma, \Lambda)$.[18] Proposition 1 states that, in order to induce the agent to choose a high level of risk prevention, this reduction in her continuation utility must be at least as large as $X_t b$. This is because $X_t b$ reflects the attractiveness of the private benefits obtained by the agent when shirking. To reason in size-adjusted terms, let $h_t = \frac{H_t}{X_t}$. The incentive compatibility condition (14) under which $\Lambda_t = \lambda$ then rewrites as:

$$h_t \geq b. \qquad (15)$$

It is convenient to introduce the notation $W_{t^-}(\Gamma, \Lambda) = \lim_{s \uparrow t} W_s(\Gamma, \Lambda)$ to denote the left-hand limit of the process $W(\Gamma, \Lambda)$ at $t > 0$. While $W_t(\Gamma, \Lambda)$ is the continuation utility of the agent at time t after observing whether or not there was a loss at time t, $W_{t^-}(\Gamma, \Lambda)$ is

---

[18]In full generality, one should also allow for jumps in the transfer process. For incentive reasons, it is however never optimal to pay the agent when a loss occurs. Moreover, it will turn out that the optimal transfer process is absolutely continuous, so that payments do not come in lump-sums. To ease the exposition, we therefore rule out jumps in the transfer process in the body of the paper. The possibility of such jumps is explicitly taken into account below in the verification theorem.

the continuation utility of the agent evaluated before such knowledge is obtained.[19] Observe that, while the process $W(\Gamma, \Lambda)$ is $\mathcal{F}^N$–adapted, the process $W_{\cdot-}(\Gamma, \Lambda) = \{W_{t-}(\Gamma, \Lambda)\}_{t \geq 0}$ is $\mathcal{F}^N$–predictable. Combining the fact that the continuation utility of the agent must remain positive according to the limited liability constraint, with the fact that it must be reduced by an amount $H_t(\Gamma, \Lambda)$ if there is a loss at time t according to (13), one must have

$$W_{t-}(\Gamma, \Lambda) \geq H_t(\Gamma, \Lambda) \tag{16}$$

for all $t \in [0, \tau)$. To simplify notation, we shall drop the arguments Γ and Λ in the remainder of the paper.

4 Optimal Contracting with Maximal Risk Prevention

While in the previous section we considered general effort processes, in the present section we characterize the optimal contract that induces maximal risk prevention, that is $\Lambda_t = \lambda$ for all $t \in [0, \tau)$. This is in line with most of the literature on the principal-agent model, which offers more precise insights into how to implement given courses of actions at minimal cost than into which course of actions is, all things considered, optimal for the principal.[20] In Section 6.1, however, we shall provide sufficient conditions under which it is optimal for the principal to request maximal risk prevention from the agent. The optimal contract that we derive in this section can be described with the help of two state variables: the size of the project, resulting from past downsizing and investment decisions, and the continuation utility of the agent, reflecting future payment decisions. To build intuition, we first provide a heuristic derivation of the principal's value function and of the main features of the optimal contract. Next, we verify that this candidate value function is indeed optimal, and we fully characterize the optimal contract.

4.1 A Heuristic Derivation

In this heuristic derivation, we suppose that transfers are absolutely continuous with respect to time, and that no payment is made after a loss, that is:

$$dL_t = X_t \ell_t 1_{\{dN_t = 0\}} dt \tag{17}$$

where

$$\ell_t \geq 0 \tag{18}$$

---

[19]$W_{0-}(\Gamma, \Lambda)$ is defined by (6).

[20]See for instance Laffont and Martimort (2002, Chapters 4 and 8) for a recent overview of that literature.

for all $t \geq 0$. Here $\{\ell_t\}_{t \geq 0}$, is assumed to be an $\mathcal{F}^N$–predictable process representing the size-adjusted transfer flow to the agent. We will later verify that this conjecture is correct at the optimal contract. Now consider project size. Downsizing is suboptimal in the first-best, and, as we will later verify, it remains so in the second-best as long as no losses occur. After losses, however, downsizing may prove necessary in the second-best. This reflects that, for incentive purposes, it is necessary to reduce the agent's continuation utility after each loss by an amount that is proportional to her private benefits from shirking. The latter, in turn, are proportional to the size of the project. When the continuation utility of the agent is low, the incentive compatibility constraint is compatible with the limited liability constraint only if the size of the project is itself low enough.

To see this more precisely, suppose that, at the outset of time t, the size of the project is X_t and the continuation utility of the agent is W_{t-}. If there is a loss at time t, the agent's continuation utility must be reduced from W_{t-} to $W_t = W_{t-} - X_t h_t$. At this point, the question arises whether this loss implies that the project should be downsized. Denote by $X_{t+} = \lim_{s \downarrow t} X_s \in [0, X_t]$ the size of the project just after time t. Since effort is still required from the agent, Proposition 1 implies that, if there were a second loss, arbitrarily close to the first, the continuation utility of the agent would have to be reduced further by at least $X_{t+} b$. This would be consistent with limited liability only if $W_{t-} - X_t h_t \geq X_{t+} b$, or, equivalently, letting $w_t = \frac{W_{t-}}{X_t}$ and $x_t = \frac{X_{t+}}{X_t}$, if

$$\frac{w_t - h_t}{b} \geq x_t. \tag{19}$$

Hence, downsizing is necessary after the first loss, that is $x_t < 1$, whenever the initial size-adjusted continuation utility w_t of the agent is so low that $\frac{w_t - h_t}{b} < 1$.

We are now ready to characterize the evolution of the continuation value $F(X_t, W_{t-})$ of the principal. Since the principal discounts the future at rate r, his expected flow of value at time t is given by

$$rF(X_t, W_{t-}). \tag{20}$$

This must be equal to the sum of the expected instantaneous cash-flows and of the expected rate of change in his continuation value. The former are equal to the expected net cash-flow from the project, minus the cost of investment and the expected payment to the agent. By (4) and (17), this yields

$$X_t[\mu - \lambda C - g_t c - \ell_t(1 - \lambda dt)]. \tag{21}$$

14

To evaluate the expected rate of change in the principal's continuation value, we use the dynamics (3) of the project's size along with that of the agent's continuation utility, setting $\Lambda_t = \lambda$ in (13). Applying the change of variable formula for processes of bounded variation, which is the counterpart of Itô's formula for these processes, this yields[21]

$$[\rho W_{t-} + X_t(\lambda h_t - \ell_t)]F_W(X_t, W_{t-}) + X_t g_t F_X(X_t, W_{t-})$$
$$- \lambda[F(X_t, W_{t-}) - F(X_t x_t, W_{t-} - X_t h_t)]. \tag{22}$$

The first term arises because of the drift of $W_{.-}$, the second corresponds to investment, and the third reflects the possibility of jumps in the project's size and in the agent's continuation utility due to losses. Adding (22) to (21), identifying to (20), and letting dt go to 0, we obtain that the value function of the principal satisfies the Hamilton–Jacobi–Bellman equation

$$rF(X_t, W_{t-}) = X_t(\mu - \lambda C) + \max\left\{-X_t\ell_t + [\rho W_{t-} + X_t(\lambda h_t - \ell_t)]F_W(X_t, W_{t-})\right.$$
$$+ X_t g_t[F_X(X_t, W_{t-}) - c] \tag{23}$$
$$\left. - \lambda[F(X_t, W_{t-}) - F(X_t x_t, W_{t-} - X_t h_t)]\right\},$$

where the maximization in (23) is over the set of controls (g_t, h_t, ℓ_t, x_t) that satisfy constraints (5), (15), (18) and (19).

To get more insight into the structure of the solution, we impose further restrictions on the function F, that we will later check to be satisfied at the optimal contract. First, because of constant returns to scale, it is natural to require F to be homogenous of degree 1,

$$F(X, W) = XF\left(1, \frac{W}{X}\right) = Xf\left(\frac{W}{X}\right)$$

for all $(X, W) \in \mathbb{R}_{++} \times \mathbb{R}_+$. Intuitively, f maps the size-adjusted continuation utility w_t of the agent into the size-adjusted continuation value of the principal. Second, we require f to be globally concave. This property, which will be formally established in the verification theorem below, has the following economic interpretation. As argued above, while downsizing is inefficient in the first-best, it is necessary in the second-best to provide incentives to the agent when w_t is low. When this is the case, the principal's value reacts strongly to bad performance because the latter significantly raises the risk of costly downsizing. By contrast, when w_t is large, bad performance has a more limited impact on downsizing risk. This greater sensitivity to shocks when w_t is low than when it is large is reflected in the concavity of the

---

[21]See for instance Dellacherie and Meyer (1982, Chapter VI, Section 92).

size-adjusted value function f. Finally, we set

$$f(w) = \frac{f(b)}{b} w$$

for all $w \in [0, b]$. This is just by convention, and to simplify the notation, since, by (14) and (16), w_t never enters the interval $[0, b)$.

We can now derive several properties of the optimal controls in the Hamilton–Jacobi–Bellman equation. Optimizing with respect to ℓ_t and using the homogeneity of F yields

$$f'(w_t) = F_W(X_t, W_{t-}) \geq -1, \tag{24}$$

with equality only if $\ell_t > 0$. Intuitively, the left-hand side of (24) is the increase in the principal's continuation value due to a marginal increase in the agent's continuation utility, while the right-hand side is the marginal cost to the principal of making an immediate payment to the agent. It is optimal to delay payments as long as they are more costly than utility promises, that is, as long as the inequality in (24) is strict. The concavity of f implies that this is the case when w_t is below a given threshold, which we denote by w^p. The optimal contract thus satisfies the following property.

Property 1 *Payments to the agent are made only if her size-adjusted continuation utility is at least w^p. The payment threshold w^p satisfies*

$$f'(w^p) = -1. \tag{25}$$

In the first-best, all payments to the agent would be made at time 0, as she is more impatient than the principal. By contrast, in the second-best, payments must be delayed and made contingent to a long enough record of good performance, in order to provide incentives to the agent. Since f is concave, it follows from (24) and (25) that $f'(w) = -1$ for all $w \geq w^p$. If one were to start from that region, the optimal contract would entail the immediate payment of a lump-sum $w - w^p$ to the agent, counterbalanced by a drop of her size-adjusted continuation utility to w^p.

