Dynamic Financial Contracting with Persistent Private Information*

SHIMING FU[†] R. VIJAY KRISHNA[‡]

14th September 2018

Abstract

This paper studies a dynamic agency model where the agent privately observes the firm's cash flows that are subject to persistent shocks. We characterize the policy dynamics and implement the optimal contract by financial securities. Because bad performance distorts investors' beliefs downward, the agent has less incentive to misrepresent private information. The optimal compensation to the agent is less than what he can divert and is convex in firm performance. The firm's current and expected future credit line limits both drop after bad performance. As private information becomes more persistent, (i) the agent is compensated more by stock options and less by equity; (ii) firm credit limits vary more with performance history; (iii) the firm is financially constrained for longer periods of time. Moreover, in contrast to the iid case, investment can be efficient in the constrained firm, and is varying with performance in the unconstrained firm. The qualitative as well as quantitative features of our contract are also in greater consonance with the empirical evidence.

Keywords: Dynamic agency, Markovian information, Payout, Capital structure, Investment.

JEL Classification: D82, D86, D92, G32, G35.

(∗) We thank Hengjie Ai, Marina Halac, Adriano Rampini, Philipp Sadowski, Lukas Schmid, S. Viswanathan, Neil Wallace, Yuzhe Zhang, and seminar participants at Duke, NYU, Rochester, Johns Hopkins, Georgia State, Columbia/Duke/MIT/Northwestern IO Theory Conference, Midwest Economic Theory Meeting, PSU-Cornell Macro Conference, Econometric Society NASM, Finance Theory Group Conference, SED Annual Meeting for helpful comments. Krishna is grateful to the National Science Foundation for financial support through grant SES-1132193. This work began while Krishna was visiting Duke University. He is most grateful to that institution for its hospitality.

(†) Shanghai University of Finance and Economics <fushiming@gmail.com>

(‡) Florida State University <rvk3570@gmail.com>

1. Introduction

There is considerable evidence that funding for firms, especially young firms, is far from efficient, and that firms must grow over time into their optimal size. In particular, financing constraints greatly affect firm size, growth, and a young firm's prospects.[1] Jensen and Meckling (1976) initiated a large literature that has focused on the conflicting interests of investors and agents — that is, agency problems — as a key friction that constrains firm financing and investment. Agency problems arise because agents have more information about their own actions, and consequently, about firm behavior, than do outside investors. The owners of the firm therefore design securities and firm policies to mitigate such agency conflicts.

An influential recent literature analyzes agency problems in dynamic contexts.[2] For instance, Clementi and Hopenhayn (2006) and DeMarzo and Fishman (2007b) characterize the optimal long-term contract when agency frictions are involved. These models provide joint predictions about how firm policies are designed and evolve in dynamic environments. Moreover, these models typically assume that the agent's private information about firm behavior is iid over time, in part because considering persistence has proved challenging in this class of models. However, in practice, firms' profitability and other economic behavior exhibit high autocorrelation. For instance, Gomes (2001) calibrated that autocorrelation of firm productivity shocks is 0.62 at the annual frequency. Many other studies show even higher numbers, which suggests that private information about firm behavior may well be persistent.

This paper studies the implications of persistent private information for firm compensation, capital structure, investment, and growth dynamics. The firm in our model consists of a risk-neutral agent and a risk-neutral principal (representing investors of the firm). The agent has the operating expertise, while the principal provides funds to launch the firm and finance risky investments over time. The agency problem is that the cash flows from investment projects are privately observed by the agent, and the agent can divert cash flows for consumption. Persistence in this setting means that firm cash flows are subject to positively correlated shocks which follow a two-state Markov process.

One might expect that persistence of productivity shocks would exacerbate the agency problem. On the contrary, we find that it alleviates the agency problem via an

(1) Gertler and Gilchrist (1994) find that manufacturing significantly declines in small firms when monetary policy tightens. Beck, Demirgüç-Kunt and Maksimovic (2005) demonstrates that financial constraints create obstacle to the growth of small firms, and that small firms benefit most from financial development.

(2) A related literature of dynamic firm financing considers limited enforcement issues, for instance, Albuquerque and Hopenhayn (2004) and Rampini and Viswanathan (2013).

additional channel that is not present in the iid case. The agent can always misreport firm performance and divert any cash flow. In the iid case, the misreport will not affect any party's belief about future shocks. But in the persistent case, it distorts downwards the principal's belief about the firm's future prospects, so that it is optimal for the principal to scale down future investment, leading to smaller future information rents for the agent. Accordingly, the agent has less incentive to misrepresent firm performance today, and is paid less than what he can divert.

Our model shows that the qualitative and quantitative features of the optimal contract are sensitive to the level of persistence. As noted above, this is because persistent private information forces the agent to weigh the tradeoffs between truthtelling versus diverting cash today and receiving smaller information rents in the future, rents that become smaller as persistence increases. Indeed, if persistence becomes sufficiently high, then the principal can *ignore* the incentive constraint completely — *private information is no longer relevant!* However, distortions from the efficient level will still persist until the agent earns enough equity to abnegate the limited liability constraints. That is, because the Principal will still want to punish the Agent after poor performance (especially if the preceding period had good performance), his only impediment is the Agent's limited liability constraint.

In terms of qualitative features, as private information becomes persistent, the agent's compensation becomes strictly convex (instead of linear) in performance. The pay-performance sensitivity increases to incorporate a dynamic information rent in addition to the current cash flow. Moreover, the payout boundary, and the investment of the unconstrained (mature) firm both vary with performance. If persistence is sufficiently high, the constrained (young) firm needs to experience consecutive good shocks to become financially unconstrained, and therefore can stay in small size for longer time than in the iid case. Also, if persistence is sufficiently high, the agent only gets paid when consecutive good performances are observed. In Section 1.1 below, we elaborate on these and other features of our contract, their relation to stylized facts, and how some of these features cannot be reconciled by a model with iid private information.

The paper also provides an implementation of the optimal contract in terms of financial securities. In the implementation, the agent holds equity and stock options. The firm is financed by a credit line with limit contingent on the firm's performance history. Both the current and the expected future credit limits drop immediately after bad performance. As persistence increases, the agent is compensated more by option payoff, and the principal holds more equity. The firm's credit line limit varies more with performance history as persistence goes up. On the contrary, the agent is compensated purely by equity, and the credit line limit is a constant in the iid case.

To show the above results, a tractable recursive formulation is needed since the approach used in iid settings is no longer sufficient. The contract in our model promises the agent utilities contingent on performance today and tomorrow. In particular, the contract promises two different pairs of continuation utilities from tomorrow onward according to good or bad performance today. This approach is shown to be convenient in finding the proper domain of the problem and in characterizing the contract properties.

We now proceed as follows. Section 1.1 summarizes some stylized facts about firm financing and our model implications. Section 1.2 reviews the relevant theoretical literature. Section 2 introduces the model, while Section 3 discusses our recursive formulation. While Section 4 descries the full information benchmark, Section 5 shows the properties of the financially constrained and unconstrained firm. Section 6 describes our implementation of the optimal contract. Section 7 concludes. All proofs can be found in the appendices.

1.1. Stylized Facts and Model Implications

We now discuss some stylized facts about financial contracts, their uses in firms, and what theoretical models have to say about them. Throughout, we first describe the relevant stylized facts, and then discuss how our model with persistence is able to explain these facts while the iid model makes predictions inconsistent with the empirical observations.

Compensation and Stock Options. Empirically, stock options are a popular way of compensating executives and employees. 71% of the 250 largest US companies in the Standard & Poor's 500 Index use stock options as incentive grant.[3] According to Larcker (2008), payment from stock options accounts for 27% (the largest component) of CEO compensation in the top 4000 US companies. Bergman and Jenter (2007) also show that stock options plans are the most common method for employee compensation below the executive rank.

In dynamic settings, the variation of future information rents provides incentives for the agent to report cash flow truthfully. In the iid case, the expected information rent becomes constant in the long-run. The only way to provide incentives is to pay the agent the amount of cash he can divert. Thus, compensation is linear in firm performance.[4]

With persistent information, the principal's belief about future prospects is always downgraded after a bad performance, leading to smaller expected information rents. Hence the agent has less incentive to misreport and the principal can carve out some

(3) See the Frederic W. Cook survey of long-term incentive grant http://www.fwcook.com/.

(4) Both DeMarzo and Sannikov (2006) and DeMarzo and Fishman (2007b) assume that agent can divert a fraction $\lambda \leq 1$ of the firm's cash flow and hence is compensated this amount by cash when the firm pays off its credit line (or short term debt). Following Clementi and Hopenhayn (2006), our model corresponds to the case where $\lambda = 1$, although it can readily be extended to the case of $\lambda < 1$.

cash flow. Since total firm revenues vary with historical shocks due to persistence but the amount retained by the principal is constant, a larger fraction is paid to the agent as revenue increases, resulting in a strictly convex compensation structure. If information is highly persistent, the principal can carve out all the firm revenue when it is low. The agent gets paid only when consecutive good performances are observed. So a report of a bad shock today in our model has a longer-lasting impact on compensation (affecting both today and tomorrow) than in the iid setting (only affecting today). The convex compensation scheme is implemented by granting the agent equity and stock options. As information becomes more persistent, the option payoff accounts for a larger portion of total compensation.

Credit Line with Contingent Limit. Credit lines are an important tool for firm financing and liquidity management. According to Demiroglu and James (2011), draw-downs of credit lines account for 75% of bank lending to firms and 63% of corporate debt. Empirically, cash flow based covenants[5] are typically written into the credit line contract. Also, most credit lines have *material adverse change* (MAC) clauses which permit lenders to withhold funds if a borrower's credit quality deteriorates significantly. According to Sufi (2009), a covenant violation is associated with a 15% to 25% drop in the availability of total line of credit.

In our model with persistence, the firm is financed by a credit line with limit contingent on compliance with a cash flow covenant. On the contrary, the credit line has a fixed limit in the iid case. In the short-run, the firm's available credit (credit limit minus draw-down) provides incentive for the agent to report cash flow truthfully. Persistent information implies that the variations in firm available credit has to be strictly greater than the current cash flow to incorporate the dynamic component of the information rent. Hence the firm's credit limit has to be adjusted according to performance history. Otherwise, the variation of available credit will always be equal to cash flow. Moreover, if the firm in our model violates the cash flow covenant, its current credit limit will immediately drop, and its expected future credit limit will also be reduced. These results are in consonance with the provision of credit lines in practice, as described above.

Firm Size and Growth Dynamics. Hurst and Pugsley (2011) show that most small firms stay small for a long time and are therefore likely to be old firms. Because of persistence, the firm in our model is more likely to receive many bad shocks in a row and become more financially constrained. Moreover, investments after bad performance will be kept low, slowing down firm growth and extending the stage of the firm being

(5) According to Demiroglu and James (2011), coverage and debt-to-cash flow covenants are the two most common financial covenants. Coverage covenants require that a borrower's coverage ratio (typically the ratio of EBITDA to fixed charges or interest expenses) remain above a minimum and debt-to-cash flow covenants restrict borrowing if the ratio of debt-to-cash flows exceeds a preset maximum.

constrained. Indeed, numerical experiments show that the amount of time that the firm remains constrained increases as we move farther away from the iid case.

Investment. Kaplan and Zingales (1997) show that less financially constrained firms exhibit greater investment to cash flow sensitivity. In our model, besides the financial slack of the firm, investment is also determined by investors' belief about the likelihood of a good shock in the future, which in turn depends on the current performance. Investment therefore varies (at efficient levels) with cash flow even when the firm is unconstrained. In contrast, if cash flow shocks are iid, investment is only determined by the credit availability of the firm. When the firm is young and constrained, its investment is very sensitive to cash flow, while investment becomes constant (at the efficient level) when the firm is unconstrained. Such dynamics cannot be reconciled with the evidence.

1.2. Related Theoretical Literature

Our work builds on a literature that studies the financing of firms under asymmetric information, typically assuming that the agent can divert cash flows without the principal's knowledge.[6] An early and seminal paper in this literature is Bolton and Scharfstein (1990), who study a two-period model, where the threat of early termination provides incentives in the first period. Fully dynamic versions of CFD models are Clementi and Hopenhayn (2006), Biais et al. (2007), and DeMarzo and Fishman (2007a), where the latter two emphasize the implementation of the optimal contract via standard securities. All these papers, regardless of time horizon, consider iid shocks to the output process. We consider the same discrete time economic environment as Clementi and Hopenhayn (2006), except that we allow for persistence in the shocks to the output process.[7]

Infinite horizon (iid) screening models were first studied by Thomas and Worrall (1990), who introduce recursive methods to such problems, and show that by using the utility promised to the agent as a state variable, the optimal contract can be reduced to a Markov decision process for the principal. Although the literature on firm financing has focused on the iid case, there is nonetheless a literature on dynamic screening

(6) Such models are therefore referred to as *cash flow diversion* (CFD) models.

(7) A related model is that of Quadrini (2004) who allows for persistent shocks that are observable (eg, via the business cycle), although the agent's private information is conditionally independent. In Quadrini (2004), the agent eventually becomes the residual claimant of the firm, and so the optimal compensation does not require the use of stock options. In particular, in the long run in Quadrini (2004), the agent's future expected rents do not depend on her report. Put differently, persistence of the public shock does *not* alleviate the severity of the agency problem, while persistent private information does, and also reconciles many empirical findings. The short run qualitative predictions of his model are also different from ours. Moreover, adding persistent public shocks to our model would not qualitatively change any of our predictions.

with Markovian types. The recursive approach is emphasized by Fernandes and Phelan (2000), who note that promised utility alone is inadequate in the Markovian case. To recursively formulate the problem, they use two *ex ante* promised utilities, one from truth-telling and the other from lying. Although we also use a vector of promised utilities, they are *interim*, contingent on the production shock in the period. Our state variables make it easier to specify the domain for the principal's dynamic programme, allowing for an analytical characterization.

Doepke and Townsend (2006) use the methods of Fernandes and Phelan (2000) in an environment that has both hidden states and hidden actions. They focus on how to reduce constraints by imposing off-path utility bounds and how to numerically solve their model. Zhang (2009) studies risk sharing with persistent private information in a continuous time setting, extending the techniques of Fernandes and Phelan (2000) to continuous time.

Tchistyi (2014) studies an environment similar as ours except that his model has finite periods and no investment. Although Tchistyi (2014) uses only one ex ante promised utility, a time varying functional has to be defined that transforms the agent's on-path utility to the utility from lying. Kapička (2013) uses a first-order approach to study an environment with continuum of states that are persistent. If the information rent is monotone, Kapička (2013) shows that the state variables can be reduced to two numbers: continuation utility and marginal continuation utility. However, the validity of the first-order approach is hard to verify.

Battaglini (2005) considers the problem of a principal sells some quantity of a good to a consumer, whose valuation for the good follows a two-point Markov process. Pavan, Segal and Toikka (2014) and Eső and Szentes (2013) study mechanism design in a dynamic quasilinear environment where the agent's persistent private information are described by 'impulse response functions'. The agent in these models is financially unconstrained and is only subject to a participation constraint at each moment in time, while our agent has no cash of her own and is protected by limited liability. Without a limited liability constraint, these models predict that the principal's expected payoff from implementing an allocation is the same as if she could observe the agent's orthogonalized private information after the initial period. Of course, this is not so in our environment. Indeed, it is precisely the inability of the principal to extract future information rents that makes our model economically interesting. Grillo and Ortner (2018) use the 'orthogonal decomposition' techniques of Eső and Szentes (2013) and Pavan, Segal and Toikka (2014) to consider a two-period model of procurement with limited liability constraints. Crucially, they allow for a continuum of cost realisations in each period, and find that the optimal contract exhibits path dependence. In contrast, Krasikov and Lamba (2018) consider an infinitely repeated procurement problem where the costs of the agent follow a two-state Markov process. Their basic results (regarding the shape of the domain, the region where the firm is unconstrained, and the qualitative properties of the value function) are similar to ours. Of course, the procurement setting introduces some other special features.

While not directly related to the principal-agent literature, Halac and Yared (2014) consider the problem of a government that has time-inconsistent preferences. The government privately observes shocks that follow a two-state Markov process. Using the same techniques as in this paper, Bloedel and Krishna (2014) study the question of immiseration in a problem of risk-sharing where the agent's taste shock follows a Markov process. Independently, Guo and Hörner (2014) use the same techniques to study mechanism design without monetary transfers.

2. Model

A principal with deep pockets has access to an investment opportunity. In order to avail herself of this opportunity, she needs the managerial skills of an agent. The agent has no funds to operate the project and is therefore dependent on the principal's funds for operational costs. Time is discrete, the horizon is infinite, the principal and agent are both risk neutral, and both discount the future at a common rate of $\delta \in (0,1)$.

The project, which we shall also refer to as the *firm*, requires an investment $k_t \geq 0$ in period t. Capital depreciates completely, and so cannot be carried over to subsequent periods. The return on capital is random, and is $\theta(s)R(k)$, where $s \in S := \{b, g\}$ and $\theta \in \{0,1\}^S$ has $\theta(b) = 0$ and $\theta(g) = 1$.[8] The function $R : \mathbb{R}_+ \to \mathbb{R}_+$ is increasing, strictly concave, and continuously differentiable, with $R(0) = 0$ and satisfies the usual Inada conditions, ie, $\lim_{k\downarrow 0} R'(k) = \infty$, and $\lim_{k\uparrow\infty} R'(k) = 0$.

The states $s \in S$ evolves according to a Markov process, which means θ is a $\{0,1\}$-valued random process. Thus, θ represents a random production shock; a return of $R(k)$ occurs if the shock is ***g***ood, while a return of 0 occurs if the shock is ***b***ad. Conditional on today's shock being $s \in \{b, g\} =: S$, the probability of having a good shock tomorrow is p_s. The transition probabilities for the production shocks are given by the matrix

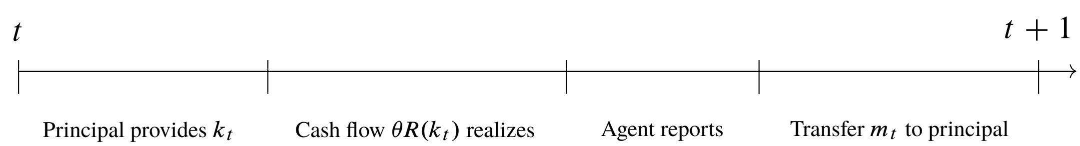
$$\begin{array}{ccc} & & \text{Tomorrow} \\ & & \begin{array}{cc} b & \quad g \end{array} \\ \text{Today} & \begin{array}{c} b \\ g \end{array} & \begin{pmatrix} 1 - p_b & p_b \\ 1 - p_g & p_g \end{pmatrix} \end{array}$$

(8) More generally, we could consider the case where $\theta(b) \in (0,1)$. Then, the output can be written as $[\theta(b) + \hat{\theta}(s)]R(k)$, where $\hat{\theta}(b) := 0$ and $\hat{\theta}(g) = \theta(g) - \theta(b)$, and $\hat{\theta}$ represents the unobservable productivity shock. For notational ease, we consider the normalisation where $\hat{\theta}(g) = 1$. All our results go through with more general values of $\theta(b)$ (and also $\theta(g)$).

We shall assume that $\Delta := p_g - p_b \geq 0$, ie, the states are positively correlated, and refer to such a Markov process as being *persistent*. The case where $\Delta = 0$ corresponds to the iid case. We also assume that $p_b, p_g \in (0, 1)$, which ensures that the Markov process has a unique ergodic measure and has neither absorbing nor transient sets. Obviously, the efficient investment $\bar{k}_s$ following a shock s is the unique k that solves $p_s R'(k) = 1$.

The agency problem arises because (i) the principal cannot observe the output while the agent can, and (ii) the agent is cash constrained and is protected by limited liability. The combination of these twin assumptions gives rise to a non-trivial contracting problem. The public history at time t consists of the investments the principal has made and the amount of cash that the agent has transferred back to her in all prior periods. A *contract* conditions investment and cash transfers in any period on the public history. We assume throughout that the agent cannot save cash made available to him in any period. In other words, all saving is done on behalf of the agent by the principal as part of the contract.

Figure 1: *Timing*



The timing runs as follows: At the beginning of time, at $t = 0$, the principal offers the agent an infinite horizon contract that he may accept or reject. If he rejects the offer, the principal and agent go their separate ways, and their interaction ends. If the contract is accepted, it is executed. The agent can leave at any time to an outside option worth 0 without further penalty. The principal fully commits to the contract. As mentioned before, the only significant difference between our model and that studied in Clementi and Hopenhayn (2006) is that we allow for persistence in the production shocks, while they restrict attention to the case where production shocks are iid.[9]

(9) There is one other, minor, difference. Clementi and Hopenhayn (2006) allow for the project to be scrapped at any time for a value that is divided between the principal and the agent according to some formula that is history dependent and optimally chosen. For simplicity, we set the scrap value to zero. Our principal results go through in the case of a positive scrap value, albeit with some straightforward modifications. In particular, the properties of the mature firm are independent of the existence or level of a scrap value (as long as the scrap value of the firm does not exceed the value of the firm when it is run efficiently).

