Dynamic Financial Contracting with Persistent Private Information *

Shiming Fu† R. Vijay Krishna‡

November, 2016

Abstract

This paper studies a dynamic agency model in which the agent privately observes the firm's cash flows that are subject to persistent shocks. Because bad performance distorts investors' belief downward, the agent has less incentive to misreport and is compensated less than what she can divert. We show that promising agent utilities contingent on performance today and tomorrow is effective in providing incentive and characterizing the problem recursively. The optimal contract can be implemented by equity, stock options, credit line with contingent limits, and long-term debt. 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; (iii) the firm experiences longer time of being financially constrained. 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 more in line with empirical evidences.

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. The most recent version of the paper is available here.

(†) Simon School of Business, University of Rochester <shiming.fu@simon.rochester.edu>

(‡) 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 and have explained much about how firm policies designed and evolve in dynamic environments. However, they 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.

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. This suggests that private information about firm behavior may well be persistent.

In this paper, we analyze a model where private information about the firm's profitability is persistent. Our main finding is that persistent private information affects incentives via an additional channel that is not present in the iid case. To see this, consider the effect of bad performance today. This distorts downwards the principal's belief about future prospects, so that it is optimal for the principal to scale down future investment, leading to smaller future information rent. Accordingly, the agent has less incentive to misrepresent firm performance, which alleviates the agency problem for the principal. We show this new mechanism has distinct implications for firm compensation, capital structure, investment, and growth dynamics, and can reconcile stylized facts that are inconsistent with iid models.

(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).

Our firm consists of a risk-neutral agent and risk-neutral investors. The agent has the expertise to operate the firm but does not have funds. The investors provide funds to launch the firm and finance the firm's risky investment 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.

When private information is iid, it is well known that promising the agent utilities contingent on performance today is sufficient to induce truth telling. However, this approach is no longer sufficient in an environment with persistent private information, because after a misreport, investors and the agent will have different beliefs about the probabilities of success tomorrow, a fact that the agent can potentially use to her advantage. Nevertheless, we show that it suffices to promise 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 provide the correct incentives and allows us to provide a tractable recursive formulation of the firm's problem.

By characterizing the short-run as well as the long-run properties of the optimal contract, we show that its qualitative and quantitative features are sensitive to the level of persistence. The young (constrained) firm does not issue payout and may need to experience consecutive good shocks to become financially unconstrained. So the young firm can stay small for longer time than in the iid case. In the long run, the firm becomes mature (unconstrained) in finite time and starts to pay out. But investment can nonetheless vary with revenue shocks. The agent is paid less than what she can divert, and if persistence is high, she only gets paid when consecutive good performances are observed.

We also provide an implementation of the optimal contract in terms of financial securities including equity, stock options, credit line, and long-term debt. The agent holds equity and stock options. The firm is financed by a credit line with limit contingent on compliance with a cash flow covenant. As persistence increases, the agent holds less equity stake (investors hold more) and is compensated more by option payoffs. The credit line limit varies more with performance as persistence goes up.

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 of the problem. Section 4 describes the full information benchmark, the optimal contract, as well as its long-run behavior. It also describes qualitative properties that are independent of the degree of persistence. Sections 5 and 6 describe the optimal contract for the financially constrained and unconstrained firm, respectively, and emphasizes how persistence affects the qualitative and quantitative properties of the contract. Section 7 describes our implementation of the optimal contract while Section 8 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 shows that stock options plan is the most common method for employee compensation below the executive rank.

In dynamic settings, the variation of future information rents provide incentives for the agent to report cash flow truthfully. In the iid case, expected information rent becomes constant when the firm is unconstrained. The only way to provide incentives in the long-run is to pay the agent exactly the amount of cash she can divert. Thus, compensation is linear in firm performance.[4] With persistent information, investors' belief about future prospects is downgraded after bad performance even in a financially unconstrained firm, leading to smaller future information rents. Hence the agent has less incentive to misreport and is paid strictly less than what she can divert.

Indeed, if information is highly persistent, Theorem 3 shows that the agent gets paid only when consecutive good performances are observed. Otherwise, investors retain all the firm's revenue. Thus, the pay-performance relation is strictly convex in our model. This convex compensation scheme can be implemented (see Theorem 4) by

(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 $\lambda < 1$.

granting the agent equity and stock options. As information becomes more persistent, the agent holds a smaller equity stake, and 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), drawdowns 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.

Our implementation of the optimal contract (in Theorem 4) highlights a key feature of our model: The firm is financed by a credit line with limit contingent on maintaining a cash flow covenant. Persistent information implies that the variations in continuation utility has to be strictly greater than the current cash flow to incorporate the dynamic component of the information rent. Due to the enlarged pay-performance sensitivity (which disappears in the iid case), the firm's credit limit has to be adjusted according to performance history in order to track the contract's evolution. On the contrary, iid models are usually implemented by a credit line with a fixed limit. Moreover, in our implementation, if the firm 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 in size for a long time and therefore are 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 constrained. Indeed, numerical experiments (see Figure 5) show that the amount of time that the firm remains constrained increases as we move farther away from the iid case, and that the firm therefore remains at an inefficient size for a longer time.

Investment. Kaplan and Zingales (1997) show that less financially constrained firms exhibit greater investment to cash flow sensitivity. As we show in Proposition 5.3, besides the financial slack of the firm, investment is also determined by investors'

(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.

belief about the likelihood of 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, which is consistent with the findings described above. 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.

If shocks are persistent, the constrained firm can temporarily see investment at an efficient (first best) level before it becomes unconstrained, while in the iid case, investment is always distorted downward before the firm is unconstrained. We show that this temporary first best investment corresponds exactly to the incentive constraint being slack, which is only possible if persistence if sufficiently high.

Consistent with the implications from the iid models, our model also predicts (see Proposition 5.2) that payouts are delayed until the firm is unconstrained, an implication that is robust for all levels of persistence. Delaying payments is optimal because it effectively relaxes the limited liability constraint. In other words, delaying compensation eases the provision of incentive and therefore helps improve future investment efficiency.

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]

(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

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 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 (2013) studies an environment similar as ours except that his model has finite periods and no investment. Although Tchistyi (2013) 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

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.

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.

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, both the principal and agent are risk neutral, and both discount the future at a common rate of $\delta \in (0, 1)$.

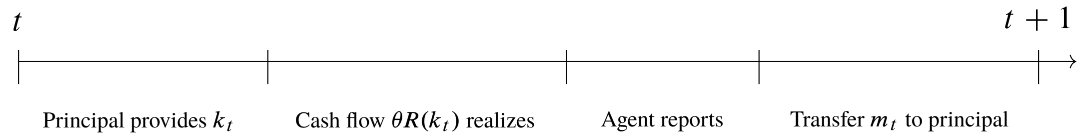
The project, which we shall also refer to as the firm, requires an investment $k_t \geq 0$ in period t. Capital depreciates completely, and so cannot be carried over to subsequent periods. The return on capital is random, and is $\theta(s)R(k)$, where $s \in S := \{b, g\}$ and $\theta \in \{0, 1\}^S$ has $\theta(b) = 0$ and $\theta(g) = 1$. 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 is given by the matrix

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

We shall assume that $\Delta := p_g - p_b \geq 0$, ie, the states are positively correlated. We shall refer to such a Markov process as being *persistent*. The case where $\Delta = 0$ corresponds

Figure 1: *Timing*



to the iid case.[8] 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.

The agency problem arises because (i) the principal cannot observe the output while the agent can, and (ii) the agent is cash constrained and so is protected by *limited liability*. These are the twin frictions of the model. If the agent were not cash constrained, the principal could simply sell him the firm. If the output were observable, the principal would pay the agent the 'minimum wage' of 0, ie, offer him just enough to stay with the firm, while she retains all the revenue. Thus (and this is true for any degree of persistence), it is the *combination* of limited liability constraints and privately observed cash flow that gives rise to a non-trivial contracting problem.

The cumulative information available to the principal 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 (conditional on positive output) in any period on all previous cash transfers by the agent, and all previous investments by the principal. 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.

The timing runs as follows: At the beginning of time, at $t = 0$, the principal offers the agent in infinite horizon contract that she may accept or reject. If she 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]

(8) However, while $\Delta > 0$ implies the process is persistent, the *levels* of p_b and p_g, and not just their difference, are also relevant for the evolution of investment etc in the optimal contract. A further discussion of degrees of persistence, especially as it pertains to contract structure, relies on more notation and terminology, and so is postponed to Section 2.1 which provides a categorisation of degrees of persistence.

(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

2.1. Degrees of Persistence

Recall that $\Delta := p_g - p_b > 0$ amounts to shocks being positively correlated. However, the same value of Δ can correspond to different combinations of p_b and p_g. Indeed, 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. It is easy to see from equations [4.1] and [4.2], for instance, that such a change would reduce firm surplus while leaving Δ intact.

Therefore, when we speak of levels of persistence, it is important to keep both Δ as well as p_g (or p_b) in mind. Towards that end, recall that $\bar{k}_g$ is the unique k that solves $p_g R'(k) = 1$, and is the efficient amount of investment following a good shock. Now define the functions $\hat{k}(p_g)$ and $\psi(p_g)$ as

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

We show in Appendix G that $\psi(p_g) \in [0, p_g]$ for all $p_g \in [0,1]$ and that it is well defined. With these definitions in hand, we can partition the space $\Pi := \{\mathbf{p} \in [0,1]^2 : p_b \leq p_g\}$ by defining $B_\ell := \{\mathbf{p} \in \Pi : \Delta < \psi(p_g)\}$ and $B_h := \{\mathbf{p} \in \Pi : \Delta \geq \psi(p_g)\}$.

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 (which corresponds to positive and zero output respectively), 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 recursive contracts without loss of generality (in that there is no loss of social surplus or utility to either principal or agent). These recursive contracts are described next.

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).

3.1. Recursive Contracts

Before we describe the recursive formulation of the principal's problem, it is useful to reconsider the recursive formulation in the special case where private information is iid over time.[10] As is well known, in this case it suffices to index contracts by *promised utility* v, which is the agent's expected lifetime utility from the contract. Upon entering a period with a specific level of promised utility v, the agent's report will result in a new level of promised utility w beginning the next period.

However, such an approach is inadequate in the Markovian case because the agent now has private information about her preferences over future streams of utility. (Recall that today's shock dictates the probability distribution over tomorrow's shocks, and today's shock is only known to the agent.) If the agent lies in a period and enters the period with promised utility v from the principal's point of view, the principal and agent have differing beliefs about the probability of shocks, and hence, about the amount of utility (in the form of information rent) that will accrue to the agent. Thus, the agent will assess a different expected utility from the stream of future cash flows than v, which may violate incentive compatibility and promise keeping. Notice that this is not an issue in the iid case.

In our Markovian setting, the principal instead indexes contracts by a pair of *interim* or *contingent* utilities and the previous period's reported shock, which determines the principal's beliefs about the current period's shock. The contingent utilities are represented by a vector $\mathbf{v} = (v_b, v_g)$ with the interpretation that if the agent reports a bad shock in the current period, she will get v_b utiles, but receives v_g utiles if she reports a good shock.

To see why it suffices to use contingent utilities, suppose the agent enters the period with promised contingent utilities $\mathbf{v} = (v_b, v_g)$. Then, promise keeping constraints and the incentive constraint [IC] only depends 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), her lifetime interim utility is still v_g, and truth-telling 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 contingent utility streams. In other words, contingent promised utilities are common knowledge after *every* history, and can therefore serve as state variables for our recursive formulation.[11] Thus, our formulation ensures that the agent cannot benefit

(10) Because this approach is now well understood, our description will be informal.

(11) 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 she has lied in the last period. Notice that both the promised and threat-point utilities

from multiple deviations and, moreover, truth telling is optimal after every history.

