Dynamic Financial Contracting with Persistent Private Information*

Shiming Fu†　　R. Vijay Krishna‡

April 2015

Abstract

We study a setting in which a principal provides an agent with funds to operate a firm. The agent is liquidity constrained and privately observes the revenue of the firm. In other words, the agent can divert revenue to projects that benefit only him. Time is discrete and infinite, and revenues are subject to stochastic shocks that are persistent over time. As in the iid case, financing constraints emerge endogenously as part of the optimal contract so that investment is inefficient, rents are backloaded, and the agent doesn't receive dividend payments until the firm matures. In the optimal contract, the firm matures with probability one, from which point on, investment is always efficient. However, in contrast to the iid case, persistence implies that investment in a mature firm is not constant, the debt-to-equity ratio of a mature firm is not constant, compensation in a mature firm is via a combination of cash and equity, and, most importantly, the agent is not the residual claimant in a mature firm, predictions which are consistent with empirical evidence.

(∗) We thank Adriano Rampini, Philipp Sadowski, S. Vish Viswanathan, and Neil Wallace for helpful comments. Krishna is grateful to the National Science Foundation for financial support through grant SES-1132193.
(†) Duke University <shiming.fu@duke.edu>
(‡) Duke University <rvk3570@gmail.com>

Contents

1. Introduction

Firms, especially young firms, typically consist of two parties, one of whom has the expertise to operate the firm but not the funds, and another who has the funds, but not necessarily the time or expertise. In such a situation, it is the latter party, the principal, who provides the financial resources necessary to launch the firm and keep it running, while the former party, the agent, actively manages the firm. However, there is considerable evidence[1] that funding for young firms in particular is far from efficient, that firms face financing constraints, and that a firm must, over time, grow into its optimal size. In particular, financing constraints greatly affect firm size, firm growth, and more generally, a young firm's prospects.

Agency problems are one, and quite possibly, a significant source of the aforementioned financing constraints. There are many models that describe how agency problems in an infinite horizon setting affect firm size and other aspects of the firm such as the evolution of its debt and equity. A quintessential member of this class is Clementi and Hopenhayn (2006)[2] which features an infinitely repeated interaction between a borrower (the agent) and a lender (the principal). The project in question yields cash returns that increase in the amount of capital invested, although returns are subject to iid shocks. Both the borrower and lender are risk neutral, but it is only the borrower who can observe cash flows, ie, it is only the agent who can observe the shocks to the returns. In particular, the borrower can divert cash for his own consumption.[3] This is the source of the agency problem. Notice that with complete information, investment in each period is efficient. Thus, it is precisely the combination of the agency problem and the liquidity constraints that lead to the particular features of the optimal contract. In this setting, firm size corresponds to the capital invested, firm value is the expected discounted cash flows, equity is merely the agent's share of firm value, and debt is therefore the lender's share of firm value.

In such an environment, Clementi and Hopenhayn (2006) show that the optimal contract exhibits history dependence, backloads all rents (payments) to the agent, has investments that are positively correlated over time, and increases the agent's equity with high cash flows and reduces it with low ones. Their model also predicts that if the project is not scrapped (which happens if the agent gets sufficiently many bad shocks), then the firm eventually matures and stays in

(1) See, for instance, Gomes (2001) and Gertler and Gilchrist (1994) who note that financing constraints are a significant factor in firm behaviour.

(2) Other works that also consider dynamic principal-agent problems in the context of investment in firms are DeMarzo and Fishman (2007) and Biais et al. (2007). These are reviewed in Section 2 below. While there are some minor differences between these works and Clementi and Hopenhayn (2006), the main results, and indeed, the incentives driving the results in all these papers are the same. A synthesis of the main ideas in an iid setting, in both discrete as well as continuous time, is presented in Biais, Mariotti and Rochet (2013).

(3) This is isomorphic to a moral hazard model where the agent gets a private benefit from shirking, and where the principal always wants the agent to exert effort.

operation forever. It is remarkable that such a parsimonious model is able to generate such a rich set of predictions that match the data and stylised facts about firms. However, the model makes some other predictions that are harder to account for. For instance, the iid model predicts (and this is true for essentially all iid models, some of which are described below in Section 2) that (below, the 'I' stands for 'iid')

(I-i) A young firm's initial level of capital and equity are dependent only on the perceived prospects of the firm, which are constant over time.

(I-ii) If the firm *matures*, its size remains constant at the efficient level, ie, investment eventually becomes constant over time.

(I-iii) When the firm matures, the agent becomes the residual claimant of the firm — in the terminology of Jensen and Meckling (1976), the agent is the *owner* of the firm because he retains all cash flows.

(I-iv) The agent's compensation scheme is a high-water mark contract — he gets dividends only if revenues are high.

(I-v) Mature firms don't compensate their employees via the issuance of equity, ie, the equity levels of the agent in a mature firm remain constant over time.

(I-vi) A mature firm's debt-to-equity ratio, and hence the firm's β, which are central to valuations of the firm, remain constant over time.

It is important to note that these are *joint* predictions, ie, (I-i)–(I-vi) are predicted to hold simultaneously. None of these predictions are observed in the data, certainly not simultaneously, nor do they resemble stylised facts about firms.

In this paper, we augment the model of Clementi and Hopenhayn (2006) by allowing the shocks to revenues to be positively correlated over time.[4] We then show that in such an environment, where revenues are subject to persistent shocks, the optimal contract, just as it did in the iid case, exhibits history dependence, backloads all rents (payments) to the agent, and has investments that are positively correlated over time (even more so than in the iid case), but most importantly, does not predict (I-i)–(I-vi).

We show that when the production shocks are persistent, the firm eventually matures, just as in the iid case. However, with persistence in the shocks to revenues, we find that (below, the 'P' stands for 'persistent')

(4) Gomes (2001) shows that the assumption that shocks to revenue is positively correlated over time is consistent with the data.