3. Contracts

A dynamic contract conditions investments and cash transfers on the history of all previous cash transfers and investments. By the Revelation Principle, we may equivalently think of the agent as reporting the current shock as being *g*ood or *b*ad, so that a sequence of reports now constitutes a history. Of course, contracts can be written so as to condition on entire histories of reports. But, as it turns out, we may restrict attention to a class of simpler, recursive contracts without loss of generality (in that there is no loss of utility to either principal or agent). These recursive contracts are described next.

3.1. Recursive Contracts

In our Markovian setting, the concern of representing the problem recursively is that the agent may lie in a period knowing that the principal and agent will have different beliefs in the subsequent period, and use that information to his advantage. To deal with this issue, the principal can index contracts by a pair of *contingent* (or *interim*) promised utilities and the previous period's reported shock. The contingent utilities are a vector $\mathbf{v} = (v_b, v_g)$ with the interpretation that if the agent reports a bad shock in the current period, he will get v_b (lifetime) utiles, while he receives v_g utiles if he reports a good shock.

To see why the approach of contingent utilities is valid, suppose the agent enters the period with contingent promised utility vector $\mathbf{v}$. Then, the promise keeping and the incentive compatibility constraints only depend on the current period's shock. Thus, and this is crucial, even if the agent lied in the last period, contingent on today's shock being g (say), his lifetime interim utility is still v_g, and reporting g truthfully is still optimal. Moreover, contingent on the lie in the last period, even though the principal and agent disagree about the probability of shocks in the current period, they nevertheless agree about the value of cash streams moving forward. In other words, contingent promised utilities are common knowledge after *every* history, and can therefore serve as state variables for our recursive formulation.[10] Thus, our formulation ensures that the agent cannot benefit from multiple deviations and, moreover, truth-telling is optimal after *every* history.

(10) Fernandes and Phelan (2000) have a slightly different formulation, where the state variables are promised utility and a *threat-point* utility, where the latter evaluates the agent's expected utility from cash streams if he has lied in the last period. Notice that both the promised and threat-point utilities are ex ante utilities, while our contingent utilities are interim in nature. Apart from this difference, the two approaches are essentially identical. Nevertheless, we shall see below that contingent utilities allow for a more tractable formulation, because the constraint set does not depend on the previous reported shock which significantly simplifies the procedure of finding the recursive domain.

3.2. Constraints

Given a pair of contingent utilities $\mathbf{v} = (v_b, v_g) \in \mathbb{R}^2$ with last period's shock being s (at least as far as the principal believes), the principal chooses an investment policy $k \in \mathbb{R}$, transfers $m_b, m_g \in \mathbb{R}$, and continuation contingent utilities $\mathbf{w}_b = (w_{bb}, w_{bg}) \in \mathbb{R}^2$ and $\mathbf{w}_g = (w_{gb}, w_{gg}) \in \mathbb{R}^2$ subject to promise keeping, incentive, and limited liability constraints. The promise keeping constraints are

$$\text{[PK}_b\text{]} \qquad v_b = -m_b + \delta\,\mathbb{E}^b[\mathbf{w}_b]$$

$$\text{[PK}_g\text{]} \qquad v_g = R(k) - m_g + \delta\,\mathbb{E}^g[\mathbf{w}_g]$$

where $\mathbb{E}^s[\mathbf{w}] = (1 - p_s)w_b + p_s w_g$ for $s = b, g$, represents the agent's expected utility from the vector $\mathbf{w} = (w_b, w_g) \in \mathbb{R}^2$ when the current shock is s. Persistence implies that, in general, $\mathbb{E}^b[\mathbf{w}] \neq \mathbb{E}^g[\mathbf{w}]$.

Clearly, the only incentive constraint that needs to be considered is when the agent incorrectly reports the state as being $b$ad rather than $g$ood,[11] which is written as

$$\text{[IC]} \qquad v_g \geq R(k) - m_b + \delta\,\mathbb{E}^g[\mathbf{w}_b]$$

where $\mathbb{E}^g[\mathbf{w}_b]$ represents the agent's expected continuation utility from reporting the current shock as bad when it is actually good. The limited liability constraints are

$$\text{[LL]} \qquad m_g \leq R(k) \quad \text{and} \quad m_b \leq 0$$

Throughout we ignore the feasibility constraint that $k \geq 0$ without further comment (because of the Inada condition $R'(0+) = \infty$). Using the promise keeping constraints [PK$_b$] and [PK$_g$], the incentive constraint [IC] can be written somewhat more simply as[12]

$$\text{[IC*]} \qquad v_g - v_b \geq R(k) + \delta\Delta(w_{bg} - w_{bb})$$

Clearly, $R(k)$ is the *iid* (or static) information rent because it it only depends on today's choice of investment and exists regardless of the degree of persistence. The term $\Delta(w_{bg} - w_{bb})$ is the *Markovian* (or dynamic) information rent. The constraint [IC*] crystallizes the effect of Markovian shocks. As we shall see below (in Theorem 1), we must necessarily have that the Markovian information rent is nonnegative, so that with persistence, the incentive constraint is tighter.

(11) We assume throughout that there is *no hidden borrowing or saving*, so the agent cannot pretend to have high output when it is actually low.

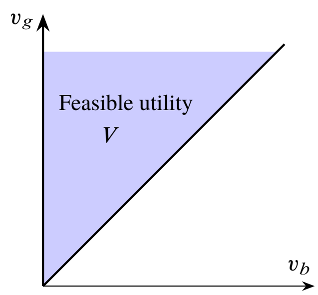
(12) Given the promise keeping constraints [PK$_b$] and [PK$_g$], the constraints [IC] and [IC*] are equivalent. In what follows we shall work with both constraints, while being explicit about which version of the incentive constraint is under consideration.

3.3. Recursive Domain

We now describe more carefully the set that indexes our recursive contracts and thus serves as the domain (or state space) for the principal's problem. Formally, we say that the tuple $(k, m_i, \mathbf{w}_i)_{i=b,g}$ *implements* (v_b, v_g) if $(k, m_i, \mathbf{w}_i)$ satisfies the incentive compatibility, promise keeping, and limited liability constraints.

Given that cash flows for the agent are always non-negative, the only feasible choices of $\mathbf{w}_i$ must lie in $\mathbb{R}^2_+$. However, even with the restriction that $\mathbf{w}_i \in \mathbb{R}^2_+$, not every $\mathbf{v} \in \mathbb{R}^2_+$ is implementable.[13] To serve as the domain for a recursive problem, the set of feasible contingent utilities ($\mathbf{v}$) that can be implemented must have the property that the continuation contingent utilities ($\mathbf{w}_i$) must also lie in this feasible set. In other words, what is required is a set $V \subset \mathbb{R}^2_+$ such that for any $\mathbf{v} \in V$, there exists a collection $(k, m_i, \mathbf{w}_i)$ that implements $\mathbf{v}$ and has $\mathbf{w}_i \in V$ for $i = b, g$. Such a set V always exists — take, for instance, $V = \{\mathbf{0}\}$. However, there exists a much larger (indeed, the largest) and non-trivial set, as described next, with this property. This largest set will serve as the domain for our recursive formulation of the principal's problem.[14]

Figure 2: *The Recursive Domain V*



Theorem 1. *There exists a largest set $V \subset \mathbb{R}^2_+$ (relative to the partial order of set inclusion) such that every $\mathbf{v} \in V$ is implemented by some $(k, m_i, \mathbf{w}_i)$ with $\mathbf{w}_i \in V$ for $i = b, g$. In particular, $V := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq v_b\}$.*

Notice that the domain V is independent of s, ie, continuation contingent utilities may be chosen from V regardless of the particular shock reported in one period. In particular, for each $(\mathbf{v}, s) \in V \times S$, let

[3.1] $$\Gamma(\mathbf{v}) := \{(k, m_i, \mathbf{w}_i) : (k, m_i, \mathbf{w}_i) \text{ implements } \mathbf{v} \text{ and } \mathbf{w}_i \in V\}$$

(13) Indeed, we show in Lemma A.1 that no $\mathbf{v} = (v_b, 0)$ with $v_b > 0$ is implementable.

(14) Theorem 1 is an analogue of Lemma 2.2 in Fernandes and Phelan (2000), but without their compactness assumptions. In the terminology of Abreu, Pearce and Stacchetti (1990), V is *self-generating*. Indeed, the proof of Theorem 1 consists of showing that V is the (largest) fixed point of an appropriate mapping.

that is, $\Gamma(\mathbf{v})$ denotes the set of feasible contractual variables $(k, m_i, \mathbf{w}_i)$ that satisfy [PK_b], [PK_g], [IC], and [LL] and have $\mathbf{w}_i \in V$. Because $\Gamma(\mathbf{v})$ is independent of s, our notation reflects this fact.

3.4. Optimal Contracts

We now consider the problem of maximizing firm surplus (which, in this setting, is precisely the social surplus). Let $Q(\mathbf{v}, s)$ denote the surplus when the previous period's shock was s, and when the agent enters the period with contingent utility $\mathbf{v}$. It is easy to see that Q must satisfy the Bellman equation

$$\text{[VF]} \qquad Q(\mathbf{v}, s) = \max_{(k, m_i, \mathbf{w}_i)} \big[-k + p_s R(k) + \delta\, \mathbb{E}^s[Q(\mathbf{w}_i, i)] \big]$$

subject to $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v})$. An *optimal contract* is the optimal policy in the firm's problem [VF]. Our first result establishes the existence of both the firm's value function as well as the optimal contract, and also establishes some of the value function's properties.

Proposition 3.1. The firm surplus is the unique function $Q : V \times S \to \mathbb{R}$ that satisfies [VF]. The function Q is concave, continuously differentiable,[15] and supermodular in $\mathbf{v}$ for each s. The optimal contract $(k, m_i, \mathbf{w}_i)$ is continuous in $(\mathbf{v}, s)$.

A nonstandard property that is particular to the Markovian case is that $Q(\cdot, s)$ is supermodular, ie, v_g and v_b are complementary instruments for the firm. For a fixed v_g, increasing v_b reduces the downside risk to the firm, because the smaller v_b is, the lower the size of the firm, which reduces future information rents. On the other hand, increasing v_b tightens the incentive constraint [IC*]. Supermodularity of Q is the observation that this second effect is less pronounced than the first when v_g is higher.

4. Efficiency

Can the optimal contract ever feature perpetual efficient investment and when do incentive constraints not matter? To answer these questions, it is useful to consider the first best benchmark, where there are no incentive constraints, because the output is observed by the principal. Then the investment level in each period is $\bar{k}_s$. Payment to the agent can be structured arbitrarily over time as long as it satisfies promise keeping. The efficient

(15) In what follows, we shall denote $\partial Q/\partial v_b$ by Q_b and $\partial Q/\partial v_g$ by Q_g.

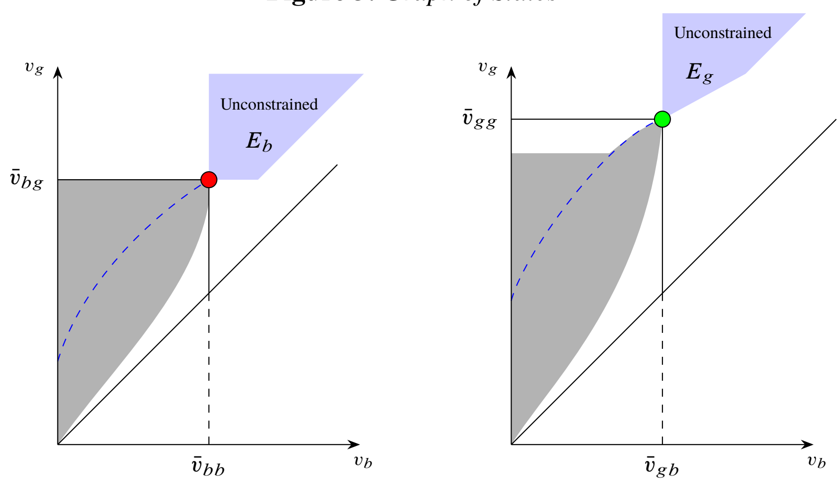
level of firm surpluses in state b, g denoted as $\bar{Q}(b)$, $\bar{Q}(g)$ are jointly determined by

$$\bar{Q}(b) = -\bar{k}_b + p_b R(\bar{k}_b) + \delta\, \mathbb{E}^b[\bar{Q}(s)] \tag{4.1}$$

$$\bar{Q}(g) = -\bar{k}_g + p_g R(\bar{k}_g) + \delta\, \mathbb{E}^g[\bar{Q}(s)] \tag{4.2}$$

The function $\bar{Q}(s)$ represents an upper bound for the firm surplus in state s, and entails *perpetual* efficient investment and production.

Figure 3: *Graph of States*†



$s = b$: bad shock last period $s = g$: good shock last period

†The pair of contingent utilities lie in the left (right) picture if a bad (good) shock occurred in the last period. The light blue areas are the states where the firm is financially unconstrained. The gray areas are the states that satisfy $Q_b, Q_g > 0$. The blanks below the gray areas (above the 45 degree line) are the states where $Q_b < 0$. The blanks above the gray area (below $\bar{v}_{sg}$) are the states where $Q_g = 0$. The areas below the blue dashed curves are the states where current investment is inefficient.

We are now in a position to describe when incentive constraints no longer matter in the optimal contract. The next proposition shows that there exist subsets of the domain V where the constrained problem reaches its upper bound, namely $\bar{Q}$, and the firm is effectively unconstrained. In this sense, these subsets are the *efficient* or *unconstrained* sets of contingent utilities, and are denoted as $E_s := \{\mathbf{v} \in V : Q(\mathbf{v}, s) = \bar{Q}(s)\}$.

Proposition 4.1. For each $s = b, g$, the sets E_s defined above are non-empty, closed, convex, and satisfy the following.

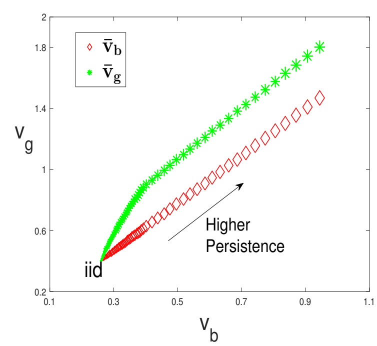
(a) Under the optimal contract, $k(\mathbf{v}, s) = \bar{k}_s$, and $\mathbf{w}_i(\mathbf{v}, s) \in E_i$, for each $\mathbf{v} \in E_s$ and $i = b, g$.

(b) There exists $\bar{\mathbf{v}}_s \in E_s$ such that $\mathbf{v} \in E_s$ implies $\mathbf{v} \geq \bar{\mathbf{v}}_s$. And $\bar{\mathbf{v}}_b \leq \bar{\mathbf{v}}_g$ with strict inequality if and only if $\Delta > 0$.

(c) For each $\mathbf{v} \in E_s$, $v_g - v_b \geq R(\bar{k}_s) + \delta\Delta \max\left[\frac{\delta \bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right]$.

The light blue areas in Figure 3 are the unconstrained sets of contingent utility. Intuitively, $\bar{\mathbf{v}}_s$ (the red and green dots in the graph) is the smallest level of contingent utility needed so that financing constraints no longer bind and perpetual efficient investments are achieved. The contingent utility in E_s must satisfy that (i) v_b is sufficiently high, and (ii) $v_g - v_b$ is sufficiently high. Requirement (ii) is peculiar to the persistent setting, because by the incentive constraint [IC*] the difference $v_g - v_b$ must be sufficiently large to permit efficient investment. Part (c) of Proposition 4.1 indicates precisely the minimal size of this difference.

Figure 4: *Threshold Contingent Utilities*



†The red diamond line indicates $\bar{\mathbf{v}}_b$. The green star line indicates $\bar{\mathbf{v}}_g$. Larger marker size indicates higher persistence level. So persistence increases from the south west corner to the north east corner. Parameters used in the picture are: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

The above result also shows that as long as persistence is positive, contingent utilities vary with production shocks and $\bar{\mathbf{v}}_b < \bar{\mathbf{v}}_g$. In particular, the continuation utility contingent on bad shock does not depend on history, ie, $\bar{v}_{bb} = \bar{v}_{gb}$. However, the continuation utility contingent on good shock does ($\bar{v}_{bg} < \bar{v}_{gg}$). The distance between $\bar{\mathbf{v}}_b$ and $\bar{\mathbf{v}}_g$ is non-zero because cash compensation is affected by the previous period's shock. As shown in Figure 4, $\bar{\mathbf{v}}_g - \bar{\mathbf{v}}_b$ increases as persistence becomes higher (by increasing p_g while fixing p_b). However, when shocks are iid, $\bar{\mathbf{v}}_g - \bar{\mathbf{v}}_b$ collapses to $\mathbf{0}$. In this case, we can also verify that the *ex ante* expected utility precisely converges to the value in Clementi and Hopenhayn (2006) as Δ converges to zero.[16]

5. Policy Characterization

The agency problem that arises due to private information is the source of the firm's financing constraint. The optimal contract determines how the financing constraint evolves

(16) Since $\bar{\mathbf{v}}_b = \bar{\mathbf{v}}_g = \bar{\mathbf{v}}$ and $p_b = p_g = p$ in the iid case, we can show that the *ex ante* continuation utility $\mathbb{E}^p[\bar{\mathbf{v}}]$ is precisely $pR(\bar{k})/(1-\delta)$, just as in Clementi and Hopenhayn (2006).

over time, as well as the firm's compensation and investment policies. We shall characterize these aspects in turn. The analysis restricts attention to *maximal rent* contracts, which are contracts where payments are made as soon as possible.[17]

As shown below, the features of the optimal contract may depend on how persistent the shocks are.[18] To facilitate the analysis, we partition the space $\Pi := \{\mathbf{p} \in (0,1)^2 : p_b \leq p_g\}$ into the low-persistence and the high-persistence subspaces in the following way. For each value of p_g, we first define a cutoff persistent level $\psi(p_g)$ as

$$\psi(p_g) := p_g - \frac{1}{R'(\hat{k}_g)}, \quad \text{where} \quad \hat{k}_g = R^{-1}\left(\frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g}\right)$$

Then the low-persistence subspace is defined to be $B_\ell := \{\mathbf{p} \in \Pi : \Delta < \psi(p_g)\}$, and the high-persistence subspace is $B_h := \{\mathbf{p} \in \Pi : \Delta \geq \psi(p_g)\}$.

5.1. Contract Evolution

When the agent is hired, the choice of the initial contingent utility vector determines the expected payoffs of the principal and the agent throughout the lifetime of the firm. We assume that the principal has all the bargaining power. Given the principal's ex ante belief about success μ, she chooses the *initial contingent utility* vector $\mathbf{v}^0_\mu$ that maximizes her expected profit $P(\mathbf{v}, \mu) := Q(\mathbf{v}, \mu) - \mathbb{E}^\mu[\mathbf{v}]$,[19] where we abuse notation to let $Q(\mathbf{v}, \mu)$ be the value of firm surplus under the measure $(\mu, 1-\mu)$ over $\{g, b\}$.

Given the optimal initial choice of $\mathbf{v}^0_\mu$, the process $\mathbf{v}^{(t)}$ evolves in a set $H := \{(\mathbf{v}, s) : Q_b(\mathbf{v}, s), Q_g(\mathbf{v}, s) > 0\}$ that lies below the thresholds $\bar{\mathbf{v}}_s$. These are precisely the states where marginally raising promised utilities (either v_g or v_b or both) increases firm surplus. This set is plotted as the gray shaded areas in Figure 3.

Theorem 2. *Let $\mathbf{w}_b, \mathbf{w}_g$ be the optimal contingent utilities at the state $(\mathbf{v}, s)$. The optimal contract evolves as follows:*

(a) *Initiation: for any $\mu \in (0,1)$, $(\mathbf{v}^0_\mu, \mu) \in H$, and $P(\mathbf{v}^0_\mu, \mu)$ is increasing in μ, strictly so if, and only if, $\Delta > 0$.*

(b) *Transition: for any $(\mathbf{v}, s) \in H$, $\mathbf{w}_b \in H$, and $\mathbf{w}_g \in H$ or $\mathbf{w}_g = \bar{\mathbf{v}}_g$.*

(17) In other words, in a maximal rent contract, if the Principal is indifferent between compensating the agent with cash or future promises, then she always chooses cash.

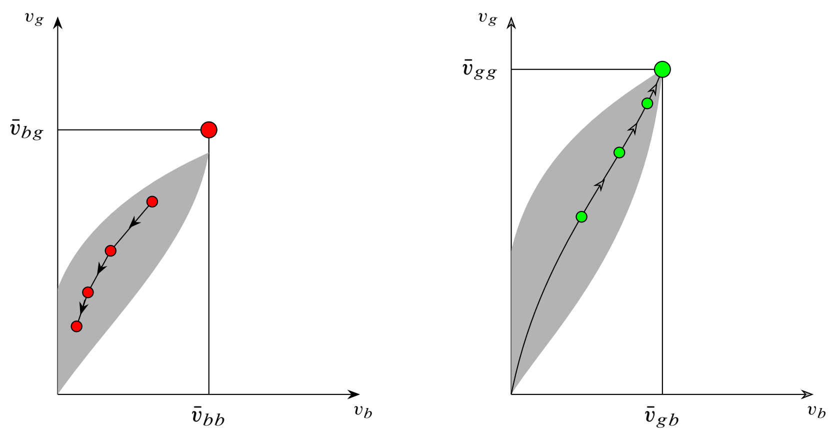
(18) When we speak of degrees of persistence, it is important to keep both Δ as well as p_g (or p_b) in mind. For instance, lowering both p_b and p_g by the same amount leaves Δ unchanged, but can change the structure of the optimal contract, because the probability of success now, and in the future, is uniformly lower.