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 $k(\mathbf{v}, s) \in \mathbb{R}$, transfers $m_b(\mathbf{v}, s), m_g(\mathbf{v}, s) \in \mathbb{R}$, 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

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

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

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, 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.[12] 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

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

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 degree of persistence. Thus,

---

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, and so render themselves more suitable for the application considered in this paper.

(12) We assume throughout that there is *no hidden borrowing*, so the agent cannot pretend to have high output when it is actually low. Similarly, we assume that the agent cannot save output or cash transfers to him, so that there is *no hidden saving*.

$\Delta(w_{bg} - w_{bb})$ is the *Markovian* (or dynamic) information rent. Clearly, the constraint [IC*] crystallizes the effect of Markovian shocks. As we shall see below (in Theorem 1), we must necessarily have $w_{bg} \geq w_{bb}$, which implies $\Delta(w_{bg} - w_{bb}) \geq 0$, so that with persistence, the incentive constraint is tighter. To see this more starkly, suppose $\mathbf{w}_b$ is feasible for a low value of Δ. Then, keeping k fixed, an increase in Δ renders $\mathbf{w}_b$ infeasible insofar as it leads to a violation of [IC*].

It is easy to see that 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. Having described the state variables and constraints for our recursive formulation, 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.

3.3. Recursive Domain

Given that cash flows for the agent are always non-negative, it is clear that any vector of contingent utilities that can be realized must also be non-negative. But our other constraints impose additional restrictions on the feasible pairs of contingent utilities (v_b, v_g). 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.[13]

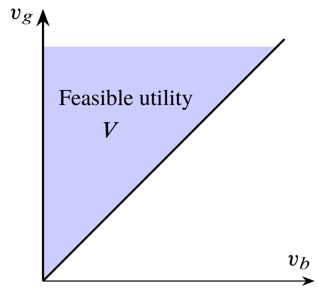
As noted above, because cash flows are non-negative, the only feasible choices of $\mathbf{w}_i$ must lie in $\mathbb{R}^2_+$. However, even with the restriction that $\mathbf{w}_i \in \mathbb{R}^2_+$, not every $\mathbf{v} \in \mathbb{R}^2_+$ is implementable. Indeed, we show in Lemma A.1 that no $\mathbf{v} = (v_b, 0)$ with $v_b > 0$ is implementable.

To serve as the domain for a recursive problem, the set of feasible utilities that can be implemented must have the property that the contingent continuation utilities 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, a largest), non-trivial set, as described next, with this property, and this set will serve as the domain for our recursive formulation of the principal's problem.[14]

(13) Strictly speaking, our notion of implementability should also include a transversality condition to ensure that promised utilities are actually delivered. For instance, to show formally that contingent utilities can never be negative because of limited liability constraints, one must use a transversality argument. However, because all the contractual variables considered in this paper will actually lie in a compact set, we eschew references to transversality conditions.

(14) Theorem 1 is an analogue of Lemma 2.2 in Fernandes and Phelan (2000), but without their compact-

Figure 2: *The Recursive Domain V*



Theorem 1. *There exists a largest set $V \subset \mathbb{R}^2_+$ 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, $V := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq v_b\}$ (see figure 2).*

A proof is in Appendix A. Notice that the domain V is independent of s, ie, continuation contingent utiles 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] $$\Gamma(\mathbf{v}, s) := \{(k, m_i, \mathbf{w}_i) : (k, m_i, \mathbf{w}_i) \text{ implements } \mathbf{v} \text{ and } \mathbf{w}_i \in V\}$$

that is, $\Gamma(\mathbf{v}, s)$ 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 $\Gamma(\mathbf{v}, s)$ is independent of s, we shall, when there is no cause for confusion, denote this set by $\Gamma(\mathbf{v})$.

A recursive contract $(k, m_i, \mathbf{w}_i)$ therefore conditions current investment and contingent transfers on the current state $(\mathbf{v}, s)$ and also determines the continuation state $(\mathbf{w}_i, i)$ for $i = b, g$ beginning in the next period. We now consider optimal contracts.

4. Optimal Contracts and the First Best

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 of the firm[15] when the previous period's shock was s, and when the agent enters the period with contingent utility $\mathbf{v} = (v_b, v_g)$. It is easy to see that $Q(\mathbf{v}, s)$ must satisfy the Bellman equation

[VF] $$Q(\mathbf{v}, s) = \max_{(k, m_i, \mathbf{w}_i)} \left[-k + p_s\big(R(k) + \delta Q(\mathbf{w}_g, g)\big) + (1 - p_s)\delta Q(\mathbf{w}_b, b)\right]$$

ness 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.

(15) Note that because both principal and agent are risk neutral, social surplus and hence firm surplus is invariant to monetary transfers.

subject to $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$, where $\Gamma(\mathbf{v}, s)$ is defined in [3.1] above. An *optimal contract* is the optimal policy $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v})$ in [VF] above.

In addition to the constrained problem in [VF], it is also useful to consider the *first best* problem, where there are no incentive constraints, ie, where the output is observed by the principal. Then, the efficient investment level $\bar{k}_s$ in each period is the unique k that solves $p_s R'(k) = 1$. Payment to the agent can be structured arbitrarily over time as long as it satisfies promise keeping. This implies that the efficient level of firm surplus in state s is $\bar{Q}(s)$, $s = b, g$, where

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

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

These two equations allow us to explicitly calculate $\bar{Q}(b)$ and $\bar{Q}(g)$. The function $\bar{Q}(s)$ represents an upper bound for the value of the firm in state s, and entails perpetual efficient investment and production.

4.1. Efficient Sets

Let $E_s := \{\mathbf{v} \in V : Q(\mathbf{v}, s) = \bar{Q}(s)\}$ denote the *efficient* or *unconstrained* sets of contingent utilities. Thus, E_s represents all levels of promised utilities where the constrained problem reaches its upper bound, and the firm is effectively unconstrained.

Investment in E_s is (perpetually) efficient because E_s comprises of all points $(\mathbf{v}, s)$ such that neither [IC] nor [LL] are *active*.[16]

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

(a) There exists $\bar{\mathbf{v}}_s \in E_s$ such that $\mathbf{v} \in E_s$ implies $\mathbf{v} \geq \bar{\mathbf{v}}_s$.

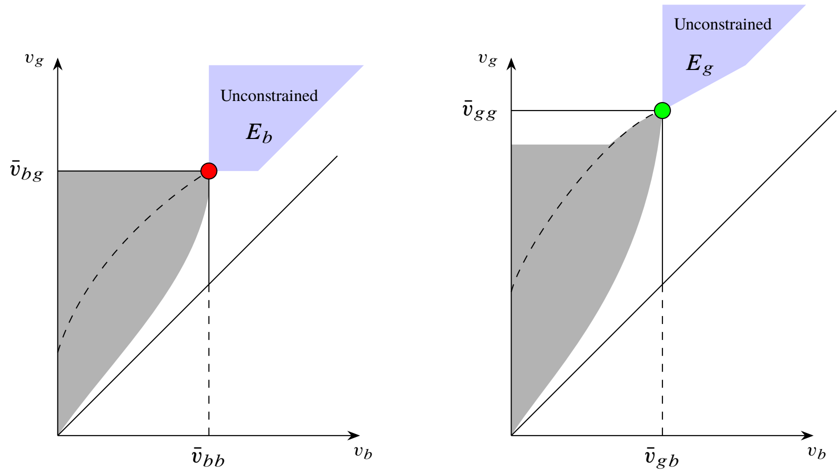
(b) For each $\mathbf{v} \in E_s$, $v_g - v_b \geq R(\bar{k}_s) + \delta\Delta \max\left[\frac{\delta \bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right]$.

(c) 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 $i = b, g$.

The proof is in Appendix B. Intuitively, $\bar{\mathbf{v}}_s$ represents the lowest levels of contingent utility that the agent must have in order to obtain *perpetual* efficient investment in state s. In other words, $\bar{\mathbf{v}}_s$ is the smallest level of contingent utility needed so that financing constraints no longer bind. Indeed, if $\mathbf{v} < \bar{\mathbf{v}}_s$, then $\mathbf{v} \notin E_s$. That output must be perpetually efficient follows from the fact that $Q|_{E_s} = \bar{Q}(s)$; if investment were ever

(16) A constraint *binds* if it holds with equality. It is *active* if removing the constraint results in a strict improvement, ie, if its Lagrange multiplier is positive.

Figure 3: *Graph of States*†



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

†The pictures plot regions in the domain with different properties. The pair of continuation utilities locate in the left (right) picture if bad (good) shock occurs last period. The light blue areas are states where the firm is financially unconstrained. The gray shaded areas are states where current investment is inefficient. The gray and dotted areas are states that satisfies $Q_b, Q_g > 0$.

to be inefficient at some date and after some history, then the present discounted value of the firm would be strictly less than the value of the efficient firm, namely $\bar{Q}(s)$.

Lemma G.6 in Appendix G explicitly describes $\bar{\mathbf{v}}_s$, which in turn gives us the unconstrained sets E_s explicitly. For now, it suffices to note that investment is efficient and firm surplus is thereby maximized if (i) v_b is sufficiently high, and (ii) $v_g - v_b$ is sufficiently high. Requirement (ii) is peculiar to our formulation in terms of contingent utilities. This says that the difference $v_g - v_b$ must also be sufficiently great, because this relaxes the incentive constraint [IC*], thereby permitting efficient investment. Part (b) of Proposition 4.1 indicates precisely the minimal size that the difference $v_g - v_b$ needs to achieve, in order for $\mathbf{v}$ to be in E_s. It is clear that (i) is a necessary property, because if contingent utility v_b is very low, then by [PK_b], the vector $\mathbf{w}_b$ must be small and, perforce, $w_{bg} - w_{bb}$ must also be very small. But this violates property (ii) in the subsequent period in the event of a bad shock.

4.2. Firm's Problem

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

Theorem 2. *The firm's discounted surplus is the unique function $Q : V \times S \to \mathbb{R}$ that satisfies* [VF]. *The function Q is concave, continuously differentiable, and supermodular in $\mathbf{v}$ for each s. The optimal contract $(k, m_i, \mathbf{w}_i)$ is continuous in $(\mathbf{v}, s)$. Moreover,*

(a) $Q_g(\mathbf{v}, s) \geq 0$ *and $Q_g + Q_b$ is a non-negative martingale that converges in finite time almost surely to* 0*.*

(b) *Given the sets E_s defined above, the optimal contract is such that such that $\mathbf{v} \in E_s$ if, and only if, $Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s) = 0$, and under the optimal contract, the induced $\mathbf{v}$ process converges to E_s in finite time almost surely.*

The proof is in Appendix E. The existence, uniqueness, concavity and differentiability properties of the surplus function Q are standard, as is the continuity of the policy function. Some further properties of the value function Q are described in Appendix B. A key 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.

Theorem 2 ensures that Q is differentiable everywhere, so the directional derivative in the direction $(1, 1)$, denoted as $\mathsf{D}_{(1,1)} Q(\mathbf{v}, s) := Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s)$, is well defined. Theorem 2 says that the long-run properties of the contract can be discerned by focusing attention on the directional derivative $Q_b + Q_g$, which is a non-negative martingale.[17] The Martingale Convergence Theorem — see, for instance, Theorem 2 on p 517 of Shiryaev (1995) — ensures that $\mathsf{D}_{(1,1)} Q$ converges to a non-negative and integrable random variable. The *proof* of Theorem 2 shows that along almost every path, if the value of $\mathsf{D}_{(1,1)} Q$ is always positive, then it must vary after 'good-good' and 'good-bad' shock pairs. (In other words, the martingale 'splits' after good shocks. Contrast this with the iid case, where the martingale splits whenever it is non-zero.) Because these shock pairs occur infinitely often almost surely, the martingale must converge to 0 almost along every path. Proposition 4.1 now implies that the induced contingent utilities will eventually lie in the unconstrained set E_s.