(P-i) Initial levels of debt for young firms are greater if the perceived likelihood of success is high, and this perception varies over time.

(P-ii) The firm eventually matures, though it can still vary in size over time.

(P-iii) The agent is not the residual claimant of the mature firm: For low levels of persistence, the firm makes a state independent debt payment to the principal, and the agent keeps the remainder of the revenue, while for high levels of persistence, the principal keeps all the revenue if investment is low, and gets paid some of the revenue if investment is high.

(P-iv) The agent's compensation scheme in a mature firm is a high-water mark only if persistence is sufficiently high or in the (non-generic) iid case.

(P-v) In a mature firm, the agent's equity position is adjusted upwards or downwards depending on whether revenues are high or low.

(P-vi) After the firm matures, its valuation, its debt-to-equity ratio (and hence its β) vary over time.

We show that all these features emerge as part of the optimal contract. Moreover, all these predictions are in accord with the data and with stylised facts about firms, as we discuss in Section 8. Our predictions about firm financing also, therefore, contradict the 'pecking order' theory of Myers (1984) and Myers and Majluf (1984), which states that only distressed firms should issue equity.[5]

The predictions that the optimal contract exhibits history dependence, backloads all rents (payments) to the agent, and has investments that are positively correlated over time regardless of the degree of persistence suggest that these are robust features of contracts. History dependence, comes about because in any dynamic interaction, using future behaviour as a (screening) instrument allows the principal to achieve more than she can in a static setting. Thus, the optimal contract will have 'memory'.

Rents are backloaded because both the principal and agents are risk neutral,[6] which means that the agent is indifferent between receiving cash now or in the future. The principal, in turn, prefers to use equity as a screening device instead of cash (at least initially) because equity is something that she can adjust in the future, as it stays within the relationship, while cash is immediately spent by the agent, and is hence out of the principal's control. Finally, positive correlation

(5) See Fama and French (2005) who argue that the data contradicts the pecking order theory of firm financing and capital structures.

(6) Kaplan and Strömberg (2003) show that risk sharing is not a significant concern in venture capital funding, so the assumption of risk neutrality is a reasonable one.

of investments occurs because after a success, the agent's equity levels go up, which relaxes incentive or financing constraints, thereby allowing the principal to increase her investment. Needless to say, the degree of persistence in investment depends on the degree of persistence in the technological shock process.

The main differences between the model with persistence and the iid model are that with persistence (i) dynamic information rents need to be paid to the agent, and (ii) optimal investment, ie, first-best or full information investment, varies over time, because investment levels are determined by the principal's belief about the likelihood of a good shock, which in turn depends on the previous period's realised shock. Therefore, with privately observed shocks, rents must vary over time even if investment is efficient.

The properties of the optimal contract depend on the degree of persistence (and this also true for a mature firm). To paraphrase Tolstoy, all models with iid shocks are alike, but every Markovian model with persistence is Markovian in its own way. Indeed, the setting with high persistence is markedly different from the setting with low persistence. With high persistence, there will be greater positive correlation in investment levels, and the optimal compensation scheme (as before, the agent is paid once the firm matures) is a high-water mark contract. Our model therefore predicts that high-water mark contracts are optimal only when there is sufficient persistence in the private information (technological shock) process. Moreover, we highlight a hitherto unremarked advantage of high-water mark contracts: they allow for adjustments to the agent's equity without explicitly having to concentrate or dilute the agent's equity holdings. In other words, high-water mark contracts provide the principal with a particularly simple mechanism to adjust (implicit) equity holdings and pay dividends.[7] This is consistent with our prediction that compensation in a mature firm involves equity adjustments for all (positive) levels of persistence.

Unlike the iid case, the agent's expected continuation equity is no longer sufficient as a state variable for the optimal contract. The Markovian nature of shocks implies that the agent has private information about how he ranks streams of continuation utilities which is the source of dynamic information rents. Following Fernandes and Phelan (2000), we formulate our problem recursively,

(7) This suggests that empirical studies of the issuance of equity, such as Fama and French (2005), underestimate the issuance of equity because they only look for explicit issuance, and ignore mechanisms such as high-water mark contracts that implicitly adjust equity levels.

using a vector of contingent equities as the state variable, where contingent equity is the agent's lifetime expected equity contingent on the present period's revenue being good or bad.[8] Roughly put, our approach makes it somewhat easier to specify the domain for the firm's dynamic programming problem. Moreover, we are also able to uncover a natural martingale which guides us through the dynamics of the contract.

In the next section we review the relevant literature. The model is formally introduced in Section 3. Section 4 discusses strategies and optimal sequential contracts. We then introduce in Section 4.2 our approach using contingent equity for dealing with the problem of persistent private information. A recursive approach requires a suitable domain, a somewhat non-trivial exercise here, which is addressed in Section 4.3, while Section 4.4 describes the firm's value as the solution to a Bellman equation. Section 5 describes the optimal contract for the young firm as it pertains to compensation, capital advancement, and the evolution of equity, while Section 7 looks at the optimal contract for a mature firm. Section 6 shows that every firm will eventually mature (we ignore the possibility that the firm will be scrapped), while Section 8 discusses empirical implications and stylised facts. Section 9 discusses some extensions, such as the possibility of scrapping the firm, the nature of path dependence in the optimal contract, and the impact, for a mature firm, of persistence, on debt, equity, and firm value. Section 10 concludes, and all proofs are in the appendices.

2. Related 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.[9] 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 (2007), where the latter two emphasise the implementation of the optimal contract via simple contracts. Another noteworthy accomplishment of

(8) As we discuss below, our choice of state variables is subtly different from Fernandes and Phelan (2000). Details are in 4.2, but also see footnotes 12 and 15.
(9) Such models are therefore referred to as *cash flow diversion* (CFD) models.