(19) The principal can always choose the initial state so that firm surplus is maximized, but this gives the agent too much rent.

(c) *Convergence: the induced* $\mathbf{v}$ *process converges to* $\{\bar{\mathbf{v}}_b, \bar{\mathbf{v}}_g\}$ *in finite time almost surely.*

(d) *High-persistence: if persistence is high* ($\mathbf{p} \in B_h$) *and a bad shock occurs, the firm needs at least two consecutive good shocks to become unconstrained.*

Figure 5: *Evolution of States*



$s = b$: bad shock last period $\quad$ $s = g$: good shock last period

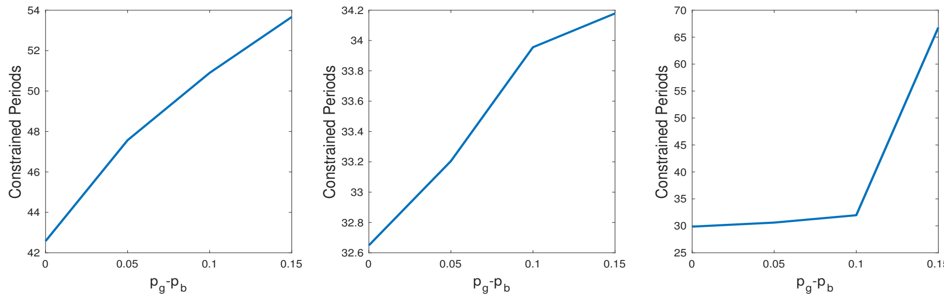
Part (a) of the theorem says that the principal gets a higher expected profit at the initiation of the contract if ex ante information indicates a greater chance of success ($s_0 = g$). According to part (b), the contingent utilities are always below their thresholds $\bar{\mathbf{v}}_s$ as long as firm financing is constrained. After a good shock, the optimal contingent utility vector $\mathbf{w}_g$ always lies on a particular curve of V.[20] An instance of this curve is illustrated in the right panel of Figure 5. The supermodularity of Q implies that w_{gg} and w_{gb} are complements. So the contingent utility vector will move up along the curve after two or more consecutive good shocks. The evolution of the contract after bad shocks is somewhat more subtle. Note that w_{bb} cannot always be larger than v_b if bad shocks keep occurring. Otherwise, the contract will converge to the efficient set E_b first which is impossible. Indeed, if w_{bb} is smaller than v_b, then the contingent utility vector will move toward the origin following subsequent bad shocks (Lemma E.4 in Appendix).

The induced contingent utilities will eventually lie in the unconstrained set E_s (part (c)), and this convergence occurs in finite time almost surely. In other words, the firm becomes financially unconstrained in finite periods. The long-run properties of the contract is implied by the directional derivative $Q_b + Q_g$ which is a non-negative martingale. This martingale must converge to 0 almost along every path, since if the martingale is always positive, then it must vary after 'good-good' and 'good-bad' shock pairs

(20) This is because the policy $\mathbf{w}_g$ is only a function of v_g, as seen in the program [VF].

which occur infinitely often almost surely. In the maximum rent contract, the contingent utility vector eventually cycles between $\bar{\mathbf{v}}_b$ and $\bar{\mathbf{v}}_g$ in the long-run.

Figure 6: *Average time spent being constrained*†



†At each persistence level, we simulate 10,000 sample paths. The figures plot the average time, over 10,000 simulations, that the firm spends in going from the initial state $(\mathbf{v}_b^0, b)$ to the absorbing state $(\bar{\mathbf{v}}_g, g)$. In all figures, $\delta = 0.9$, and $R(k) = 1.2\sqrt{k}$. In the left panel, $p_b = 0.2$, and p_g increases from 0.2 to 0.35. In the middle panel, $p_b = 0.4$, and p_g increases from 0.4 to 0.55. In the right panel, $p_b = 0.6$, and p_g increases from 0.6 to 0.75.

The convergence pattern exhibits different properties as the Markov process governing the shocks moves further away from being iid. In particular, it necessarily takes *two or more consecutive* good shocks for the firm to grow out of the financially constrained stage in the high-persistence case (part (d) of the theorem). In other words, after a bad shock occurs, the firm needs at least two good shocks in a row to become unconstrained. This result is unique to the Markovian case. In addition, Figure 6 shows our numerical experiments regarding the time periods that the firm spends in the constrained stage. The figure plots the average constrained time versus the persistence measure. In all panels, the firm spends an increasingly longer time being financially constrained as shocks move further away from being iid. These firm growth patterns in our model indicate that the persistent environment can potentially better match the empirical evidence that small (constrained) firms are actually old in age, as observed by Hurst and Pugsley (2011).

5.2. Compensation and Investment

We now characterize the compensation and the investment policies at all states induced by the optimal contract. According to the previous results, these states always lie in the set $\bar{H} := H \cup (\bar{\mathbf{v}}_b, b) \cup (\bar{\mathbf{v}}_g, g)$. The optimal contract stipulates that the principal gets a transfer m_i when shock i is reported. To ease notation, we denote the rest of the firm's cash flow which is the agent's compensation as $e_i := \theta(i)R(k) - m_i$. Accordingly, compensation in the unconstrained firm is denoted by $\bar{e}_{si} = e_i(\bar{\mathbf{v}}_s, s)$.

Theorem 3. *For any state $(\mathbf{v}, s) \in \bar{H}$, the compensation policy satisfies the following.*

(a) *No pay for failure:* $e_b(\mathbf{v}, s) = 0$.
(b) *Pay for success when* v_g *is large enough:* $e_g(\mathbf{v}, s) > 0$ *if, and only if,* $v_g > \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$.
(c) *At low persistence, the unconstrained firm always pays the agent for success:*

$$0 < \bar{e}_{bg} \leq \bar{e}_{gg} \;\; \textit{if} \;\; \mathbf{p} \in B_\ell; \quad \bar{e}_{bg} < \bar{e}_{gg} \;\; \textit{if and only if} \;\; \Delta > 0$$

(d) *At high persistence, the unconstrained firm pays the agent only for consecutive successes:*

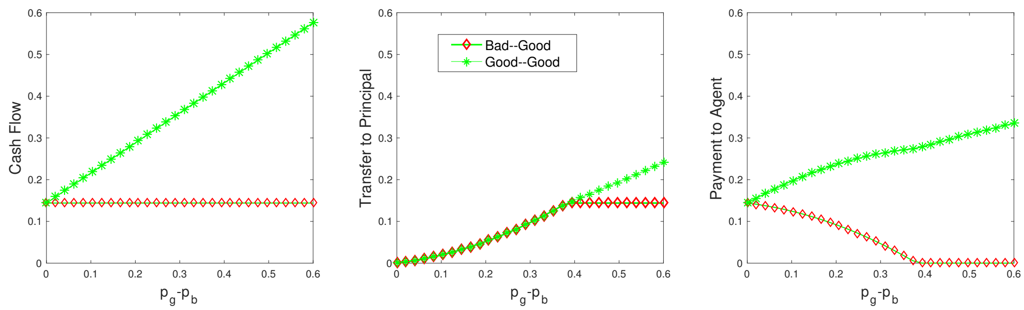
$$0 = \bar{e}_{bg} < \bar{e}_{gg} \;\; \textit{if} \;\; \mathbf{p} \in B_h$$

Parts (a) and (b) of the theorem provide necessary conditions for the agent to receive compensation are that (i) he reports a positive cash flow, and (ii) contingent on the good shock, the promised utility v_g is sufficiently large. This happens when the present good shock sends the firm to the unconstrained stage. In other words, the firm is *one step* away from being unconstrained before the good shock occurs. Thus, all payments are *back loaded*, which is a robust property that holds regardless of the degree of persistence.

A striking feature of the compensation policy is that the cash payments to the agent as well as the transfers to the principal both vary with the degree of persistence. Figure 7 plots the split of cash flows between the principal and the agent as a function of persistence. It is clear that in the iid case, the transfer to the principal is always zero and the agent gets the entire cash flow, which makes the agent the residual claimant of the firm. However, for any positive degree of persistence ($\Delta > 0$), the principal always gets some (non-zero) part of the cash flow even though the agent can (in principle) divert everything. Moreover, as persistence increases, the principal gets larger amounts of the firm's cash flow. The intuition is that if the agent misreports a good shock to be a bad one, investment in the subsequent period will be optimally chosen as $\bar{k}_b$ instead of $\bar{k}_g$. The low investment size reduces the agent's future information rent, making him less willing to lie today. Hence, the principal can carve out a transfer today by rationally committing to reduce future information rent if a bad shock is ever reported.

As we move away from the iid case, the shock history starts to matter for compensation. The agent is always compensated more for consecutive successes. In particular, the agent gets a larger pay when the firm cash flow is $R(\bar{k}_g)$ (after a 'good-good' shock pair) than when the firm cash flow is $R(\bar{k}_b)$ (after a 'bad-good' shock pair). Thus, as persistence becomes strictly positive, compensation changes from linear to strictly convex in performance. Indeed, if persistence is sufficiently large, or $\mathbf{p} \in B_h$ ($\Delta > 0.4$ in Figure 7), then the principal obtains all the cash flow (if any), contingent on the previous shock being bad. The agent receives cash pay only when two successes occur in a row. In this sense, a bad performance in the last period not only punished the agent then (no pay last period) but also punishes the agent today (low payoff in spite of success).

Figure 7: *Cash Flow, Transfer, and Compensation in Unconstrained Firm*†



†The pictures plot cash flow (left panel), transfer (middle panel), and compensation (right panel) when a good shock occurs contingent on the previous shock being bad or good. The cash flow always equals the sum of transfer and compensation. Parameters used in the pictures are: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

The investment policy is monotone in the state variables. If the firm had a good shock in the last period, it has a higher probability of generating positive cash flow, which makes it optimal to invest more today. Moreover, larger v_g or smaller v_b relaxes the incentive constraint, leading to more efficient investment.

Proposition 5.1. The investment policy at any $(\mathbf{v}, s) \in H$ has the following properties:
(a) $k(\mathbf{v}, s)$ increases in s and in v_g, but decreases in v_b;
(b) $k(\mathbf{w}_g, g) < \bar{k}_g$ if $\mathbf{w}_g \notin E_g$;
(c) $k_b(\mathbf{w}_b, b) \leq \bar{k}_b$, and if $\mathbf{p} \in B_\ell$ then $k(\mathbf{w}_b, b) < \bar{k}_b$.

When the firm is financially constrained, a key feature of the iid model is that investment is always inefficient, because investment size determines the static information rent which adds to the cost of financing. In our model with persistence, the investment is below the efficient level after good shocks. However, investment can be temporarily efficient after a bad shock even though the firm is not fully unconstrained. This can happen in the high-persistence case ($\mathbf{p} \in B_h$). In the left panel of figure 3, the points in H that immediately surround $\bar{\mathbf{v}}_b$ all have efficient investment ($\bar{k}_b$). Because policies are continuous, the optimal contract can possibly evolve to this region. In that case, the firm investment will stay at $\bar{k}_b$ until a good shock occurs.

To understand this result, observe first that when persistence is high, $\bar{k}_b$ is necessarily small, and $\mathbb{E}^g[\bar{\mathbf{v}}_g]$ is much larger than $\mathbb{E}^g[\bar{\mathbf{v}}_b]$. Now, consider the incentive constraint. If cash flow $R(k)$ realizes and the agent lies, he can divert $R(k)$ and obtain expected utility (from tomorrow onward) $\mathbb{E}^g[\mathbf{w}_b] \simeq \mathbb{E}^g[\bar{\mathbf{v}}_b]$. If the agent reports truthfully, he gets expected utility $\mathbb{E}^g[\mathbf{w}_g] \simeq \mathbb{E}^g[\bar{\mathbf{v}}_g]$. Because persistence is high, firm surplus is very volatile, and as noted above, $\mathbb{E}^g[\bar{\mathbf{v}}_g]$ is much larger than $\mathbb{E}^g[\bar{\mathbf{v}}_b]$. So even at the efficient investment $\bar{k}_b$, the agent strictly prefers truth-telling, because a lie results in a loss in continuation utility that is much larger than the current cash flow that he can divert.[21]

6. Implementation

This section provides an implementation of the optimal contract by a set of standard financial securities whose definitions can be found in the Appendix. Though the implementation is not unique, we follow a similar structure used in the dynamic financial contracting literature such as DeMarzo and Fishman (2007b). The focus here is to highlight the changes when deviating from the iid environment. We first show below how the implementation works by laying out its key elements. Then we argue and state in Theorem 4 that this implementation does not give the agent any strategies that are strictly more profitable than truth-telling.

Table 1: *Firm's Balance Sheet*

Assets	Liabilities
NPV of Investments	Debt
Compensating Balance	Stock Options
	Equity

Because securities can be held by widely dispersed investors or intermediaries, it is not necessary to rely on a single principal to execute the contract. In this sense, the principal simply represents a group of firm stakeholders. At time zero, the firm is financed by issuing equity, long-term debt, and by obtaining a credit line. The firm also issues equity and stock options to the agent as compensation. The firm's debts are associated with a cash-flow covenant[22] that determines the debt coupon[23] and credit limit. The amount of credit drawn down by the firm from the credit line is charged an interest rate $r = 1/\delta - 1$. To establish the credit line, investors also require a cash deposit (compensating balance) that pays a fixed interest of $\bar{k}_g$ each period. The firm defaults if it ever exhausts the credit line or does not pay the long-term debt coupon. In a nutshell, the firm has a balance sheet shown in Table 1.

(21) As shown in the Appendix, the incentive compatibility constraint [IC] is active when $\mathbf{p} \in B_\ell$. However, when $\mathbf{p} \in B_h$, the incentive constraint is slack and the limited liability constraint [LL] becomes active.

(22) A *debt covenant* is a clause in the debt contract that requires the borrower to maintain certain financial terms. For example, cash flow to debt ratio and leverage ratio are commonly stipulated terms. The *cash-flow covenant* here is a debt covenant that requires the firm to maintain positive cash flows.

(23) A *coupon* payment on a debt is the periodic (typically annual) interest payment that the bondholder receives from the bond's issue date until it matures.

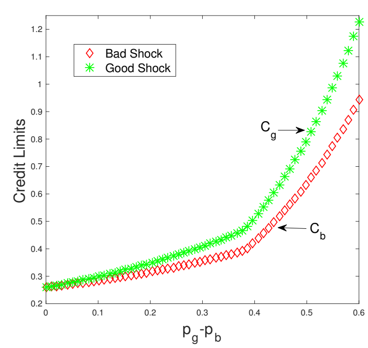
The first step of the implementation is to find an alternative way of representing the state variables $(\mathbf{v}, s)$. The variable s simply indicates whether the firm complied with its debt covenant $(s = g)$ or not $(s = b)$ in the previous period. The promised utilities $\mathbf{v}$ can be represented by the firm's available funds: the amount that can be withdrawn from the credit line plus any output. The agent can simply withdraw these funds to compensate himself at any time.[24] Suppose, starting a certain period, the firm has balance M on its credit line. If the firm violates the covenant $(i = b)$, it has a credit limit of C_b and generates zero output. If the firm complies with the covenant $(i = g)$, it has credit limit C_g and generates $R(k)$ cash flow. Hence, we get the following relations

$$v_i = C_i - M + \theta(i)R(k(\mathbf{v}, s)), \quad i = b, g \tag{6.1}$$

Obviously, we cannot uniquely pin down C_b, C_g, and M for a pair of promised utilities (v_b, v_g). Instead, we normalize C_b to be $\bar{v}_b$ after any history, and jointly determine C_g and M by [6.1]. Accordingly, it is equivalent to use (C_g, M, s) as state variables and treat firm policies as functions of the new state. Moreover, the constraints imposed on $\mathbf{v}$ in the contract translate to restrictions on credit limits as stated in the following result.

Proposition 6.1. Given any history, the firm has a lower credit limit contingent on violating the cash flow covenant, ie, $C_g \geq C_b$. Moreover, $C_g > C_b$ if and only if $\Delta > 0$. In the unconstrained stage, the credit limit gap $(C_g - C_b)$ strictly increases in Δ when fixing p_b.

Figure 8: *Credit Limits of the Unconstrained Firm*†



†The picture plots how contingent credit limits of the unconstrained firm change with persistence levels. Parameters used in the picture are: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

(24) Note that while in the contract the agent cannot immediately obtain the promised utilities, he can withdraw this amount of funds and default here. But Theorem 4 will show that the agent gets no greater payoff from defaulting than from operating the firm.

Note that this result does not depend on the normalization of C_b. Instead, it reflects a distinct feature of the persistent model. It's well known that in the iid case a credit line with constant limit suffices to implement the optimal contract, as described in DeMarzo and Fishman (2007b). This is because if the agent repays the cash flow $R(k)$, then the credit line balance is reduced by the same amount. The firm has $R(k)$ more of available funds (the agent's discounted future payoff). Therefore, repaying the credit line offers the agent the same utility as diverting output. This mechanism won't work in the persistent case because the agent also obtains a Markovian information rent on top of the cash flow $R(k)$. So the credit limits need to be adjusted to implement the optimal contract. Figure 8 plots how credit limits and their gap vary with persistence in the unconstrained firm, reflecting the result of Proposition 6.1.

Given this representation of the contract state, we can derive a security and compensation design that implements the contract policies. Without further delay, we state the main result here, illustrating all the detailed steps in the following subsections.

Theorem 4. *The optimal contract can be implemented by equity, long-term debt, credit line, and stock options with the following stipulations.*

(i) The agent holds a fraction λ of the outstanding equity, and is granted stock options with strike price K each period, where λ and K are endogenous and specified in Proposition 6.2 below.

(ii) Contingent on whether the firm violates the covenant ($i = b$) or not ($i = g$), it receives an immediate credit limit C_i and pays a coupon c_i today, while it receives a contingent credit limit pair (C'_{ib}, C'_{ig}) tomorrow.[25] *The values of C'_{ig} and c_i are specified in Lemma 6.3 below.*

It is incentive compatible for the agent to truthfully disclose cash flow and use it to pay back the credit line balance. Once the balance is fully repaid, the firm starts paying out.

6.1. Equity and Stock Options

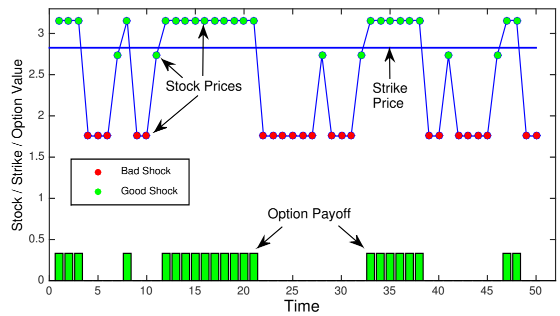
The optimal compensation is implemented by a combination of equity and stock option payoffs, the total of which is the firm's payout. We normalize the firm's outstanding equity share to be one and define its price z as the present value of future payouts.

At the beginning of each period the agent is granted the option to purchase one share at price K, which expires at the end of the period. This option pays the agent $\max\{z - K, 0\}$. The rest of the firm's payout is issued as a dividend split between the agent and investors according to their equity holdings: the agent holds $\lambda \in [0, 1]$. The implementation is essentially to design λ and K such that the agent's security payoff matches his compensation in the optimal contract.

(25) The first (second) subscript of the credit limits denotes what happens today (tomorrow).

Proposition 6.2. The agent gets the same payoff from security holdings as from the optimal contract if $\lambda = \frac{\bar{e}_{bg}}{R(\bar{k}_b)}$ and $K = \bar{z}_{bg} + \bar{m}_{gg} - \bar{m}_{bg}$. As long as $\Delta > 0$, the agent is not the residual claimant of the firm, ie, $\lambda < 1$. Given either p_b or p_g, the agent's equity holding and equity payoff both decrease in Δ, but the fraction of the option payoff in the compensation increases in Δ.

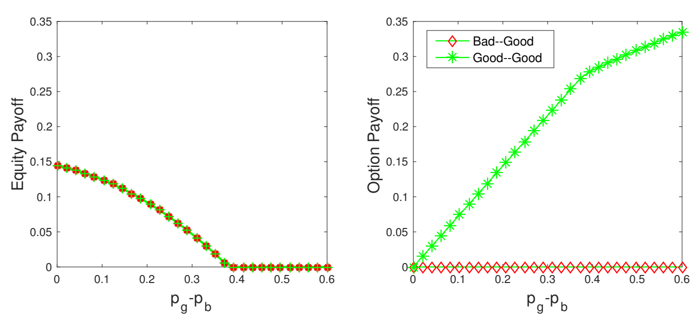
Figure 9: *Stock Prices and Option Payoffs.*†



†The picture plots stock prices, the strike price, and the option payoff on a simulated path when the firm is unconstrained. Parameters used in the simulation: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.3$, $p_g = 0.7$.