While the conclusion of (martingale) convergence is also drawn in the iid case, there is an important difference. In the iid case, if the value function is strictly concave, there is a one-to-one relationship between the derivative and ex ante promised utility. In the case of persistence, knowing $\mathsf{D}_{(1,1)} Q(\mathbf{v}, s) = c$ for some $c > 0$ does not pin down

(17) The counterpart to this in the iid model, where ex ante promised v is the state variable, is the fact that the derivative of the value function is a martingale. This observation was first made by Thomas and Worrall (1990).

$\mathbf{v}$. Instead, it only gives us a set of points (typically a curve in V) where the directional derivative is c. This makes the convergence argument more subtle.

Theorem 2 also states that the firm becomes efficient in finite time almost surely. This implies that the value of $\mathsf{D}_{(1,1)}\, Q$ cannot stay positive but very close to zero for infinite amounts of time. If that happens, the contingent utility vector $\mathbf{w}_g$ must necessarily be in a small neighborhood below $\bar{\mathbf{v}}_g$ after a good shock. An additional good shock will then send continuation contingent utility vector to $\bar{\mathbf{v}}_g$ where the martingale is zero. Thus, all convergence occurs in finite time almost surely.

In the sequel, we will restrict attention to *maximal rent* contracts, which are contracts where payments are made as soon as possible.[18] In the maximal rent contract, the contingent utility eventually cycles between the two points $(\bar{\mathbf{v}}_g, g)$ and $(\bar{\mathbf{v}}_b, b)$. This property is another major difference between iid shocks and those with persistence. In the former, $\bar{\mathbf{v}}_b = \bar{\mathbf{v}}_g$, and so firm dynamics do not exist in the long run. However, when shocks are persistent, the existence of non-trivial dynamics even after the firm is unconstrained leads to interesting short- and long-run properties of the optimal contract that we now examine.

We first describe the properties of the optimal contract in the early stages of the contract, when the firm is constrained, ie, prior to it reaching the threshold levels of contingent utility.

5. Financially Constrained Firm

The agency problem that arises due to private information causes the firm's financing constraint. The optimal contract determines how the financing constraint evolves over time, as well as the firm's compensation and investment policies. We shall characterize these aspects in turn for the financially constrained firm.

5.1. Contract Evolution

When the firm is formed, 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, or equivalently, that entrepreneurs compete for the principal's funds. Given the principal's ex-ante belief about success $s_0 = s$, she chooses the *initial contingent utility* vector $\mathbf{v}_s^0$ that maximizes her expected profit $P(\mathbf{v}, s) := Q(\mathbf{v}, s) - \mathbb{E}^s[\mathbf{v}]$. The principal can always choose the initial

(18) 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.

state so that firm surplus is maximized, but this gives the agent too much rent. Instead, it is optimal for the principal to choose $\mathbf{v}_s^0 \in H := \{(\mathbf{v}, s) : Q_b(\mathbf{v}, s) > 0,\ Q_g(\mathbf{v}, s) > 0\}$, which is the set of states where marginally raising promised utilities (either v_g or v_b or both) increases firm surplus. (The set H is plotted as the gray shaded areas in figure 3.)

Proposition 5.1. Let $\mathbf{w}_b, \mathbf{w}_g$ be the optimal contingent utilities at the state $(\mathbf{v}, s)$. The optimal contract evolves in the following way:

(a) $(\mathbf{v}_s^0, s) \in H$, and $P(\mathbf{v}_g^0, g) \geq P(\mathbf{v}_b^0, b)$ with strict inequality iff $\Delta > 0$;

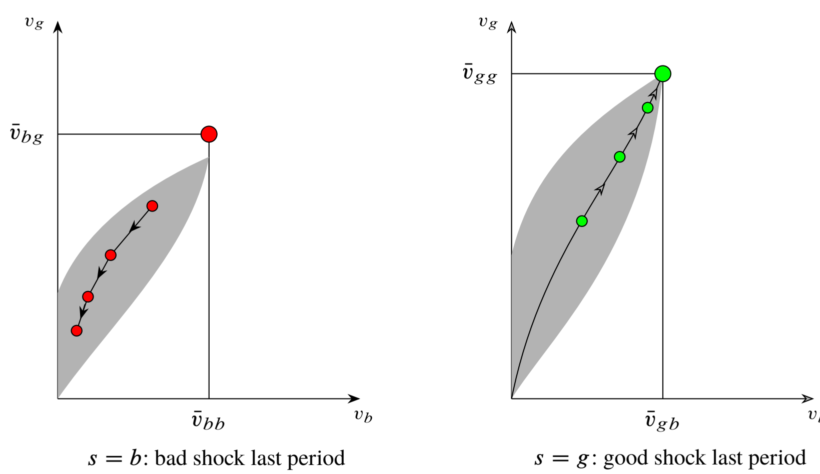
(b) for any $(\mathbf{v}, s) \in H$, $\mathbf{w}_b \in H$, and $\mathbf{w}_g \in H$ or E_g;

(c) $w_{gg} > v_g$, $\mathbf{w}_g$ increases in v_g, and the contingent utility vector strictly increases after two or more consecutive good shocks;

(d) 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.

The proposition is proved in Appendix F.1. It shows that the principal gets a higher expected profit if ex ante information indicates a greater chance of success ($s_0 = g$). In venture capital investments, the dependence of initial expected profits on the surrounding business climate is a subject that has received much attention because these investments are volatile and correlated with the business cycle, as described in Gompers et al. (2008).

Figure 4: *Evolution of States*

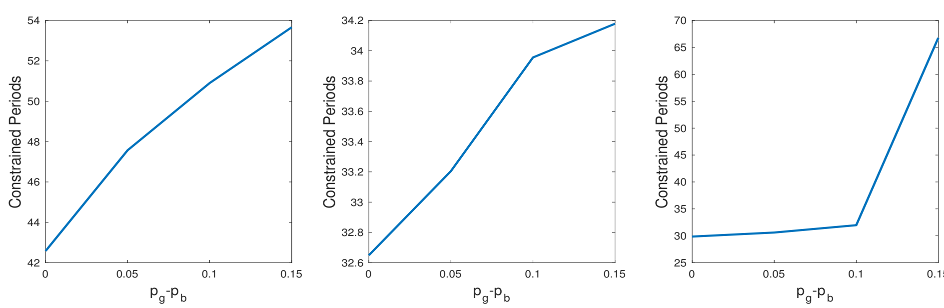


The contingent utilities are always below their thresholds as long as firm financing is constrained because H lies below $\bar{\mathbf{v}}_s$ (see Lemma 3.4 of the Supplemental Appendices). After a good shock, the optimal contingent utility vector $\mathbf{w}_g$ always lies on a particular curve (a one-dimensional subspace) of V. This is because the policy $\mathbf{w}_g$ is only a function of v_g, as seen in the program [VF]. (An instance of this curve is illustrated in the right panel of Figure 4.) Moreover, the supermodularity of Q implies that w_{gg} and w_{gb} are complements. So both of them become larger as v_g increases. Because w_{gg} is greater than v_g, 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 when shocks are persistent. The above result shows that w_{bb} cannot always be larger than v_b if bad shocks keep occurring. If it is, the contract will converge to the efficient set E_b first (instead of E_g), 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 F.2 in Appendix) which is intuitive as it represents a loss in contingent utility for the agent.

Markovian shocks are also different in that the convergence pattern has different properties as shocks move further away from the iid case. This is because the optimal contract with persistence exhibits stronger path dependence. After a bad shock, the contingent utility vector will lie in the left panel of Figure 4 where the contingent utilities are relatively smaller. Another bad shock is more likely to occur because of persistence. So the contingent utilities are kept low. Even if the second shock is good, the firm will not recover immediately, because v_g is small and the [PK_g] constraint restricts contingent utilities to move up. In particular, it necessarily takes *two or more consecutive* good shocks for the firm to become efficient when persistence is high. 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.

Figure 5: *Average time spent being constrained*[†]



[†]At each persistence level, we simulate 10,000 sample paths. The figures plot the average time periods that the firm spends in the 10,000 simulations 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.

In addition, Figure 5 shows our numerical experiments regarding the time periods that the firm spends in the constrained stage. The three panels plot the average constrained time versus persistence at different values of p_b. As p_b increases, it's easier for the firm to grow out of the constrained stage in the iid case ($p_g - p_b = 0$), because good shocks are more likely to occur. In all panels, the firm spends increasingly longer time being financially constrained as the Markov process governing the shocks moves further away from being iid. In particular, the right panel shows that the constrained time is significantly longer when $p_b = 0.6$ and p_g increases from 0.6 to 0.75. 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 optimal compensation and investment policy at different points of the set H. (By Proposition 5.1, if $\mathbf{v} \in H$, then $\mathbf{w}_i(\mathbf{v}, s) \in H$, ie, if contingent utilities lie in H, then regardless of the shock, continuation contingent utilities also lie in H, unless there is a transition to E_g.) The optimal contract stipulates that the principal gets a transfer m_i when shock i is reported. The rest of the firm's cash flow, $\theta(i)R(k) - m_i$, is the agent's compensation.

Proposition 5.2. For any state $(\mathbf{v}, s) \in H$, the compensation policy satisfies:
(a) no pay for failure: $m_b(\mathbf{v}, s) = 0$;
(b) pay for success when v_g is large enough: $R(k(\mathbf{v}, s)) - m_g(\mathbf{v}, s) > 0$ if, and only if, $v_g > \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$.

Propositions from this Section are proved in Appendix F.2. Proposition 5.2 says that a necessary condition for the agent to receive compensation is that she reports a positive cash flow. Moreover, the agent receives compensation if, and only if, $v_g \geq \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. This value is the *one-step* boundary in the sense that a good shock in the present period will send the firm to the unconstrained stage if v_g is above this boundary. Thus, all payments are *back loaded* until the firm becomes financially unconstrained. This is a robust property that holds regardless of the degree of persistence.

The presence of private information implies that investment may well be inefficient, because investment size determines the static information rent which adds to the cost of financing. In figure 3, the gray shaded areas below the dashed lines indicate all the states in H with under-investment.

Proposition 5.3. 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 $k(\mathbf{w}_b, b) < \bar{k}_b$.

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 today, which makes it optimal to invest more today. Moreover, larger v_g or smaller v_b relaxes the incentive constraint.

A key feature of the iid model is that investment is always inefficient for a financially constrained firm. In our model with persistence, the investment is below the efficient level after good shock. However, investment can be temporarily efficient after bad shock though the firm is not fully unconstrained. Lemma F.5 in the Appendix shows that $k(\mathbf{w}_b, b) = \bar{k}_b$ can happen when persistence is high ($\mathbf{p} \in B_h$). In the left panel of figure 3, the points in H that immediately surround $\bar{\mathbf{v}}_b$ all have efficient investment ($\bar{k}_b$). Since policies are continuous, the optimal contract can evolve to the region. In that case, the firm investment will stay at $\bar{k}_b$ until a good shock occurs. To understand this result, let us consider the incentive constraint. If cash flow $R(k)$ realizes and the agent lies, she 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, she gets expected utility $\mathbb{E}^g[\mathbf{w}_g] \simeq \mathbb{E}^g[\bar{\mathbf{v}}_g]$. When persistence is high, since firm surplus is very volatile, $\mathbb{E}^g[\bar{\mathbf{v}}_g]$ is much larger than $\mathbb{E}^g[\bar{\mathbf{v}}_b]$. So even investing the efficient amount $\bar{k}_b$ still satisfies the incentive constraint. In other words, the agent strictly prefers truth-telling in such case.