Biais et al. (2007) is that they take a continuous time limit of a discrete time model, which allows them to derive some additional properties of the optimal contract, and coincides with the continuous time cash-flow diversion model of DeMarzo and Sannikov (2006). A major difference between the present paper and Biais et al. (2007) and DeMarzo and Fishman (2007) is that these papers don't allow for investment in capital while we do. On the other hand, capital in these models does not depreciate, while capital in our model depreciates perfectly at the end of every period.

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. (Another difference, and this is purely for simplicity, is that unlike Clementi and Hopenhayn (2006), our base model doesn't allow for the project to be scrapped.) Formally, our model is one of dynamic screening in agency contracts.

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 emphasised by Fernandes and Phelan (2000), who note that promised utility alone is inadequate in the Markovian case, and introduce the idea of a *threat-point* utility as an additional state variable. While Fernandes and Phelan (2000) describe the formulation of the problem, the generality of their environment precludes a clean analysis of the optimal contract. It is worth emphasising that although our approach, which also uses a vector of promised utilities as a state variable, is subtly different from that of Fernandes and Phelan (2000). Their vector of promised utilities is from an *ex ante* perspective, and consists of the ex ante promised utility from truth-telling, and the ex ante threatened utility from lying. (The Markovian nature of types means that these are the only cases that need be considered.) Our vector of promised utilities is *interim*, contingent on the production shock in the period. While the differences are minor, our state variable is easier to interpret in the present setting, as the agent's contingent equity.

Battaglini (2005) considers the problem of a seller (the principal) selling some quantity of a good to a consumer, whose valuation for the good follows a two-point Markov process. He derives a number of results, most of which don't have counterparts here. For instance, he shows that the high type is always treated efficiently, while the distortion in the quantity for the low type declines monotonically over time (along all histories). This is not true for the corresponding variable in our model, namely investment. The key difference between our model and that of Battaglini (2005) is that our agent has no cash and is protected by limited liability, while the agent in Battaglini (2005) has unlimited cash reserves, and only has a participation constraint that must be respected at each point in time. Indeed, it is the lack of liquidity constraints that drive his results. In particular, in Battaglini (2005), if the agent has a low valuation at any point in time, he ends up paying his entire future expected surplus to the principal, which in turn ends up reducing the problem to one with a single state variable. It is precisely the inability of the principal to extract future information rents that makes our model economically interesting (even without Markovian types).

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. To study the question of optimal fiscal policy, Halac and Yared (2014) also use a vector of promised utilities as in Fernandes and Phelan (2000). 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. Some of the technical aspects of that paper are different because they allow for unbounded utilities but not monetary transfers, and because risk aversion implies that rents cannot be backloaded. Independently, Guo and Hörner (2014) use the same techniques to study mechanism design without monetary transfers. Their environment is different from Bloedel and Krishna (2014) in that they consider bounded utilities and bounded consumption, which substantially alters the long-run behaviour of the contract.

3. 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 the common rate $\delta \in (0, 1)$.

The project, which we shall also refer to as the firm, requires an investment $k_t \geq 0$ in every period. Capital depreciates completely, and so cannot be carried over to subsequent periods. The return on capital is random, and is either $R(k)$ or 0, where $R : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing, strictly concave, and continuously differentiable function with $R(0) = 0$, $\lim_{k\downarrow 0} R'(k) = \infty$, and $\lim_{k\uparrow\infty} R'(k) = 0$. We shall say a return of $R(k)$ occurs if the random production shock is ***g***ood, while a return of 0 occurs the random production shock is ***b***ad. The probability of having a good shock today is p_s where $s \in \{b, g\} =: S$ was the shock in the last period. The production shocks follow a Markov process with transition probabilities

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

where the probability of a good shock in the next period given a bad shock today is p_b, and the probability of a good shock in the next period given a good shock today is p_g. We shall assume that $\Delta := p_g - p_b \geq 0$, ie, the Markov process is *persistent*. The case where $\Delta = 0$ corresponds to the iid case. We also assume that $p_b, p_g \in (0, 1)$, which ensures that the Markov process has a unique ergodic measure and has neither absorbing nor transient sets.

The agency problem arises because (i) the principal cannot observe the output while the agent can, and (ii) the agent is cash constrained. The agent's lack of funds implies that he needs, as mentioned above, working capital from the principal. We assume, therefore, that the principal cannot extract payments in excess of revenues from the agent, so the agent 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 '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* (formally defined below) 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 savings are 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 he may accept or reject. If he rejects the offer, the principal and agent go their separate ways, and their interaction ends. If the contract is accepted, it is executed. The agent can leave at any time to an outside option worth 0 without further penalty. The principal fully commits to the contract. As mentioned before, in terms of the evolution of the state, 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. There is one other, minor, difference. Clementi and Hopenhayn (2006) allow for the project to be scrapped at any time for a value of S, divided between the principal and the agent according to some formula that is history dependent and optimally chosen. For simplicity, we set the scrap value to zero. Our principal results go through in the case of a positive scrap value, albeit with some straightforward modifications. In particular, the properties of the mature firm are independent of the existence or level of a scrap value.

4. Contracts

A 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 production as being *good* or *bad* (which corresponds to positive and zero output respectively), so that a sequence of reports now constitutes a history. The set of states is $\{b, g\} =: S$, so that a *private* history is a sequence of output states $s^t := (s_1, \ldots, s_t) \in S_t$ that is only observed by the agent. A *reporting strategy* for the agent is a function $\tilde{s}_t : S_{t-1} \times S \to S$, ie, $\tilde{s}_t(h^{t-1}, s) \in \{b, g\}$, where $h^{t-1} := (\tilde{s}_1, \ldots, \tilde{s}_{t-1})$ is a *public* history of reports by the agent. Such a history represents public information available at the beginning of period t. Let H_{t-1} denote the collection of all such period-t public histories. We are now in a position to describe contracts.