The option payoff is positive only when the firm is unconstrained. In this stage, the firm pays off its credit line and issues all cash flows as payouts which cycle among 0, $R(\bar{k}_b)$, and $R(\bar{k}_g)$. The stock prices also cycle among three levels $\bar{z}_{sb}$, $\bar{z}_{bg}$, and $\bar{z}_{gg}$, the values of which are derived in Lemma G.1 of the Appendix. Since the designed strike price K falls between $\bar{z}_{bg}$ and $\bar{z}_{gg}$, the options have a positive payoff only after 'good-good' shock pairs (when the stock price is the highest). Figure 9 plots the stock prices and the option payoffs from a simulated path.

Figure 10: *Security Payoffs to Agent* †



†The pictures plot the agent's security payoffs when a good shock occurs today contingent on the previous shock being bad or good. Parameters used in the pictures are: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

Proposition 6.2 also shows how the compensation structure changes with persistence levels. In the iid case, the agent is paid entirely by equity and he also holds all the firm equity. As long as persistence is positive, the agent is always paid by a combination of equity and stock options. The agent is not the residual claimant of the firm because the investors also hold equity. These results highlight the sensitive implications of the iid environment. Moreover, as persistence (ie, $\Delta = p_g - p_b$) increases, the agent is paid more by stock options and less by equity. Figure 10 shows these patterns in an example plot by comparing the two sources of compensation.

6.2. Evolution of Credit Limit and Balance

Thus far, we have implemented the optimal payout via securities. Now we consider the design of credit line limits and debt coupons to implement the evolution of contingent utilities. The credit line balance evolves according to withdrawals and repayments. One the one hand, the credit line is withdrawn to invest, to pay debt coupons, and to issue payouts (outflows). On the other hand, firm outputs and interest payments on the compensating balance (inflows) are used to pay back the credit line balance.

Suppose the current state is (C_g, M, s), then the credit line balance starting next period contingent on whether the firm violates covenant or not today will be

$$M_i = (1 + r)[M + k + c_i + d_i - \theta(i)R(k) - \bar{k}_g] \tag{6.2}$$

where c_i is the coupon payment, d_i is the payout, and $\bar{k}_g$ is the interest received. The new state starting next period will be (C'_{ig}, M_i, i). Investors design C'_{ig} and c_i so that the firm's available funds always match the contingent utilities in the optimal contract.

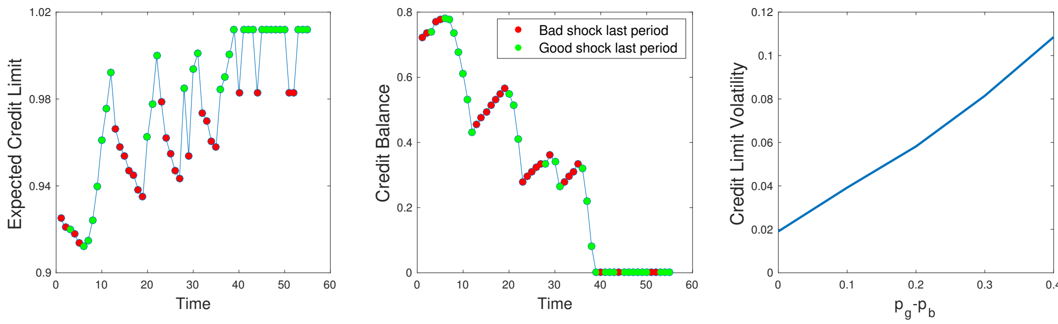
Lemma 6.3. Given any state (C_g, M, s), the firm's available funds evolve in the same way as contingent utilities w_i in the optimal contract under the following mechanism:

$$C'_{ig} = w_{ig} + M_i - R(k(w_i, i)) \tag{6.3}$$

$$c_i = \frac{1}{1 + r}(\bar{v}_b - w_{ib}) - M + \bar{k}_g - k + \theta(i)R(k) - d_i \tag{6.4}$$

Given the proposed mechanism, the agent has no incentive to divert the firm's available funds or outputs. Note that credit limits C'_{ig} and coupon payments c_i (always positive) in Lemma 6.3 are functions of (C_g, M, s). From the above mechanism ([6.4] and [6.3]), we know that if the agent diverts all the available funds and defaults at state (C_g, M, s), he gets a utility of either $\bar{v}_b - M$ or $C_g - M + R(k)$. However, by the relation specified in [6.1], these two values are exactly the equilibrium contingent utilities that the agent gets by operating the firm.[26] Therefore, the agent has no incentive to default. Moreover, since these contingent utilities satisfy the incentive compatibility constraint [IC], the agent has no incentive to lie about cash flows either.

Figure 11: *Evolution of Credit Limits and Balance*†



†The left panel plots the evolution of expected credit limit on a simulated sample path. The middle panel plots the evolution of credit account balance on the same sample path. The right panel plots how credit limit volatility varies with persistence. At each persistence level, we simulate 10,000 sample paths from the initial state $(\mathbf{v}_b^0, b)$ to the absorbing state $(\bar{\mathbf{v}}_g, g)$. The right panel plots the average of sample standard deviations over the 10,000 simulated paths. In all pictures, $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, and $p_b = 0.3$. In the left and middle panels, $p_g = 0.7$. In the right panel, p_g increases from 0.3 to 0.7.

Figure 11 simulates how the credit limit and balance evolve throughout the firm's life cycle. The left panel plots the expected credit limit (defined as $\mathbb{E}^s[C_s]$) from a simulated sample path. The expected credit limit measures the firm's future credit condition given a performance history. After periods of good performances, the firm tends to have higher expected credit limit moving forward. On the contrary, the firm's credit condition deteriorates after a sequence of bad performances. The right panel shows that the volatility of the firm's credit limit significantly increases as persistence becomes larger. In other words, the credit limit varies more with firm performance as persistence increases. This is because the gap between credit limit is driven by the magnitude of the Markovian rent which is larger as persistence increases.

7. Conclusion

In this paper, we explore the question of how a firm is financed when its cash flows are privately observed by an agent who operates the firm. The new ingredient of our model is that firm cash flows are subject to persistent and privately observed shocks. Many studies have already shown that adopting this assumption is crucial if we are to quantify dynamic agency models. Persistent and private information about current cash

(26) From the optimal contract, contingent utilities are the amount that the agent gets by obtaining security payoffs until the firm credit line is fully repaid.

flow implies that the agent is also better informed about the firm's future, information valuable to investors. We show that if investors can design the long term contract optimally, then the agent actually has less incentive to misrepresent firm performance than in the iid case. In other words, the agency problem becomes less severe.

We show that promising agent utilities contingent on performance today and tomorrow is effective in providing incentive and formulating the problem recursively. With this recursive approach, we can analytically characterize firm policies and show that they depend crucially on the degree of persistence.

When the firm is initiated in the optimal contract, it faces financing constraints and hence investment cannot be always efficient. Incentive is provided exclusively through adjustments to agent's continuation utilities until reaching some thresholds. After that the firm is no longer financially constrained, its investments are forever optimal, and the agent may get cash pay. Depending on persistent level, investment may be temporarily efficient (after bad shocks) before the firm becomes fully unconstrained. Moreover, depending on persistence, the firm may need to receive a sequence of good shocks in a row in order to reach the unconstrained stage, which means the firm may be stuck in the constrained stage for longer time than in the iid case.

By identifying the appropriate martingale, we show that the firm converges to the unconstrained stage in finite time with probability one. When it becomes unconstrained, its investments cycle between the efficient levels according to its shocks. The agent gets cash payments that are less than what he can divert. In the case of high persistence, the agent may not even get cash pay after good shocks. This also implies investors hold more stake of the firm than in the iid case.

An implementation of the optimal contract using financial instruments highlights the distinct features of our model with persistent private information. The agent is paid by holding equity and stock options. As information becomes more persistent, option payments accounts for larger fraction of compensation. After bad shocks, the credit line limit will drop immediately and the future expected credit line limit will also be reduced. The recursive approach allows us to characterize the firm's policies in this environment and helps us better understand implications of persistent private information in firm financing, investment, compensation, and growth.

Appendix[27]

A. Proofs from Section 3

We begin with a demonstration that not all points $\mathbf{v} \in \mathbb{R}^2_+$ are implementable.

Lemma A.1. Let $\mathbf{v} = (v_b, 0)$ where $v_b > 0$. Then, such a $\mathbf{v}$ is not implementable with $\mathbf{w} \in \mathbb{R}^2_+$.

Proof. Notice [PK_g] requires that

$$0 = R(k) - m_g + \delta\, \mathbb{E}^g[\mathbf{w}_g]$$

By [LL], we know that $R(k) - m_g \geq 0$, and by assumption, $\mathbf{w}_g \in \mathbb{R}^2_+$, which implies $\mathbb{E}^g[\mathbf{w}_g] \geq 0$. Therefore, it must be that $R(k) = m_g$, and $\mathbf{w}_g = (0, 0)$. Now notice that by [IC], we obtain

$$0 \geq R(k) - m_b + \delta\, \mathbb{E}^g[\mathbf{w}_b]$$

As noted above, $\mathbf{w}_b \in \mathbb{R}^2_+$, and $R(k) \geq 0$. By [LL], we also have $m_b \leq 0$, which implies $0 \geq R(k) - m_b + \delta\, \mathbb{E}^g[\mathbf{w}_b] \geq 0$, ie, $R(k) = m_b = k = 0$ and $\mathbf{w}_b = (0, 0)$. Therefore, by [PK_b], we must have $v_b = -m_b + \delta\, \mathbb{E}^b[\mathbf{w}_b] = 0$. But this contradicts our assumption that $v_b > 0$. Thus, $(v_b, 0)$ with $v_b > 0$ is not implementable, or equivalently, is infeasible. $\square$

We now present the proof of Theorem 1. It is easy to see that the set of contingent utilities $\mathbf{v} \in \mathbb{R}^2_+$ that can be implemented by $(k, m_i, \mathbf{w}_i)$ with $\mathbf{w}_i \in \mathbb{R}^2_+$ is a closed and convex cone. Therefore, in our search for a suitable domain, it suffices to restrict attention to closed and convex cones.

Recall (from Section 3.3) that the tuple $(k, m_i, \mathbf{w}_i)_{i=b,g}$ *implements* (v_b, v_g) if $(k, m_i, \mathbf{w}_i)$ satisfies the incentive compatibility, promise keeping, and limited liability constraints. Let $\mathcal{K}$ denote the space of closed and convex cones that are subsets of $\mathbb{R}^2_+$ and are of the form $\{\mathbf{v} \in \mathbb{R}^2_+ : v_g \geq \alpha v_b\}$ for some $\alpha \in [0, 1]$. Thus, members of Φ are cones that contain points above the diagonal and are bounded below by a ray (from the origin) below the diagonal.

Following Abreu, Pearce and Stacchetti (1990), we define the operator $\Phi : \mathcal{K} \to \mathcal{K}$ as follows: for $C \in \mathcal{K}$, let

$$\Phi(C) := \left\{\mathbf{v} \in \mathbb{R}^2_+ : \exists\, (k, m_i, \mathbf{w}_i) \text{ that implements } \mathbf{v} \text{ and has } \mathbf{w}_i \in C,\ i = b, g\right\}$$

(27) For ease of exposition, several technical statements and their proofs are left to the Supplementary Appendix, referred to as SA in what follows.

In other words, $\Phi(C)$ consists of all implementable contingent utilities $\mathbf{v}$ wherein the continuation contingent utilities $\mathbf{w}_i$ lie in the set C. Clearly, any recursive program must only consider contingent utilities $\mathbf{v}$ that lie in a set C such that C is a fixed point of Φ, so that all present contingent utilities as well as future continuation contingent utilities lie in the same set. Essentially, Theorem 1 delineates such a set.

Proof of Theorem 1. We shall first show that Φ is well defined, ie, it maps closed and convex cones to closed and convex cones. Towards this end, let $\alpha \in [0,1]$, and define $C_\alpha := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq \alpha v_b\}$. Clearly, $C_\alpha \in \mathcal{K}$.

Let $\mathbf{v} \in \mathbb{R}^2_+$ be such that $(k, m_i, \mathbf{w}_i)$ implements $\mathbf{v}$ with the restriction that $\mathbf{w}_i \in C_\alpha$. The set of all such $\mathbf{v}$ is precisely the set $\Phi(C_\alpha)$.

By [PK$_b$], we obtain

$$\begin{aligned} v_b &= -m_b + \delta\,\mathbb{E}^b[\mathbf{w}_b] \geq -m_b + \delta[p_b\alpha w_{bb} + (1-p_b)w_{bb}] \\ &\geq -(1-p_b+p_b\alpha)m_b + \delta[(1-p_b)w_{bb} + p_b\alpha w_{bb}] \\ &= (1-p_b+p_b\alpha)(\delta w_{bb} - m_b) \end{aligned}$$

where the first inequality follows from the assumption that $w_{bg} \geq \alpha w_{bb}$ (recall that $\mathbf{w}_b \in C_\alpha$), and the second follows from [LL] which requires that $m_b \leq 0$, and from the fact that $1 - p_b(1-\alpha) \leq 1$. This implies

$$m_b - \delta w_{bb} \geq -v_b/(1-p_b+p_b\alpha)$$

Notice that [PK$_b$] can be written as $\delta p_b(w_{bg} - w_{bb}) = v_b + (m_b - \delta w_{bb})$, which implies (using the inequality displayed above) that

$$\begin{aligned} \delta(w_{bg} - w_{bb}) &\geq \frac{v_b}{p_b}\left[1 - \frac{1}{1-p_b+p_b\alpha}\right] \\ &= -v_b\left[\frac{1-\alpha}{1-p_b+p_b\alpha}\right] \end{aligned} \tag{A.1}$$

Plugging this into [IC*], we obtain

$$\begin{aligned} v_g &\geq v_b + R(k) + \delta\Delta(w_{bg} - w_{bb}) \\ &\geq v_b\left[1 - \frac{(1-\alpha)\Delta}{1-p_b+p_b\alpha}\right] \\ &= v_b\left[\frac{1-p_g+p_g\alpha}{1-p_b+p_b\alpha}\right] \\ &=: \alpha' v_b \end{aligned}$$

where the first inequality is merely [IC*] and the second inequality follows from [A.1] and the fact that $R(k) \geq 0$.

Thus, if continuation contingent utilities $\mathbf{w}_i$ are constrained to lie in the set C_α, then the set of implementable $\mathbf{v}$ must lie in the set $C_{\alpha'}$, where $\alpha' = (1 - p_g(1-\alpha))/(1 - p_b(1-\alpha))$. This demonstrates that $\Phi(C_\alpha) \subset C_{\alpha'}$.

To see that any $\mathbf{v} \in C_{\alpha'}$ can be implemented by some $(k, m_i, \mathbf{w}_i)_{i=b,g}$ with $\mathbf{w}_i \in C_\alpha$, consider the following implementation. Set $k = 0$, $m_g = -v_g$, and $\mathbf{w}_g = \mathbf{0}$. This ensures [$\text{PK}_g$] holds. Also set $m_b = 0$, $w_{bg} = \alpha w_{bb}$, and $w_{bb} = v_b/[\delta(1 - p_b + \alpha p_b)]$. Clearly, [$\text{PK}_b$] is satisfied because $\delta\,\mathbb{E}^b[\mathbf{w}_b] = \delta w_{bb}[1 - p_b + \alpha p_b] = v_b$. Notice also that $v_g \geq \alpha' v_b = \left[\frac{1-p_g+\alpha p_g}{1-p_b+\alpha p_b}\right] \delta w_{bb}(1 - p_b + \alpha p_b) = \delta w_{bb}(1 - p_g + \alpha p_g) = \delta\,\mathbb{E}^g[\mathbf{w}_b]$, which is precisely [IC], and where the first inequality is because $\mathbf{v} \in C_{\alpha'}$ with the subsequent equalities following from the implementation values. This establishes that $\Phi(C_\alpha) = C_{\alpha'}$, and in particular, shows that Φ is well defined.

We claim that if $\alpha \in [0, 1)$, then $\alpha' > \alpha$. To see this, notice that

$$\begin{aligned} & \alpha' = \frac{1 - p_g(1-\alpha)}{1 - p_b(1-\alpha)} > \alpha \\ \text{iff} \quad & 1 - p_g(1-\alpha) > \alpha - \alpha p_b(1-\alpha) \\ \text{iff} \quad & (1-\alpha)(1-p_g) > -\alpha p_b(1-\alpha) \\ \text{iff} \quad & (1-p_g) > -\alpha p_b \end{aligned}$$

which always holds because $p_b, p_g \in (0,1)$ and $\alpha \in [0,1)$. Therefore, for any such $\alpha \in [0,1)$, $\Phi(C_\alpha) = C_{\alpha'} \subsetneq C_\alpha$. Notice that $\Phi^n(C_0) = \bigcap_{k \leq n} \Phi^k(C_0) = C_{\alpha_n}$, where $\Phi^n(C_0) := \Phi\big(\Phi^{n-1}(C_0)\big)$, $\alpha_n = \frac{1-p_g(1-\alpha_{n-1})}{1-p_b(1-\alpha_{n-1})}$, and $\alpha_0 = 0$. This means iterating the operator Φ from $C_0 = \mathbb{R}^2_+$ induces a strictly increasing sequence $(\alpha_n)_{n=0}^\infty \in [0,1)$, and a corresponding sequence of strictly nested sets C_{α_n}. It is easy to see that $\lim_{n\to\infty} \alpha_n = 1$, and therefore, $\lim_{n\to\infty} \Phi^n(C_0) = C_1 = V$.

To see that $V := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq v_b\}$ is a fixed point of Φ, we apply the operator Φ to V. Take any contingent continuation utility $\mathbf{v} \in V$, and consider the policy $(k, m_i, \mathbf{w}_i)$ such that $k = 0$, $\mathbf{w}_i = \mathbf{0}$, and $m_i = -v_i$ for $i = b, g$. Clearly, [PK_g] and [PK_b] are satisfied. Incentive compatibility [IC] holds because $v_g \geq -m_b = v_b$, and $\mathbf{v} \in V$. Thus, V is the largest fixed point of Φ, which completes the proof. □

Proof of Proposition 3.1. The existence (and uniqueness) of the function Q, its concavity, differentiability, supermodularity, continuity of the optimal contract as well as certain other properties are established in Theorem 1 of SA via routine arguments. □

B. Proofs from Section 4

Lemma B.1. The efficient surpluses of the firm are

$$\bar{Q}(b) = \frac{1 - p_g\delta}{(1-\delta)(1-\Delta\delta)}[p_b R(\bar{k}_b) - \bar{k}_b] + \frac{p_b\delta}{(1-\delta)(1-\Delta\delta)}[p_g R(\bar{k}_g) - \bar{k}_g]$$
$$\bar{Q}(g) = \frac{(1 - p_g)\delta}{(1-\delta)(1-\Delta\delta)}[p_b R(\bar{k}_b) - \bar{k}_b] + \frac{1-\delta+p_b\delta}{(1-\delta)(1-\Delta\delta)}[p_g R(\bar{k}_g) - \bar{k}_g]$$

Proof. We simply solve equations [4.1] and [4.2] that jointly determine $\bar{Q}(s)$. $\square$

To properly show the existence of threshold contingent utilities, we define the unconstrained sets alternatively using the derivatives of Q and show they are equivalent to sets E_s defined in the text. This alternative definition has the following procedure. First, we fix the level of v_b and find the smallest value of v_g at which Q_g becomes zero. This defines a cutoff curve as functions of v_b along which Q_g is zero. Second, we find the smallest value of v_b at which Q_b becomes zero along the defined cutoff curve.

Lemma B.2. There exists an increasing function $f_s : \mathbb{R}_+ \to \mathbb{R}_+$ such that (i) $f_s(v_b) > v_b$ for all $v_b \geq 0$, (ii) $Q_g(\mathbf{v}, s) = 0$ if $v_g \geq f_s(v_b)$, and (iii) $Q_g(\mathbf{v}, s) > 0$ if $v_b \leq v_g < f_s(v_b)$.

Proof. First, we show that for any $v_b \geq 0$, there exists some $v_g > v_b$ such that $Q_g(\mathbf{v}, s) = 0$. Suppose not. Then there is some $\hat{v}_b \geq 0$ such that $Q_g((\hat{v}_b, \hat{v}_g), s) > 0$ for all $\hat{v}_g \geq \hat{v}_b$. The supermodularity of Q further implies that $Q_g((v'_b, v'_g), s) \geq Q_g((\hat{v}_b, v'_g), s) > 0$ for any $\mathbf{v}' \in V$ with $v'_b > \hat{v}_b$. But this is a contradiction with Theorem 1 (b) of SA.

So for any $v_b \geq 0$, we can define

$$f_s(v_b) := \min\{x \geq v_b : Q_g((v_b, x), s) = 0\} \tag{B.1}$$

which proves (i), (ii) and (iii).

To see that $f_s(\cdot)$ is increasing, take any v_b and v'_b such that $0 \leq v_b < v'_b$. Supermodularity of Q implies $0 \leq Q_g[(v_b, f_s(v'_b)), s] \leq Q_g[(v'_b, f_s(v'_b)), s] = 0$. So we have $f_s(v_b) \leq f_s(v'_b)$ by the definition of $f_s(\cdot)$. $\square$

Lemma B.3. $Q_b[(v_b, f_s(v_b)), s] \geq 0$ for any v_b, and equals zero at some v_b.