6. Financially Unconstrained Firm

An unconstrained firm is one that has efficient investment and pays a dividend to its equity holders. The structure of the unconstrained firm is stark in the iid case: the agent retains all the output, and firm investment and capital structure are constant over time. These properties are seldom seen in practice. As we shall see below, when private information displays persistence, none of these conclusions hold.

In the unconstrained firm, it is easy to see that the investment $\bar{k}_s$ and the state variables $(\bar{\mathbf{v}}_s, s)$ all depend on the previous period's production shock s. Moreover, the optimal transfers $\bar{m}_{si} = m_i(\bar{\mathbf{v}}_s, s)$ in general depend on both the current and the previous shocks. The explicit values of $\bar{m}_{si}$ and $\bar{\mathbf{v}}_s$ are pinned down by the active constraints. Clearly, the promise keeping conditions [PK_b] and [PK_g] must hold. As shown in the Appendix, the incentive compatibility constraint [IC] is active when persistence level is low, or $\mathbf{p} \in B_\ell$. However, when persistence is high, or $\mathbf{p} \in B_h$, the incentive constraint is slack and the limited liability constraint [LL] becomes active.

Theorem 3. *In the unconstrained firm, compensation and contingent utility are sensitive to persistence levels and shocks. The agent gets higher compensation for consecutive successes than a single success. In particular,*

(a) *no pay for failure:* $\bar{m}_{sb} = 0$*;*

(b) *at low persistence, always pay the agent for success:*

$$0 < R(\bar{k}_b) - \bar{m}_{bg} < R(\bar{k}_g) - \bar{m}_{gg}, \quad \textit{if} \;\; \mathbf{p} \in B_\ell$$

(c) *at high persistence, pay the agent only for consecutive successes:*

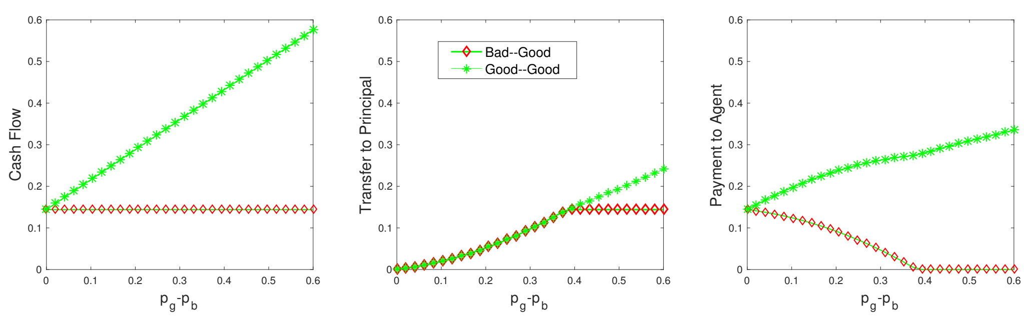
$$0 = R(\bar{k}_b) - \bar{m}_{bg} < R(\bar{k}_g) - \bar{m}_{gg}, \quad \textit{if} \;\; \mathbf{p} \in B_h$$

(d) *the agent gets larger contingent utilities after success*

$$\bar{\mathbf{v}}_b < \bar{\mathbf{v}}_g \;\; \textit{if and only if} \;\; \Delta > 0$$

The proof is in Appendix G. A common feature of firm policies is that they all vary with the degree of persistence. Figure 6 plots the split of cash flows between principal and 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 positive 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.

Figure 6: *Cash Flow, Transfer and Compensation in Unconstrained Firm*†

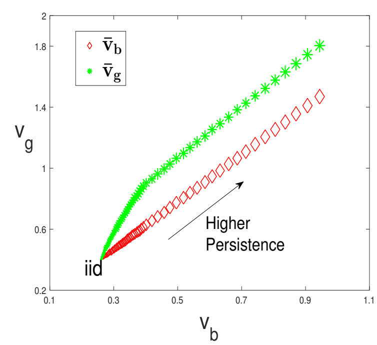


†The pictures plot cash flow (left panel), transfer (middle panel), and compensation (right panel) when a good shock occurs contingent on the previous shock being bad or good. The cash flow always equals the sum of transfer and compensation. Parameters used in the pictures are: $\delta = 0.9$, $R(k)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

The 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 'good-good' shock pair) than when the firm cash flow is $R(\bar{k}_b)$ (after 'bad-good' shock pair). So compensation transforms from linear to strictly convex in performance, as persistence deviates from zero. Indeed, if persistence is sufficiently large, or $\mathbf{p} \in B_h$ ($\Delta > 0.4$ in Figure 6), 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 spit of success).

Figure 7: *Threshold Contingent Utilities*



†The picture shows how the threshold contingent utilities $\bar{\mathbf{v}}_b$ and $\bar{\mathbf{v}}_g$ change with persistence. 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: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

Theorem 3 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, or $\bar{v}_{bb} = \bar{v}_{gb}$. But the continuation utility contingent on good shock does, or $\bar{v}_{bg} < \bar{v}_{gg}$. 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 7, $\bar{\mathbf{v}}_g - \bar{\mathbf{v}}_b$ increases as persistence becomes higher (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.[19]

(19) 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

7. Implementation

This section implements the optimal contract by a set of standard financial securities. 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. At time zero, the firm is financed by issuing debt, equity, and by obtaining a credit line. Both the long-term debt as well as the credit line are associated with a cash-flow covenant. The agent is compensated with firm equity and stock options. Moreover, the agent controls the use of the firm's credit and its payout policy. The firm defaults if it ever exhausts the credit line or does not make the coupon payment. We now illustrate the basic features of these instruments.

Equity: equity holders receive dividend payments made by the firm.

Stock options: option holders can buy the firm's stock at the specified strike price and then sell it back to the firm at the market value.

Debt covenant: a clause in the debt contract that requires the borrower (the firm) to maintain certain financial terms. For example, cash flow to debt ratio and leverage ratio are commonly stipulated terms. The covenant associated with the credit line and the long-term debt in this implementation requires the firm to maintain a positive cash flow.

Credit line: revolving credits provided to the firm with the covenant. The credit line limits are adjusted each period based on the history of covenant violations. The credit account balance is the amount of debt drawn down by the firm from the credit line and is charged an interest rate $r = 1/\delta - 1$ each period.

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

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

Table 1: *Firm's Balance Sheet*

Assets	Liabilities
NPV of Investments	Debt
Compensating Balance	Equity
	Stock Options

utility $\mathbb{E}^p[\bar{\mathbf{v}}]$ is precisely $pR(\bar{k})/(1-\delta)$, just as in Clementi and Hopenhayn (2006).

Theorem 4. *The optimal contract can be implemented by equity, stock options, credit line, long-term debt, and compensating balance such that*

(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 7.2 below.

(ii) The compensating balance pays interest $\bar{c}$ to the firm in each period. Contingent on whether the firm violates the covenant or not, it receives an immediate credit limit C_b or C_g and makes a coupon payment c_b or c_g today, and it receives a contingent credit limit pair (C'_{bb}, C'_{bg}) or (C'_{gb}, C'_{gg}) tomorrow.[20] *The values of $\bar{c}$, c_i, and (C'_{ib}, C'_{ig}) are specified in Lemma 7.3 below.*

It is incentive compatible for the agent to truthfully report cash flow and use it to pay back credit balance. Once the balance is fully repaid, payout d specified in [7.2] *below is issued.*

According to the above implementation, the firm has a balance sheet illustrated in Table 1. Although the implementation is not unique, this mechanism highlights the distinct features of a model with persistent private information. The option payoff to the agent reflects the convex pay-performance relation. The investors' equity holding reflects their stake in the firm. The firm's credit limit is contingent on covenant, so that the evolution of the firm's available credit matches that of contingent utilities in the contract.

7.1. Credit Line with Contingent Limit

In the implementation, we allow the agent to control how to use the firm's available funds. In particular, the agent can withdraw funds from the credit line to invest, repay coupon, issue payout, or simply compensate himself. Because the agent can always divert all credits and default, the firm's available credit has to match the promised utilities in the contract in order to provide incentives for the agent to not deviate from the optimal policies.

The first step is to find an alternative way of representing the state variables of the contract. The binary variable s simply indicates whether the firm complied with its credit line covenant or not in the previous period. Suppose, starting a certain period, the firm has a credit balance M. If the firm violates the covenant ($i = b$), its credit limit is immediately adjusted to C_b. If the firm complies with the covenant ($i = g$), it has credit limit C_g and generates cash flow $R(k)$. Because the agent can always divert the firm's

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

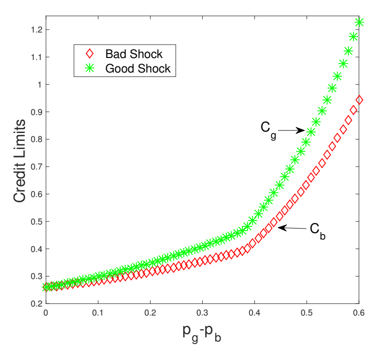
available credits, her contingent payoffs are

$$v_i = C_i - M + \theta(i)R(k(\mathbf{v}, s)) \tag{7.1}$$

An immediate result is that the firm's credit limits fluctuate with performance throughout the life cycle of the firm, as long as persistence is positive. On the contrary, the firm always has a constant credit limit in the iid case. When persistence is positive, the credit limit is reduced if the firm violates the cash flow covenant. The amount of credit gap, $C_g - C_b$, also endogenously varies with history. In the unconstrained stage, this credit gap is shown to be increasing with persistence level when fixing p_b.

Proposition 7.1. Given any performance history, the firm has 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.

In the iid case, a credit line with constant limit simply works. This is because if the agent repays the cash flow $R(k)$, then the credit account balance is reduced and the available credit is raised by the same amount. The agent has $R(k)$ more of discounted future payoff and therefore does not have incentive to divert the cash. This mechanism won't work in the persistent case because the agent also obtains the Markovian information rent on top of the cash flow $R(k)$. So in order to deter the agent from misbehaving, investors need to adjust the firm's credit limits. Eliminating persistence, our implementation becomes a similar mechanism as in DeMarzo and Fishman (2007b).

Obviously, we cannot uniquely pin down C_b, C_g, and M for a given state $(\mathbf{v}, s)$. One way of implementing the contract is to normalize $C_b = \bar{v}_{sb}$ after any history. Then

[7.1] for $i = b, g$ jointly determine C_g and M for any $(\mathbf{v}, s)$. Figure 8 plots how C_b, C_g vary with persistence in the unconstrained firm. It's clearly seen that the credit gap monotonically enlarges as persistence level increases. By the normalization, it is equivalent to use (C_g, M, s) as state variables. We can then rewrite the policies in the optimal contract as functions of the new state variables and denote them as $(\tilde{k}, \tilde{m}_i, \tilde{\mathbf{w}}_i)$. Moreover, contingent on whether the firm violates the covenant or not today, the next period's state variables are either (C'_{bg}, M_b, b) or (C'_{gg}, M_g, g).

7.2. Equity and Stock Options

To implement the optimal compensation, we normalize the firm's outstanding equity share to be one and define its (cum-dividend) stock price z as the present value of future payouts. The firm payout d consists of dividend and option payments. Stock prices can be recursively expressed as $z_{si} = d_i + \delta\,\mathbb{E}^i[\mathbf{z}'_i]$, where $s, i \in S$ are the previous and the current shock respectively, $d_i \geq 0$ is the current payout, and $\mathbf{z}'_i = (z'_{ib}, z'_{ig})$ is the next period's stock prices.