Suppose that w_t is below the threshold w^p, implying that $\ell_t = 0$. Then, using the homogeneity of F, one can rewrite (23) as follows:

$$rf(w_t) = \mu - \lambda C + \max \Big\{ (\rho w_t + \lambda h_t) f'(w_t) + g_t[f(w_t) - w_t f'(w_t) - c]$$

$$- \lambda \Big[f(w_t) - x_t f \Big(\frac{w_t - h_t}{x_t} \Big) \Big] \Big\}. \tag{26}$$

16

Since f is concave and vanishes at 0, the mapping $x_t \mapsto x_t f\left(\frac{w_t - h_t}{x_t}\right)$ is increasing. It is thus optimal to let x_t be as high as possible in (26), reflecting that downsizing is costly since the project is profitable. Using (19) along with the fact that $x_t \leq 1$ then leads to the second property of the optimal contract.

Property 2 *The optimal downsizing policy if there is a loss at time t is given by*

$$x_t = \frac{w_t - h_t}{b} \wedge 1. \tag{27}$$

This property of the optimal contract reflects that downsizing is imposed only as the last resort, in order to maintain the consistency between the incentive compatibility constraint and the limited liability constraint. Using our convention that f is linear over $[0, b]$, one can substitute (27) into (26) to obtain

$$rf(w_t) = \mu - \lambda C + \max\left\{(\rho w_t + \lambda h_t)f'(w_t) + g_t[f(w_t) - w_t f'(w_t) - c]\right. \tag{28}$$
$$\left. - \lambda[f(w_t) - f(w_t - h_t)]\right\}.$$

The concavity of f implies that it is optimal to let h_t be as low as possible in (28), which according to the incentive compatibility condition (15) leads to the third property of the optimal contract.

Property 3 *The sensitivity to losses of the agent's continuation utility is given by*

$$h_t = b. \tag{29}$$

Intuitively, (29) reflects that because the principal's continuation value is concave in the agent's continuation utility, it is optimal to reduce the agent's exposure to risk by letting h_t equal the minimal value consistent with her exerting effort.

Finally turn to investment decisions. Note that the size-adjusted social value of the project, $f(w) + w$, is increasing in w until w^p and flat afterwards. A necessary and sufficient condition for investment to ever take place is that the maximal size-adjusted social value of the project be larger than the unit cost of investment:

$$f(w^p) + w^p > c. \tag{30}$$

If (30) does not hold, the value of investment is lower than its cost, and hence it is suboptimal to invest. Thus, as will be checked below in the verification theorem, there is some investment in the optimal contract only if c is not too high. Optimizing in (28) with respect to g_t under

constraint (5), we obtain that $g_t = \gamma$ if

$$f(w_t) - w_t f'(w_t) > c, \tag{31}$$

and $g_t = 0$ otherwise. The left-hand side of (31) is the marginal benefit of an additional capacity unit, while the right-hand side is the unit cost of investment. Scale expansion is optimal when the former is greater than the latter. In that case, because of the linearity in the technology, size grows at the maximal feasible rate γ. The concavity of f implies that (31) holds when w_t is above a given threshold, which we denote by w^i. The optimal contract thus satisfies the following property.

Property 4 *Investment takes place, at rate γ, if and only if the size-adjusted continuation utility of the agent is above w^i. The investment threshold w^i satisfies*

$$w^i = \inf \{ w > b \,|\, f(w) - w f'(w) > c \}. \tag{32}$$

In the first-best, because of condition (1), investment always takes place at the maximal rate γ. By contrast, in the second-best, if c is not too low, this is the case only if a long enough record of good performance has been accumulated. This is because increasing the size of the project raises the private benefits from shirking and thus worsens the moral hazard problem. This jeopardizes incentives, except if the agent has enough at stake to still prefer high effort, that is, only if w_t is large enough. An important alternative scenario arises whenever c is low enough. In that case, inequality (31) is satisfied for all $w_t > b$, so that $w^i = b$ and it is always optimal to invest, even in the second-best. Formally, this is reflected in the fact that the function f is not differentiable at b, with $f'_-(b) = \frac{f(b)}{b} > f'_+(b)$, so that $f(b) - b f'_+(b) > c$ for c close enough to 0.

The dynamics of the principal's size-adjusted continuation value depends on whether or not there is investment. In the no investment region $(b, w^i]$, one has

$$rf = \mu - \lambda C + \mathcal{L}f, \tag{33}$$

where the delay differential operator $\mathcal{L}$ is defined by

$$\mathcal{L}f(w) = (\rho w + \lambda b) f'(w) - \lambda [f(w) - f(w - b)]. \tag{34}$$

In the investment region $(w^i, w^p]$, one has

$$(r - \gamma) f = \mu - \lambda C - \gamma c + \mathcal{L}_\gamma f, \tag{35}$$

18

where the delay differential operator $\mathcal{L}_\gamma$ is defined by

$$\mathcal{L}_\gamma f(w) = [(\rho - \gamma)w + \lambda b]f'(w) - \lambda[f(w) - f(w - b)]. \tag{36}$$

Comparing equations (35) and (36) to equations (33) and (34) reveals that, besides the decrease γc in the size-adjusted cash-flow, the impact of investment at rate γ is comparable to that of a decrease γ in both the principal's and the agent's discount rates. Intuitively, this reflects that investment makes delaying payments less costly, because the total size of the pie grows while the players wait. Thus, although incentive considerations imply that both investment and payments should be delayed relative to the first-best, investment takes place before payments do, as stated now.

Property 5 *The investment threshold w^i is strictly lower than the payment threshold w^p.*

This follows from evaluating (31) at w^p, which yields $f(w^p) - w^p f'(w^p) > c$ because of (25) and (30). While investment takes place in a region where the size-adjusted social value of the project is strictly increasing, payments are made to the agent when the size-adjusted social value of the project reaches its maximum, so that it is inefficient to delay payments any longer. At the payment threshold w^p, transfers are constructed in such a way that the agent's continuation utility stays constant until there is a loss. That is, they are set to the highest level consistent with the size-adjusted social value remaining at its maximum. This level can be computed as follows. Setting $\Lambda_t = \lambda$ in (13) and making use of (17) and Property 3, we obtain that

$$dW_t = (\rho W_t + X_t \lambda b)dt - X_t b dN_t - X_t \ell_t 1_{\{dN_t = 0\}}dt. \tag{37}$$

Suppose now that $w_t = w^p$, so that the size-adjusted social value of the project is at its maximum, and that $dN_t = 0$, so that there is no loss at time t. Then $W_t = X_t w^p$ and $dX_t = X_t \gamma dt$. Substituting in (37), we obtain the following property.

Property 6 *The size-adjusted transfer flow if there is no loss at time t is*

$$\ell_t = [(\rho - \gamma)w^p + \lambda b]1_{\{w_t = w^p\}}. \tag{38}$$

According to (38), when payments are made at the payment threshold w^p, they come in a steady flow in size-adjusted terms until a loss occurs.

4.2 The Verification Theorem

We now show that the above heuristic characterization does correspond to the optimal

contract. To do this, we first show that there exists a size-adjusted value function f such that Properties 1 to 6 hold.

Proposition 2 *Suppose that*

$$\mu - \lambda C > (\rho - r)b\left(2 + \frac{r}{\lambda}\right). \tag{39}$$

Then there exists a constant $\bar{c} > 0$ such that if

$$c < \bar{c}, \tag{40}$$

the delay differential equation

$$
\begin{cases}
f(w) = \frac{f(b)}{b}\, w & \text{if } w \in [0, b], \\[2mm]
rf(w) = \mu - \lambda C + \mathcal{L}f(w) & \text{if } w \in (b, w^i], \\[2mm]
(r - \gamma)f(w) = \mu - \lambda C - \gamma c + \mathcal{L}_\gamma f(w) & \text{if } w \in (w^i, w^p], \\[2mm]
f(w) = f(w^p) + w^p - w & \text{if } w \in (w^p, \infty)
\end{cases}
\tag{41}
$$

has a maximal solution f, where the thresholds w^p and w^i are endogenously determined by (25) and (32), with $w^p > w^i$, and the operators $\mathcal{L}$ and $\mathcal{L}_\gamma$ are defined as in (34) and (36). The function f is globally concave and continuously differentiable except at b.

If condition (39) did not hold, the solution would be degenerate, with downsizing taking place after each loss. This would arise because the private benefits from shirking would be large relative to the expected cash-flow from the project, making the agency problem very severe. Condition (40) ensures that the investment cost is low enough so that there are circumstances in which it is optimal to increase the size of the project. If condition (40) did not hold, the optimal contract would be similar to that described above, except that there would be no investment region. The threshold value $\bar{c}$ corresponds to the maximum of the size-adjusted social value of the project arising in this no investment situation.

The next step of the analysis is to show that the function constructed in Proposition 2 yields the maximal value that can be obtained by the principal, and to explicitly construct the optimal contract. To do so, fix an initial project size X_0 and an initial expected utility W_{0-} for the agent, and consider the processes $\{w_t\}_{t \geq 0}$ and $\{l_t\}_{t \geq 0}$ solutions to

$$w_t = w_0 + \int_0^{t^-} \left\{ [(\rho - \gamma 1_{\{w_s > w^i\}})w_s + \lambda b]\, ds - b\left(\frac{w_s - b}{b} \wedge 1\right) dN_s - dl_s \right\}, \tag{42}$$

20

$$l_t = (w_0 - w^p) \vee 0 + \int_0^t [(\rho - \gamma)w^p + \lambda b] 1_{\{w_{s+} = w^p\}} \, ds \tag{43}$$

for all $t \geq 0$, where $w_0 = \frac{W_{0-}}{X_0}$, and w^i and w^p are defined as in Proposition 2. For the moment, we simply take these processes as given. Yet, consistent with the heuristic derivation of Section 4.1, it will eventually turn out in equilibrium that, at any time t, w_t is the initial size-adjusted continuation utility of the agent, while l_t represents cumulative size-adjusted transfers up to and including time t. The following proposition is central to our results.

Proposition 3 *Under conditions (39) and (40), the optimal contract $\Gamma = (X, L, \tau)$ that induces maximal risk prevention and delivers the agent an initial expected discounted utility W_{0-} given initial firm size X_0 is as follows:*

(i) *The project is downsized by a factor $\frac{w_{T_k} - b}{b} \wedge 1$ at any time T_k at which there is a loss. Moreover, the size of the project grows at rate γ as long as $w_t > w^i$, and at rate 0 otherwise. As a result of this, the size of the project is*

$$X_t = X_0 \prod_{k=1}^{N_{t-}} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \exp\left(\int_0^t \gamma 1_{\{w_s > w^i\}} \, ds \right) \tag{44}$$

at any time $t \geq 0$.[22]

(ii) *The flow of transfers to the agent is $X_t[(\rho - \gamma)w^p + \lambda b]$ as long as $w_t = w^p$ and no loss occurs. As a result of this, the cumulative transfers to the agent are*

$$L_t = X_0 l_0 + \int_0^t X_s \, dl_s \tag{45}$$

at any time $t \geq 0$.[23]

(iii) *Liquidation occurs with probability 0 on the equilibrium path:*

$$\tau = \infty, \tag{46}$$

   P–*almost surely.*

The value for the principal of this contract is $F(X_0, W_{0-}) = X_0 f\left(\frac{W_{0-}}{X_0}\right)$, with f constructed as in Proposition 2.