Proof. Note that the definition of $f_s(\cdot)$ means $Q_g[(v_b, f_s(v_b), s] = 0$ for any $v_b \geq 0$. Hence, $Q_b[(v_b, f_s(v_b)), s] = \mathsf{D}_{(1,1)}[(v_b, f_s(v_b)), s] \geq 0$. Moreover, Theorem 1 (b) of SA implies that there exists some $\hat{\mathbf{v}}_s \in V$ such that $Q_b(\hat{\mathbf{v}}_s, s) = Q_g(\hat{\mathbf{v}}_s, s) = 0$. By the definition in [B.1], $f_s(\hat{v}_{sb}) \leq \hat{v}_{sg}$. Then $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] \leq Q_b(\hat{\mathbf{v}}_s, s) = 0$ by the supermodularity of Q. So it has to be $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] = 0$. $\square$

By Lemma B.3, we can define the threshold contingent utilities as

[B.2] $$\bar{v}_{sb} = \min\{v_b \geq 0 : Q_b((v_b, f_s(v_b)), s) = 0\} \quad \text{and} \quad \bar{v}_{sg} = f_s(\bar{v}_{sb})$$

From Lemma 3.1 of SA, $Q_b((0, f_s(0)), s) = \infty$, which implies $\bar{v}_{sb} > 0$. Next, we define the efficient sets of contingent utilities as

[B.3] $$E_s^* := \{\mathbf{v} \in V : v_b \geq \bar{v}_{sb}, v_g \geq f_s(v_b)\}$$

Lemma B.4. $\mathbf{v} \in E_s^*$ if and only if $Q_b(\mathbf{v}, s) = Q_g(\mathbf{v}, s) = 0$. Moreover, $E_s^* = E_s := \{\mathbf{v} \in V : Q(\mathbf{v}, s) = \bar{Q}(s)\}$, and E_s is closed and convex.

Proof. We shall first prove the 'if' part. To see this, let $\mathbf{v} \in V$ such that $Q_b(\mathbf{v}, s) = 0 = Q_g(\mathbf{v}, s)$. Because Q is concave and continuously differentiable, Q achieves its maximum value at any such point.

If $v_g < f_s(v_b)$, then the definition of f_s in [B.1] implies $Q_g(\mathbf{v}, s) > 0$, a contradiction. If $v_b < \bar{v}_{sb}$ and $v_g \geq f_s(v_b)$, then $0 < Q_b[(v_b, f_s(v_b)), s] \leq Q_b(\mathbf{v}, s)$. The first inequality is by the definition of $\bar{v}_{sb}$ in [B.2], and the second inequality is because the supermodularity of Q. Hence, $\mathbf{v} \in E_s^*$.

To see the 'only if' part, consider any $\mathbf{v} \in E_s^*$. By the definition of E_s^*, $v_g \geq f_s(v_b)$. Hence, by the definition of f_s, $Q_g(\mathbf{v}, s) = 0$. Then we know $Q_b(\mathbf{v}, s) = \mathsf{D}_{(1,1)}(\mathbf{v}, s) \geq 0$. Moreover, the concavity of Q implies $Q_b(\mathbf{v}, s) \leq Q_b((\bar{v}_{sb}, v_g), s)$. So if we can show $Q_b((\bar{v}_{sb}, v_g), s) = 0$, then we can also conclude $Q_b(\mathbf{v}, s) = 0$. Now take any $\hat{v}_g \geq \bar{v}_{sg}$ and any sufficiently small $\varepsilon > 0$. Monotonicity of f_s implies

$$f_s(\bar{v}_{sb} - \varepsilon) \leq f_s(\bar{v}_{sb}) = \bar{v}_{sg} \leq \hat{v}_g$$

So by the definition of f_s, $Q_g\big((\bar{v}_{sb}, \hat{v}_g), s\big) = Q_g\big((\bar{v}_{sb} - \varepsilon, \hat{v}_g), s\big) = 0$. Since these relations hold for arbitrary $\hat{v}_g \geq \bar{v}_{sg}$, we have

$$Q((\bar{v}_{sb}, v_g), s) = Q(\bar{\mathbf{v}}, s) \;\text{ and }\; Q((\bar{v}_{sb} - \varepsilon, v_g), s) = Q((\bar{v}_{sb} - \varepsilon, \bar{v}_{sg}), s)$$

which further implies

$$\begin{aligned} Q_b((\bar{v}_{sb}, v_g), s) &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\big[Q((\bar{v}_{sb}, v_g), s) - Q((\bar{v}_{sb} - \varepsilon, v_g), s)\big] \\ &= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}\big[Q((\bar{v}_{sb}, \bar{v}_{sg}), s) - Q((\bar{v}_{sb} - \varepsilon, \bar{v}_{sg}), s)\big] \\ &= Q_b(\bar{\mathbf{v}}_s, s) = 0 \end{aligned}$$

where the last equality is by the definition of $\bar{v}_{sb}$. Therefore, we have $Q_b(\mathbf{v}, s) = 0$, as claimed.

Fix any $\mathbf{v} \in E_s^*$. The above result means that $Q(\mathbf{v}, s) = \max_{\tilde{\mathbf{v}} \in V} Q(\tilde{\mathbf{v}}, s)$. Lemma 1.2 of SA also shows that there exist $\hat{\mathbf{v}}_s \in V$ such that $Q(\hat{\mathbf{v}}_s, s) = \bar{Q}(s)$. Hence, $Q(\mathbf{v}, s) = \bar{Q}(s)$, and $E_s^* = E_s$. The concavity and continuity of Q now imply that E_s is closed and convex, which completes the proof. □

Proof of Proposition 4.1. It is immediate from Lemma B.4 that E_s is non-empty, closed, and convex. To prove part (a), take any $\mathbf{v} \in E_s$ and observe that because Q is concave, $Q_b(\mathbf{v}, s) = Q_g(\mathbf{v}, s) = 0$ implies $Q(\mathbf{v}, s)$ achieves its maximum $\bar{Q}(s)$. By the definition of $Q(\mathbf{v}, s)$ and $\bar{Q}(s)$ we have

$$\begin{aligned}
\bar{Q}(s) &= -\bar{k}_s + p_s[R(\bar{k}_s) + \delta \bar{Q}(g)] + (1 - p_s)\delta \bar{Q}(b) \\
&= -k(\mathbf{v}, s) + p_s[R(k(\mathbf{v}, s)) + \delta Q(\mathbf{w}_g(\mathbf{v}, s), g)] + (1 - p_s)\delta Q(\mathbf{w}_b(\mathbf{v}, s), b) \\
&= Q(\mathbf{v}, s)
\end{aligned}$$

which implies that we must have $k(\mathbf{v}, s) = \bar{k}_s$, $Q(\mathbf{w}_g(\mathbf{v}, s)) = \bar{Q}(g)$, and $\bar{Q}(b) = Q(\mathbf{w}_b(\mathbf{v}, s))$. Lemma B.4 then implies $\mathbf{w}_g(\mathbf{v}, s) \in E_g$ and $\mathbf{w}_b(\mathbf{v}, s) \in E_b$.

The first part of (b) follows immediately from the definition of E_s^* in [B.3] and from Lemma B.4 which proves that $E_s^* = E_s$. The second part of (b) is shown in the proof of Lemma F.6 below.

To prove part (c), we shall establish some intermediate claims.

(i) Let $\mathbf{v} \in E_b$. We show that $v_g - v_b \geq R(\bar{k}_b) + \delta\Delta\frac{R(\bar{k}_b)}{1-\delta\Delta}$ is necessary to obtain efficient firm surplus at $(\mathbf{v}, b)$. Let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at $(\mathbf{v}, b)$. We know $k = \bar{k}_b$ from part (c). The constraint [IC*] at $(\mathbf{v}, b)$ implies $v_g - v_b \geq R(\bar{k}_b) + \delta\Delta(w_{bg} - w_{bb}) \geq R(\bar{k}_b)$. Moreover, we must also have $\mathbf{w}_b \in E_b$. Otherwise, $Q(\mathbf{v}, b)$ will be smaller than the first best surplus $\bar{Q}(b)$, a contradiction. The constraint [IC*] at $(\mathbf{w}_b, b)$ implies that $w_{bg} - w_{bb} \geq R(\bar{k}_b)$. So we have $v_g - v_b \geq R(\bar{k}_b) + \delta\Delta R(\bar{k}_b)$. Repeating this procedure we obtain $v_g - v_b \geq (1 + \delta\Delta + \delta^2\Delta^2 + \ldots)R(\bar{k}_b) = R(\bar{k}_b) + \delta\Delta\frac{R(\bar{k}_b)}{1-\delta\Delta}$.

(ii) Let $\mathbf{v} \in E_g$ and let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at $(\mathbf{v}, g)$. Similar argument shows that $k = \bar{k}_g$, and $\mathbf{w}_b \in E_b$. So [IC*] at $(\mathbf{v}, g)$ implies that $v_g - v_b \geq R(\bar{k}_g) + \delta\Delta(w_{bg} - w_{bb}) \geq R(\bar{k}_g) + \delta\Delta\frac{R(\bar{k}_b)}{1-\delta\Delta}$, because $w_{bg} - w_{bb} \geq \frac{R(\bar{k}_b)}{1-\delta\Delta}$ from the first step.

(iii) Take any $\mathbf{v} \in E_s$ for $s = b, g$. Let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at $(\mathbf{v}, s)$. We show that $w_{bg} - w_{bb} \geq \frac{\delta\bar{v}_{bg} - v_b}{\delta(1-p_b)}$. Suppose not. Then we find that

$$w_{bg} - w_{bb} < \frac{\delta\bar{v}_{bg} - v_b}{\delta(1 - p_b)} = \frac{\delta\bar{v}_{bg} + m_b - \delta[p_b w_{bg} + (1 - p_b)w_{bb}]}{\delta(1 - p_b)} \tag{B.4}$$

where the equality is from [PK$_b$]. Rearranging [B.4] we get $w_{bg} < \bar{v}_{bg} + \frac{m_b}{\delta} \leq \bar{v}_{bg}$. This means $\mathbf{w}_b \in V \setminus E_b$ by part (a). Hence, $Q(\mathbf{w}, b) < \bar{Q}(b)$, implying $Q(\mathbf{v}, s) < \bar{Q}(s)$, a contradiction with $\mathbf{v} \in E_s$. Since $k = \bar{k}_s$, [IC*] at $(\mathbf{v}, s)$ implies that $v_g - v_b \geq R(\bar{k}_s) + \delta\Delta\frac{\delta\bar{v}_{bg} - v_b}{\delta(1-p_b)}$.

Combining (i) to (iii) we conclude that part (c) must hold. □

C. Auxiliary Problem

To proceed with the proofs of claims in Section 5 and beyond, it is convenient to define an auxiliary problem as follows:

$$\text{[P3]} \qquad \Psi(y,s) = \max_{x_g \geq x_b \geq 0} \delta Q(\mathbf{x}, s)$$
$$\text{subject to} \quad y \geq \delta(p_s x_g + (1-p_s)x_b)$$

where $y \geq 0$ and $s = b, g$. To ease notation, let $\mathbf{x}(y,s)$ be a solution to problem [P3]. In addition, we define the directional derivative of the surplus function at any feasible state $(\mathbf{v}, s)$ to be $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) := Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s)$.

Lemma C.1. Function $\Psi(y,s)$ defined in [P3] has the following properties:
(a) $\Psi(y,s)$ is continuously differentiable, increasing, and concave in y.
(b) $\Psi_y(y,s) = \mathsf{D}_{(1,1)}\, Q(\mathbf{x}(y,s))$.
(c) When $y \geq \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$, $\Psi(y,s) = \delta \bar{Q}(s)$, $\Psi_y(y,s) = 0$, and $\mathbf{x}(y,s) \in E_s$.
(d) When $y < \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$, $\Psi(y,s)$ is strictly increasing in y, and the constraint of [P3] is active, ie, the Lagrange multiplier for the constraint is strictly positive (which implies that the constraint holds as an equality).

Proof. (a) By the Theorem of the Maximum, $\Psi(\cdot, s)$ is continuous. Continuous differentiability follows from an argument similar to that used to prove that Q is continuously differentiable, ie, the Benveniste-Scheinkman Theorem — see Theorem 4.11 of Stokey, Lucas and Prescott (1989). Notice that raising y always relaxes the constraint in problem [P3], so we have $\Psi_y(y,s) \geq 0$. Moreover, because [P3] has a concave objective (shown in Theorem 1 (a) of SA) and a convex constraint, $\Psi(y,s)$ is concave in y.

(b) Let γ_s be the Lagrange multiplier of [P3]. The first order conditions and the envelope condition for problem [P3] are

$$\text{[C.1]} \qquad Q_b(\mathbf{x}(y,s),s) = (1-p_s)\gamma_s \quad \text{and} \quad Q_g(\mathbf{x}(y,s),s) = p_s\gamma_s$$
$$\text{[C.2]} \qquad \Psi_y(y,s) = \gamma_s$$

It is easy to see from [C.1] and [C.2] that $\Psi_y(y,s) = \mathsf{D}_{(1,1)}\, Q(\mathbf{x}(y,s),s)$.

(c) When $y \geq \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$, $\mathbf{x}(y,s) = \bar{\mathbf{v}}_s$ is feasible in [P3]. Because $Q(\cdot, s)$ reaches its upper bound $\bar{Q}(s)$ at $\bar{\mathbf{v}}_s$, we must have $\Psi(y,s) = \delta\bar{Q}(s)$. Obviously, $\Psi_y(y,s) = 0$, because $\Psi(y,s)$ is a constant. Proposition 4.1 implies that $\mathbf{x}(y,s) \in E_s$.

(d) Suppose there exists some $\tilde{y}_s < \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$ such that $\Psi_y(\tilde{y}_s, s) = 0$. Concavity of Ψ implies that $\Psi(\tilde{y}_s, s) = \Psi(\bar{y}_s, s) = \delta\bar{Q}(s)$, where $\bar{y}_s = \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$. From Proposition 4.1, we know $\mathbf{x}(\tilde{y}_s, s) \geq \bar{\mathbf{v}}_s$. However, this implies $\tilde{y}_s \geq \delta\, \mathbb{E}^s[\bar{\mathbf{v}}_s]$ by the constraint of

[P3], a contradiction. Therefore, $\Psi(y,s)$ is strictly increasing in y when $y < \delta\,\mathbb{E}^s[\bar{\mathbf{v}}_s]$. The envelope condition is $\Psi_y(y,s) = \gamma_s$. So when $y < \delta\,\mathbb{E}^s[\bar{\mathbf{v}}_s]$, we know $\gamma_s > 0$ and therefore the constraint of [P3] is active, and by complementary slackness, holds as an equality. □

Recall that $\mathbf{w}_i(\mathbf{v},s)$ (for $i = b,g$) is the continuation contingent utility policy in the optimal contract, which is the optimal policy for the value function Q in [VF]. Also, let $\alpha(\mathbf{v},s)$ be the Lagrange multiplier for the incentive compatibility constraint [IC].

Lemma C.2. For any $(\mathbf{v},s) \in V \times S$, the optimal policy $\mathbf{w}_i(\mathbf{v},s)$ satisfies
(a) $\mathbf{w}_g(\mathbf{v},s)$ is a solution to problem [P3] at (v_g,g), and is independent of v_b and s.
(b) $\mathbf{w}_g(\mathbf{v},s) \in E_g$ when $v_g \geq \delta\,\mathbb{E}^s[\bar{\mathbf{v}}_s]$.
(c) If $\alpha(\mathbf{v},s) = 0$, then $\mathbf{w}_b(\mathbf{v},s)$ is a solution to problem [P3] at (v_b,b).

Proof. Part (a) is a direct implication of Lemma 1.8 in SA. Part (b) simply follows from the result (c) in Lemma C.1, because $\mathbf{w}_g(\mathbf{v},s)$ is a solution to problem [P3] at (v_g,g). Part (c) holds simply because when [IC] does not bind at $(\mathbf{v},s)$, $\mathbf{w}_b(\mathbf{v},s)$ has to solve the problem without [IC]. □

Lemma C.3. The function $\Psi(\cdot,s)$ is strictly concave on the set $[0,\delta\,\mathbb{E}^s[\bar{\mathbf{v}}_s])$.

Proof. Let $\bar{y}_s = \delta\,\mathbb{E}^s[\bar{\mathbf{v}}_s]$. Take any distinct $\hat{y},\tilde{y} \in [0,\bar{y}_s)$. From Lemma C.1, the constraint for Problem [P3] binds at both $y = \hat{y}$ and $y = \tilde{y}$. So $\hat{y} \neq \tilde{y}$ implies $\mathbf{x}(\hat{y},s) \neq \mathbf{x}(\tilde{y},s)$. Moreover, from equations [C.1] and [C.2], we can see that

$$\min[Q_b(\mathbf{x}(y,s),s),\, Q_g(\mathbf{x}(y,s),s)] > 0$$

when $y < \bar{y}_s$, because $\Psi_y(y,s) > 0$. This means $(\mathbf{x}(\hat{y},s),s), (\mathbf{x}(\tilde{y},s),s) \in H$. Hence,

$$\begin{aligned}
\theta\Psi(\hat{y},s) + (1-\theta)\Psi(\tilde{y},s) &= \delta[\theta Q(\mathbf{x}^*(\hat{y},s),s) + (1-\theta)Q(\mathbf{x}^*(\tilde{y},s),s)] \\
&< \delta Q[\theta\mathbf{x}^*(\hat{y},s),s) + (1-\theta)\mathbf{x}^*(\tilde{y},s),s] \\
&\leq \Psi(\theta\hat{y} + (1-\theta)\tilde{y},s)
\end{aligned}$$

where the second line is implied by Proposition 2.5 of SA. □

D. Optimal Conditions and Directional Derivative

In this section, we first derive the first order conditions an envelope conditions that are necessary and sufficient for the firm's maximization problem [VF]. We then show the directional derivative of the surplus function is a nonnegative martingale and examine how it evolves in the optimal contract.

In what follows, $\eta_g(\mathbf{v}, s)$ and $\eta_b(\mathbf{v}, s)$ are the Lagrange multipliers for the promise keeping constraints [PK_g] and [PK_b], $\alpha(\mathbf{v}, s)$ is the Lagrange multiplier for the incentive compatibility constraint [IC], and $\mu_b(\mathbf{v}, s)$ and $\mu_g(\mathbf{v}, s)$ are the multipliers for the liquidity constraints [LL] when the current period's state is reported to be b or g respectively. This leads us to the first order conditions

$$
\begin{aligned}
&[\text{FOC}k] && R'(k) = 1/[p_s - \eta_g(\mathbf{v}, s) + \mu_g(\mathbf{v}, s)] \\
&[\text{FOC}w_{bb}] && (1 - p_s) Q_b(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s)(1 - p_b) + \alpha(\mathbf{v}, s)(1 - p_g) \\
&[\text{FOC}w_{bg}] && (1 - p_s) Q_g(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s) p_b + \alpha(\mathbf{v}, s) p_g \\
&[\text{FOC}w_{gb}] && p_s Q_b(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s)(1 - p_g) - \alpha(\mathbf{v}, s)(1 - p_g) \\
&[\text{FOC}w_{gg}] && p_s Q_g(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s) p_g - \alpha(\mathbf{v}, s) p_g
\end{aligned}
$$

In addition, we also have the following envelope conditions

$$
\begin{aligned}
&[\text{En}v_b] && Q_b(\mathbf{v}, s) = \eta_b(\mathbf{v}, s) \\
&[\text{En}v_g] && Q_g(\mathbf{v}, s) = \eta_g(\mathbf{v}, s)
\end{aligned}
$$

We show in the following that the directional derivative is a nonnegative martingale. In addition, it must split, ie, it goes down after a good shock and goes up after a bad shock, if the previous period had a good shock.

Lemma D.1. The process $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s)$ induced by the optimal contract is a nonnegative martingale.

Proof. Take any $(\mathbf{v}, s) \in V \times S$. Adding the first order conditions [$\text{FOC}w_{bb}$] to [$\text{FOC}w_{gg}$], and using envelope conditions [$\text{En}v_b$] and [$\text{En}v_g$] to substitute $\eta_i(\mathbf{v}, s)$, we get

$$
[\text{D}.1] \qquad (1 - p_s)\, \mathsf{D}_{(1,1)}\, Q[\mathbf{w}_b(\mathbf{v}, s), b] + p_s\, \mathsf{D}_{(1,1)}\, Q[\mathbf{w}_g(\mathbf{v}, s), g] = \mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s)
$$

Moreover,

$$
\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) = \lim_{\varepsilon \downarrow 0} \frac{Q(\mathbf{v} + \varepsilon(1, 1), s) - Q(\mathbf{v}, s)}{\varepsilon} \geq 0
$$

where the inequality is by Theorem 1 (e) of SA. Hence, the process $\mathsf{D}_{(1,1)}\, Q$ is a nonnegative martingale. $\square$

Lemma D.2. Suppose the optimal contract starts at $(\mathbf{v}, s)$ and evolves to the state $(\mathbf{w}_g, g)$ satisfying $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) > 0$ after a good shock. Then the directional derivative goes down after another good shock and goes up after another bad shock, ie, $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g^g, g) < \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g)$ and $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_b^g, b) > \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g)$.