The agent is compensated by equity and stock options. First, the agent holds a fraction $\lambda \in [0, 1]$ of the firm equity, and the investors hold the remaining $1-\lambda$ fraction. Second, at the beginning of each period the agent is granted the option to purchase the firm stock at price K which expires at the end of the period. So the option pays $\max\{z - K, 0\}$ to the agent each period. The rest of the firm's payout is issued as a dividend split between the agent and investors according to their equity holdings. The optimal compensation is implemented by designing λ, K such that the agent's total security payoff matches her payment in the contract. That is, at any state (C_g, M, s), the following holds for any $s, i \in S$

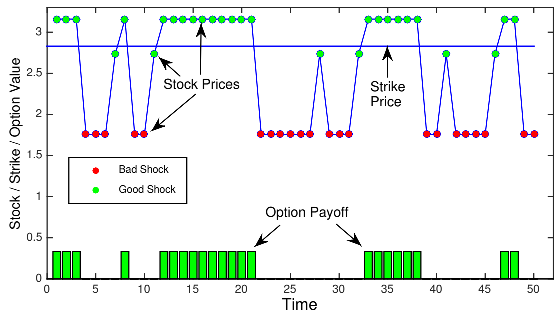
[7.2] $$\lambda d_i + (1-\lambda)\max\{z_{si} - K, 0\} = \theta(i)R(\tilde{k}) - \tilde{m}_i$$

Proposition 7.2. The agent gets the same payoff from security holdings as from the optimal contract if $\lambda = \frac{R(\bar{k}_b)-\bar{m}_{bg}}{R(\bar{k}_b)}$ 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 option payoff in compensation increases in Δ.

Given this compensation design, Figure 9 plots the stock prices and the option payoffs from a simulated path of the unconstrained firm. 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 option has positive payoff only after 'good-good' shock pairs when the stock price reaches its highest level. So the option holding highlights the convex feature of compensation in the persistent environment.

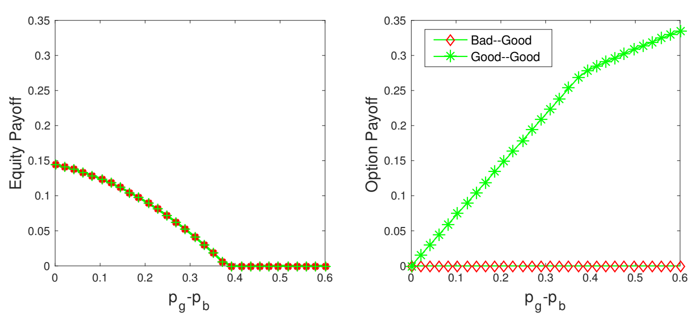
Figure 9: *Stock Prices and Option Payoffs.*†



†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: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.3$, $p_g = 0.7$.

Proposition 7.2 also shows how the compensation structure changes with persistence levels. In the iid case, the agent is paid entirely by equity and she 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 implication of the iid environment. Moreover, as persistence level 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.

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 zare: $\delta = 0.9$, $R(k) = 1.2\sqrt{k}$, $p_b = 0.2$, and p_g increases from 0.2 to 0.8.

7.3. Evolution of Credit Limit and Balance

The credit balance evolves according to withdraws and repayments. Suppose that the agent does not divert firm cash or credit. The credit line is withdrawn to make investment, to pay coupons of the long-term debt, and to issue payouts. Cash flow and interest payment are used to pay back credit balance. If the current state is (C_g, M, s), then the credit balance starting next period contingent on shock i occurs today will be

[7.3] $$M_i = (1+r)[M + \tilde{k} + c_i + d_i - \theta(i)R(\tilde{k}) - \bar{c}]$$

where c_i is the coupon payment (outflow), d_i is the optimal payout (outflow) that satisfies [7.2], and $\bar{c}$ is the interest payment (inflow). The new state will be (C'_{ig}, M_i, i). To provide incentive for the agent to not deviate, investors will design c_i, $\bar{c}$, and C'_{ig} so that the firm's available credit after any history always matches the contingent utility in the optimal contract.

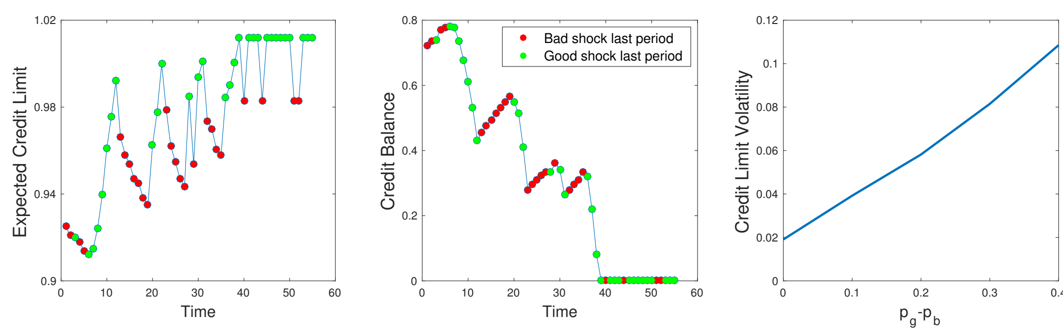
Lemma 7.3. Suppose the agent does not divert cash or credit and choose the optimal payout d_i at any state (C_g, M, s). Then the firm's available credit evolves in the same way as contingent utility in the optimal contract under the following mechanism:

[7.4] $$\bar{c} = \bar{k}_g, \quad c_i = x_i + \bar{k}_g - \tilde{k} + \theta(i)R(\tilde{k}) - d_i$$

[7.5] $$C'_{ig} = \tilde{w}_{ig} + M_i - R(\tilde{k}_i)$$

where $x_b = \frac{1}{1+r}[p_b(\tilde{w}_{bg} - \tilde{w}_{bb}) - r\bar{v}_b]$, $x_g = \frac{1}{1+r}[(1-p_b)(\tilde{w}_{bg} - \tilde{w}_{bb}) - r\bar{v}_b]$, and $\tilde{k}_i = k(\tilde{\mathbf{w}}_i, i)$.

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.

The coupon payment c_i and credit limit C'_{ig} in Lemma 7.3 are functions of (C_g, M, s) constructed from the policies of the optimal contract. Under the proposed mechanism, c_i is always positive and the firm's available credit always satisfies [7.1]. By truthfully reporting cash flow and by obtaining security payoffs until the firm credit account is fully repaid, the agent always gets the same contingent utilities, as in the optimal contract. Since these utilities satisfy the incentive compatible constraint, the agent has no incentive to lie about cash flow. Moreover, if the agent diverts all available credits and defaults at state (C_g, M, s), she gets either $\bar{v}_b - M$ or $C_g - M + R(\tilde{k})$. However, by [7.1], these values are exactly the equilibrium contingent utilities. Therefore, the agent has no incentive to divert all the credits either. Theorem 4 formally shows these arguments.

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.

8. 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 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 she 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 [21]

A. Recursive Domain

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. □

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 $\mathbb{R}^2_+$. Following Abreu, Pearce and Stacchetti (1990), we define the operator $\Phi : \mathcal{K} \to \mathcal{K}$ as follows: for $C \in \mathcal{K}$, let

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

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.

(21) For ease of exposition, several technical statements and their proofs are left to the Supplementary Appendix which are short for SA in the references below.

Proof of Theorem 1. It is easy to see that Φ is well defined, that is, Φ maps closed and convex cones to closed and convex cones. Let $\alpha \in [0,1]$, and define $C_\alpha := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq \alpha v_b\}$. 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

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

where the inequality follows from the assumption that $w_{bg} \geq \alpha w_{bb}$, 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

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

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 $\alpha' = (1-p_g(1-\alpha))/(1-p_b(1-\alpha))$. In particular, any $\mathbf{v} \in C_{\alpha'}$ can be implemented by $(k, m_i, \mathbf{w}_i)$ with $\mathbf{w}_i \in C_\alpha$ for $i = b, g$.

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

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

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 $\Phi^n(C_0) = \bigcap_{k\le n} \Phi^k(C_0) = C_{\alpha_n}$, 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 $V := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \ge v_b\}$ is a fixed point of Φ, we apply the operator Φ to V. Take any continuation utility $\mathbf{v} \in V$, and consider the policy $(k, m_i, \mathbf{w}_i)$ that satisfies $R(k) = m_g = v_g - v_b$, $m_b = 0$, $w_{ig} = w_{ib} = v_i/\delta$. Since $(k, m_i, \mathbf{w}_i)$ implements $\mathbf{v}$ and $\mathbf{w}_i \in V$, we must have $\mathbf{v} \in \Phi(V)$, which means $V = \Phi(V)$. By construction, V is the largest fixed point of Φ, which completes the proof. □

B. Some Proofs from Section 4

Proposition B.1. The firm's discounted surplus is the unique function $Q : V \times S \to \mathbb{R}$ that satisfies [VF]. The function Q is concave, continuously differentiable, and supermodular in $\mathbf{v}$ for each s. The optimal contract $(k, m_i, \mathbf{w}_i)$ is continuous in $(\mathbf{v}, s)$. Moreover, $Q_g(\mathbf{v}, s) \ge 0$, and $\mathsf{D}_{(1,1)}(\mathbf{v}, s) \ge 0$.

Proof. 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. Here, we shall only establish the claims pertaining to the partial derivatives.

Because it is feasible to invest zero in each period, make transfers $m_g = -v_g$, $m_b = -v_b$ (compensate the agent) in the first period, and make no transfer in all subsequent periods, we must have $Q(\mathbf{v}, s) \ge 0$ for any $(\mathbf{v}, s) \in V \times S$. Moreover, the surplus $Q(\mathbf{v}, s)$ is bounded above by the efficient surplus $\bar{Q}(s)$. So let $M = \max\{\bar{Q}(g), \bar{Q}(b)\}$, then $Q(\mathbf{v}, s) \le M$.

Take any $(\mathbf{v}, s) \in V \times S$. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at $(\mathbf{v}, s)$, $\mathbf{v}' = \mathbf{v} + (0, \varepsilon)$ for any $\varepsilon > 0$, and $m'_g = m_g - \varepsilon$. Then the policy

$$(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}_g) \in \Gamma(\mathbf{v}', s)$$

because the specified change in states and policy increase both sides of [PK_g] by ε and only increase the left hand side of [IC*]. Moreover, because the repayment m_g does not appear in the objective of [VF], we must have $Q(\mathbf{v}', s) \geq Q(\mathbf{v}, s)$, implying $Q_g(\mathbf{v}, s) \geq 0$. Moreover

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

where the inequality is by Theorem 1 (e) of SA.

The results that $Q(\mathbf{v}, g) \geq Q(\mathbf{v}, b)$ and that Q is concave and supermodular are all shown in Theorem 1 of SA. □

Lemma B.2. 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 $\bar{Q}(s)$. □

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.3. 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

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

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 $f_s(\cdot)$. □

Lemma B.4. $Q_b[(v_b, f_s(v_b)), s] \geq 0$ 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 $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] = 0$. □

By Lemma B.4, 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.5. $\mathbf{v} \in E_s^*$ 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.5 that E_s is non-empty, closed, and convex. Then, part (a) follows immediately from the definition of E_s^* in [B.3] and from Lemma B.5 which proves that $E_s^* = E_s$.

To prove part (b), 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 $v_g - v_b \geq (1 + \delta\Delta + \delta^2\Delta^2 + \dots)R(\bar{k}_b) = R(\bar{k}_b) + \delta\Delta\frac{R(\bar{k}_b)}{1-\delta\Delta}$.

(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

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

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 it is necessary to satisfy $v_g - v_b \geq R(\bar{k}_s) + \delta\Delta \max\left[\frac{\delta \bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right]$ for any $\mathbf{v} \in E_s$.