4.1. Sequential Contracts

A *sequential* contract is a collection of functions $k_t : h^{t-1} \to \mathbb{R}_+$ and $m_t : h^{t-1} \times \{b, g\} \to \mathbb{R}$ for $t = 1, 2, \ldots$. Here, $k_t(h^{t-1})$ specifies the investment in period t conditional on the public history at the end of period $t-1$, while $m_t(h^{t-1}, \tilde{s}_t)$ specifies the transfer of cash from the agent to the principal in period t conditional on the (reported) output state $\tilde{s}_t$ in period t and the public history of reports h^{t-1}. The net cash flow for the agent at time t is $R(k_t(h^{t-1})) - m_t(h^{t-1}, \tilde{s}_t)$. As noted before, the agent cannot save any of this cash. A sequential contract is *feasible* if for all t, $m_t(h^{t-1}, g) \leq R(k_t(h^{t-1}))$ and $m_t(h^{t-1}, b) \leq 0$. In other words, a sequential contract is feasible if it respects the agent's limited liability constraints.

Given a contract (k_t, m_t), the expected utility (in terms of expected discounted cash flows) for the agent from a reporting strategy $(\tilde{s}_t)$ is $V_a((s_t), (\tilde{s}_t), h^0)$, where (s_t) is a sequence of output states observed only by the agent. The contract (k_t, m_t) is *incentive compatible* if truth-telling is an optimal reporting strategy, ie, if $V_a((s_t), (s_t), h^0) \geq V_a((s_t), (\tilde{s}_t), h^0)$ for all alternative reporting strategies $(\tilde{s}_t)$.

Thus, the goal of the principal is to maximise her utility — consisting of expected discounted cash flows — by choosing a sequential contract subject to the contract being feasible and incentive compatible. Unfortunately, working in the space of all sequential contracts is difficult, to say the least. We now show that the principal's problem has a recursive formulation that can be fruitfully employed.

4.2. 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 of iid states.[10] Consider a history h^{t-1} before the beginning of period t and a report $\tilde{s}$ in period t. Let $w_{\tilde{s}}$ be the agent's lifetime continuation utility upon a report of $\tilde{s} \in S$, so that the agent enters the next period expecting a lifetime utility of $w_{\tilde{s}}$. Notice that because states are iid, the agent's preferences over continuation problems are common knowledge — and in particular, are independent of the true state in period t — which implies that by choosing $w_{\tilde{s}}$ suitably, the principal can incentivise the agent to report truthfully.

Consider now the equivalence class of all histories h^{τ} such that prior to (observing and) reporting the state in period $\tau + 1$, the agent's *ex ante* expected utility is v. Since the agent's expected utility beginning with these histories is constant, any optimal contract contract will also deliver the principal the same expected utility conditional on these histories. Therefore, we may restrict attention to contracts that are constant on any such equivalence class. But this implies that we can let promised utility (prior to entering a period) be a state variable in a recursive problem, and then use standard dynamic programming techniques to derive the optimal contract. Put differently, because states are iid, continuation promised utility is sufficient as an instrument to screen the agent.[11]

However, such an approach is inadequate in the Markovian case precisely because the agent has private information about his preferences over future streams of cash. (Recall that today's state dictates the probability distribution over tomorrow's states, and today's state is only known to the agent.) If the agent has lied in period $t - 1$, and if the principal has promised utility v (in the form of future cash flows), the agent will assess a different expected utility from the stream of future cash flows than v, and more importantly, the principal does not have enough instruments to successfully screen the agent according to his information.

To screen the agent, the principal needs more instruments. Consider some

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

(11) In a two-period model with iid states, Bolton and Scharfstein (1990) use the threat of termination as a proxy for continuation utility to provide incentives for truthful revelation in the first period. Needless to say, if the contract is not terminated in the first period, there cannot be truthful revelation in the second period.

history h^{t-1}, and suppose the agent reports state $\tilde{s} \in S$ in period t. Rather than give the agent some expected continuation utility, the principal provides a pair of *interim* or *contingent* utilities beginning in the next period, that are conditional on the state in the next period. Such a vector of utilities is $\mathbf{w}_{\tilde{s}} := (w_{\tilde{s}b}, w_{\tilde{s}g})$. If the true state in period t is s, the agent's expected utility from such a pair of contingent utilities (obtained by reporting $\tilde{s}$) is $(1 - p_s)w_{\tilde{s}b} + p_s w_{\tilde{s}g}$. Thus, in spite of preferences over continuation problems not being common knowledge, by using contingent continuation utilities appropriately, the principal can provide the agent with the right incentives so as to induce truth-telling.

Thus, after a history h^t, the agent enters period $t + 1$ being promised a pair $\mathbf{v} = (v_b, v_g)$ of contingent utilities. Let us consider the equivalence class of all histories such that after any history in this class, the vector (v_b, v_g) of continuation expected utilities are identical. On this equivalence class, the principal's expected utility must again be constant, and so we may take the vector (v_b, v_g) to be our state variable, along with the previous period's report. Notice that even if the agent has lied in the last period, we are now able to write down incentive constraints in a meaningful way.

There is one significant difference here from the iid case that needs comment. Suppose the agent enters the period with promised contingent utilities $\mathbf{v} = (v_b, v_g)$. If states were iid, his expected utility from this pair is independent of his reports in the past, and in particular, does not depend on whether he lied in the last period. However, in the Markovian case, his expected utility from this pair depends on his belief about the probabilities of the good and bad state today, which in turn depends on yesterday's state, which he may not have reported truthfully. But, and this is crucial, even if the agent lied yesterday, contingent on today's shock being s, his lifetime interim utility is still v_s. This is because past information is now rendered payoff irrelevant (which follows directly from the assumption that states follow a Markov process). Thus, our formulation ensures that the agent cannot benefit from double deviations.