Proof. We shall show in two cases that $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g^g, g) < \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g)$. The conclusion that $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_b^g, b) > \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g)$ then simply follows from the martingale equation [D.1] at the state $(\mathbf{w}_g, g)$.

First, we consider the case of $w_{gg} < \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. From [$\mathrm{PK}_g$], we know that $v_g \leq \delta\, \mathbb{E}^g[\mathbf{w}_g] \leq \delta w_{gg} < w_{gg}$, implying $v_g < w_{gg} < \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. By Lemma C.3, $\Psi_y(w_{gg}, g) < \Psi_y(v_g, g)$, because $\Psi(\cdot, g)$ is strictly concave on $(0, \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g])$. By Lemmas C.1 and C.2, we see that $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) = \Psi_y(v_g, g)$ and $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g^g, g) = \Psi_y(w_{gg}, g)$. So the conclusion follows.

Second, we consider the case of $w_{gg} \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. From part (b) of Lemma C.2, we know that $\mathbf{w}_g^g \in E_g$. So the left hand side of [$\mathrm{FOC}w_{gg}$] at $(\mathbf{w}_g, g)$ is zero implying $\eta_g(\mathbf{w}_g, g) = \alpha(\mathbf{w}_g, g)$. The condition $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) > 0$ implies that $(\mathbf{w}_g, g) \in H$ by Lemma E.1. Hence, $\alpha(\mathbf{w}_g, g) = Q_g(\mathbf{w}_g, g) > 0$ by the definition of set H. And equation [3.5] of SA simply means $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g^g, g) < \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g)$. □

E. Proof of Theorem 2

We will prove the various parts of Theorem 2 in turn.

E.1. Proof of Theorem 2 (b) and (c): Transition and Convergence

Define the process $(\mathbf{v}^{(t)}, s_{t-1})_{t=0}^{\infty}$ to be the states induced by the optimal contract starting at some $(\mathbf{v}^{(0)}, s_{-1}) \in V \times S$. In the high persistence case the optimal contract may need enough good shocks (at least two) to reach efficient sets. So to establish the result that efficient sets are achieved in finite time, we need to consider sequences with two good shocks in a row happen infinitely often. We begin with some preliminary lemmata.

Lemma E.1. Let $\mathbf{w}_i = \mathbf{w}_i(\mathbf{v}, s)$ for any $(\mathbf{v}, s) \in V \times S$. We must have either $(\mathbf{w}_g, g) \in H$ or $\mathbf{w}_g \in E_g$. For any $(\mathbf{v}, s) \in H$, we must have $(\mathbf{w}_b, b) \in H$.

Proof. Fix $(\mathbf{v}, s) \in V \times S$. From a little manipulation of the first order conditions [$\mathrm{FOC}w_{gb}$] and [$\mathrm{FOC}w_{gg}$], we get $p_g Q_b(\mathbf{w}_g, g) = (1 - p_g) Q_g(\mathbf{w}_g, g)$. So we must either have $\min[Q_b(\mathbf{w}_g, g), Q_g(\mathbf{w}_g, g)] > 0$, or have $Q_b(\mathbf{w}_g, g) = 0 = Q_g(\mathbf{w}_g, g)$. The former case means $(\mathbf{w}_g, g) \in H$ by definition of H, while the latter case means that $\mathbf{w}_g \in E_g$.

Now consider any $(\mathbf{v}, s) \in H$. From [$\mathrm{FOC}w_{bb}$] and [$\mathrm{FOC}w_{bg}$] we know

$$(1 - p_s) Q_b(\mathbf{w}_b, b) \geq (1 - p_b) Q_b(\mathbf{v}, s) > 0$$
$$(1 - p_s) Q_g(\mathbf{w}_b, b) \geq p_b Q_b(\mathbf{v}, s) > 0$$

The first inequality in both lines is because $\alpha(\mathbf{v}, s) \geq 0$. Hence $(\mathbf{w}_b, b) \in H$. □

Notice that Lemma E.1 establishes part (b) of Theorem 2. We now proceed to establish part (c).

Lemma E.2. The non-negative martingale $\left(\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})\right)_{t=0}^{\infty}$ converges to 0 almost surely.

Proof. By Lemma D.1, the process $\left(\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})\right)_{t=0}^{\infty}$ is a nonnegative martingale. Then, the Martingale Convergence Theorem — see, for instance, Theorem 1 and Corollary 3 on pp 508–509 of Shiryaev (1995) — ensures $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges almost surely to a non-negative and integrable random variable.

Because $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ is a non-negative martingale, it follows immediately that if $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\tau)}, s_{\tau-1}) = 0$ for some (possibly random) time τ, then we must have $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ at all subsequent times $t \geq \tau$. Therefore, in order to show that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges to 0 almost surely, it suffices to consider the dates and states where $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) > 0$.

Consider a path where good shocks occur infinitely many times and let $a = \lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$. Suppose also that $a > 0$. Then, following the observation in the previous paragraph, we find that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) > 0$ for all $t \geq 0$. Lemma E.1 then implies $\mathbf{w}_g(\mathbf{v}^{(t)}, s_{t-1}) \in H$ for all $t \geq 0$.

Now consider the subsequence on this path that has only good shocks. This subsequence stays in H forever. Moreover, by Lemma 3.4 (a) of SA, this subsequence is bounded and must have a convergent subsequence $(\mathbf{v}^{(\tau_t)}, g)_{t=0}^{\infty}$ with limit $(\hat{\mathbf{v}}, g)$. By definition, $s_{\tau_t - 1} = g$, and it is either the case that $s_{\tau_t} = b$ infinitely often (ie, the 'good-bad' shock pair occurs infinitely often), or the case that $s_{\tau_t} = g$ infinitely often (ie, the 'good-good' shock pair occurs infinitely often) along this subsequence.

In the former case, let $(\mathbf{v}^{(\tau_t')}, g)_{t=0}^{\infty}$ be a subsequence of $(\mathbf{v}^{(\tau_t)}, g)_{t=0}^{\infty}$ where $s_{\tau_t'-1} = g$ and $s_{\tau_t'} = b$. This means $\mathbf{v}^{(\tau_t'+1)} = \mathbf{w}_b(\mathbf{v}^{(\tau_t')}, g)$. The continuity of $\mathbf{w}_b(\cdot, g)$ implies $\lim_{t\to\infty} \mathbf{v}^{\tau_t'+1} = \mathbf{w}_b(\hat{\mathbf{v}}, g)$. Moreover, because $\mathsf{D}_{(1,1)}\, Q$ is continuous,

$$\lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\tau_t')}, g) = \mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, g) = a > 0 \quad \text{and}$$
$$\lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\tau_t'+1)}, b) = \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_b(\hat{\mathbf{v}}, g), b) = a > 0$$

But now recall Lemma D.2, which implies that if $\mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, g) > 0$, then it must be that $\mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, g) < \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_b(\hat{\mathbf{v}}, g), b)$, which contradicts the displayed equations above.

In the latter case, we can use the same argument to show that $\mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, g) = \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g(\hat{\mathbf{v}}, g), g) = a$. But this case also contradicts Lemma D.2. Therefore, we must have $a = 0$.

Finally, by Proposition 4.2 of SA, paths with only finitely many good shocks have measure zero, so we have $\lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ almost surely. □

Proof of Theorem 2 (c). Lemma E.2 shows that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s)$ is a martingale that converges to 0 almost surely. All that remains is to show that this convergence occurs in finite time almost surely.

Towards this end, let us consider a path with the property that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) > 0$ for all $t < \infty$ and where the 'good-good' shock pair occurs infinitely often. Take a subsequence $(\mathbf{v}^{(\gamma_t)}, s_{\gamma_t-1})_{t=0}^{\infty}$ such that $s_{\gamma_t-1} = s_{\gamma_t} = g$. From Lemma E.1, $(\mathbf{v}^{(\gamma_t)}, s_{\gamma_t-1}) \in H$ for all $t < \infty$ along this path.

Let $\varepsilon > 0$ be some sufficiently small number. Note the set $C := \{\mathbf{v} \in V : (\mathbf{v}, g) \in \mathrm{cl}(H), v_g \leq \bar{v}_{gg} - \varepsilon\}$ is a compact set that. Also, we know from Lemma 3.4 (b) of SA that $(\bar{\mathbf{v}}_s, s)$ are the only points in $\mathrm{cl}(H)$ that have zero directional derivative. The continuity of $\mathsf{D}_{(1,1)}\, Q(\cdot, g)$ implies that there exists some $\eta > 0$ such that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, g) \geq \eta$ for all $\mathbf{v} \in C$. Moreover, because $\lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\gamma_t)}, g) = 0$ by Lemma E.2, there must exist some T such that $\mathbf{v}^{(\gamma_t)} \notin C$ for any $t \geq T$. In other words, $\bar{v}_{gg} - v_g^{(\gamma_t)} < \varepsilon$ for all $t \geq T$. But this means $(\mathbf{v}^{(\gamma_T)}, s_{\gamma_{T-1}}) \in A_{1,g}$, because ε is sufficiently close to zero. And since $s_T = g$ by construction, we must have $(\mathbf{v}^{(\gamma_T+1)}, g) \in E_g$ and $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\gamma_T+1)}, g) = 0$, a contradiction.

In Proposition 4.2 of SA, we show that paths with infinitely many 'good-good' shocks occur with probability one. Hence, the argument above shows $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ for some $t < \infty$ almost surely. Let $\tau = \min\{t : \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0\}$. From Proposition 4.1, we know $(\mathbf{v}^{(\gamma_t)}, s_{t-1}) \in E_{s_{t-1}}$ for any $t \geq \tau$. So in a maximum rent contract, the contingent utilities $\mathbf{v}^{(\gamma_t)} = \bar{\mathbf{v}}_{s_{t-1}}$ for all $t \geq \tau$. □

E.2. Proof of Theorem 2 (a) and (d): Initiation and High Persistence

We begin with a preliminary lemma.

Lemma E.3. $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$ if and only if $\mathbf{w}_g(\mathbf{v}, s) \in E_g$.

Proof. Suppose $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. From Lemma C.1 (c), $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. That is the contingent utility reaches the unconstrained set after a good shock. Now suppose $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. From Proposition 4.1 (a), $w_{gi}(\mathbf{v}, s) \geq \bar{v}_{gi}$ for $i = b, g$. Then by [PK_g] at $(\mathbf{v}, s)$, we know $v_g \geq \delta\, \mathbb{E}^g[\mathbf{w}_g(\mathbf{v}, s)] \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. □

Proof of Theorem 2 (a) and (d). Let $(\mu, 1-\mu)$ be a probability measure over $\{g, b\}$. Abusing notation, we denote by $Q(\mathbf{v}, \mu)$ the value of firm surplus when the initial measure over states is given by $(\mu, 1-\mu)$. Then, given such a μ, the initial contingent utility vector $\mathbf{v}_\mu^0$ maximises $P(\mathbf{v}, \mu) := Q(\mathbf{v}, \mu) - \mathbb{E}^\mu[\mathbf{v}]$, where $\mathbf{v} \in V$. The first order conditions for this initialization problem are

[E.1] $$Q_g(\mathbf{v}_\mu^0, \mu) = \mu \quad \text{and} \quad Q_b(\mathbf{v}_\mu^0, \mu) = 1 - \mu$$

which implies $(\mathbf{v}^0_\mu, \mu) \in H$ by (abusing) the definition of the set H. Then Lemma E.1 implies that the optimal contract can only leave H after a good shock and before that it stays in H which establishes the claim about initiation.

To see the rest of part (a), let $\mu_1 < \mu_2$ and suppose $(\mu_j, 1-\mu_j)$ are probability measures over $\{g, b\}$ for $j = 1, 2$. Let $(k, m_i, \mathbf{w}_i)_{i=b,g}$ be the optimal policy at $(\mathbf{v}^0_{\mu_1}, \mu_1)$. Then,

$$\begin{aligned} P(\mathbf{v}^0_{\mu_1}, \mu_1) =& -k + \mu_1 R(k) + \delta\mu_1[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] \\ &+ \delta Q(\mathbf{w}_b, b) - \mathbb{E}^{\mu_1}[\mathbf{v}^0_{\mu_1}] \\ \leq& -k + p_b\left\{R(k) + \delta[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (v^0_{\mu_1 g} - v^0_{\mu_1 b})\right\} \\ &+ \delta[Q(\mathbf{w}_b, b) - \mathbb{E}^b(\mathbf{w}_b)] \end{aligned}$$

The inequality is because [PK_b] at $(\mathbf{v}^0_{\mu_1}, \mu_1)$ implies $-v^0_{\mu_1 b} \leq -\mathbb{E}^b[\mathbf{w}_b]$. So we can rewrite the above inequality as

$$\begin{aligned} & p_b\left\{R(k) + \delta[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (v^0_{\mu_1 g} - v^0_{\mu_1 b})\right\} \\ & \geq P(\mathbf{v}^0_b, b) - \delta[Q(\mathbf{w}_b, b) - \mathbb{E}^b(\mathbf{w}_b)] + k > 0 \end{aligned} \tag{E.2}$$

The second inequality is because $\delta < 1, \mathbf{w}_b \in V$. Moreover, because $\mathbf{v}^0_{\mu_j} \in V$ for $j = 1, 2$, the optimality in choosing $\mathbf{v}^0_{\mu_2}$ implies

$$\begin{aligned} & P(\mathbf{v}^0_{\mu_2}, \mu_2) - P(\mathbf{v}^0_{\mu_1}, \mu_1) \\ & \geq Q(\mathbf{v}^0_{\mu_1}, \mu_2) - \mathbb{E}^{\mu_2}(\mathbf{v}^0_{\mu_1}) - [Q(\mathbf{v}^0_{\mu_1}, \mu_1) - \mathbb{E}^{\mu_1}(\mathbf{v}^0_{\mu_1})] \\ & = Q(\mathbf{v}^0_{\mu_1}, \mu_2) - Q(\mathbf{v}^0_{\mu_1}, \mu_1) - \Delta(v^0_{\mu_1 g} - v^0_{\mu_1 b}) \\ & \geq \Delta\left\{R(k) + \delta[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (v^0_{\mu_1 g} - v^0_{\mu_1 b})\right\} \geq 0 \end{aligned}$$

The first inequality in the last line is implied by $(k, m_i, \mathbf{w}_i)_{i=b,g} \in \Gamma(\mathbf{v}^0_{\mu_1})$. The second inequality in the last line is by [E.2] and is strict iff $\Delta > 0$, which establishes part (a).

Finally, we establish part (d). When $\mathbf{p} \in B_h$, Lemma F.6 below verifies that the contingent utility thresholds satisfy $\bar{v}_{bg} = \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$. Now consider starting at state $(\mathbf{v}, s) \in H$ and a bad shock occurs. We know from Lemma E.1 that $\mathbf{w}_b(\mathbf{v}, s) \in H$. Moreover, from Lemma 3.4 in SA, we know any $(\mathbf{v}, s) \in H$ satisfies $\mathbf{v} \ll \bar{\mathbf{v}}_s$. Hence, $w_{bg}(\mathbf{v}, s) < \bar{v}_{bg} = \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$. Then it's easy to see from Lemma F.5 that $\mathbf{w}_g[\mathbf{w}_b(\mathbf{v}, s), b] \notin E_g$. This means the contingent utility vector will not reach the unconstrained set E_g after another good shock. But because the firm reaches the unconstrained stage only after a good shock, it is only possible to become unconstrained after another two or more good shocks in a row. $\square$

We end with a lemma that describes the behaviour of continuation utility after consecutive bad shocks.

Lemma E.4. Take any $(\mathbf{v}, s) \in V \times S$. Let $\mathbf{w}_b = \mathbf{w}_b(\mathbf{v}, s)$ and $\mathbf{w}_b^b = \mathbf{w}_b[\mathbf{w}(\mathbf{v}, s), b]$. If $w_{bb} \leq v_b$, then $w_{bb}^b \leq w_{bb}$.

Proof. Note that Lemma E.1 implies $(\mathbf{w}_b, b)$ and $(\mathbf{w}^b, b)$ both lie in the set H. We will verify in Theorem 3 below that $m_b(\mathbf{v}, s) = m_b(\mathbf{w}_b, b) = 0$, which further implies (by [PK$_b$])

$$\delta\, \mathbb{E}^b[\mathbf{w}_b^b] = w_{bb} \leq v_b = \delta\, \mathbb{E}^b[\mathbf{w}_b] \tag{E.3}$$

Suppose $w_{bb}^b > w_{bb}$. Then the above display means $w_{bg} > w_{bg}^b$. The strict concavity and supermodularity of Q in H then imply $Q_b(\mathbf{w}_b^b, b) < Q_b(\mathbf{w}_b, b)$. However, from [FOC$w_{bb}$], we know $Q_b(\mathbf{w}_b^b, b) \geq Q_b(\mathbf{w}_b, b)$ because $\alpha(\mathbf{w}, b) \geq 0$. This is a contradiction. Thus, we must have $w_{bb}^b \leq w_{bb}$. □

F. Proofs from Section 5

The result in this section first shows that the optimal contract always stay in the set H. Then depending on where the state lies in H, we characterize the firm policies. Lemma 3.4 of SA shows some useful properties of the set H: first, $\mathbf{v} \ll \bar{\mathbf{v}}_s$ for any $(\mathbf{v}, s) \in H$; second, $(\bar{\mathbf{v}}_s, s)$ are the only states in the closure of H that satisfy $\mathsf{D}_{(1,1)}(\mathbf{v}, s) = 0$. Lemma 2.3 of SA shows that the surplus function $Q(\mathbf{v}, s)$ is strictly concave (in v_b, v_g) for any state $(\mathbf{v}, s) \in H$.

The condition derived in Lemma 1.6 of SA regarding Lagrange multipliers also holds for Lagrange multipliers in problem [VF]. This is because the function P in [P1] (defined in SA) satisfies all the properties of the function $Q(\mathbf{v}, s)$ in [VF]. Therefore we know

$$\eta_b(\mathbf{v}, s) + \alpha(\mathbf{v}, s) - \mu_b(\mathbf{v}, s) \geq 0,\ m_b(\mathbf{v}, s)[\eta_b(\mathbf{v}, s) + \alpha(\mathbf{v}, s) - \mu_b(\mathbf{v}, s)] = 0 \tag{F.1}$$

$$\eta_g(\mathbf{v}, s) - \alpha(\mathbf{v}, s) - \mu_g(\mathbf{v}, s) = 0 \tag{F.2}$$

We can also use [F.2] to rewrite [FOCk] as:

$$R'(k(\mathbf{v}, s)) = 1/[p_s - \alpha(\mathbf{v}, s)] \tag{FOCk^*}$$

F.1. Compensation Policy and Proofs from Section 5.2

We partition the transition probability space Π into B_ℓ and B_h by defining $\psi(p_g)$ and $\hat{k}_g$ as

$$\psi(p_g) = p_g - \frac{1}{R'(\hat{k}_g)} \tag{F.3}$$

$$\hat{k}_g = R^{-1}\left(\frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g}\right) \tag{F.4}$$

The pair $(\varphi(p_g), \hat{k}_g)$ is unique for a given p_g.

Lemma F.1. The functions ψ and $\hat{k}_g$ are continuous. For any $p_g > 0$, we have $0 < \psi(p_g) < p_g$, and $\lim_{p_g \downarrow 0} \psi(p_g) = p_g$.

Proof. It is clear from [F.4] that $\hat{k}_g$ is continuous and $\hat{k}_g(0) = 0$. Continuity of ψ for all $p_g > 0$ follows immediately from [F.3]. Because $R'(0) = \infty$, we have $\lim_{p_g \downarrow 0} \psi(p_g) = p_g$. From [F.4], $\hat{k}_g < \bar{k}_g$, since $\frac{\delta p_g}{1+\delta p_g} < 1$. Concavity of R then implies that $R'(\hat{k}_g) > R'(\bar{k}_g)$. Then $\psi(p_g)R'(\hat{k}_g) = p_g R'(\hat{k}_g) - p_g R'(\bar{k}_g) > 0$, implying that $0 < \psi(p_g) < p_g$. □

Lemma F.2. The partition of transition probabilities can be characterized as follows:
(a) $\mathbf{p} \in B_\ell$ iff $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$.
(b) $\mathbf{p} \in B_h$ iff $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] \geq R(\bar{k}_b)$; the inequality is strict iff $\Delta > \psi(p_g)$.

Proof. To see part (a), notice that if $\mathbf{p} \in B_\ell$, then $p_b R'(\hat{k}_g) > 1 = p_b R'(\bar{k}_b)$. Hence, $\mathbf{p} \in B_\ell$ implies $R'(\hat{k}_g) > R'(\bar{k}_b)$. The strict concavity of R then implies $\hat{k}_g < \bar{k}_b$. From [F.4], $R(\hat{k}_g) = \dfrac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} < R(\bar{k}_b)$. We can rearrange the last inequality to obtain $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$.