As shown in the proof of Proposition 3, the optimal contract entails at any time t a

---

[22]By convention, $\prod_\emptyset = 1$.

[23]Observe from (42) and (43) that $w_{t+} = w^p$ if and only if $w_t = w^p$ and there is no loss at time t.

continuation utility $W_t = \lim_{s \downarrow t} X_s w_s$ for the agent. The process W obtained in this way satisfies (13) with $\Lambda_t = \lambda$ and $H_t = X_t b$, and thus induces maximal risk prevention. As conjectured in Section 4.1, the optimal contract involves two state variables, the size of the project, X_t, and the size-adjusted continuation utility of the agent, w_t, or, equivalently, her beginning-of-period continuation utility $W_{t-} = X_t w_t$.

The main features of the optimal contract are also in line with the heuristic derivation of Properties 1 to 6. First consider transfers, as given by (43) and (45). If $w_0 > w^p$, an initial lump-sum is immediately distributed to the agent. Then, at time $t > 0$, transfers take place if and only if $w_t = w^p$ and there is no loss, and they are constructed in such a way that the agent's size-adjusted continuation utility stays constant until a loss occurs. This is consistent with Properties 1 and 6.

Next consider the size of the project, as given by (44). The first term on the right-hand side of (44) is the initial size of the firm. The second term on the right-hand side of (44) reflects downsizing, which takes place only after losses occur at the random times T_k and when $\frac{w_{T_k} - b}{b} < 1$. This is consistent with Properties 2 and 3. The third term on the right-hand side of (44) reflects that investment takes place, at rate γ, if and only if $w_t > w^i$. This is consistent with Properties 4 and 5.

Finally consider the size-adjusted continuation utility of the agent, as given by (42). Its dynamics is somewhat complicated, as it reflects the joint effect of direct changes in the agent's continuation utility and indirect changes due to the variations in the project's size. It follows from (42) that, if a loss occurs at a time T_k such that $w_{T_k} \geq 2b$, no downsizing takes place, and the size-adjusted continuation utility of the agent drops by an amount b. This is consistent with Property 3. By contrast, if a loss occurs at a time T_k such that $b \leq w_{T_k} < 2b$, the project is downsized by a factor $\frac{w_{T_k} - b}{b}$, and the size-adjusted continuation utility of the agent drops by an amount $w_{T_k} - b$. Thus in any case, the sensitivity to losses of the agent's size adjusted utility is $(w_{T_k} - b) \wedge b$.

It should be emphasized that liquidation plays virtually no role in the optimal incentive contract, as reflected by (46). Indeed, as can be seen from (42), $w_t = \frac{W_{t-}}{X_t}$ always remains strictly greater than b. As a result of this, W_t, which is in the worst case equal to $W_{t-} - X_t b$ if there is a loss at time t, always remains strictly positive.[24] This is in sharp contrast with the Brownian models studied by DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin and Rochet (2007) and Sannikov (2008), in which the optimal contract relies crucially on liquidation and involves no downsizing. Admittedly, even in the context of our Poisson

---

[24]Exceptions arise only with probability 0, for instance if $W_{0-} = X_0 b$ and there is a loss at time 0.

model, an alternative way to provide incentives to the agent in case of bad performance would be to allow for randomly liquidating the project, as is customary in discrete-time models (see for instance Clementi and Hopenhayn (2006) or DeMarzo and Fishman (2007a, 2007b)). But in contrast with what happens in Brownian models, liquidation would then necessarily have to be both *stochastic* (as it would depend on the realization of a lottery at each potential liquidation time) and *unpredictable* (as it would take place only after a loss). Modeling liquidation in this way would allow to achieve essentially the same outcome as under downsizing. This would however be both less tractable analytically, and less conducive to a realistic implementation of the optimal contract.

Proposition 3 describes the optimal contract for a given initial project size X_0 and a given initial expected discounted utility W_{0-} for the agent. In Biais, Mariotti, Rochet and Villeneuve (2009, Section D.3), we examine how these are determined at time 0 whenever the principal is competitive. That is, we look for a pair (X_0, W_{0-}) that maximizes utilitarian welfare under the constraint that the principal breaks even on average. As soon as f takes strictly positive values, it is optimal to start operating the project at full scale, $X_0 = 1$.[25] When the participation constraint of the principal is slack, the contract is initiated at the payment threshold $w_0 = w^p$, so that the agent is immediately compensated. By contrast, when the participation constraint of the principal binds, it is necessary to initiate the contract at a lower level $w_0 < w^p$, so that it is optimal to wait before compensating the agent.

5   Firm Size Dynamics

In this section, we build on the above analysis to study size dynamics under maximal risk prevention. Because of downsizing and investment, the scale of operations varies over time in the optimal contract. These variations can be interpreted as the dynamics of firm size. Our model generates a rich variety of possible paths for such dynamics. Over its life cycle, the firm can grow, stagnate or decline.

To illustrate this point, consider the following typical path, depicted in Figure 1. Firm size starts at the level X_0. As long as there is no loss, the size-adjusted continuation utility of the agent rises, and eventually reaches the investment threshold w^i. From this point on, investment takes place at rate γ and the firm grows. However, if a loss occurs at time T_k, the size-adjusted continuation utility of the agent drops from w_{T_k} to $w_{T_k^+} = (w_{T_k} - b) \vee b$. If this lower utility level is below w^i, investment stops and firm size remains constant. Furthermore, if $w_{T_k^+} < 2b$ and there is another loss shortly afterwards, downsizing is necessary.

---

[25]Otherwise it is optimal to let $X_0 = W_{0-} = 0$ and not to operate the project.

While, in the short-run, firm size can grow, stagnate or decline, it is unclear how it is likely to behave in the long run. Will downsizing bring it down to 0? Or will the firm grow indefinitely thanks to investment? To address this issue, we study the limit as t goes to ∞ of the average growth rate of the firm until time t. For simplicity set X_0 to 1. Then Proposition 3 implies that this average growth rate is equal to

$$\frac{\ln(X_t)}{t} = \frac{1}{t}\left[\sum_{k=1}^{N_{t-}} \ln\left(\frac{w_{T_k} - b}{b} \wedge 1\right) + \int_0^t \gamma 1_{\{w_s > w^i\}}\, ds\right]. \tag{47}$$

Now, let $\boldsymbol{\mu^w}$ be the unique invariant measure associated to the process $\{w_{T_k}\}_{k \geq 1}$ of the agent's size-adjusted continuation utility just before losses, let $\boldsymbol{\mu^{w+}}$ be the unique invariant measure associated to the process $\{w_{T_k^+}\}_{k \geq 1}$ of the agent's size-adjusted continuation utility just after losses, and let $\boldsymbol{\lambda}$ be the exponential distribution random variable with parameter λ. Then, using an appropriate law of large numbers for Markov ergodic processes, one can derive the following result.[26]

Proposition 4 *Under conditions (39) and (40), the long-run growth rate of the firm is*

$$\lim_{t \to \infty} \frac{\ln(X_t)}{t} = \lambda \int_{[b,2b)} \ln\left(\frac{w - b}{b}\right) \boldsymbol{\mu^w}(dw)$$

$$+ \gamma\left[1 - \lambda \int_{[b,w^i) \times \mathbb{R}_+} (t_{w,w^i} \wedge s)\, \boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}(dw, ds)\right], \tag{48}$$

$\mathbf{P}$*–almost surely, where*

$$t_{w,w^i} = \frac{1}{\rho} \ln\left(\frac{\rho w^i + \lambda b}{\rho w + \lambda b}\right)$$

is the time it takes for the agent's size-adjusted continuation utility to reach w^i when starting from $w \in [b, w^i)$, if there are no losses in the meanwhile.

The first term of the right-hand size of (48) reflects the impact of downsizing. Downsizing takes place when losses occur, which is more likely if the intensity λ of the loss process N is high, and when the size-adjusted continuation utility of the agent lies in the region $[b, 2b)$ where downsizing cannot be avoided whenever a loss occurs.

The second term of the right-hand size of (48) reflects the impact of investment. The latter takes place, at rate γ, when the size-adjusted continuation utility of the agent is above the investment threshold w^i. The term within brackets multiplying γ on the right-hand

---

[26]See for instance Stout (1974, Theorem 3.6.7).

side of (48) is the frequency with which the size-adjusted continuation utility of the agent is above w^i. To build intuition about this term, consider the time interval $(T_k, T_{k+1}]$ between two consecutive losses. There is no investment during this time interval as long as the size-adjusted continuation utility of the agent stays below w^i. The probability of that event depends on the value of the continuation utility of the agent at the beginning of the time interval, $w_{T_k^+}$, as well as on the length $T_{k+1} - T_k$ of this time interval. This is why there is a double integral in (48), with respect to the invariant measures $\boldsymbol{\mu^{w+}}$ and $\boldsymbol{\lambda}$ of these two independent random variables. The interpretation of the term in parentheses inside the double integral is that there is no investment during $(T_k, T_{k+1}]$ if $T_{k+1} - T_k < t_{w_{T_k^+}, w^i}$, that is, if a loss occurs before the size-adjusted continuation utility of the agent had the time to reach w^i starting from $w_{T_k^+} < w^i$.

To gain more insights into the long-run behavior of the size of the firm, consider for tractability the case where c is small, so that

$$f(b) - bf'_+(b) > c. \tag{49}$$

In that case, the optimal contract stipulates that investment should continuously take place at rate γ, and we obtain the following result.