It goes without saying that the equivalence of the proposed recursive formulation, with what we refer to as *contingent* promised utilities, and the sequential contract needs proof. However, the proof is very similar to the proof offered in Theorem 2.1 of Fernandes and Phelan (2000) and so is omitted.[12]

(12) 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

Given a pair of contingent utilities $\mathbf{v} = (v_b, v_g) \in \mathbb{R}^2$ with the last period being in state s (at least as far as the principal believes), the principal chooses a capital advancement policy $k(\mathbf{v}, s) \in \mathbb{R}$, transfers $m(\mathbf{v}, s, \tilde{s}) \in \mathbb{R}$, 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 the following promise keeping constraints:

$$[\mathrm{PK}_b] \qquad v_b = -m_b + \delta[(1-p_b)w_{bb} + p_b w_{bg}]$$

$$[\mathrm{PK}_g] \qquad v_g = R(k) - m_g + \delta[(1-p_g)w_{gb} + p_g w_{gg}]$$

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

$$\begin{aligned} & R(k) - m_g + \delta[(1-p_g)w_{gb} + p_g w_{gg}] \\ [\mathrm{IC}] \qquad & \geq R(k) - m_b + \delta[(1-p_g)w_{bb} + p_g w_{bg}] \end{aligned}$$

The limited liability constraints are

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

Throughout we impose the feasibility constraint that $k \geq 0$ without comment. Using the promise keeping constraints [PK_b] and [PK_g], the incentive constraint [IC] can be written somewhat more simply as

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

On the right hand side of the constraint [IC*], $R(k)$ is the *static* information rent while $\Delta(w_{bg} - w_{bb})$ is the *dynamic* information rent which is 0 in the iid case, ie, if $\Delta = 0$. Thus, the constraint [IC*] crystallises the effect of Markovian states. If production shocks are iid, $\Delta = 0$ and [IC*] reduces to $v_g - v_b \geq R(k)$. As we shall see below, 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.

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 domain for the principal's problem.

utility from cash streams if he has lied in the last period. Notice that both the promised and threat-point utilities are ex ante utilities, while our contingent utilities are interim in nature. Apart from this difference, the two approaches are essentially identical. Nevertheless, we shall see below that contingent utilities are somewhat easier to interpret, and so render themselves more suitable for the application considered in this paper; see also footnote 15.

4.3. A 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 realised must also be non-negative. But our other constraints impose even more restrictions on the feasible (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. To see this, suppose $\mathbf{v} = (v_b, 0)$, where $v_b > 0$. Then, [PK$_g$] requires that

$$0 = R(k) - m_g + \delta\big[(1 - p_g)w_{gb} + p_g w_{gg}\big]$$

By [LL], we know that $R(k) - m_g \geq 0$, and by assumption, $\mathbf{w}_g \in \mathbb{R}^2_+$, which implies $(1 - p_g)w_{gb} + p_g w_{gg} \geq 0$. Therefore, it must be that $R(k) = m_g$, and $\mathbf{w}_g = (0, 0)$. Now notice that by [IC], we obtain

$$0 \geq R(k) - m_b + \delta\big[(1 - p_g)w_{bb} + p_g w_{bg}\big]$$

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 $0 \geq R(k) - m_b + \delta\big[(1 - p_g)w_{bb} + p_g w_{bg}\big] \geq 0$, ie, $R(k) = m_b = k = 0$ and $\mathbf{w}_b = (0, 0)$. Therefore, by [PK$_b$], we must have $v_b = -m_b + \delta[(1 - p_b)w_{bb} + p_b w_{bg}] = 0$. But this contradicts our assumption that $v_b > 0$. Thus, $(v_b, 0)$ with $v_b > 0$ is not implementable, or equivalently, is infeasible.

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),

(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.

non-trivial set, as described next, that will serve as the domain for our recursive formulation of the principal's problem.[14]

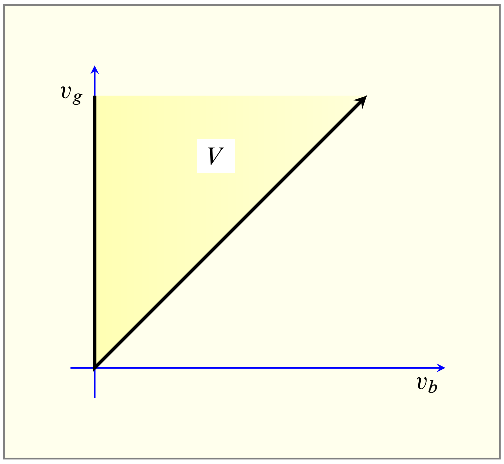

Figure 1: *Recursive Domain V*

Proposition 4.1. 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 1).

In what follows, for each $(\mathbf{v}, s) \in V \times S$, let

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

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 $\Gamma(\mathbf{v})$.

4.4. Optimal Contracts

An optimal contract is a solution to the principal's problem. However, instead of solving the principal's maximisation problem, we note that transfers are linear, and so don't affect social surplus. Therefore, we shall consider recursive the

(14) Proposition 4.1 is an analogue of Lemma 2.2 in Fernandes and Phelan (2000). In the terminology of Abreu, Pearce and Stacchetti (1990), V is *self-generating*. Indeed, the proof of Proposition 5.1 consists of showing that V is the (largest) fixed point of an appropriate mapping.

problem of maximising firm value (which in this setting is precisely the social surplus), and use the agent's contingent utilities (and the previous period's state) as state variables for our dynamic program. Towards this end, we let $Q(\mathbf{v}, s)$ denote the value of the firm when the previous period's shock was s, and when the agent enters the period with contingent utility $\mathbf{v}$.