Finally, to see part (c), 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.5 then implies $\mathbf{w}_g(\mathbf{v}, s) \in E_g$ and $\mathbf{w}_b(\mathbf{v}, s) \in E_b$. □

C. Auxiliary Problem

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

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

where $y \geq 0$ and $s = b, g$. To ease notation, let $\mathbf{x}(y, s)$ be a solution to problem [P3].

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 part (a) of Theorem 1) and a convex constraint, $\Psi(y, s)$ is concave in y.

(b) Lemma 3.1 implies that any $\mathbf{x}(\mathbf{v}, s)$ must lie in the interior of V. Let γ_s be the Lagrange multiplier of [P3]. The first order conditions and the envelope condition for problem [P3] are

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

[C.2] $$\Psi_y(y,s) = \gamma_s$$

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. □

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 in 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

$$\text{[FOC}k\text{]} \qquad R'(k) = 1/[p_s - \eta_g(\mathbf{v}, s) + \mu_g(\mathbf{v}, s)]$$

$$\text{[FOC}w_{bb}\text{]} \qquad (1-p_s)Q_b(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s)(1-p_b) + \alpha(\mathbf{v}, s)(1-p_g)$$

$$\text{[FOC}w_{bg}\text{]} \qquad (1-p_s)Q_g(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s)p_b + \alpha(\mathbf{v}, s)p_g$$

$$\text{[FOC}w_{gb}\text{]} \qquad p_s Q_b(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s)(1-p_g) - \alpha(\mathbf{v}, s)(1-p_g)$$

$$\text{[FOC}w_{gg}\text{]} \qquad p_s Q_g(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s)p_g - \alpha(\mathbf{v}, s)p_g$$

In addition, we also have the following envelope conditions

$$\text{[Env}_b\text{]} \qquad Q_b(\mathbf{v}, s) = \eta_b(\mathbf{v}, s)$$

$$\text{[Env}_g\text{]} \qquad Q_g(\mathbf{v}, s) = \eta_g(\mathbf{v}, s)$$

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 $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) = Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s)$ induced by the optimal contract is a nonnegative martingale.

Proof. Take any $(\mathbf{v}, s) \in V \times S$. Adding the first order conditions [FOCw_{bb}] to [FOCw_{gg}], and using envelope conditions [Env$_b$] and [Env$_g$] 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, since $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) \geq 0$, the process $\mathsf{D}_{(1,1)}\, Q$ is a nonnegative martingale. □

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 [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 [FOCw_{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, $\alpha(\mathbf{w}_g, g) = Q_g(\mathbf{w}_g, g) > 0$. And equation [3.5] 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

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 lemma.

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 [FOCw_{gb}] and [FOCw_{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 [FOCw_{bb}] and [FOCw_{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 $(\mathbf{w}_b, b) \in H$. □

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.** We have established the existence of the value function Q as well as the relevant properties in Proposition B.1. 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 $C := \{\mathbf{v} \in V : (\mathbf{v}, g) \in \mathrm{cl}(H), v_g \le \bar{v}_{gg} - \varepsilon\}$ 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 $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, g) \ge \eta$ 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 $t \ge T$. In other words, $\bar{v}_{gg} - v_g^{(\gamma_t)} < \varepsilon$ for all $t \ge T$. 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 $t \ge \tau$. So in a maximum rent contract, the contingent utilities $\mathbf{v}^{\gamma_t} = \bar{\mathbf{v}}_{s_{t-1}}$ for all $t \ge \tau$. □

F. Proofs from Section 5: Constrained Firm

The result in this section first shows that the optimal contract always stay in the set H. Then depending on where the state locates in H, we characterize the firm policies. Lemma 3.4 of the Supplemental Appendices shows some useful properties of the set H: first, $\mathbf{v} << \bar{\mathbf{v}}_s$ 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 $(\mathbf{v}, s) \in H$.

F.1. Proofs from Section 5.1

Lemma F.1. $v_g \ge \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$ if and only if $\mathbf{w}_g(\mathbf{v}, s) \in E_g$.

Proof. Suppose $v_g \ge \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. 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), $w_{gi}(\mathbf{v}, s) \ge \bar{v}_{gi}$ for $i = b, g$. Then by [PK_g] at $(\mathbf{v}, s)$, we know $v_g \ge \delta\, \mathbb{E}^g[\mathbf{w}_g(\mathbf{v}, s)] \ge \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$. □

Proof of Proposition 5.1. Given $s \in S$, the initial contingent utility vector $\mathbf{v}_s^0$ maximizes $P(\mathbf{v}, s) := Q(\mathbf{v}, s) - \mathbb{E}^s[\mathbf{v}]$ for all $\mathbf{v} \in V$. The first order conditions are

$$\text{[F.1]} \qquad Q_g(\mathbf{v}_s^0, s) = p_s, \quad Q_b(\mathbf{v}_s^0, s) = 1 - p_s$$

which implies $(\mathbf{v}_s^0, s) \in H$ by 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 part (b).

To see the rest of part (a), let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at $(\mathbf{v}_b^0, b)$. Then,

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

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

$$\text{[F.2]} \qquad \begin{aligned} & p_b\{R(k) + \delta[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (v_{bg}^0 - v_{bb}^0)\} \\ & \quad \geq P(\mathbf{v}_b^0, b) - \delta[Q(\mathbf{w}_b, b) - \mathbb{E}^b(\mathbf{w}_b)] + k > 0 \end{aligned}$$

The second inequality is because $\delta < 1, \mathbf{w}_b \in V$. Moreover, because $\mathbf{v}_b^0 \in V$, the optimality in choosing $\mathbf{v}_g^0$ implies

$$\begin{aligned} & P(\mathbf{v}_g^0, g) - P(\mathbf{v}_b^0, b) \\ & \quad \geq Q(\mathbf{v}_b^0, g) - \mathbb{E}^g(\mathbf{v}_b^0) - [Q(\mathbf{v}_b^0, b) - \mathbb{E}^b(\mathbf{v}_b^0)] \\ & \quad = Q(\mathbf{v}_b^0, g) - Q(\mathbf{v}_b^0, b) - \Delta(v_{bg}^0 - v_{bb}^0) \\ & \quad \geq \Delta\{R(k) + \delta[Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (v_{bg}^0 - v_{bb}^0)\} \geq 0 \end{aligned}$$

The first inequality in the last line is implied by $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}_b^0, g)$. The second inequality in the last line is by [F.2] and is strict iff $\Delta > 0$.

To see part (c), let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at any $(\mathbf{v}, s) \in H$. If $v_g < \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g]$, then Proposition 5.2 implies $m_g = R(k)$. By [PK$_g$], $v_g = \delta\, \mathbb{E}^g[\mathbf{w}_g] \leq \delta w_{gg}$, which further implies $w_{gg} > v_g$. If $v_g \geq \mathbb{E}^g[\bar{\mathbf{v}}_g]$, then Lemma F.1 implies $w_{gg} \geq \bar{v}_{gg} > v_g$.

Let $\hat{\mathbf{v}} = \mathbf{v} + (0, \varepsilon)$ for some sufficiently small $\varepsilon > 0$. And let $(\hat{k}, \hat{m}_i, \hat{\mathbf{w}}_i)$ be the optimal policy at $(\hat{\mathbf{v}}, s)$. If $\hat{v}_g > \mathbb{E}^g[\bar{\mathbf{v}}_g]$, then $\hat{\mathbf{w}}_g = \bar{\mathbf{v}}_g \geq \mathbf{w}_g$. If $\hat{v}_g < \mathbb{E}^g[\bar{\mathbf{v}}_g]$, then $m_g = R(k)$ and $\hat{m}_g = R(\hat{k})$. From [PK$_g$] at $(\mathbf{v}, s)$ and $(\hat{\mathbf{v}}, s)$, either $\hat{w}_{gg} > w_{gg}$ or $\hat{w}_{gb} > w_{gb}$ or both hold. In addition, from [C.1], [C.2], and Lemma C.3, we must have $Q_g(\hat{\mathbf{w}}_g, g) < Q_g(\mathbf{w}_g, g)$ and $Q_b(\hat{\mathbf{w}}_g, g) < Q_b(\mathbf{w}_g, g)$. Now suppose $\hat{w}_{gg} \leq w_{gg}$ and $\hat{w}_{gb} > w_{gb}$. By the supermodularity and concavity of Q, we know $Q_g(\hat{\mathbf{w}}_g, g) \geq Q_g(\mathbf{w}_g, g)$, a contradiction. Similarly, $\hat{w}_{gb} \leq w_{gb}$ and $\hat{w}_{gg} > w_{gg}$ will also lead to a contradiction. So we must have $\hat{\mathbf{w}}_g \gg \mathbf{w}_g$, ie, the contingent utility vector strictly increases after two consecutive good shocks.

Finally, we establish part (d). When $\mathbf{p} \in B_h$, Lemma G.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 of the Supplemental Appendices, we know any $(\mathbf{v}, s) \in H$ satisfies $\mathbf{v} << \bar{\mathbf{v}}_s$. 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.1 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 good shock, it is only possible to become unconstrained after another two or more good shocks in a row. □

Lemma F.2. 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 locate in the set H. We will verify in Proposition 5.2 below that $m_b(\mathbf{v}, s) = m_b(\mathbf{w}_b, b) = 0$, which further implies (by [PK$_b$])

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

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.2. Investment Dynamics and Proofs from Section 5.2

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

$$\eta_b(\mathbf{v}, s) + \alpha(\mathbf{v}, s) - \mu_b(\mathbf{v}, s) \geq 0,\ m_b(\mathbf{v}, s)[\eta_b(\mathbf{v}, s) + \alpha(\mathbf{v}, s) - \mu_b(\mathbf{v}, s)] = 0 \tag{F.4}$$

$$\eta_g(\mathbf{v}, s) - \alpha(\mathbf{v}, s) - \mu_g(\mathbf{v}, s) = 0 \tag{F.5}$$

We can also use [F.5] to rewrite [FOCk] as:

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

Proof of Proposition 5.2. Let $(k, m_i, \mathbf{w}_i)$ be the optimal policy at any $(\mathbf{v}, s) \in H$.

(a) 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.4] implies $m_b = 0$.

(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.1 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

$$R(k) - m_g = v_g - \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g] \tag{F.6}$$

Hence, $R(k) - m_g > 0$. To see the 'only if ' part, suppose $R(k) - m_g > 0$. 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 $R(k) - m_g = 0$ 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.6] holds. Hence, $v_g > \delta\,\mathbb{E}^g[\bar{\mathbf{v}}_g]$. □

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

$$h_s(v_b) := \min\{y \geq v_b : k((v_b, y), s) = \bar{k}_s\} \tag{F.7}$$

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.3. 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 $w_{bb}^b, w_{bb} < \bar{v}_{bb} = \delta\,\mathbb{E}^b[\bar{\mathbf{v}}_b]$. The equality is verified in Lemma G.6 below. 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 $w^b_{bb} > w_{bb}$. 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 $w^b_{bb} \leq w_{bb}$. By the assumptions and by the monotonicity of $h_b(\cdot)$, we know $w^b_{bg} < h_b(w^b_{bb}) \leq h_b(w_{bb}) \leq w_{bg}$. Then we can obtain:

[F.8] $$Q_g[(0, w_{bg}), b] < Q_g[(0, w^b_{bg}), b] \leq Q_g(\mathbf{w}^b_b, b)$$

The first inequality is implied by the fact $Q(\mathbf{v}, b)$ is strictly concave in v_g on the set $[(0, w^b_{bg}), (0, w_{bg})]$. 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.8], 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

[F.9] $$\begin{aligned} Q_g(\mathbf{w}^b_b, b) &= \frac{p_b}{1 - p_b} Q_b(\mathbf{w}_b, b) \\ &= \frac{p_b}{1 - p_b} Q_b(\mathbf{v}, s) + \frac{p_b(1 - p_g)}{(1 - p_b)(1 - p_s)} \alpha(\mathbf{v}, s) \\ &\leq \frac{p_b}{1 - p_b} Q_b(\mathbf{v}, s) + \frac{p_g}{(1 - p_s)} \alpha(\mathbf{v}, s) = Q_g(\mathbf{w}_b, b) \end{aligned}$$

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.4. Suppose $w_{bg} \geq h_b(w_{bb})$ in Lemma F.3. Then $w^b_{bb} = w_{bb}$ and $w^b_{bg} \geq w_{bg}$.