Proposition 5 *Under conditions (39), (40) and (49), if γ is close to 0, then*

$$\lim_{t \to \infty} X_t = 0, \tag{50}$$

$\mathbf{P}$–*almost surely, while if $\gamma > \frac{\lambda^2}{\rho - r + \lambda}$, then*

$$\lim_{t \to \infty} X_t = \infty, \tag{51}$$

$\mathbf{P}$–*almost surely.*

First consider (50). In that case, the maximal feasible growth rate γ of the firm is low, so that the impact of investment is negligible. Now, to maintain incentive compatibility and limited liability, downsizing must take place when losses occur and the agent's size-adjusted continuation utility is close to its lower bound b. Because the stochastic process describing the agent's size-adjusted continuation utility is Markov ergodic over $[b, w^p]$, this situation will prevail an infinite number of times with probability 1. As a result of this, the size of the firm and the continuation utility of the agent must eventually go to 0.

Next consider (51). In that case, the maximal feasible growth rate γ of the firm is high relative to the frequency λ of losses, so that the positive effect of investment dominates the

negative effect of downsizing. Thus, in the long run, the firm becomes infinitely large. Note however that, even in this case, the long-run growth rate of the firm remains strictly lower than in the first-best, because of downsizing.

These two asymptotic results differ from the classic immiseration result of Thomas and Worrall (1990). In their model, the agent's continuation utility eventually diverges to $-\infty$. But the reason why this outcome obtains differs from the reason why, in our model, firm size goes to 0 when γ is low. Indeed, in Thomas and Worrall (1990), the period utility function of the agent is concave and unbounded below. Consequently, providing incentives is cheaper, the lower the agent's continuation utility is. This reflects the fact that the cost of obtaining a given spread in the agent's continuation utility is then lower. The principal thus has an incentive to let the agent's utility drift to $-\infty$. By contrast, in our model, the cost of incentive compatibility is high when the agent's continuation utility is low. This reflects the fact that limited liability makes it then more difficult to induce large variations in the agent's continuation utility. Yet, firm size can go to 0 if γ is low relative to λ, so that the effect of downsizing overcomes that of investment. If γ is high relative to λ, firm size goes to infinity. Now, the continuation utility of the agent is equal to her size-adjusted continuation utility, which by construction lives in $[b, w^p]$, multiplied by firm size. Hence in that case the continuation utility of the agent eventually goes to ∞, which is exactly the opposite of the immiseration result.

Proposition 5 provides parameter restrictions under which firm size X_t unambiguously goes to 0 or ∞ with probability 1 when t goes to ∞. More generally, for all parameter values, including those under which (49) does not hold, the following holds.

Proposition 6 *Under conditions (39) and (40), each of the events $\{\lim_{t\to\infty} X_t = 0\}$ and $\{\lim_{t\to\infty} X_t = \infty\}$ has either a probability 0 or 1 of occurring.*

The intuition for this result is twofold. First, as can be seen from (44), the events that firm size goes to 0 or to ∞ are tail events. That is, whether or not they occur depends on what happens in the long run, and not on what happens over any finite horizon. Second, the stochastic processes that drive the evolution of firm size satisfy a mixing property, which implies that tail events have either probability 0 or 1. Note that Proposition 6 does not assert that one of the events $\{\lim_{t\to\infty} X_t = 0\}$ and $\{\lim_{t\to\infty} X_t = \infty\}$ must occur with probability 1: both of them may have probability 0. What it rules out, for instance, is a scenario in which, with probability p, the size of the firm eventually goes to 0, while with probability $1 - p$ it eventually goes to ∞, for some $p \in (0, 1)$.

This asymptotic result sharply differs from that arising in Clementi and Hopenhayn (2006). In their long-run analysis, either the firm is eventually liquidated, or the first-best is eventually attained and the firm is never liquidated. Each of these absorbing outcomes has a strictly positive probability in the stationary distribution. This difference with our results stems from the fact that, in their model, the principal and the agent have identical discount rates, while in ours the agent is more impatient than the principal.[27] In Clementi and Hopenhayn (2006), because the principal and the agent are equally patient, it is costless to delay the agent's consumption while capitalizing it at the common discount rate. Hence, it is optimal to try and accumulate pledges to the agent until her savings are so high that she can buy the firm and implement the first-best policy. With some probability, the agent is lucky enough that such high performance is achieved and the first-best is attained. With the complementary probability, the agent is not as lucky, and liquidation eventually occurs. By contrast, in our model, delaying consumption is costly, since the agent is more impatient than the principal. It is therefore optimal to let her consume before the first-best is attained. This reduces the growth in the accumulated pledge to the agent, which, in turn, raises the risk of downsizing. Whenever the maximal investment rate is low, such downsizing eventually brings firm size to 0 with probability 1. Whenever the maximal investment rate is high, firm size tends to grow so fast that it eventually goes to ∞ in spite of downsizing. Note however that, in that case, the first-best is not attained, even in the long run, because moral hazard still slows down the rate at which the firm grows.

6 Robustness

In this section, we discuss the robustness of our results. We first provide sufficient conditions for the optimality of maximal risk prevention. Then, we briefly examine the case of non constant returns to scale.

6.1 Optimality of Maximal Risk Prevention

So far, our analysis focused on the optimal contract under maximal risk prevention. We now investigate under which circumstances it is actually optimal for the principal to require such a high level of effort from the agent. For simplicity, we conduct this analysis in the case where there is no investment, that is $\gamma = 0$.

---

[27]In Clementi and Hopenhayn's (2006) discrete-time model, unlike in our continuous-time model, identical discount rates for the principal and the agent do not preclude the existence of an optimal contract. A further difference is that they assume that capital fully depreciates from one period to the next, while there is no capital depreciation in our model.

Note that the contract characterized in Proposition 3 depends on B and $\Delta\lambda$ only through their ratio $b = \frac{B}{\Delta\lambda}$. Hence there is one degree of freedom in the parameters of the model, as one can scale B and $\Delta\lambda$ up or down while keeping b constant, leaving the optimal contract under maximal risk prevention unaffected. Intuition suggests that when $\Delta\lambda$ gets large, it is optimal to prevent losses as much as possible. To see why, observe that if a contract induced shirking during some infinitesimal time interval $[t, t + dt)$, the agent's continuation utility would not need to be affected were a loss to occur at time t. That is, one should have $H_t = 0$ in (13). Since it is optimal to make no transfers over $[t, t + dt)$ as the agent is shirking, (13) then implies that this would result in a change

$$dw_t = (\rho w_t - B)dt \tag{52}$$

in the agent's size-adjusted continuation utility. To determine whether requiring the agent to always exert effort is optimal, we compare the continuation value of the principal under maximal risk prevention to its counterpart when the agent shirks during $[t, t + dt)$ and then reverts to exerting effort. The former is greater than the latter if

$$f(w_t) \geq [\mu - (\lambda + \Delta\lambda)C]dt + e^{-rdt} f(w_t + dw_t), \tag{53}$$

where dw_t is given by (52). The first term on the right-hand side of (53) reflects the increased intensity of losses over $[t, t + dt)$ due to shirking, while the second term corresponds to the continuation value for the principal from requesting maximal risk prevention from time $t + dt$ on. Given (52), a first-order Taylor expansion in (53) leads to

$$rf(w_t) \geq \mu - (\lambda + \Delta\lambda)C + (\rho w_t - B)f'(w_t). \tag{54}$$

Unlike in (33), there is no delay term on the right-hand side of (54), because the agent's continuation utility is not sensitive to losses during the time interval $[t, t + dt)$. Maximal risk prevention is optimal if (54) holds for any value of $w_t > b$. One has the following result.

Proposition 7 *Suppose that $\gamma = 0$, and fix all the parameters of the model except B and $\Delta\lambda$, for which only the ratio $b = \frac{B}{\Delta\lambda}$ is fixed, so that an increase in B is compensated by a proportional increase in $\Delta\lambda$. Then there exists a threshold $\underline{\Delta\lambda} > 0$ such that the optimal contract involves maximal risk prevention for all $\Delta\lambda > \underline{\Delta\lambda}$.*

The intuition for this result is as follows. Both B and $\Delta\lambda$ affect the magnitude of the moral hazard problem and therefore the cost of incentives. However, under maximal risk prevention, they do so only via their ratio b; formally, this is reflected in the fact that

the function f depends on B and $\Delta\lambda$ only through b. Now, while an increase in $\Delta\lambda$ makes shirking easier to detect, and raises the value to the principal of a high level of risk prevention effort, an increase in B leaves this value unaffected. Hence, when one keeps b and thus the cost of incentives constant, increasing $\Delta\lambda$ raises the benefit of effort for the principal without affecting its cost. As a result, when $\Delta\lambda$ is sufficiently high, it is optimal for the principal to require the agent to always exert effort.

6.2   Non Constant Returns to Scale

Our analysis relies on the assumption that there are constant returns to scale. What can be said when one relaxes this assumption? Suppose for instance that the private benefits from shirking are equal to some increasing function $B(X)$ of firm size X, and, for simplicity, keep all our other assumptions unchanged. Incentive compatibility conditions are basically the same in that extension. The continuation utility of the agent writes as:

$$W_t(\Gamma, \Lambda) = \mathbf{E}^\Lambda\left[\int_t^\tau e^{-\rho(s-t)}(dL_s + 1_{\{\Lambda_s = \lambda + \Delta\lambda\}}B(X_s)ds)\,|\,\mathcal{F}_t^N\right]1_{\{t < \tau\}},$$

and the underlying martingale is still M^Λ, so that the martingale representation theorem applies and Lemma 1 continues to hold. Similarly, Proposition 1 is essentially unchanged, except that the incentive compatibility condition under which the agent exerts effort is now

$$H_t(\Gamma, \Lambda) \geq \frac{B(X_t)}{\Delta\lambda}.$$

Suppose now that the principal wants to implement maximal risk prevention. Then, as when returns to scale are constant, it will be necessary to downsize the project after a loss if the agent's continuation utility is too low. To see this more precisely, suppose that, at the outset of time t, the size of the project is X_t and the continuation utility of the agent is $W_{t^-}(\Gamma, \Lambda)$. If there is a loss at time t, incentive compatibility requires that the continuation utility be reduced by at least $\frac{B(X_t)}{\Delta\lambda}$. Downsizing can avoided at this point only if the new level of continuation utility is high enough that it is still possible to provide incentives while satisfying the limited liability constraint, that is, if

$$W_{t^-}(\Gamma, \Lambda) - \frac{B(X_t)}{\Delta\lambda} \geq \frac{B(X_t)}{\Delta\lambda}.$$

Thus downsizing must take place whenever $W_{t^-}(\Gamma, \Lambda) \leq \frac{2B(X_t)}{\Delta\lambda}$ and there is a loss at time t. Yet, it is hard to push the analysis of the optimal contract much further without assuming constant returns to scale. Indeed, the Hamilton–Jacobi–Bellman equation now writes as:

$$rF(X_t, W_{t^-}) = X_t(\mu - \lambda C) + \max \left\{ -X_t \ell_t + \left[\rho W_{t^-} + \lambda \frac{B(X_t)}{\Delta \lambda} - X_t \ell_t \right] F_W(X_t, W_{t^-}) \right.$$

$$+ X_t g_t [F_X(X_t, W_{t^-}) - c]$$

$$\left. - \lambda \left[F(X_t, W_{t^-}) - F\left(X_t x_t, W_{t^-} - \frac{B(X_t)}{\Delta \lambda} \right) \right] \right\},$$

and it is difficult to study this delay partial differential equation when $B(X)$ is not a linear function of X.