The vector $\mathbf{v} \in V$ is the agent's contingent utility. We shall interpret $\mathbf{v}$ as his *contingent equity*; *contingent* because v_b, for instance, is his equity contingent on today's shock being bad, and *equity* because v_s is the present discounted expected cash flow that will accrue to the agent, contingent on today's shock being $s \in S$.[15]

Because $Q(\mathbf{v}, s)$ is the value of the firm and $\mathbf{v}$ is the agent's contingent equity, the principal's utility is $Q(\mathbf{v}, s) - (1 - p_s)v_b - p_s v_g$, which therefore represents the *debt* that the firm carries. We interpret the principal's utility as debt because it is a senior claim on the cash flow relative to the agent, whose equity in the firm is a junior claim on the cash flow.

In what follows, we shall denote $\partial Q/\partial v_b$ by Q_b and $\partial Q/\partial v_g$ by Q_g. An *optimal* contract is a solution to the firm's recursive maximisation problem. We shall denote an optimal contract by $(k, m, \mathbf{w})$, with the understanding that k depends exclusively on $(\mathbf{v}, s)$, while m and $\mathbf{w}$ depend on $(\mathbf{v}, s)$ as well as the current period's state. Our first result establishes the existence of the firm's value function, as well as some of its properties. It also shows that an optimal contract exists by virtue of being the policy function for the firm's value function.

Theorem 1. *The firm's discounted value under an optimal contract* $(k, m_i, \mathbf{w}_i)$ *is given by a unique, concave, and continuously differentiable function* $Q : V \times S \to \mathbb{R}$ *that satisfies*

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

subject to $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$. *The contract* $(k, m, \mathbf{w})$ *is continuous in* $(\mathbf{v}, s)$. *Moreover, the following are true:*

(a) $Q(\mathbf{0}, s) = 0$, $Q_g((v, v), s) = \infty$ *and* $Q_b((0, v), s) = \infty$ *for all* $v \geq 0$.
(b) *There exists* $M > 0$ *such that* $0 \leq Q(\mathbf{v}, s) \leq M$ *for all* $(\mathbf{v}, s)$.

(15) Thus, in our setting, it is natural to interpret the interim contingent promised utilities as contingent equity. While one may take an ex ante view of promised equity (as in the iid setting), it is somewhat less clear how one should interpret ex ante threat point utility in terms of equity.

(c) $Q_g(\mathbf{v}, s) \geq 0$ *for all* $(\mathbf{v}, s) \in V \times S$*, though* $Q_b(\mathbf{v}, s)$ *is sometimes negative.*
(d) $Q((v_b, v_g), s)$ *is increasing in* v_g *and* $Q(\cdot, g) \geq Q(\cdot, b)$.
(e) $Q(\mathbf{w}_g(\mathbf{v}, s), g) \geq Q(\mathbf{w}_b(\mathbf{v}, s), b)$.
(f) $Q(\cdot, s)$ *is supermodular in* $\mathbf{v}$ *for all* $s \in S$.

The existence, uniqueness, concavity and differentiability properties of the surplus function Q are standard, as is the continuity of the policy function. By never investing and immediately paying the agent his expected promised utilities, which effectively shuts down the firm, the principal can always reduce the value of the firm to 0, which is then a lower bound on the value of the firm. If the principal could operate the firm herself or equivalently, if she could observe cash flows, she would invest the efficient amount $\bar{k}_s$ (which solves $p_s R'(\bar{k}_s) = 1$) in each period, and retain all cash flows. The resulting expected (discounted) cash flows clearly provide an upper bound to the value of the firm.

It is easy to see that $\Gamma(\mathbf{0}, s) = \{(k = 0, m_i = 0, \mathbf{w}_i = \mathbf{0}) : i = b, g\}$. In other words, with promised contingent utilities of $(0, 0)$, the principal can neither invest in the current period nor promise utilities in the future. Therefore, $Q(\mathbf{0}, s) = 0$.

The values of the partial derivatives deserve comment, because they arise fundamentally from the structure of the incentive constraints, and because we are working with contingent promised utilities as opposed to (ex ante) promised utilities which arise in the iid case. Part (a) of Theorem 1 says that $Q_g(\mathbf{0}, s) = \infty$. To understand this, notice that if we consider contingent promises of $(0, \varepsilon)$, the principal can ensure production of $R(k) = \varepsilon$ by promising a move to the state $\big(\varepsilon/(\delta p_g)(1, 1), g\big)$ in the event of a success, and $\big((0, 0), b\big)$ in the event of a failure. Intuitively, this allows the principal to increase production by a small amount in every period with positive probability. The proof shows that the marginal value of this increase to the firm is actually infinite, and relies crucially on the assumption that $R'(0) = \infty$.

Similarly, consider the claim in part (a) that $Q_b((0, v), s) = \infty$. To understand this, notice that in the event of zero output, [PK_b] requires that $\mathbf{w}_b = \mathbf{0}$, so the state in the following period is $(\mathbf{0}, b)$. But the argument above now shows that because $Q_g(\mathbf{0}, g) = \infty$, we must also have $Q_b((0, v), s) = \infty$. The other partial derivatives are established in a similar fashion.

The main observations with regards to contingent (promised) utilities are the following: First, part (c) says that increasing v_g is always beneficial to firm value. Second, part (d) states that an increase in the contingent utility v_b does not always increase firm value. If v_b is very low relative to v_g, there is little effect on [IC*], but we can raise $\mathbf{w}_b$ as dictated by [PK_b]. If v_b is sufficiently high, then the primary impact of raising v_b is on tightening [IC*]. Thus, increasing v_b may be beneficial, but it can also reduce the value of the firm because it constrains feasible levels of output. Third, as part (e) notes, success in the present period increases firm value in the next period. And fourth, in part (f) which says that for each s, $Q(\mathbf{v}, s)$ is supermodular in $\mathbf{v}$, ie, v_b and v_g 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 in the next period. On the other hand, increasing v_b tightens the incentive constraint [IC*]. Intuitively, this second effect is not less pronounced when v_g is higher, so $Q(\mathbf{v}, s)$ is supermodular in $\mathbf{v}$.