Conversely, if we know $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$, then $R(\bar{k}_b) > \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} = R(\hat{k}_g)$ by [F.4]. Hence, $\bar{k}_b > \hat{k}_g$ implies $p_b R'(\hat{k}_g) > p_b R'(\bar{k}_b) = 1$. From [F.3] we have $\Delta < \psi(p_g)$. Part (b) is established by similar arguments. □

Lemma F.3. At state $(\bar{\mathbf{v}}_s, s)$, the constraints [IC] and [LL] for $s = g$ cannot both hold as strict inequalities.

Proof. Suppose not. Then at state $(\bar{\mathbf{v}}_s, s)$, we have

$$\bar{v}_{sg} - \bar{v}_{sb} > R(\bar{k}_s) + \delta\Delta(\bar{v}_{bg} - \bar{v}_{bb}), \quad R(\bar{k}_s) > \bar{m}_{sg}$$

So there exist some $\varepsilon_{IC}, \varepsilon_{LL} > 0$ such that

$$(\bar{v}_{sg} - \varepsilon_{IC}) - \bar{v}_{sb} = R(\bar{k}_s) + \delta\Delta(\bar{\mathbf{v}}_{bg} - \bar{\mathbf{v}}_{bb}), \quad R(\bar{k}_s) = \bar{m}_{sg} + \varepsilon_{LL}$$

Let $\varepsilon := \min[\varepsilon_{IC}, \varepsilon_{LL}]$, and using [PK$_b$] at state $(\bar{\mathbf{v}}_s, s)$, notice that

$$(\bar{v}_{sg} - \varepsilon) - \bar{v}_{sb} \geq R(\bar{k}_s) + \delta\Delta(\bar{\mathbf{v}}_{bg} - \bar{\mathbf{v}}_{bb}), \quad R(\bar{k}_s) \geq \bar{m}_{sg} + \varepsilon$$

with at least one of them holding as equality.

Let $\hat{\mathbf{v}}_s = (\bar{v}_{sb}, \bar{v}_{sg} - \varepsilon)$, and $\hat{m}_{sg} = \bar{m}_{sg} + \varepsilon$. Then the policy $(\bar{k}_s, \bar{m}_{sb}, \hat{m}_{sg}, \mathbf{w}_b = \bar{\mathbf{v}}_b, \mathbf{w}_g = \bar{\mathbf{v}}_g) \in \Gamma(\hat{\mathbf{v}}_s)$. The contracts at $(\bar{\mathbf{v}}_s, s)$ and at $(\hat{\mathbf{v}}_s, s)$ only differ in their transfers, which implies that $Q(\hat{\mathbf{v}}_s, s) = \bar{Q}(s)$. However, $\hat{\mathbf{v}}_s < \bar{\mathbf{v}}_s$ by construction, which implies that $\hat{\mathbf{v}}_s \notin E_s$, by Proposition 4.1. This forms a contradiction because the definition of E_s implies $Q(\hat{\mathbf{v}}_s, s) < \bar{Q}(s)$.

□

Lemma F.4. The threshold contingent utilities satisfy: $\bar{v}_{bb} = \bar{v}_{bg}$, $\bar{v}_{gb} \leq \bar{v}_{gg}$; the inequality is strict if and only if $\Delta > 0$.

Proof. The right hand side of [PK$_b$] at $(\bar{\mathbf{v}}_s, s)$ is not contingent on s because $\bar{m}_{sb} = 0$. So we must have $\bar{v}_{sb} = \delta\,\mathbb{E}^b[\bar{\mathbf{v}}_b]$, which implies that $\bar{v}_{gb} = \bar{v}_{bb}$. Using this relation and $\bar{m}_{sb} = 0$, we can rewrite [IC] at $(\bar{\mathbf{v}}_s, s)$ as

$$\bar{m}_{sg} \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) \tag{F.5}$$

Moreover, subtracting [PK$_g$] at state $(\bar{\mathbf{v}}_b, b)$ from [PK$_g$] at state $(\bar{\mathbf{v}}_g, g)$, we can get:

$$(\bar{m}_{gg} - \bar{m}_{bg}) = R(\bar{k}_g) - R(\bar{k}_b) - (\bar{v}_{gg} - \bar{v}_{bg}) \tag{F.6}$$

Suppose $\bar{v}_{gg} < \bar{v}_{bg}$. Then [F.6] implies that $0 \leq R(\bar{k}_g) - R(\bar{k}_b) < \bar{m}_{gg} - \bar{m}_{bg}$. So we must have $\bar{m}_{bg} < \bar{m}_{gg} \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) < 0$, by [F.5]. However, this means $\bar{m}_{bg} < R(\bar{k}_b)$ and $\bar{m}_{bg} < \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$. In other words, [IC] and [LL] for g both hold as strict inequality at $(\bar{\mathbf{v}}_b, b)$, a contradiction of Lemma F.3.

If $\Delta = 0$, then we must have $\bar{v}_{gg} = \bar{v}_{bg}$. Otherwise, [F.6] implies $\bar{m}_{gg} < \bar{m}_{bg} \leq R(\bar{k}_g)$. And [F.5] implies $\bar{m}_{gg} < \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$, a contradiction with Lemma F.3 at state $(\bar{\mathbf{v}}_g, g)$. If $\Delta > 0$, then we must have $\bar{v}_{gg} > \bar{v}_{bg}$. Otherwise, [F.6] and [F.6] imply that $\bar{m}_{bg} < \bar{m}_{gg} \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) = 0 < R(\bar{k}_b)$, a contradiction with Lemma F.3 at state $(\bar{\mathbf{v}}_b, b)$. □

Lemma F.5. $v_g \geq \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$ if and only if $\mathbf{w}_g(\mathbf{v}, s) \in E_g$.

Proof. Suppose $v_g \geq \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$. From Lemma C.1 (c), $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. That is the contingent utility reaches the unconstrained set after a good shock. Now suppose $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. From Proposition 4.1 (a), $w_{gi}(\mathbf{v}, s) \geq \bar{v}_{gi}$ for $i = b, g$. Then by [PK$_g$] at $(\mathbf{v}, s)$, we know $v_g \geq \delta\,\mathbb{E}^g[\mathbf{w}_g(\mathbf{v}, s)] \geq \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$. □

Proof of Theorem 3. Let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at any $(\mathbf{v}, s)$.

(a) Suppose $(\mathbf{v}, s) \in H$. If $\mu_b(\mathbf{v}, s) > 0$, then complementary slackness implies $m_b = 0$. If $\mu_b(\mathbf{v}, s) = 0$, then $\eta_b(\mathbf{v}, s) + \alpha(\mathbf{v}, s) - \mu_b(\mathbf{v}, s) > 0$, because $\eta_b(\mathbf{v}, s) > 0$ and $\alpha(\mathbf{v}, s) \geq 0$. Then [F.1] implies $m_b = 0$.
Moreover, since $(\bar{\mathbf{v}}_b, b)$ and $(\bar{\mathbf{v}}_g, g)$ both locate in the closure of set H, policy continuity then implies that $e_b(\mathbf{v}, s) = m_b(\mathbf{v}, s) = 0$ for any $(\mathbf{v}, s) \in \bar{H}$.

(b) Note that in the maximum rent contract $\mathbf{w}_g = \bar{\mathbf{v}}_g$ whenever $\mathbf{w}_g \in E_g$. So we conclude from Lemma F.5 that $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$ if and only if $\mathbf{w}_g = \bar{\mathbf{v}}_g$ in the maximum rent contract.
To see the 'if' part, suppose $v_g > \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. Since $\mathbf{w}_g = \bar{\mathbf{v}}_g$, [$\text{PK}_g$] implies

$$e_g = R(k) - m_g = v_g - \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g] \tag{F.7}$$

Hence, $e_g > 0$. To see the 'only if ' part, suppose $e_g > 0$. Then we must have $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. Suppose not. From Lemma C.2 (a), $\mathbf{w}_g$ is a solution of [P3] at (v_g, g). Then from Lemma C.1 (d), the constraint of [P3] at (v_g, g) must bind, which implies $e_g = 0$ by [PK_g] at $(\mathbf{v}, s)$, a contradiction. Moreover, the conclusion $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$ further implies that $\mathbf{w}_g = \bar{\mathbf{v}}_g$ in the maximum rent contract. Then [PK_g] implies [F.7] holds. Hence, $v_g > \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$.

(c) First, suppose $\bar{e}_{gg} = 0$. From [F.6], $\bar{e}_{bg} = \bar{v}_{bg} - \bar{v}_{gg} < 0$, which violates [LL] for b. So we must have $\bar{e}_{gg} > 0$. From Lemma F.3, [IC] must hold as equality at $(\bar{\mathbf{v}}_g, g)$. Therefore, from display [F.5], we must have $\bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$.
Next, take $\mathbf{p} \in B_\ell$. Suppose $\bar{e}_{bg} = 0$, then by [F.6] we have

$$\bar{v}_{gg} - \bar{v}_{bg} = R(\bar{k}_g) - \bar{m}_{gg} = R(\bar{k}_g) - \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$$

Rearrange to obtain

$$\delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) = \frac{\delta p_g R(\bar{k}_g)}{1 + \delta p_g} < R(\bar{k}_b) = \bar{m}_{bg}$$

The inequality is from Lemma F.2 when $\mathbf{p} \in B_\ell$. Combining this display with [F.5], we have $\bar{m}_{bg} < \bar{m}_{bg}$, a contradiction. Hence, $\bar{e}_{bg}, \bar{e}_{gg} > 0$. Then Lemma F.3 and display [F.5] imply that $\bar{m}_{bg} = \bar{m}_{gg}$. Therefore, $0 < \bar{e}_{bg} \leq \bar{e}_{gg}$, and $\bar{e}_{bg} < \bar{e}_{gg}$ if and only if $\Delta > 0$.

(d) Take $\mathbf{p} \in B_h$ and suppose $\bar{e}_{bg} > 0$. By Lemma F.3, [IC] must hold as equality at $(\bar{\mathbf{v}}_b, b)$. This means $\delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) = \bar{m}_{bg} < R(\bar{k}_b)$ by [F.5]. But this is a contradiction with Lemma F.2 which concludes that $\delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) \geq R(\bar{k}_b)$ when $\mathbf{p} \in B_h$. Therefore, $0 = \bar{e}_{bg} < \bar{e}_{gg}$. □

Lemma F.6. The threshold levels of contingent utility are as follows.

(a) for $\mathbf{p} \in B_\ell$,

$$\bar{\mathbf{v}}_b = (\bar{v}_{bb}, \bar{v}_{bg}) = \left(\frac{\delta p_b \bar{v}_{bg}}{1 - \delta(1 - p_b)}, \frac{[1 - \delta(1 - p_b)] R(\bar{k}_b)}{(1 - \delta)(1 - \delta \Delta)} \right)$$

$$\bar{\mathbf{v}}_g = (\bar{v}_{gb}, \bar{v}_{gg}) = \left(\frac{\delta p_b \bar{v}_{bg}}{1 - \delta(1 - p_b)}, \frac{\delta(p_g - \delta \Delta) \bar{v}_{bg}}{1 - \delta(1 - p_b)} + R(\bar{k}_g) \right)$$

(b) for $\mathbf{p} \in B_h$,

$$\bar{\mathbf{v}}_b = (\bar{v}_{bb}, \bar{v}_{bg}) = \left(\frac{\delta p_b \bar{v}_{bg}}{1 - \delta(1 - p_b)}, \frac{\delta p_g [1 - \delta(1 - p_b)] R(\bar{k}_g)}{(1 + \delta p_g)(1 - \delta)(1 - \delta \Delta)} \right)$$

$$\bar{\mathbf{v}}_g = (\bar{v}_{gb}, \bar{v}_{gg}) = \left(\frac{\delta p_b \bar{v}_{bg}}{1 - \delta(1 - p_b)}, \bar{v}_{bg} + \frac{R(\bar{k}_g)}{1 + \delta p_g} \right)$$

Proof. From [PK_b] and [PK_g] at $(\bar{\mathbf{v}}_b, b)$ and $(\bar{\mathbf{v}}_g, g)$ respectively, we can get

[F.8] $$\bar{v}_b = \delta \, \mathbb{E}^b [\bar{\mathbf{v}}_b]$$

[F.9] $$\bar{v}_{bg} = R(\bar{k}_b) - \bar{m}_{bg} + \delta \, \mathbb{E}^g [\bar{\mathbf{v}}_g]$$

[F.10] $$\bar{v}_{gg} = R(\bar{k}_g) - \bar{m}_{gg} + \delta \, \mathbb{E}^g [\bar{\mathbf{v}}_g]$$

where $\bar{v}_{sb} = \bar{v}_b$, because $\bar{m}_{sb} = 0$.

(a) When $\mathbf{p} \in B_\ell$, Theorem 3 (c) and Lemma F.3 imply that [IC] must hold as equality at both $(\bar{\mathbf{v}}_b, b)$ and $(\bar{\mathbf{v}}_g, g)$. Then by [F.5],

[F.11] $$\bar{m}_{bg} = \bar{m}_{gg} = \delta p_g (\bar{v}_{gg} - \bar{v}_{bg})$$

Combining [F.8] [F.9] [F.10] [F.11] and eliminating $\bar{m}_{sg}$, we can obtain the specified solutions in the Lemma.

(b) When $\mathbf{p} \in B_h$, Theorem 3 (d) and Lemma F.3 imply:

[F.12] $$\bar{m}_{bg} = R(\bar{k}_b), \quad \bar{m}_{gg} = \delta p_g (\bar{v}_{gg} - \bar{v}_{bg})$$

Combining [F.8] [F.9] [F.10] [F.12] and eliminating $\bar{m}_{sg}$, we can obtain the specified solutions in the Lemma. □

F.2. Investment Dynamics and Proofs from Section 5.2

To characterize the investment dynamics, we show in Lemma 3.5 and 3.6 of SA that given v_b and s, there exists a cutoff value $h_s(v_b)$ such that optimal investment is efficient if and only if $v_g \geq h_s(v_b)$. Formally, we can define

[F.13] $$h_s(v_b) := \min\{y \geq v_b : k((v_b, y), s) = \bar{k}_s\}$$

The function $h_s(\cdot)$ is also shown to be strictly increasing. In addition, we show in the following lemma that whenever the contract evolves to the efficient investment region (but still constrained) after a bad shock it will stay in that region until a good shock occurs.

Lemma F.7. For any $(\mathbf{v}, s) \in H$, let $\mathbf{w}_b = \mathbf{w}_b(\mathbf{v}, s)$ and $\mathbf{w}_b^b = \mathbf{w}_b[\mathbf{w}_b(\mathbf{v}, s), b]$. If $w_{bg} \geq h_b(w_{bb})$, then $w_{bg}^b \geq h_b(w_{bb}^b)$.

Proof. From Lemma E.1, $\mathbf{w}_b, \mathbf{w}_b^b \in H$ and hence $w_{bb}^b, w_{bb} < \bar{v}_{bb} = \delta\, \mathbb{E}^b[\bar{\mathbf{v}}_b]$. Now take $w_{bg} \geq h_b(w_{bb})$ as given and suppose $w_{bg}^b < h_b(w_{bb}^b)$. In the following two cases, we show this assumption always leads to contradictions. First, consider the case of $w_{bb}^b > w_{bb}$. We can obtain:

$$\begin{aligned} Q_b(\mathbf{w}_b^b, b) &\leq Q_b[(w_{bb}^b, h_b(w_{bb}^b)), b] \\ &< Q_b[(w_{bb}, h_b(w_{bb})), b] \leq Q_b(\mathbf{w}_b, b) \end{aligned}$$

The first and the last inequality is by the supermodularity of Q and the assumptions. The second inequality is by Lemma 3.8 of SA and $w_{bb}^b > w_{bb}$. However, [FOCw_{bb}] at state $(\mathbf{w}_b, b)$ implies $Q_b(\mathbf{w}_b^b, b) = Q_b(\mathbf{w}_b, b)$ because $\alpha(\mathbf{w}_b, b) = 0$, which forms a contradiction with above inequality.

Second, consider the case of $w_{bb}^b \leq w_{bb}$. By the assumptions and by the monotonicity of $h_b(\cdot)$, we know $w_{bg}^b < h_b(w_{bb}^b) \leq h_b(w_{bb}) \leq w_{bg}$. Then we can obtain:

$$Q_g[(0, w_{bg}), b] < Q_g[(0, w_{bg}^b), b] \leq Q_g(\mathbf{w}_b^b, b) \qquad \text{[F.14]}$$

The first inequality is implied by the fact $Q(\mathbf{v}, b)$ is strictly concave in v_g on the set $[(0, w_{bg}^b), (0, w_{bg})]$. The second inequality is by the supermodularity of Q. Moreover, Lemma 3.7 of SA implies $Q_g[(0, w_{bg}), b] = Q_g(\mathbf{w}_b, b)$. Combining this with [F.14], we find that $Q_g(\mathbf{w}_b, b) < Q_g(\mathbf{w}_b^b, b)$. However, from the first order conditions we can derive

$$\begin{aligned} Q_g(\mathbf{w}_b^b, b) &= \frac{p_b}{1 - p_b} Q_b(\mathbf{w}_b, b) \\ &= \frac{p_b}{1 - p_b} Q_b(\mathbf{v}, s) + \frac{p_b(1 - p_g)}{(1 - p_b)(1 - p_s)} \alpha(\mathbf{v}, s) \\ &\leq \frac{p_b}{1 - p_b} Q_b(\mathbf{v}, s) + \frac{p_g}{(1 - p_s)} \alpha(\mathbf{v}, s) = Q_g(\mathbf{w}_b, b) \end{aligned} \qquad \text{[F.15]}$$

The first equality is from [FOCw_{bg}] at state $(\mathbf{w}_b, b)$. The second and the last equalities are from [FOCw_{bb}] and [FOCw_{bg}] at state $(\mathbf{v}, s)$ respectively. This forms a contradiction. □

Corollary F.8. Suppose $w_{bg} \geq h_b(w_{bb})$ in Lemma F.7. Then $w_{bb}^b = w_{bb}$ and $w_{bg}^b \geq w_{bg}$.

Proof. Suppose $w_{bb}^b \neq w_{bb}$. To ease notation, we denote $\hat{\mathbf{w}}_b = (w_{bb}, h_b(w_{bb}))$ and $\hat{\mathbf{w}}_b^b = (w_{bb}^b, h_b(w_{bb}^b))$.

Because $w_{bg} \geq h_b(w_{bb})$ by assumption and $w_{bb}^b \geq h_b(w_{bb}^b)$ by Lemma F.7, we know from Lemma 3.7 that $Q_b(\mathbf{w}_b, b) = Q_b(\hat{\mathbf{w}}_b, b)$ and $Q_b(\mathbf{w}_b^b, b) = Q_b(\hat{\mathbf{w}}_b^b, b)$. Since $w_{bb}^b \neq w_{bb}$, Lemma 3.8 of SA then implies $Q_b(\hat{\mathbf{w}}_b^b, b) \neq Q_b(\hat{\mathbf{w}}_b, b)$, which further means $Q_b(\mathbf{w}_b, b) \neq Q_b(\mathbf{w}_b^b, b)$. This forms a contradiction with [FOCw_{bb}] at state $(\mathbf{w}_b, b)$ which implies $Q_b(\mathbf{w}_b^b, b) = Q_b(\mathbf{w}_b, b)$ because $\alpha(\mathbf{w}_b, b) = 0$. Thus, $w_{bb}^b = w_{bb}$.

Now suppose $w_{bg}^b < w_{bg}$. The strict concavity of $Q(\mathbf{v}, s)$ in v_g on the set $[\mathbf{w}_b^b, \mathbf{w}_b]$ implies $Q_g(\mathbf{w}_b^b, b) > Q_g(\mathbf{w}_b, b)$ which violates [F.15]. Thus, $w_{bg}^b \geq w_{bg}$. $\square$

Proof of Proposition 5.1. (a) Take any $\mathbf{v} \in V$. To ease notation, let $k = k(\mathbf{v}, g)$,$\mathbf{w}_i = \mathbf{w}_i(\mathbf{v}, g)$, and denote $\hat{k} = k(\mathbf{v}, b)$, $\hat{\mathbf{w}}_i = \mathbf{w}_i(\mathbf{v}, b)$. We show here that $k \geq \hat{k}$ and show in Lemma 2.4 of SA that k decreases in v_b, and strictly increases in v_g.

Suppose $\hat{k} > k$. Optimality at $(\mathbf{v}, g)$ implies

$$\begin{aligned} &-k + p_g[R(k) + \delta Q(\mathbf{w}_g, g)] + (1-p_g)\delta Q(\mathbf{w}_b, b) \\ &\geq -\hat{k} + p_g[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1-p_g)\delta Q(\hat{\mathbf{w}}_b, b) \end{aligned} \tag{F.16}$$

From part (a) of Lemma C.2, $Q(\mathbf{w}_g, g) = Q(\hat{\mathbf{w}}_g, g)$. Then [F.16] implies $Q(\mathbf{w}_b, b) > Q(\hat{\mathbf{w}}_b, b)$. Moreover, optimality at $(\mathbf{v}, b)$ implies

$$\begin{aligned} &-\hat{k} + p_b[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1-p_b)\delta Q(\hat{\mathbf{w}}_b, b) \\ &\geq -k + p_b[R(k) + \delta Q(\mathbf{w}_g, g)] + (1-p_b)\delta Q(\mathbf{w}_b, b) \end{aligned} \tag{F.17}$$

Add [F.16] [F.17] and rearrange to get:

$$0 > \delta[Q(\hat{\mathbf{w}}_b, b) - Q(\mathbf{w}_b, b)] \geq R(\hat{k}) - R(k) > 0$$

which forms a contradiction.