7 Empirical Implications

While, in the first-best, firms in our model should always invest, in the second-best the optimal contract stipulates that firms can invest only after a long enough record of good performance, at least when the unit cost of investment is not too low.[28] Such clauses are consistent with the empirical results of Kaplan and Strömberg (2004), who find that venture capital funding for new investment is contingent on financial and non-financial milestones. They also find that such conditioning is more frequent when the proxy for agency problems is more severe.

In our model, the optimal contract specifies that after good performance agents will be compensated, while after bad performance the firm will be partial liquidated. This is in line with the contractual clauses documented by Kaplan and Strömberg (2003). The circumstances in which downsizing takes place in the optimal contract can be interpreted as financial distress. This is in line with the empirical findings of Denis and Shome (2005), who report that financially distressed firms are often downsized.

The implications for downsizing of our model are also consistent with stylized facts for financial firms such as banks or insurance companies. Indeed, prudential regulation implies that the scale at which these firms are allowed to operate is proportionate to their capital. After losses, capital drops. Hence, in such circumstances, financial firms should be downsized if their capital before the loss was close to the regulatory requirement. This is similar to our optimal contract, provided one interprets W as a proxy for capital, which is natural, since both increase after good performance and decrease after bad performance.

In our model, small firms tend to be below the investment threshold. They are thus likely to be exposed to financial constraints on investment, as documented by Beck, Demirgüç-Kunt and Maksimovic (2005). Our model also predicts that small firms are relatively more fragile,

---

[28]Throughout this section, we assume that $f(b) - bf'_+(b) < c < \bar{c}$, so that $w^i > b$.

since a few negative shocks are enough to drive them into the zone where further losses would trigger downsizing. Conversely, large firms that have enjoyed long periods of sustained investment are more likely to have long records of good performance, which pushes them away from that zone. Overall, the probability of downsizing is decreasing in firm size. This is in line with the empirical findings of Dunne, Roberts and Samuelson (1989), who report that failure rates decline with increases in firm or plant size. Note however that the same logic implies that, according to our model, large firms should tend to have higher growth rates than smaller ones, while data suggest that on average the opposite is true, see Evans (1987a, 1987b) and Dunne, Roberts and Samuelson (1989). Interestingly, though, Dunne, Roberts and Samuelson (1989) find that this pattern is reversed in the case of multiplant firms: mean growth rates for plants owned by such firms tend to increase with size, reflecting that the tendency for growth rates of plants to decline with size is outweighed by a substantial fall in their failure rates. This evidence suggests that our analysis is particularly relevant for multiplant firms. A further testable implication of our model is that downsizing decisions should typically be followed by relatively long periods during which no investment takes place, corresponding to the time it takes for the firm to reach the investment threshold again and resume growing.

Gabaix and Landier (2008) note that different theoretical explanations have been offered for variations in CEO pay. While some analyses emphasize incentive problems, Gabaix and Landier (2008) propose to focus on firm size. Empirically, they find that CEO pay increases with firm size. Consistent with these results, our incentive theoretic analysis implies that the size of the firm and the compensation of the agent ought to be positively correlated: after a long stream of good performance, the scale of operations is large, and so are the payments to the agent. Conceptually, our analysis suggests that explanations based on size should not be divorced from explanations based on incentives, and that investment and managerial compensation are complementary incentive instruments, in line with the empirical findings of Kaplan and Strömberg (2003).

8 Conclusion

This paper analyzes the dynamic moral hazard problem arising when agents with limited liability must exert costly unobservable effort to reduce the likelihood of large but relatively infrequent losses. We characterize the optimal downsizing, investment and compensation policies and provide explicit formulae for firm size and its asymptotic growth rate.

Our analysis generates policy and managerial implications for the prevention of large

risks. Under moral hazard, managers' compensation should be based on long-term track records, and it should be reduced after accidents or large losses by an amount that increases with the private benefits from shirking and the extent to which shirking is difficult to detect. Losses in our model are negative externalities, since they affect society beyond the managers' or the firms' ability to pay for the damages they cause. It is therefore natural to think of the optimal dynamic contract as a regulatory tool. For instance, in the context of financial institutions, our analysis suggests that, to prevent large losses, the compensation of bank and insurance managers should be regulated. Moreover, changes in expected compensation are not sufficient to induce an appropriate level of risk prevention: in addition, downsizing and investment decisions should be made contingent on performance.

Our analysis also generates implications for firm size dynamics. Simon and Bonini (1958) and Ijiri and Simon (1964) analyze the link between the stochastic process according to which firms grow and the size distribution of firms. While these early works do not rely on the characterization of optimal investment policies, they have been embedded within the neoclassical framework, see for instance Lucas (1978) or Luttmer (2007, 2008). In these models, firm growth is limited by technology. In Lucas (1978) managerial skills are assumed to exhibit diminishing returns to scale, while in Luttmer (2008) it is assumed that, when ideas are replicated, their quality depreciates. Our modeling framework offers an opportunity to revisit these issues in a context where the endogenous limits to firm growth result from moral hazard. A key issue in models of the size distribution of firms is whether Gibrat's law holds, that is, whether firm growth is independent of firm size. This is not the case in our model, since firm size and downsizing and investment decisions are correlated in the optimal contract, being all functions of the agent's size-adjusted continuation utility process. It would be interesting, in further research, to analyze the implications of our analysis for the size distribution of firms.

Appendix: Sketches of Proofs

In this appendix, we shall merely outline the structure of the proofs. The interested reader will find complete proofs in the supplement to this paper (Biais, Mariotti, Rochet and Villeneuve (2009)). All the references thereafter made to sections and auxiliary results correspond to this supplementary document.

Proof of Lemma 1 (sketch). The predictable representation (12) of the martingale $U(\Gamma, \Lambda)$ follows from Brémaud (1981, Chapter III, Theorems T9 and T17). The factor

$e^{-\rho s}$ in (12) is just a convenient rescaling. ∎

Proof of Proposition 1 (sketch). The proof extends Sannikov's (2008, Proposition 2) arguments to the case where output is modeled as a point process. ∎

Proof of Proposition 2 (sketch). It turns out to be more convenient to work with the size-adjusted social value function, defined by $v(w) = f(w) + w$ for all $w \geq 0$. Just as f, the function v is linear over $[0, b]$. From (33) and (35), one has

$$rv(w) = \mu - \lambda C - (\rho - r)w + \mathcal{L}v(w) \tag{55}$$

for all $w \in (b, w^i]$, and

$$(r - \gamma)v(w) = \mu - \lambda C - \gamma c - (\rho - r)w + \mathcal{L}_\gamma v(w) \tag{56}$$

for all $w \in (w^i, w^p]$. The investment threshold w^i satisfies

$$w^i = \inf\{w > b \mid v(w) - wv'(w) > c\}, \tag{57}$$

while the payment threshold w^p satisfies

$$v'(w^p) = 0. \tag{58}$$

Finally, v is constant and equal to $v(w^p)$ over $[w^p, \infty)$. The proof consists of two main parts. In the first part of the proof (Section C.1), we first suppose that investment is not feasible, that is $\gamma = 0$. This allows us to pin down the constant $\bar{c}$ in (40), and provides key insights into the properties of the solution to (41) in the no investment region $(b, w^i]$. In the second part of the proof (Section C.2), we suppose that investment is feasible, that is $\gamma > 0$, and we use the results of the first part of the proof to solve (41).

Part 1 In the no investment case, we look for the maximal solution to (55) that satisfies (58) at some payment threshold. Note that the only unknown parameter is the slope of that solution over $[0, b]$. To determine that slope, we use the following shooting method. For each $\beta \geq 0$, denote by v_β the function that is linear with slope β over $[0, b]$ and then satisfies (55) over (b, ∞). One can show that v_β can be decomposed over $\mathbb{R}_+$ as $u_1 + \beta u_2$, where u_2 is a positive function with strictly positive derivative.[29] This implies that the derivatives of the functions $(v_\beta)_{\beta \geq 0}$ are strictly increasing with respect to β (Proposition C.1.1). We then prove that the ratio $-\frac{u_1'}{u_2'}$ attains a maximum β_0 over (b, ∞), which implies that v_{β_0} is the

---

[29]The functions u_1, u_2 and v_β are continuously differentiable except at b.

maximal function in the family $(v_\beta)_{\beta \geq 0}$ whose derivative has a zero in (b, ∞) (Proposition C.1.2). Thus v_{β_0} is the desired maximal solution. Let $w_{\beta_0}^p$ be the first point at which v_{β_0}' vanishes. The last step of the proof then consists in showing that v_{β_0} is concave over $[0, w_{\beta_0}^p]$, and strictly so over $[b, w_{\beta_0}^p]$ (Proposition C.1.3). As explained in the text, the cost threshold $\bar{c}$ below which investment is valuable is $v_{\beta_0}(w_{\beta_0}^p)$. For $c > \bar{c}$, the size-adjusted social value function is $v_{\beta_0} \wedge v_{\beta_0}(w_{\beta_0}^p)$ (Section D.2, Remark).