As noted above, an optimal contract is a solution to the firm's recursive maximisation problem [VF]. Theorem 1 says that an optimal contract $(k, m, \mathbf{w})$ exists and that it is continuous in $(\mathbf{v}, s)$, but says little more. In the next section, we shall study in greater detail the structure of the optimal contract.

Before analysing the structure of the contract, it is worthwhile to consider the set of contingent utilities that ensure perpetual efficient production. Intuitively, there exist *threshold* levels of contingent equity $\bar{\mathbf{v}}^s$ so that once the agent reaches these threshold levels, private information no longer matters.

The following proposition describes sets of contingent equity that are above the thresholds $\bar{\mathbf{v}}^s$. Before doing so, let us consider the first best (efficient) value of the firm. This is precisely the case where there are no agency problems and the principal operates the firm. Then, the efficient surplus level in state s is $\bar{Q}(s)$, $s = b, g$, where

$$\bar{Q}(b) = -\bar{k}_b + p_b\big[R(\bar{k}_b) + \delta\bar{Q}(g)\big] + (1 - p_b)\delta\bar{Q}(b) \tag{4.1}$$

$$\bar{Q}(g) = -\bar{k}_g + p_g\big[R(\bar{k}_g) + \delta\bar{Q}(g)\big] + (1 - p_g)\delta\bar{Q}(b) \tag{4.2}$$

and where $\bar{k}_s$ solves $p_s R'(\bar{k}_s) = 1$. These two equations allow us to explicitly calculate $\bar{Q}(b)$ and $\bar{Q}(g)$. What is relevant for us is that $\bar{Q}(s)$ represents an upper bound for the value of the firm in state s, and that it entails efficient production.

Proposition 4.2. For each $s = b, g$, there exist threshold levels of equity $\bar{\mathbf{v}}^s = (\bar{v}_b^s, \bar{v}_g^s) \in V$ where $\bar{\mathbf{v}}^b \leq \bar{\mathbf{v}}^g$, although $\bar{v}_b^b = \bar{v}_b^g$, and closed sets $E_s \subset V$, called the

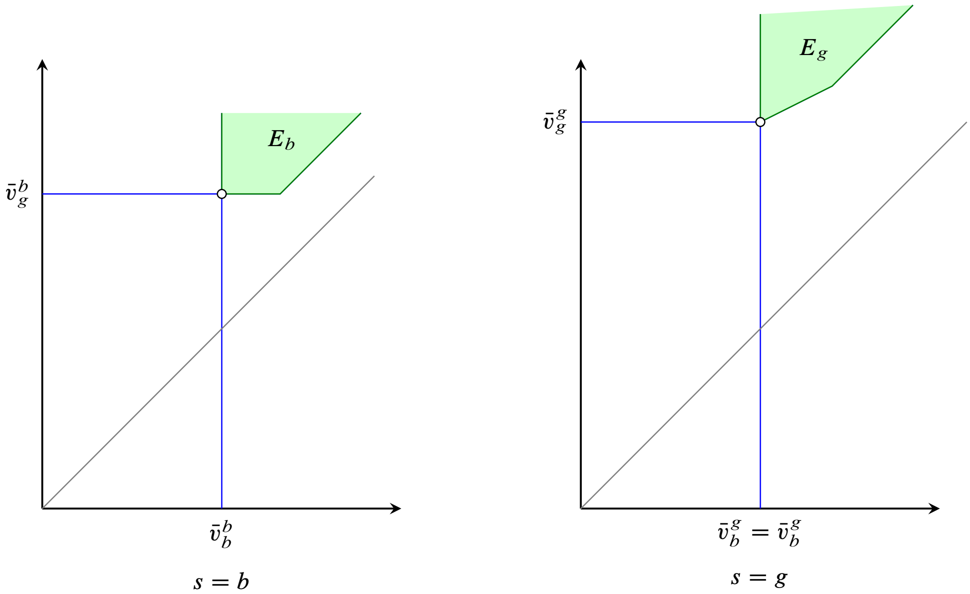

Figure 2: *Efficient Sets*

efficient sets, defined as

$$E_b := \{\mathbf{v} \geq \bar{\mathbf{v}}^b : v_g - v_b \geq R(\bar{k}_b)/(1-\delta\Delta)\}$$

$$E_g := \left\{\mathbf{v} \geq \bar{\mathbf{v}}^g : v_g - v_b \geq R(\bar{k}_g) + \delta\Delta \max\left[\frac{\delta\bar{v}^g_b - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right]\right\}$$

with the following properties:

(a) For each $\mathbf{v} \in E_s$, $Q(\mathbf{v}, s) = \bar{Q}(s)$, and $k(\mathbf{v}, s) = \bar{k}_s$.
(b) For each $\mathbf{v} \in E_s$, $Q_b(\mathbf{v}, s) = Q_g(\mathbf{v}, s) = 0$.
(c) For any $\mathbf{v} \in V \setminus E_s$, $Q(\mathbf{v}, s) < \bar{Q}(s)$.

Intuitively, $\bar{\mathbf{v}}^s$ represents the lowest levels of contingent equity that the agent must have in order to obtain *perpetual* efficient production in state s. In other words, $\bar{\mathbf{v}}^s$ is the smallest level of contingent equity 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 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 7.1 in Section 7 below explicitly describes $\bar{\mathbf{v}}^s$, which in turn gives us the sets E_s explicitly. For now, what is important is the observation that we get efficient investment and firm value is thereby maximised if (i) levels of contingent equity in state s are sufficiently high, reflected in the requirement that $\mathbf{v} \geq \bar{\mathbf{v}}^s$, and (ii) v_g is sufficiently greater than v_b. It is clear that (i) is a necessary property, because if contingent equity is too low, then by [IC*], production cannot be efficient. 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 production.