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

Because $w_{bg} \geq h_b(w_{bb})$ by assumption and $w^b_{bb} \geq h_b(w^b_{bb})$ by Lemma F.3, 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 $w^b_{bb} \neq w_{bb}$, 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, $w^b_{bb} = w_{bb}$.

Now suppose $w^b_{bg} < w_{bg}$. 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.9]. Thus, $w^b_{bg} \geq w_{bg}$. □

Proof of Proposition 5.3. (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 $\hat{k} = k(\mathbf{v}, b), \hat{\mathbf{w}}_i = \mathbf{w}_i(\mathbf{v}, b)$. Suppose $\hat{k} > k$. Since $\Gamma(\mathbf{v}, g) = \Gamma(\mathbf{v}, b)$, optimality at $(\mathbf{v}, g)$ implies

[F.10] $$\begin{aligned} &-k + p_g[R(k) + \delta Q(\mathbf{w}_g, g)] + (1 - p_g)\delta Q(\mathbf{w}_b, b) \\ &\geq -\hat{k} + p_g[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1 - p_g)\delta Q(\hat{\mathbf{w}}_b, b) \end{aligned}$$

From part (a) of Lemma C.2, $Q(\mathbf{w}_g, g) = Q(\hat{\mathbf{w}}_g, g)$. Then [F.10] implies $Q(\mathbf{w}_b, b) > Q(\hat{\mathbf{w}}_b, b)$. Moreover, optimality at $(\mathbf{v}, b)$ implies

$$\begin{aligned} & -\hat{k} + p_b[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1 - p_b)\delta Q(\hat{\mathbf{w}}_b, b) \\ & \quad \geq -k + p_b[R(k) + \delta Q(\mathbf{w}_g, g)] + (1 - p_b)\delta Q(\mathbf{w}_b, b) \end{aligned} \tag{F.11}$$

Add [F.10] [F.11] 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. We show in Lemma 2.4 of SA that investment $k(\mathbf{v}, s)$ decreases in v_b, and strictly increases in v_g.

To see part (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^*].

We will verify in Lemma G.6 that $\Delta < \psi(p_g)$ implies $\bar{v}_{bb} = \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$, and $\Delta > \psi(p_g)$ implies $\bar{v}_{bb} > \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$. We now establish (c).

Consider the former case where $\mathbf{p} \in B_\ell$, or $\Delta < \psi(p_g)$. 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.4, 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 $w_{bb}(\mathbf{v}, s) \geq \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)} = \bar{v}_{bb}$. However, this is a contradiction with $(\mathbf{w}_b, b) \in H$ which requires $w_{bb} < \bar{v}_{bb}$. □

Lemma F.5. 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. Given $\Delta > \psi(p_g)$, Theorem 3 in section 7 will verify that at $(\bar{\mathbf{v}}_b, b)$ 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: Unconstrained Firm

In this section, we show the level of threshold contingent utilities and repayments of the unconstrained firm.

For a given p_g, we define a pair $(\varphi(p_g), \hat{k}(p_g))$ on $(0, 1]$ implicitly by

$$\varphi(p_g) R'[\hat{k}(p_g)] = 1 \tag{G.1}$$

$$R[\hat{k}(p_g)] = \frac{\delta p_g R(\bar{k}_g)}{1 + \delta p_g} \tag{G.2}$$

where $\bar{k}_g$ is defined as the unique k that satisfies $p_g R'(k) = 1$. It is clear that for a given p_g, the pair $(\varphi(p_g), \hat{k}(p_g))$ is unique.

Lemma G.1. The functions $\hat{k}$ and φ are continuous. For any $p_g > 0$, we have $0 < \varphi(p_g) < p_g$, and $\varphi(p_g)$ increases in p_g. Moreover, $\lim_{p_g \downarrow 0} \varphi(p_g) = 0$.

Proof. It is clear from [G.2] that $\hat{k}$ is continuous and $\hat{k}(0) = 0$. Continuity of φ for all $p_g > 0$ follows immediately from [G.1]. Because $R'(0) = \infty$, we see that $\lim_{p_g \downarrow 0} \varphi(p_g) = 0$.

From [G.2], $\hat{k}(p_g) < \bar{k}_g$, since $\frac{\delta p_g}{1+\delta p_g} < 1$. Concavity of R then implies that $R'[\hat{k}(p_g)] > R'(\bar{k}_g)$. From [G.1], $\varphi(p_g) R'[\hat{k}(p_g)] = p_g R'(\bar{k}_g) = 1$, implying that $0 < \varphi(p_g) < p_g$. Moreover, $\frac{\delta p_g R[\bar{k}_g]}{1+\delta p_g}$ increases in p_g. Hence, $\hat{k}(p_g)$ and $\varphi(p_g) = 1/R'(\hat{k}(p_g))$ both increase in p_g. □

Using the cutoff function $\varphi(p_g)$, we can partition the space of transition probability $\{(p_b, p_g) : 0 < p_b \leq p_g < 1\}$ as $B_\ell = \{\mathbf{p} : \varphi(p_g) < p_b \leq p_g\}$ and $B_h = \{\mathbf{p} : p_b \leq \varphi(p_g)\}$.

Lemma G.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 $p_b < \varphi(p_g)$.

Proof. To see part (a), notice that from [G.1], we find $\varphi(p_g) R'[\hat{k}(p_g)] = p_b R'(\bar{k}_b) = 1$. Then $\mathbf{p} \in B_\ell$ implies $R'[\hat{k}(p_g)] > R'(\bar{k}_b)$. The strict concavity of R then implies

$\hat{k}(p_g) < \bar{k}_b$. From [G.2], $R[\hat{k}(p_g)] = \dfrac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} < R(\bar{k}_b)$. 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 $R(\bar{k}_b) > \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} = R[\hat{k}(p_g)]$ by [G.2]. Hence, $\bar{k}_b > \hat{k}(p_g)$ implies $R'[\hat{k}(p_g)] > R'(\bar{k}_b)$. But since $\bar{k}_b$ is the unique solution of $p_b R'(\bar{k}_b) = 1$, we know $\varphi(p_g) > p_b$, ie, $\mathbf{p} \in B_\ell$. Part (b) is established by similar arguments. □

Now, we can define $\psi(p_g) := p_g - \varphi(p_g)$ for a given p_g and rewrite the above partitions as $B_\ell = \{\mathbf{p} : \Delta < \psi(p_g)\}$ and $B_h = \{\mathbf{p} : \Delta \geq \psi(p_g)\}$.

Lemma G.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 [$\text{PK}_b$] 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 $(\bar{k}_s, \bar{m}_{sb}, \hat{m}_{sg}, \mathbf{w}_b = \bar{\mathbf{v}}_b, \mathbf{w}_g = \bar{\mathbf{v}}_g) \in \Gamma(\hat{\mathbf{v}}_s, s)$. 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 part (a) of Proposition 4.1. This forms a contradiction because the definition of E_s implies $Q(\hat{\mathbf{v}}_s, s) < \bar{Q}(s)$. □

Lemma G.4. At state $(\bar{\mathbf{v}}_s, s)$, the agent gets zero payment when a bad shock occurs in the current period, ie, $\bar{m}_{sb} = 0$.

Proof. Suppose $\bar{m}_{sb} < 0$. Then there exists some $\varepsilon > 0$ such that $\bar{m}_{sb} + \varepsilon < 0$. From [$\text{PK}_b$] at state $(\bar{\mathbf{v}}_s, s)$ we obtain

$$\bar{v}_{sb} - \varepsilon = -(\bar{m}_{sb} + \varepsilon) + \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_b]$$

From [IC*] at state $(\bar{\mathbf{v}}_s, s)$ we find

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

Let $\hat{\mathbf{v}}_s = (\bar{v}_{sb} - \varepsilon, \bar{v}_{sg})$ and $\hat{m}_{sb} = \bar{m}_{sb} + \varepsilon$. Then the policy $(\bar{k}_s, \hat{m}_{sb}, \bar{m}_{sg}, \mathbf{w}_b = \bar{\mathbf{v}}_b, \mathbf{w}_g = \bar{\mathbf{v}}_g) \in \Gamma(\hat{\mathbf{v}}_s, s)$. 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 constructions, which implies that $\hat{\mathbf{v}}_s \notin E_s$, by Proposition 4.1 (a). Hence, $Q(\hat{\mathbf{v}}_s, s) < \bar{Q}(s)$, a contradiction. □

Lemma G.5. The threshold contingent utilities satisfy: $\bar{v}_{bb} = \bar{v}_{bg}$, $\bar{v}_{gb} < \bar{v}_{gg}$.

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

[G.3] $$\bar{m}_{sg} \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$$

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

[G.4] $$(\bar{m}_{gg} - \bar{m}_{bg}) = R(\bar{k}_g) - R(\bar{k}_b) - (\bar{v}_{gg} - \bar{v}_{bg})$$

Suppose $\bar{v}_{gg} \leq \bar{v}_{bg}$. Then [G.3] implies that $\bar{m}_{gg}, \bar{m}_{bg} \leq 0$. In addition, [G.4] implies that $0 < R(\bar{k}_g) - R(\bar{k}_b) \leq \bar{m}_{gg} - \bar{m}_{bg}$. So we must have $\bar{m}_{bg} < \bar{m}_{gg} \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) \leq 0$, by [G.3]. 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 G.3. □

Proof of Theorem 3. We have already shown $\bar{m}_{sb} = 0$ in Lemma G.4, which establishes part (a). So we are only left to show the conditions satisfied by $\bar{m}_{sg}$. We proceed by considering the following three cases.

(i) We first establish that $\bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) < R(\bar{k}_g)$. Suppose $\bar{m}_{gg} = R(\bar{k}_g)$. From [G.4], $\bar{m}_{bg} - R(\bar{k}_b) = \bar{v}_{gg} - \bar{v}_{bg} > 0$, which means $\bar{m}_{bg} > R(\bar{k}_b)$. But this violates [LL] for b. So we must have $\bar{m}_{gg} < R(\bar{k}_g)$. From Lemma G.3, [IC] must hold as equality at $(\bar{\mathbf{v}}_g, g)$. Therefore, from display [G.3], we must have $\bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$.

(ii) We now show that $\bar{m}_{bg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) < R(\bar{k}_b)$ when $\mathbf{p} \in B_\ell$. Suppose $\bar{m}_{bg} = R(\bar{k}_b)$, then by [G.4]

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

The last equality is from part (i) above. We then 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 G.2 when $\mathbf{p} \in B_\ell$. Combining this display with [G.3], we have $\bar{m}_{bg} < \bar{m}_{bg}$, a contradiction. Hence, $\bar{m}_{bg} < R(\bar{k}_b)$. Then by Lemma G.3 [IC] must hold as equality at $(\bar{\mathbf{v}}_b, b)$. So from [G.3], we obtain $\bar{m}_{bg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$.