Part 2 In the investment case, we look for the maximal solution to (55) and (56) that satisfies (57) and (58) at some investment and payment thresholds. As in Part 1, the only unknown parameter is the slope of v over $[0, b]$. To determine that slope, which must clearly be higher than β_0, we use the following shooting method. For each $\beta \geq \beta_0$, denote by $v_{\beta,\gamma}$ the function that is linear with slope β over $[0, b]$ and then satisfies (55) over $(b, w_\beta^i]$ and (56) over (w_β^i, ∞), where $w_\beta^i = \inf \{w > b \,|\, v_{\beta,\gamma}(w) - w v_{\beta,\gamma}'(w) > c\}$. One may have $w_\beta^i = b$, in which case the region $(b, w_\beta^i]$ is empty. We first show that $v_{\beta,\gamma}$ is well-defined, and that the threshold w_β^i belongs to $[b, w_{\beta_0}^p]$ and continuously decreases as β increases (Lemma C.2.1). Key to this result is the fact that u_2 is strictly concave over $[b, \infty)$. We then show that, in analogy with the functions $(v_\beta)_{\beta \geq 0}$, the derivatives of the functions $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ are strictly increasing with respect to β (Proposition C.2.1). The next step of the proof, which is crucial, consists in showing that there exists a maximal function $v_{\beta_\gamma,\gamma}$ in the family $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ whose derivative has a zero in (b, ∞) (Proposition C.2.2). To establish this result, we first show that the set of $\beta \geq \beta_0$ such that $v_{\beta,\gamma}'$ has a zero over (b, ∞) is a nonempty interval I that contains β_0 (Lemma C.2.2). Second, we show that I has a finite upper bound β_γ, so that $v_{\beta,\gamma}'$ has no zero in (b, ∞) when $\beta > \beta_\gamma$ (Lemma C.2.3). Third, letting $w_{\beta,\gamma}^p$ be the first point at which $v_{\beta,\gamma}'$ vanishes for any given $\beta \in I$, we show that $w_{\beta,\gamma}^p$ is strictly increasing with respect to β over I and converges to a finite limit when β converges to β_γ from below (Lemma C.2.4). Fourth, we show that the derivatives of the functions $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ vary continuously with β, which in turn implies that I contains its upper bound β_γ (Lemma C.2.5). Thus $v_{\beta_\gamma,\gamma}$ is the desired maximal solution, and $w_{\beta_\gamma,\gamma}^p$ is the first point at which $v_{\beta_\gamma,\gamma}'$ vanishes. The last step of the proof then consists in showing that $v_{\beta_\gamma,\gamma}$ is concave over $[0, w_{\beta_\gamma,\gamma}^p]$, and strictly so over $[b, w_{\beta_\gamma,\gamma}^p]$ (Proposition C.2.3). Key to this result is the fact that $\beta_\gamma > \beta_0$ and that the maximal solution v_{β_0} derived in the no investment case is concave over $[0, w_{\beta_0}^p]$ as established in Proposition C.1.3. Finally, letting $f(w) = v_{\beta_\gamma,\gamma}(w) \wedge v_{\beta_\gamma,\gamma}(w_{\beta_\gamma,\gamma}^p)$ for all $w \geq 0$ and writing $w^i = w_{\beta_\gamma}^i$ and $w^p = w_{\beta_\gamma,\gamma}^i$ to simplify notation, it is immediate to check that the triple (f, w^i, w^p) satisfies all the properties stated in Proposition 2. ∎

Proof for Proposition 3 (sketch). The argument follows somewhat standard lines in optimal control theory. In the first step of the proof, we establish that F provides an upper bound for the expected payoff that the principal can obtain from any incentive compatible contract inducing maximal risk prevention, that is:

$$F(X_0, W_{0-}) \geq \mathbf{E}\left[\int_0^\tau e^{-rt}\{X_t[(\mu - g_t c)dt - C dN_t] - dL_t\}\right] \tag{59}$$

for any contract $\Gamma = (X, L, \tau)$ inducing maximal risk prevention. For any such contract, the dynamics of the agent's continuation utility W is given by (13), for a process H that satisfies the incentive compatibility condition (14). Substituting X and L from Γ into the function F, and applying the change of variable formula for processes of locally bounded variation (Dellacherie and Meyer (1982, Chapter VI, Section 92)) yields

$$F(X_0, W_{0-}) = e^{-rT}F(X_{T+}, W_T) - \int_0^T e^{-rt}[(\rho W_{t-} + \lambda H_t)F_W(X_t, W_{t-}) - rF(X_t, W_{t-})]\, dt$$

$$- \int_0^T e^{-rt}F_X(X_t, W_{t-})\,(dX_t^{d,c} + g_t X_t\, dt)$$

$$+ \int_0^T e^{-rt}F_W(X_t, W_{t-})\, dL_t^c \tag{60}$$

$$- \sum_{t \in [0,T]} e^{-rt}[F(X_{t+}, W_t) - F(X_t, W_{t-})]$$

for all $T \in [0, \tau)$, where $X^{d,c}$ and L^c stand for the pure continuous parts of X^d and L. Imposing limited liability and incentive compatibility, along with the homogeneity of F, the concavity of f and the fact that $f'_+ \geq -1$, we show that, in expectation, the right-hand side of (60) is greater than that of (59).

In the second step of the proof, we establish that the contract described in Proposition 3 yields the principal a value $F(X_0, W_{0-})$. This contract must therefore be optimal, since, from the first step of the proof, $F(X_0, W_{0-})$ is an upper bound for the value that the principal can derive from any contract that induces maximal risk prevention. Specifically, we start from (60) and we use the properties of the contract spelled out in Properties 1 to 6, and more precisely described in Proposition 3, to show that, in expectation, the right-hand side of (60) is in this case equal to that of (59). ∎

Proof for Proposition 4 (sketch). In the first step of the proof, we establish that the process $\{w_{T_k}\}_{k\geq 1}$ is Markov ergodic and then rely on the strong law of large numbers for

Markov ergodic processes (Stout (1974, Theorem 3.6.7)) to show that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{N_{t^-}} \ln\left(\frac{w_{T_k} - b}{b} \wedge 1\right) = \lambda \int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw), \tag{61}$$

$\mathbf{P}$–almost surely. The main technical difficulty consists in proving that the integral on the right-hand side of (61) converges.

In the second step of the proof, we establish that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^{T_{N_{t^-}}} 1_{\{w_s > w^i\}}\, ds = 1 - \lambda \int_{[b,w^i)\times\mathbb{R}_+} (t_{w,w^i} \wedge s)\, \boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}(dw, ds), \tag{62}$$

$\mathbf{P}$–almost surely. The argument goes as follows. Consider for each $k \geq 1$ the integral $I_k = \int_{T_{k-1}}^{T_k} 1_{\{w_s > w^i\}}\, ds$, where $T_0 = 0$ by convention. If $w_{T_{k-1}^+} \geq w^i$, then $w_s > w^i$ for all $s \in (T_{k-1}, T_k]$, and thus $I_k = T_k - T_{k-1}$. If $w_{T_{k-1}^+} < w^i$ and $T_k - T_{k-1} \leq t_{w_{T_{k-1}^+}, w^i}$, then $w_s \leq w^i$ for all $s \in (T_{k-1}, T_k]$, and thus $I_k = 0$. Last, if $w_{T_{k-1}^+} < w^i$ and $T_k - T_{k-1} > t_{w_{T_{k-1}^+}, w^i}$, then $w_s > w^i$ for all $s \in (T_{k-1} + t_{w_{T_{k-1}^+}, w^i}, T_k]$, and thus $I_k = T_k - T_{k-1} - t_{w_{T_{k-1}^+}, w^i}$. Summing over $k = 1, \ldots, n$ and rearranging yields

$$\frac{1}{n} \int_0^{T_n} 1_{\{w_s > w^i\}}\, ds = \frac{1}{n} \sum_{k=1}^n (T_k - T_{k-1}) - \frac{1}{n} \sum_{k=1}^n \left[t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})\right] 1_{\{w_{T_{k-1}^+} < w^i\}} \tag{63}$$

for all $n \geq 1$. Since the random variables $(T_k - T_{k-1})_{k\geq 1}$ are independently and identically distributed according to the exponential distribution $\boldsymbol{\lambda}$, it follows from the strong law of large numbers that the sequence $\left(\frac{1}{n} \sum_{k=1}^n (T_k - T_{k-1})\right)_{n\geq 1}$ converges $\mathbf{P}$–almost surely to $\frac{1}{\lambda}$. Furthermore, we show that the process $\{(w_{T_{k-1}^+}, T_k - T_{k-1})\}_{k\geq 1}$ is Markov ergodic, with invariant measure $\boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}$ over $[b, w^p] \times \mathbb{R}_+$. Since the function $(w, s) \mapsto (t_{w,w^i} \wedge s)1_{\{w<w^i\}}$ is measurable, positive and bounded above by $(w, s) \mapsto s$, and hence $\boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}$–integrable, it follows from the strong law of large numbers for Markov ergodic processes (Stout (1974, Theorem 3.6.7)) that the sequence $\left(\frac{1}{n} \sum_{k=1}^n \left[t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})\right] 1_{\{w_{T_{k-1}^+} < w^i\}}\right)_{n\geq 1}$ converges $\mathbf{P}$–almost surely to

$$\int_{[b,w^p]\times\mathbb{R}_+} (t_{w,w^i} \wedge s)1_{\{w<w^i\}}\, \boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}(dw, dt) = \int_{[b,w^i)\times\mathbb{R}_+} (t_{w,w^i} \wedge s)\, \boldsymbol{\mu^{w+}} \otimes \boldsymbol{\lambda}(dw, dt).$$

Using the fact that $\frac{N_{t^-}}{t}$ converges $\mathbf{P}$–almost surely to λ as t goes to ∞ by the strong law of large numbers for the Poisson process then yields (62).

In the last step of the proof, we establish that

$$\lim_{t\to\infty} \frac{1}{t} \int_{T_{N_{t^-}}}^t 1_{\{w_s > w^i\}}\, ds = 0, \tag{64}$$

36

P–almost surely. Merging (61), (62) and (64) finally leads to (48). ∎

Proof of Proposition 5 (sketch). We first check that (49) holds uniformly in γ whenever c is close enough to 0. This implies that the expression (48) for the long-run growth rate of the firm simplifies to

$$\lim_{t\to\infty} \frac{\ln(X_t)}{t} = \lambda \int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw) + \gamma. \tag{65}$$

The remainder of the proof consists in constructing upper and lower bounds for the integral on the right-hand side of (65).