(b) Let $\mathbf{w}_g = \mathbf{w}_g(\mathbf{v}, s)$ and $\mathbf{w}_g^g = \mathbf{w}_g[\mathbf{w}_g(\mathbf{v}, s), g]$. From Lemma E.1, we know $\mathbf{w}_g \notin E_g$ implies $\mathbf{w}_g \in H$. So it must be that $Q_g(\mathbf{w}_g, g) > 0$ and $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) > 0$. First, let us consider the case $\mathbf{w}_g^g \in E_g$. Then $Q_g(\mathbf{w}_g^g, g)$ which is the left hand of [FOCw_{gg}] at state $(\mathbf{w}_g, g)$ is zero. So from the right hand side of [FOCw_{gg}] we get $\alpha(\mathbf{w}_g, g) = Q_g(\mathbf{w}_g, g) > 0$. Second, consider $\mathbf{w}_g^g \notin E_g$. Then Lemma D.2 shows that $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) > \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g^g, g)$. Hence by Lemma 3.2 (a) of SA we know $\alpha(\mathbf{w}_g, g) > 0$. Therefore we always have $k(\mathbf{w}_g, g) < \bar{k}_g$ by [FOCk^*].

(c) Consider the case where $\mathbf{p} \in B_\ell$. Take $(\mathbf{v}, s) \in H$ and let $\mathbf{w}_b = \mathbf{w}_b(\mathbf{v}, s)$, $\mathbf{w}_b^b = \mathbf{w}_b[\mathbf{w}_b(\mathbf{v}, s), b]$. Suppose $w_{bg} \geq h_s(w_{bb})$. Then from [PK$_b$] and [IC*] at state $(\mathbf{w}_b, b)$ and from Corollary F.8, we get

$$\begin{aligned} w_{bb} &\geq \delta\, \mathbb{E}^b[\mathbf{w}_b^b] \geq \delta\, \mathbb{E}^b[\mathbf{w}_b] \\ w_{bg} - w_{bb} &\geq R(\bar{k}_b) + \delta\Delta(w_{bg}^b - w_{bb}^b) \geq R(\bar{k}_b) + \delta\Delta(w_{bg} - w_{bb}) \end{aligned}$$

which together imply $w_{bb}(\mathbf{v}, s) \geq \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)} = \bar{v}_{bb}$. The last equality is from Lemma F.6 (a). However, this is a contradiction with $(\mathbf{w}_b, b) \in H$ which requires $w_{bb} < \bar{v}_{bb}$. Hence, $k(\mathbf{w}_b, b) < \bar{k}_b$ □

Lemma F.9. If $\Delta > \psi(p_g)$, then [IC] holds as strict inequality at state $(\bar{\mathbf{v}}_b, b)$.

Proof. Suppose not. Then from [F.5] and [F.12], we know

$$R(\bar{k}_b) = \bar{m}_{bg} = \bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$$

Moreover, [F.6] implies

$$\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) = R(\bar{k}_b)$$

which contradicts Lemma F.2 (b). □

Lemma F.10. If $\Delta > \psi(p_g)$, then there exists a neighborhood in the set H where the contract can evolve to after a bad shock and where investment is temporarily efficient.

Proof. Lemma F.9 shows that the constraint [IC] holds as strict inequality. From Lemma 3.4 (b) of the Supplementary Appendix, we know $(\bar{\mathbf{v}}_s, s) \in \mathrm{cl}(H)$, which means we can find a sequence $\{(\mathbf{v}^{(n)}, g)\}_{n=1}^{\infty}$ in the equilibrium region that converges to $(\bar{\mathbf{v}}_g, g)$. Since $\mathbf{w}_b(\bar{\mathbf{v}}_g, g) = \bar{\mathbf{v}}_b$ and $\mathbf{w}_b(\cdot, g)$ is continuous, there exists N such that [IC] holds as strict inequality at state $[\mathbf{w}_b(\mathbf{v}^{(n)}, g), b]$ if $n \geq N$. This means $k[\mathbf{w}_b(\mathbf{v}^{(n)}, g)] = \bar{k}_b$ if $n \geq N$. In other words, there must exit a neighborhood $F \subset \{(\mathbf{v}, b) : \mathbf{v} \in V, \bar{\mathbf{v}}_b - (\varepsilon, \varepsilon) < \mathbf{v} < \bar{\mathbf{v}}_b\} \cap H$ for some sufficient small $\varepsilon > 0$ such that $k(\mathbf{v}, b) = \bar{k}_b$ if $(\mathbf{v}, b) \in F$. Moreover, contingent utility vector can possibly evolve to this neighborhood F after a bad shock. □

G. Proofs from Section 6

Before proceeding to the proofs, we formally define all the securities used in the implementation as follows.

Compensating balance: the required cash deposit when the credit line is established. This deposit pays a fixed interest to the firm each period.

Long-term debt: a consol bond that pays floating coupon each period contingent on the history of covenant violations.

Credit line: revolving credits provided to the firm with the covenant. The limits of the credit line can be adjusted each period by the investors based on the history of covenant violations.

Stock options: the right to buy the firm's stock at the specified strike price with maturity of one period. These options are settled by cash, meaning that no new share is actually issued.

Equity: the difference between the firm value (investment NPV plus compensating balance) and the sum of firm debts and stock options. The equity holders receive dividend payments made by the firm according to their share holdings.

Proof of Proposition 6.1. Let $(\mathbf{v}, s)$ be a state induced by the optimal contract. We can simplify [6.1] for $i = b, g$ to get

$$\text{[G.1]} \qquad \begin{aligned} C_g - C_b &= v_g - v_b - R(k(\mathbf{v}, s)) \\ &\geq \delta\Delta[w_{bg}(\mathbf{v}, s) - w_{bb}(\mathbf{v}, s)] \end{aligned}$$

The inequality is by [IC*]. Hence, $C_b \leq C_g$. Because $w_{bg}(\mathbf{v}, s) > w_{bb}(\mathbf{v}, s)$ in the optimal contract, positive persistence implies $C_b < C_g$ by [G.1]. In the iid case, since [IC*] holds as equality, $C_b = C_g$.

When the firm is unconstrained, we can plug the explicit values of $\bar{v}_{sb}, \bar{v}_{sg}$ (from Lemma F.6) into [G.1] to obtain the gap of credit limit as follows. If persistence is low ($\mathbf{p} \in B_l$), we have $C_g - C_b = \frac{\delta\Delta R(\bar{k}_b)}{1-\delta\Delta}$ which does not vary with s. If persistent is high ($\mathbf{p} \in B_h$), this gap varies with s. For $s = b$, we have $C_g - C_b = \frac{\delta p_g R(\bar{k}_g)}{(1+\delta p_g)(1-\delta\Delta)} - R(\bar{k}_b)$. For $s = g$, we have $C_g - C_b = \frac{\delta^2\Delta p_g R(\bar{k}_g)}{(1+\delta p_g)(1-\delta\Delta)}$. So by fixing p_b and raising p_g, we can see that the credit limit gap increases for all the cases. □

Lemma G.1. When the firm is unconstrained, its stock prices are:

$$[\bar{z}_{sb}] \qquad \bar{z}_{bb} = \bar{z}_{gb} = \frac{p_b\delta}{(1-\delta) + \delta p_b}\bar{z}_{bg}$$

$$[\bar{z}_{bg}] \qquad \bar{z}_{bg} = \frac{R(\bar{k}_b) + p_g\delta[R(\bar{k}_g) - R(\bar{k}_b)]}{1 - p_g\delta - \frac{(1-p_g)p_b\delta^2}{1-\delta(1-p_b)}}$$

$$[\bar{z}_{gg}] \qquad \bar{z}_{gg} = \bar{z}_{bg} + R(\bar{k}_g) - R(\bar{k}_b)$$

Proof. According to our definition of stock price, we have $z_{si} = d_i + \delta\,\mathbb{E}^i[z_i']$, where the payout d_i is a function of the state (C_g, M, s), and $\mathbf{z}_i' = (z_{ib}', z_{ig}')$ is the next period's stock prices. In the unconstrained stage, the firm issues all cash flows as payouts. Denote the firm's stock prices in this stage as $\bar{\mathbf{z}}_s = (\bar{z}_{sb}, \bar{z}_{sg})$. Then we have

$$\bar{z}_{sb} = \delta\,\mathbb{E}^b[\bar{\mathbf{z}}_b], \quad \bar{z}_{sg} = R(\bar{k}_s) + \delta\,\mathbb{E}^g[\bar{\mathbf{z}}_g]$$

Obviously, $\bar{z}_{bb} = \bar{z}_{gb}$. Moreover, we can solve the three values $\bar{z}_{sb}, \bar{z}_{bg}, \bar{z}_{gg}$ jointly from the above equations. □

Proof of Proposition 6.2. First, let us consider the agent's payoff from equity and stock options and show in three possible cases that the combined payoff is equivalent to his compensation in the optimal contract, that is,

$$\lambda d_i + (1-\lambda)\max\{z_{si} - K, 0\} = e_{si} \tag{G.2}$$

Note that the specified strike price in the proposition satisfies $\bar{z}_{sb} \le \bar{z}_{bg} \le K \le \bar{z}_{gg}$. The first inequality is by [$\bar{z}_{sb}$] and the second is because $\bar{m}_{gg} \ge \bar{m}_{bg}$. The last inequality is because by [$\bar{z}_{gg}$] we have

$$\begin{aligned} \bar{z}_{gg} - K &= \bar{z}_{gg} - \bar{z}_{bg} - (\bar{m}_{gg} - \bar{m}_{bg}) \\ &= R(\bar{k}_g) - R(\bar{k}_b) - (\bar{m}_{gg} - \bar{m}_{bg}) \\ &= \bar{e}_{gg} - \bar{e}_{bg} \ge 0 \end{aligned} \tag{G.3}$$

Also note that the payout of the constrained firm is always zero and [G.2] simply holds. In the unconstrained firm, the payout is equal to any possible output, and there are three cases.

(1) Consider a bad shock occurs today ($i = b$). Then payout $d_i = 0$ and both securities have zero payoff. So [G.2] simply holds.

(2) Consider a good shock occurs today ($i = g$) and a bad shock occurs last period ($s = b$). Since $\bar{z}_{bg} \le K$ and $d_i = R(\bar{k}_b)$, the agent's option payoff is zero and equity payoff is $\lambda R(\bar{k}_b) = \bar{e}_{bg}$, according to the specified λ in the Proposition. So the total security payoff is equal to the right hand side of [G.2].

(3) Consider a good shock occurs both this and last period ($i = g, s = g$). We have $K \le \bar{z}_{gg}$ and $d_i = R(\bar{k}_g)$. There are two subcases. When $\mathbf{p} \in B_h$, we know $\bar{e}_{bg} = 0$ from Theorem 3 (d), which further implies $\lambda = 0$. So the left hand side of [G.2] simply becomes $\bar{z}_{gg} - K = \bar{e}_{gg}$, which is exactly its right hand side. When $\mathbf{p} \in B_\ell$, we have $\bar{m}_{bg} = \bar{m}_{gg}$ from the proof of Theorem 3 (c). Then the left hand side of [G.2] becomes

$$\begin{aligned} \lambda R(\bar{k}_g) + (1-\lambda)[R(\bar{k}_g) - R(\bar{k}_b)] =& R(\bar{k}_g) - R(\bar{k}_b) + \lambda R(\bar{k}_b) \\ =& R(\bar{k}_g) - R(\bar{k}_b) + \bar{e}_{bg} = \bar{e}_{gg} \end{aligned}$$

which is the same as its right hand side.

Second, let us consider how λ, equity payoff, and the fraction of option payoff in total compensation vary with persistence if we fix p_b and raise p_g, or if we fix p_g and decrease p_b. We consider two possible cases $\mathbf{p} \in B_h$ and $\mathbf{p} \in B_l$. Since λ is zero if $\mathbf{p} \in B_h$, and is equal to $1 - \delta p_g \left[\frac{R(\bar{k}_g)}{R(\bar{k}_b)} - 1\right]$ if $\mathbf{p} \in B_l$, it's easy to see that $\lambda < 1$ as long as $\Delta > 0$, and that λ decreases in Δ and strictly so in the case of $\mathbf{p} \in B_l$. Moreover, the agent's equity payoff is $[R(\bar{k}_i) - \max\{\bar{z}_{si} - K, 0\}]\lambda = \bar{e}_{bg}$ in all possible cases, which decreases in Δ, and strictly so if $\mathbf{p} \in B_l$. Lastly, the entire compensation is from stock options if $\mathbf{p} \in B_h$. In the case of $\mathbf{p} \in B_l$, the option payoff, either $R(\bar{k}_g) - R(\bar{k}_b)$ or 0, increases in Δ. So the option payoff always accounts for a higher fraction of the agent's compensation as Δ increases. $\square$

Proof of Lemma 6.3. Given state (C_g, M, s) and the specified coupon payment c_i in [6.4], we can obtain from [6.2] next period's credit balance contingent on firm violating covenant or not today is $M_i = \bar{v}_b - w_{ib}$. So if the firm violates covenant tomorrow, the available credit (contingent on firm violating covenant or not today) is $\bar{v}_b - M_i = w_{ib}$. Moreover, if the firm complies with covenant tomorrow, by [6.3] the available credit (contingent on firm violating covenant or not today) is $C'_{ig} - M_i + R(k_i) = w_{ig}$. Hence, under the designed mechanism, the firm's available credit always matches the agent's contingent utility obtained in the contract. $\square$

Proof of Theorem 4. Take any (C_g, M, s). Then [6.1] transforms the state to the corresponding $(\mathbf{v}, s)$ in the contract. First, note that the agent will never draw down credit to issue dividends, because he can obtain the same or higher utility by directly diverting from the available credit.

Second, the agent has no incentive to misreport cash flow. Suppose good shock occurs and the agent lies. He can divert the cash flow $R(k(\mathbf{v}, s))$. If the agent also diverts the available credit, his total payoff is

$$R(k) + \bar{v}_b - M = R(k) + v_b \leq v_g \tag{G.4}$$

The equality is by the definition of M and the inequality is by [IC*]. If the agent does not divert the available credits, by Lemma 6.3 his contingent utilities from next period onwards is $\mathbf{w}_b(\mathbf{v}, s)$, which means his total payoff today is $R(k) + \delta\, \mathbb{E}^g[\mathbf{w}_b] \leq v_g$ by [IC].

Third, the agent has no incentive to divert credit. By doing so, the agent gets payoffs of either $\bar{v}_b - M$ or $C_g - M + R(k)$ depending on firm violating covenant or not. According to [6.1], these values are equal to v_b or v_g. But the agent obtains the same contingent payoffs by waiting for the firm's payouts until the credit balance is fully repaid. This is because by Proposition 6.2, v_i is equal to all the future payments that the agent gets from his security holdings. $\square$

References

Abreu, Dilip, David Pearce and Ennio Stacchetti. 1990. 'Toward a theory of discounted repeated games with imperfect monitoring'. *Econometrica* 58 (5): 1041–1063. (Cit. on pp. 12, 28).

Albuquerque, Rui, and Hugo A. Hopenhayn. 2004. 'Optimal lending contracts and firm dynamics'. *Rev. Econom. Stud.* 71 (2): 285–315. (Cit. on p. 2).

Battaglini, Marco. 2005. ‘Long-Term Contracting with Markovian Consumers’. *American Economic Review* 3:637–658. (Cit. on p. 7).

Beck, Thorsten, AslI Demirgüç-Kunt and Vojislav Maksimovic. 2005. ‘Financial and Legal Constraints to Growth: Does Firm Size Matter?’ *The Journal of Finance* 60 (1): 137–177. (Cit. on p. 2).

Bergman, Nittai K., and Dirk Jenter. 2007. ‘Employee sentiment and stock option compensation’. *Journal of Financial Economics* 84:667–712. (Cit. on p. 4).

Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet. 2007. ‘Dynamic security design: convergence to continuous time and asset pricing implications’. *Review of Economic Studies* 74 (2): 345–390. (Cit. on p. 6).

Bloedel, Alexander W., and R. Vijay Krishna. 2014. ‘Misery, Persistence, and Growth’. *working paper*. (Cit. on p. 8).

Bolton, Patrick, and David S. Scharfstein. 1990. ‘A Theory of Predation Based on Agency Problems in Financial Contracting’. *American Economic Review* 80:93–106. (Cit. on p. 6).

Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. ‘A Theory of Financing Constraints and Firm Dynamics’. *Quarterly Journal of Economics* 121:229–265. (Cit. on pp. 2, 4, 6, 9, 15).

DeMarzo, Peter M., and Michael J. Fishman. 2007a. ‘Agency and Optimal Investment Dynamics’. *Review of Financial Studies* 20:151–188. (Cit. on p. 6).

– . 2007b. ‘Optimal Long-Term Financial Contracting’. *Review of Financial Studies* 20:2079–2128. (Cit. on pp. 2, 4, 21, 23).

DeMarzo, Peter, and Yuliy Sannikov. 2006. ‘Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model’. *Journal of Finance* 61:2681–2724. (Cit. on p. 4).

Demiroglu, Cem, and Christopher James. 2011. ‘The use of bank lines of credit in corporate liquidity management: A review of empirical evidence’. *Journal of Banking and Finance* 35:775–782. (Cit. on p. 5).

Doepke, Matthias, and Robert M. Townsend. 2006. ‘Dynamic mechanism design with hidden income and hidden actions’. *Journal of Economic Theory* 126:235–285. (Cit. on p. 7).

Eső, Péter, and Balázs Szentes. 2013. ‘Dynamic Contracting: An Irrelevance Result’. *working paper*. (Cit. on p. 7).

Fernandes, Ana, and Christopher Phelan. 2000. 'A Recursive Formulation for Repeated Agency with History Dependence'. *Journal of Economic Theory* 91:223–247. (Cit. on pp. 7, 10, 12).

Gertler, Mark, and Simon Gilchrist. 1994. 'Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms'. *Quarterly Journal of Economics* 109:309–340. (Cit. on p. 2).

Gomes, Joao F. 2001. 'Financing Investment'. *American Economic Review* 91 (5): 1263–1285. (Cit. on p. 2).

Grillo, Edoardo, and Juan Ortner. 2018. *Dynamic Contracting with Limited Liability Constraints*. Tech. rep. Boston University. (Cit. on p. 7).

Guo, Yingni, and Johannes Hörner. 2014. 'Dynamic Mechanisms without Money'. *Working Paper*. (Cit. on p. 8).

Halac, Marina, and Pierre Yared. 2014. 'Fiscal rules and discretion under persistent shocks'. *Econometrica* 82 (5): 1557–1614. issn: 0012-9682. (Cit. on p. 8).

Hurst, Erik, and Benjamin W. Pugsley. 2011. 'What Do Small Business Do?' *NBER Working Paper*. (Cit. on pp. 5, 18).

Jensen, Michael C., and William H. Meckling. 1976. 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure'. *Journal of Financial Economics* 3 (4): 305–360. (Cit. on p. 2).

Kapička, Marek. 2013. 'Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach'. *Review of Economic Studies* 80 (3): 1027–1054. (Cit. on p. 7).

Kaplan, Steven N., and Luigi Zingales. 1997. 'Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?' *The Quarterly Journal of Economics* 112:169–215. (Cit. on p. 6).

Krasikov, Ilia, and Rohit Lamba. 2018. *A Theory of Dynamic Contracting with Financial Constraints*. Tech. rep. Pennsylvania State University. (Cit. on p. 7).

Larcker, David F. 2008. 'Executive Compensation and Incentives'. *Stanford Graduate School of Business*. (Cit. on p. 4).

Pavan, Alessandro, Ilya Segal and Juuso Toikka. 2014. 'Dynamic Mechanism Design: A Myersonian Approach'. *Econometrica* 82 (2): 601–653. (Cit. on p. 7).

Quadrini, Vincenzo. 2004. 'Investment and Liquidation in Renegotiation-proof Contracts with Moral Hazard'. *Journal of Monetary Economics* 51:713–751. (Cit. on p. 6).

Rampini, Adriano A., and S. Viswanathan. 2013. ‘Collateral and Capital Structure’. *Journal of Financial Economics* 109:466–429. (Cit. on p. 2).

Shiryaev, Albert N. 1995. *Probability*. Verlag, New York: Springer. (Cit. on p. 38).

Stokey, Nancy L., Robert E. Lucas Jr and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press. (Cit. on p. 34).

Sufi, Amir. 2009. ‘Bank Lines of Credit in Corporate Finance: An Empirical Analysis’. *The Review of Financial Studies* 22:1057–1088. (Cit. on p. 5).

Tchistyi, Alexei. 2014. ‘Security Design with Correlated Hidden Cash Flows: The Optimal of Performance Pricing’. *working paper*. (Cit. on p. 7).

Thomas, Jonathan P., and Tim Worrall. 1990. ‘Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem’. *Journal of Economic Theory* 51:367–390. (Cit. on p. 6).

Zhang, Yuzhe. 2009. ‘Dynamic Contracting with Persistent Shocks’. *Journal of Economic Theory* 144:635–675. (Cit. on p. 7).