(iii) Next we show that $\bar{m}_{bg} = R(\bar{k}_b) \leq \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$ when $\mathbf{p} \in B_h$, and the inequality is strict if $\Delta > \psi(p_g)$. Suppose $\bar{m}_{bg} < R(\bar{k}_b)$. By Lemma G.3, [LL] for g must hold as equality at $(\bar{\mathbf{v}}_b, b)$. This means $\bar{m}_{bg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$ by [G.3]. But this is a contradiction with Lemma G.2 which concludes that $\delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) \geq R(\bar{k}_b)$ when $\mathbf{p} \in B_h$. Moreover, Lemma G.2 also implies $\bar{m}_{bg} < \delta p_g(\bar{v}_{gg} - \bar{v}_{bg})$ when $\Delta > \psi(p_g)$, or [IC] holds as inequality.

Thus, part (b) of the theorem follows from (i) and (ii) above, while part (c) follows from (i) and (iii) above. Part (d) is established in Lemma G.6 below, which completes the proof. □

Lemma G.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

$$\bar{v}_b = \delta\, \mathbb{E}^b[\bar{\mathbf{v}}_b] \tag{G.5}$$

$$\bar{v}_{bg} = R(\bar{k}_b) - \bar{m}_{bg} + \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g] \tag{G.6}$$

$$\bar{v}_{gg} = R(\bar{k}_g) - \bar{m}_{gg} + \delta\, \mathbb{E}^g[\bar{\mathbf{v}}_g] \tag{G.7}$$

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

(a) When $\mathbf{p} \in B_\ell$, Theorem 3 implies:

$$\bar{m}_{bg} = \bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) \tag{G.8}$$

From [G.6] and [G.7], we know $\bar{v}_{bg} < \bar{v}_{gg}$. Hence, $\bar{\mathbf{v}}_b < \bar{\mathbf{v}}_g$. In addition, combining [G.5] [G.6] [G.7] [G.8] and eliminating $\bar{m}_{sg}$, we can obtain the specified solution in the Lemma.

(b) When $\mathbf{p} \in B_h$, Theorem 3 implies:

$$\bar{m}_{bg} = R(\bar{k}_b), \quad \bar{m}_{gg} = \delta p_g(\bar{v}_{gg} - \bar{v}_{bg}) < R(\bar{k}_g) \tag{G.9}$$

From [G.6] and [G.7], we know $\bar{v}_{bg} < \bar{v}_{gg}$. Hence, $\bar{\mathbf{v}}_b < \bar{\mathbf{v}}_g$. In addition, combining [G.5] to [G.9] and eliminating $\bar{m}_{sg}$, we can obtain the specified solution of of $\bar{v}_b, \bar{v}_{bg}, \bar{v}_{gg}$ in the Lemma. □

H. Proofs from Section 7

Proof of Proposition 7.1. Let $(\mathbf{v}, s)$ be a state induced by the optimal contract. We can simplify [7.1] for $i = b, g$ to get

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

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 [H.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 G.6) into [H.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 H.1. When the firm is unconstrained, its stock prices are:

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

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

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

Proof. In the unconstrained stage, the firm issues all cash flows as payouts. Denote the firm's stock prices in this stage as $\bar{\mathbf{z}}_s = (\bar{z}_{sb}, \bar{z}_{sg})$. According to the definition of $z_{si} = d_i + \delta\,\mathbb{E}^i[\bar{z}'_i]$, 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 7.2. Note that the specified strike price satisfies

$$\bar{z}_{sb} \leq \bar{z}_{bg} \leq K < \bar{z}_{gg} \tag{H.2}$$

The first inequality is by [$\bar{z}_{sb}$] and the second is because $\bar{m}_{gg} \geq \bar{m}_{bg}$. And the last inequality is because by [$\bar{z}_{gg}$] we have

$$\begin{aligned}\bar{z}_{gg} - K &= \bar{z}_{gg} - \bar{z}_{bg} - (\bar{m}_{gg} - \bar{m}_{bg}) \\ &= [R(\bar{k}_g) - \bar{m}_{gg}] - [R(\bar{k}_b) - \bar{m}_{bg}] > 0\end{aligned}$$

Let us consider the agent's payoff from equity and stock options and show in all possible cases [7.2] holds. First, if a bad shock occurs today, then payout $d = 0$ and both securities have zero payoff. Second, consider a good shock occurs today and a bad shock occurs last period. Since $\bar{z}_{bg} \leq K$ and payout is $R(\bar{k}_b)$, the agent's option payoff is zero and equity payoff is $\lambda R(\bar{k}_b) = R(\bar{k}_b) - \bar{m}_{bg}$. Third, consider a good shock occurs both today and yesterday. Since $K < \bar{z}_{gg}$, the option payoff is

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

The second line is by [$\bar{z}_{gg}$]. If $\mathbf{p} \in B_h$, then $R(\bar{k}_b) = \bar{m}_{bg}$ by Theorem 3. So [H.3] reduces to $R(\bar{k}_g) - \bar{m}_{gg}$, and $\lambda = 0$. Hence, the agent's total security payoff is just $R(\bar{k}_g) - \bar{m}_{gg}$. If $\mathbf{p} \in B_\ell$, then $\bar{m}_{bg} = \bar{m}_{gg}$ by Theorem 3, and [H.3] reduces to $R(\bar{k}_g) - R(\bar{k}_b)$. Since the firm payout is $R(\bar{k}_g)$, the agent's equity payoff in this case is $\lambda R(\bar{k}_b) = R(\bar{k}_b) - \bar{m}_{gg}$. So the total security payoff is $R(\bar{k}_g) - \bar{m}_{gg}$. Hence, under the specified λ, K the agent's security payoff is always the same as her compensation in the contract.

Now 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 $1 - \delta p_g \left[\frac{R(\bar{k}_g)}{R(\bar{k}_b)} - 1\right]$ 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, since the agent's equity payoff is always $R(\bar{k}_b) - \bar{m}_{bg}$, it 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 $\mathbf{p} \in B_l$, because the option payoff, either $R(\bar{k}_g) - R(\bar{k}_b)$ or 0, increases in Δ, and because the equity payoff decreases in Δ, the option payoff accounts for a higher fraction of the agent's compensation as Δ increases. □

Proof of Lemma 7.3. Given state (C_g, M, s) and the specified coupon payment c_i in [7.4], we can obtain from [7.3] next period's credit balance contingent on firm violating covenant or not today as

$$M_b = (1+r)M - r\bar{v}_b + p_b(\tilde{w}_{bg} - \tilde{w}_{bb}), \quad M_g = \bar{v}_b - \tilde{w}_{gb}$$

So if the firm violates covenant tomorrow, the available credit (contingent on firm violating covenant or not today) is

$$\begin{aligned} \bar{v}_b - M_b &= (1+r)(\bar{v}_b - M) - p_b(\tilde{w}_{bg} - \tilde{w}_{bb}) = \tilde{w}_{bb} \\ \bar{v}_b - M_g &= \tilde{w}_{gb} \end{aligned}$$

The last equality in the first line is from rewriting [PK_b] as $(1+r)(\bar{v}_b - M) = \mathbb{E}^b[\tilde{\mathbf{w}}_b]$, because $v_b = \bar{v}_b - M$, $\delta = 1/(1+r)$, and $\tilde{m}_b = 0$. Moreover, if the firm complies with covenant tomorrow, by [7.5] the available credit (contingent on firm violating covenant or not today) is $C'_{ig} - M_i + R(\tilde{k}_i) = \tilde{w}_{ig}$. Hence, under the designed mechanism, the firm's available credit always matches the agent's contingent utility obtained in the contract. □

Proof of Theorem 4. Take any (C_g, M, s). Then [7.1] 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 she 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. She can divert the cash flow $R(\tilde{k})$, where $\tilde{k} = k(\mathbf{v}, s)$. If the agent also diverts the available credit, her total payoff is

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

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 7.3 her contingent utilities from next period onwards is $\tilde{\mathbf{w}}_b = \mathbf{w}_b(\mathbf{v}, s)$, which means her total payoff is $R(\tilde{k}) + \delta\,\mathbb{E}^g[\tilde{\mathbf{w}}_b] \le v_g$ by [IC].

Third, the agent has no incentive to divert credit. By doing so, the agent gets payoff $\bar{v}_b - M$ or $C_g - M + R(\tilde{k})$ depending on firm violating covenant or not. According to [7.1], these values are equal to v_b or v_g. But the agent obtains the same contingent payoffs by waiting to issue payout d_i until credit balance is fully repaid. This is by Proposition 7.2, v_i is equal to all the future payments that the agent gets from her security holdings. □

References

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

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

Battaglini, Marco. 2005. 'Long-Term Contracting with Markovian Consumers'. *American Economic Review* 3:637–658.

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.

Bergman, Nittai K., and Dirk Jenter. 2007. 'Employee sentiment and stock option compensation'. *Journal of Financial Economics* 84:667–712.

Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet. 2007. 'Dynamic security design: convergence to continuous time and asset pricing implications'. *Rev. Econom. Stud.* 74 (2): 345–390.

Bloedel, Alexander W., and R. Vijay Krishna. 2014. 'Misery, Persistence, and Growth'. *working paper*.

Bolton, Patrick, and David S. Scharfstein. 1990. 'A Theory of Predation Based on Agency Problems in Financial Contracting'. *American Economic Review* 80:93–106.

Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. 'A Theory of Financing Constraints and Firm Dynamics'. *Quarterly Journal of Economics* 121:229–265.

DeMarzo, Peter M., and Michael J. Fishman. 2007a. 'Agency and Optimal Investment Dynamics'. *Review of Financial Studies* 20:151–188.

– . 2007b. 'Optimal Long-Term Financial Contracting'. *Review of Financial Studies* 20:2079–2128.

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.

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.

Doepke, Matthias, and Robert M. Townsend. 2006. 'Dynamic mechanism design with hidden income and hidden actions'. *Journal of Economic Theory* 126:235–285.

Eső, Péter, and Balázs Szentes. 2013. 'Dynamic Contracting: An Irrelevance Result'. *working paper*.

Fernandes, Ana, and Christopher Phelan. 2000. 'A Recursive Formulation for Repeated Agency with History Dependence'. *Journal of Economic Theory* 91:223–247.

Gertler, Mark, and Simon Gilchrist. 1994. ‘Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms’. *Quarterly Journal of Economics* 109:309–340.

Gomes, Joao F. 2001. ‘Financing Investment’. *American Economic Review* 91 (5): 1263–1285.

Gompers, Paul, Anna Kovner, Josh Lerner and David S. Scharfstein. 2008. ‘Venture capital investment cycles: the impact of public markets’. *Journal of Financial Economics* 27:1–23.

Guo, Yingni, and Johannes Hörner. 2014. ‘Dynamic Mechanisms without Money’. *Working Paper*.

Halac, Marina, and Pierre Yared. 2014. ‘Fiscal rules and discretion under persistent shocks’. *Econometrica* 82 (5): 1557–1614. ISSN: 0012-9682.

Hurst, Erik, and Benjamin W. Pugsley. 2011. ‘What Do Small Business Do?’ *NBER Working Paper*.

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.

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.

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.

Larcker, David F. 2008. ‘Executive Compensation and Incentives’. *Stanford Graduate School of Business*.

Pavan, Alessandro, Ilya Segal and Juuso Toikka. 2014. ‘Dynamic Mechanism Design: A Myersonian Approach’. *Econometrica* 82 (2): 601–653.

Quadrini, Vincenzo. 2004. ‘Investment and Liquidation in Renegotiation-proof Contracts with Moral Hazard’. *Journal of Monetary Economics* 51:713–751.

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

Shiryaev, Albert N. 1995. *Probability*. Verlag, New York: Springer.

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

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

Tchistyi, Alexei. 2013. ‘Security Design with Correlated Hidden Cash Flows: The Optimal of Performance Pricing’. *working paper*.

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.

Zhang, Yuzhe. 2009. ‘Dynamic Contracting with Persistent Shocks’. *Journal of Economic Theory* 144:635–675.