To construct the upper bound we first define $\overline{w}^p = \frac{\mu-\lambda C}{\rho-r}$ and show that $w^p < \overline{w}^p$ uniformly in γ. We then define auxiliary processes $\{\overline{w}_t\}_{t\geq 0}$ and $\{\overline{l}_t\}_{t\geq 0}$ by

$$\overline{w}_t = w_0 + \int_0^{t^-} \left\{ (\rho\overline{w}_s + \lambda b)\,ds - b\left(\frac{\overline{w}_s - b}{b} \wedge 1\right) dN_s - d\overline{l}_s \right\},$$

$$\overline{l}_t = \max\{\overline{w}_0 - \overline{w}^p, 0\} + \int_0^t (\rho\overline{w}^p + \lambda b)\mathbf{1}_{\{\overline{w}_{s^+}=\overline{w}^p\}}\,ds$$

for all $t \geq 0$, that are independent of γ. It is easy to check that $w_t \leq \overline{w}_t$ for all $t \geq 0$ and that $\{\overline{w}_{T_k}\}_{k\geq 1}$ has a unique stationary initial distribution $\boldsymbol{\mu^{\overline{w}}}$. Furthermore

$$\int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw) \leq \int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^{\overline{w}}}(dw) < 0,$$

uniformly in γ, which yields the desired upper bound. The strict inequality follows from the fact that for each $k \geq 1$ and $w \in (b, \overline{w}^p]$, there is a strictly positive probability that $\overline{w}_{T_{k+1}} < w$ given that $\overline{w}_{T_{k+1}} \geq w$, which implies that the lower bound of the support of the stationary initial distribution $\boldsymbol{\mu^{\overline{w}}}$ of $\{\overline{w}_{T_k}\}_{k\geq 1}$ is b. Therefore, for γ close enough to 0,

$$\lambda \int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw) + \gamma < 0,$$

which establishes (50).

The lower bound is provided by the proof that $\int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw)$ converges (Section E, Proof of Proposition 4, Claim 1, Step 2). Specifically, one can show that

$$\int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw) \geq -\frac{\lambda}{\rho - r + \lambda},$$

uniformly in γ. Therefore, for $\gamma > \frac{\lambda^2}{\rho-r+\lambda}$,

$$\lambda \int_{[b,2b)} \ln\left(\frac{w-b}{b}\right) \boldsymbol{\mu^w}(dw) + \gamma > 0,$$

which establishes (51). ∎

Proof for Proposition 6 (sketch). Consider for each $k \geq 1$ the σ–fields

$$\mathcal{F}_1^k = \sigma((w_0, T_1 - T_0), (w_{T_1}, T_2 - T_1), \ldots, (w_{T_{k-1}}, T_k - T_{k-1})),$$

$$\mathcal{F}_k^\infty = \sigma((w_{T_{k-1}}, T_k - T_{k-1}), (w_{T_k}, T_{k+1} - T_k), \ldots),$$

and denote by

$$\mathcal{T} = \bigcap_{k=1}^\infty \mathcal{F}_k^\infty$$

the corresponding tail σ–field. The first step of the proof consists in showing that for each $E \in \mathcal{T}$, either $\mathbf{P}[E] = 0$ or $\mathbf{P}[E] = 1$. To establish this zero-one law, we first show that, for each $\varepsilon > 0$, there exists $n_0 \geq 1$ such that

$$\Delta(k, n, w, t, A) = \mathbf{P}[(w_{T_{k+n-1}}, T_{k+n} - T_{k+n-1}) \in A \,|\, (w_{T_{k-1}}, T_k - T_{k-1}) = (w, t)]$$

$$- \mathbf{P}[(w_{T_{k+n-1}}, T_{k+n} - T_{k+n-1} \in A)] \tag{66}$$

$$\leq \varepsilon$$

for all $k \geq 1$, $n \geq n_0$, $(w, t) \in [b, w^p] \times \mathbb{R}_+$ and $A \in \mathcal{B}([b, w^p] \times \mathbb{R}_+)$. Following Bártfai and Révész (1967, Example 2), we then show that a consequence of condition (66) is that for each $\varepsilon > 0$, there exists $n_0 \geq 1$ such that the following mixing property holds:

$$\mathbf{P}[E \,|\, \mathcal{F}_1^k] - \mathbf{P}[E] \leq \varepsilon \tag{67}$$

for all $k \geq 1$, $n \geq n_0$, and $E \in \mathcal{F}_{k+n}^\infty$, $\mathbf{P}$–almost surely. Fix some $E \in \mathcal{T}$, so that in particular $E \in \mathcal{F}_{k+n}^\infty$ for all $n \geq n_0$. Since ε is arbitrary, the mixing property (67) then implies that $\mathbf{P}[E \,|\, \mathcal{F}_1^k] \leq \mathbf{P}[E]$ for all $k \geq 1$, $\mathbf{P}$–almost surely. From Doob (1953, Chapter VII, Theorem 4.3), it follows that $\mathbf{P}[E \,|\, \bigvee_{k=1}^\infty \mathcal{F}_1^k] \leq \mathbf{P}[E]$, $\mathbf{P}$–almost surely. Since $E \in \mathcal{T} \subset \bigvee_{k=1}^\infty \mathcal{F}_1^k$, one finally has $\mathbf{P}[E] = \int_E \mathbf{P}[E \,|\, \bigvee_{k=1}^\infty \mathcal{F}_1^k] \, d\mathbf{P} \leq \int_E \mathbf{P}[E] \, d\mathbf{P} = \mathbf{P}[E]^2$. Thus either $\mathbf{P}[E] = 0$ or $\mathbf{P}[E] = 1$, as claimed.

The second step of the proof consists in showing that each of the events $\{\lim_{n\to\infty} X_{T_n} = 0\}$ and $\{\lim_{n\to\infty} X_{T_n^+} = \infty\}$ belongs to $\mathcal{T}$. First consider $\{\lim_{n\to\infty} X_{T_n} = 0\}$. Fix some $k_0 \geq 1$. For each $n \geq k_0 + 1$, one has

$$X_{T_n} = X_0 \prod_{k=1}^{N_{T_n^-}} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \exp\left(\int_0^{T_n} \gamma 1_{\{w_s > w^i\}} \, ds \right)$$

$$= X_0 \prod_{k=1}^{n-1} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right)$$

$$\exp\left(\gamma \left\{ \sum_{k=1}^{n} (T_k - T_{k-1}) - \sum_{k=1}^{n} \left[t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1}) \right] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right)$$

$$= X_{T_{k_0}} \prod_{k=k_0}^{n-1} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right)$$

$$\exp\left(\gamma \left\{ \sum_{k=k_0+1}^{n} (T_k - T_{k-1}) - \sum_{k=k_0+1}^{n} \left[t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1}) \right] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right)$$

with $\prod_\emptyset = 1$ by convention, where the second equality follows from (63) and from the fact that $N_{T_n^-} = n - 1$. Since $X_{T_{k_0}}$ is a strictly positive random variable, it follows that $\{\lim_{n \to \infty} X_{T_n} = 0\} \in \mathcal{F}_{k_0+1}^\infty$. Since k_0 is arbitrary, $\{\lim_{n \to \infty} X_{T_n} = 0\} \in \mathcal{T}$. The proof for $\{\lim_{n \to \infty} X_{T_n^+} = \infty\}$ is similar, observing that

$$X_{T_n^+} = X_{T_{k_0}^+} \prod_{k=k_0+1}^{n} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right)$$

$$\exp\left(\gamma \left\{ \sum_{k=k_0+1}^{n} (T_k - T_{k-1}) - \sum_{k=k_0+1}^{n} \left[t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1}) \right] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right)$$

and that $X_{T_{k_0}^+}$ is a finite random variable.

Finally, to conclude the proof, one verifies that $\{\lim_{t \to \infty} X_t = 0\} = \{\lim_{n \to \infty} X_{T_n} = 0\}$ and $\{\lim_{t \to \infty} X_t = \infty\} = \{\lim_{n \to \infty} X_{T_n^+} = \infty\}$. ∎

Proof of Proposition 7. Define $w_{\beta_0}^p$ as in the proof of Proposition 2. One can show that

$$r f(w_t) \geq \mu - \lambda C + (\rho w_t + \lambda b) f'(w_t) - \lambda [f(w_t) - f(w_t - b)] \tag{68}$$

for any value of $w_t > b$, with equality if $w_t \in (b, w_{\beta_0}^p]$ (Section D.2, Remark). Hence a sufficient condition for (54) to hold is that the right-hand side of (68) be larger than the right-hand side of (54), which is the case if

$$\Delta\lambda[C + b f'(w_t)] \geq \lambda[f(w_t) - f(w_t - b) - b f'(w_t)] \tag{69}$$

since $b = \frac{B}{\Delta\lambda}$. The right-hand side of (69) is positive by concavity of f, and it is bounded as f is affine over $(w_{\beta_0}^p, \infty)$. Consider next the left-hand side of (69). By (2), one has $C > b$, reflecting that maximal risk prevention is socially optimal in the first-best.[30] Since $f' \geq -1$,

---

[30]Note that in the limit first-best case, the principal's continuation value is linear in the agent's continuation utility, with a slope equal to -1. Condition (69) then reduces to $C > b$, as postulated in (2).

this implies that the mapping $C + bf'$ is positive and bounded away from 0. Since f depends on B and $\Delta\lambda$ only through their ratio b, it follows that (69) is satisfied for any value of $w_t > b$ when $\Delta\lambda$ is high enough, while B is proportionally adjusted so as to keep b constant. The result follows. $\blacksquare$

References

[1] Abreu, D., P. Milgrom, and D. Pearce (1991): "Information and Timing in Repeated Partnerships," *Econometrica*, 59, 1713–1733.

[2] Akerlof, G.A., and L.F. Katz (1989): "Workers' Trust Funds and the Logic of Wage Profiles," *Quarterly Journal of Economics*, 104, 525–536.

[3] Bártfai, P., and P. Révész (1967): "On a Zero-One Law," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 7, 43–47.

[4] Beck, T., A. Demirgüç, and V. Maksimovic (2005): "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance*, 60, 137–177.

[5] 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*, 74, 345–390.

[6] Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve (2009): "Supplement to "Large Risks, Limited Liability and Dynamic Moral Hazard"," Mimeo, Toulouse School of Economics.

[7] Brémaud, P. (1981): *Point Processes and Queues: Martingale Dynamics*, New York, Heidelberg, Berlin, Springer-Verlag.

[8] Clementi, G.L., and H. Hopenhayn (2006): "A Theory of Financing Constraints and Firm Dynamics," *Quarterly Journal of Economics*, 121, 229–265.

[9] Dellacherie, C., and P.-A. Meyer (1978): *Probabilities and Potential*, Volume A, Amsterdam, North-Holland.

[10] Dellacherie, C., and P.-A. Meyer (1982): *Probabilities and Potential*, Volume B, Amsterdam, North-Holland.

[11] DeMarzo, P.M., and M.J. Fishman (2007a): "Agency and Optimal Investment Dynamics," *Review of Financial Studies*, 20, 151–188.

[12] DeMarzo, P.M., and M.J. Fishman (2007b): "Optimal Long-Term Financial Contracting," *Review of Financial Studies*, 20, 2079–2128.

[13] DeMarzo, P.M., M.J. Fishman, Z. He, and N. Wang (2008): "Dynamic Agency and the q Theory of Investment," Mimeo, Stanford University.

[14] DeMarzo, P.M., and Y. Sannikov (2006): "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance*, 61, 2681–2724.

[15] Denis, D.K., and D.K. Shome (2005): "An Empirical Investigation of Corporate Asset Downsizing," *Journal of Corporate Finance*, 11, 427–448.

[16] Doob, J.L. (1953): *Stochastic Processes*, New York, John Wiley & Sons, Inc.

[17] Dunne, T., M.J. Roberts, and L. Samuelson (1989): "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, 104, 671–698.

[18] Evans, D.S. (1987a): "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, 35, 567–581.

[19] Evans, D.S. (1987b): "Tests of Alternative Theories of Firm Growth," *Journal of Political Economy*, 95, 657–674.

[20] Gabaix, X., and A. Landier (2008): "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123, 49–100.

[21] Gordon, R.P.E., R.H. Flin, K. Means, and M.T. Fleming (1996): "Assessing the Human Factors Causes of Acccidents in the Offshore Oil Industry," in M. Ognedal, R.J. Edwardes and J.P. Visser (Eds.), *Proceedings of the 3rd International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production*, pp. 635–644, Richardson, Society of Petroleum Engineers.

[22] Green, E. (1987): "Lending and the Smoothing of Uninsurable Income," in E.C. Prescott and N. Wallace (Eds.), *Contractual Arrangements for Intertemporal Trade*, pp. 3–25, Minneapolis, University of Minnesota Press.

[23] Hayashi, F. (1982): "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, 50, 213–224.

[24] Hollnagel, E. (2002): "Understanding Accidents–From Root Causes to Performance Variability," in J.J. Persensky, B.P. Hallbert and H.S. Blackman (Eds.), *New Century, New Trends*: *Proceedings of the 7th IEEE Conference on Human Factors and Power Plants*, pp. 1–6, New York, Intitute of Electrical and Electronics Engineers.

[25] Holmström, B., and J. Tirole (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112, 663–691.

[26] Ijiri, Y., and H.A. Simon (1964): "Business Firm Growth and Size," *American Economic Review*, 54, 77–89.

[27] Jost, P.-J. (1996): "Limited Liability and the Requirement to Purchase Insurance," *International Review of Law and Economics*, 16, 259–276.

[28] Kalesnik, V. (2005): "Continuous Time Partnerships with Discrete Events," Mimeo, University of California, Los Angeles.

[29] Kaplan, S.N., and P. Strömberg (2003), "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts," *Review of Economic Studies*, 70, 281–315.

[30] Kaplan, S.N, and P. Strömberg (2004): "Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses," *Journal of Finance*, 59, 2177–2210.

[31] Katzman, M.T. (1988): "Pollution Liability Insurance and Catastrophic Environmental Risk," *The Journal of Risk and Insurance*, 55, 75–100.

[32] Kydland, F.E., and E.C. Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.

[33] Laffont, J.-J., and D. Martimort (2002): *The Theory of Incentives*: *The Principal-Agent Model*, Princeton, Princeton University Press.

[34] Leplat, J., and J. Rasmussen (1984): "Analysis of Human Errors in Industrial Incidents and Accidents for Improvements of Work Safety," *Accident Analysis and Prevention*, 16, 77–88.

[35] Lucas, R.E., Jr. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523.

[36] Luttmer, E.G.J. (2007): "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 122, 1103–1144.

[37] Luttmer, E.G.J. (2008): "On the Mechanics of Firm Growth," Working Paper 657, Federal Reserve Bank of Minneapolis.

[38] Myerson, R.B. (2008): "Leadership, Trust, and Power: Dynamic Moral Hazard in High Office," Mimeo, University of Chicago.

[39] Phelan, C., and R.M. Townsend (1991): "Private Information and Aggregate Behaviour: Computing Multi-Period, Information-Constrained Optima," *Review of Economic Studies*, 58, 853–881.

[40] Polborn, M.K. (1998): "Mandatory Insurance and the Judgment-Proof Problem," *International Review of Law and Economics*, 18, 141–146.

[41] Quadrini, V. (2004): "Investment and Liquidation in Renegotiation-Proof Contracts with Moral Hazard," *Journal of Monetary Economics*, 51, 713–751.

[42] Rogerson, W. (1985): "Repeated Moral Hazard," *Econometrica*, 58, 597–619.

[43] Sannikov, Y. (2005): "Agency Problems, Screening and Increasing Credit Lines," Mimeo, University of California, Berkeley.

[44] Sannikov, Y. (2008): "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 75, 957–984.

[45] Sannikov, Y., and A. Skrzypacz (2007): "The Role of Information in Repeated Games with Frequent Actions," Mimeo, University of California, Berkeley.

[46] Shapiro, C., and J.E. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74, 433–444.

[47] Shavell, S. (1984): "Liability for Harm versus Regulation of Safety," *Journal of Legal Studies*, 13, 357–374.

[48] Shavell, S. (1986): "The Judgment Proof Problem," *International Review of Law and Economics*, 6, 45–58.

[49] Shavell, S. (2000): "On the Social Function and the Regulation of Liability Insurance," *Geneva Papers on Risk and Insurance*, 25, 166–179.

[50] Simon, H.A., and C.P. Bonini (1958): "The Size Distribution of Business Firms," *American Economic Review*, 48, 607–617.

[51] Spear, S.E. and S. Srivastava (1987): "On Repeated Moral Hazard with Discounting," *Review of Economic Studies*, 54, 599–617.

[52] Sraer, D., and D. Thesmar (2007): "Performance and Behavior of Family Firms: Evidence from the French Stock Market," *Journal of the European Economic Association*, 5, 709–751.

[53] Stout, W.F. (1974): *Almost Sure Convergence*, New York, San Francisco, London, Academic Press.

[54] Thomas, J. and T. Worrall (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51, 367–390.
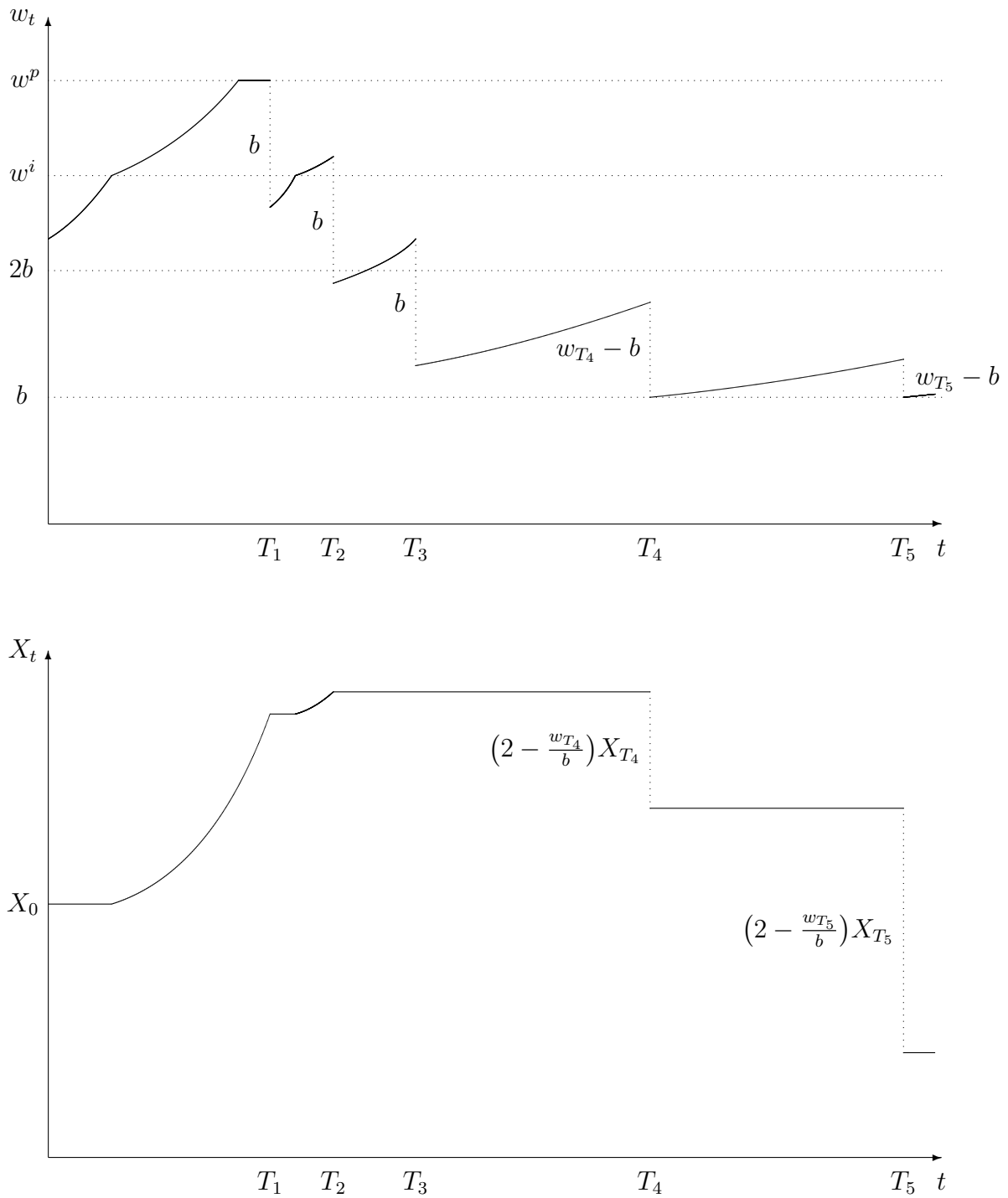
Figure 1 The top panel depicts a path for the agent's size-adjusted continuation utility. The bottom panel depicts the corresponding path for the evolution of firm size. Investment takes place as long as $w_t > w^i$. Losses occur at times T_1 to T_5. Because at T_1, T_2 and T_3, $w_{T_k} > 2b$, losses at these times induce a drop of b in continuation utility and no downsizing. By contrast, at T_4 and T_5, $w_{T_k} < 2b$, so that losses at these times induce a drop of $w_{T_k} - b$ in continuation utility, along with downsizing by an amount $X_{T_k} - X_{T_k}^+ = \left(2 - \frac{w_{T_k}}{b}\right)X_{T_k}$.