Parts (b) and (c) of Proposition 4.2 say that firm value is maximised precisely on the sets E_s. Part (a) establishes that the threshold equity levels satisfy $\bar{\mathbf{v}}^b \leq \bar{\mathbf{v}}^g$, even though $\bar{v}_b^b = \bar{v}_b^g$. Thus, the threshold equity contingent on being in a bad state today is the same, regardless of yesterday's state. Having described threshold levels of contingent equity, we now turn our attention to the properties of the optimal contract in the early stages of the contract, when the firm is young, ie, prior to reaching the threshold levels of equity.

5. Optimal Contract — The Young Firm

An important consequence of the characterisation in Theorem 1 is that the firm's value function $Q(\mathbf{v}, s)$ is the value for a concave programming problem (after an appropriate change of variables). Thus, an optimal contract is a solution to the relevant first order conditions (which are necessary and sufficient). 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], $\lambda(\mathbf{v}, s)$ 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

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

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

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

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

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

By an adaptation of Lemma B.6 in the appendix, the first order condition for optimal investment of capital can be rewritten as

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

Thus, the agency problem which arises due to private information *is* the financing constraint, and the intensity of the financing constraint is measured by $\lambda(\mathbf{v}, s)$. In addition, we also have the following envelope conditions

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

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

The optimal contract determines repayment, capital advancement, as well as the evolution of contingent equity. We shall consider these in turn in the early stages of the contract, when the firm is young.

5.1. Optimal Repayment

The following proposition gives us some insight into the nature of the contract when contingent equities are below the threshold levels. Before stating the proposition, let us define the set

$$A_{1,s} := \left\{(\mathbf{v}, s) \in V \times S : v_b < \bar{v}_b^s,\ \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b^g] \leq v_g < \bar{v}_g^s\right\}$$

Proposition 5.1. For any optimal contract $(k, m, \mathbf{w})$, suppose $\mathbf{v} < \bar{\mathbf{v}}^s$. Then, (i) $m_b(\mathbf{v}, s) = 0$, and (ii) $(\mathbf{v}, s) \notin A_{1,s}$ implies $m_g(\mathbf{v}, s) = R(k(\mathbf{v}, s))$. Moreover, $\mathbf{v} \in A_{1,s}$ if, and only if, $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. Finally, there exist *maximal rent* contracts such that the agent's contingent equity stakes are never greater than $\bar{\mathbf{v}}^s$ following the shock s, and for $\mathbf{v} \in A_{1,s}$, $v_g > \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b^b]$ implies $R(k(\mathbf{v}, s)) > m_g$.

Proposition 5.1 says that we may regard $A_{1,s}$ as a *one-step away* set in the sense that it contains all contingent equity levels such that a good shock in the present period will send the agent's equity to the efficient set E_g, and in a maximal rent contract, to the threshold contingent equity level $\bar{\mathbf{v}}^g$. It also says that if the agent's contingent equity levels are sufficiently low (and in particular, are outside the one-step away set $A_{1,s}$), then incentives are provided exclusively through adjustments to contingent equity. On the other hand, if $m_g(\mathbf{v}, s) > 0$, then it must be that $\mathbf{w}_g(\mathbf{v}, s) = \bar{\mathbf{v}}^g$.

Put differently, Proposition 5.1 says that all rents are *back loaded* — the principal initially keeps all the revenue from production, and only eventually does the agent get a share of the proceeds. The proposition also shows that there exists a useful class of contracts, namely, the maximal rent contracts. These contracts have the feature that they involve the earliest possible payment to the agent, which results in contingent equity levels never rising above $\bar{\mathbf{v}}^s$ in state s.

The most important feature of Proposition 5.1 is that the back loading of rents holds regardless of the degree of persistence. In particular, it also holds in the iid case; see, for instance, Proposition 3 of Clementi and Hopenhayn (2006).

This suggests that back loading of rents is a property that does not depend on the persistence (or lack thereof) of the process generating the shocks. Indeed, there is a more fundamental property at play here. The principal is a monopolist and incentivises the agent by investing inefficient amounts of capital and making promises of equity. By withholding cash payments and instead by adjusting contingent equity levels, the principal ensures that whatever utility accrues to the agent (in the form of contingent equity) stays within the relationship, and is therefore available for the principal to use in the future (to draw down or raise). If however, the principal makes a cash payment to the agent, then that cash is lost forever because the agent does not save. In other words, when the agent's contingent equity is low (which results in inefficient production), it is cheaper for the principal to adjust contingent equity rather than make a cash payment. We note that this is no longer true once contingent equity reaches a threshold level.[16]

The optimal contract has another feature that should be remarked on. Because all incentives are provided via dilutions or concentrations of the agent's contingent equity, the optimal contract also features *debt forgiveness* as well as *debt rollover*.

5.2. Evolution of Contingent Equity

If $\mathbf{v} \notin A_{1,s}$, ie, if $\mathbf{v}$ is not in the one-step away set and if $\mathbf{v} < \bar{\mathbf{v}}^s$, Proposition 5.1 says that the agent does not enjoy any instantaneous rents. But the promise keeping constraints [PK_b] and [PK_g] imply that $v_b = \delta[(1 - p_b)w_{bb} + p_b w_{bg}]$ and

(16) Back loading of rents also crucially depends on the twin assumptions that the principal and agent are risk neutral and that they share a common discount factor.

$v_g = \delta[(1 - p_g)w_{gb} + p_g w_{gg}]$. Notice that $\mathbf{w}_g$ does not enter any of the other constraints, so in determining the optimal contract, we can first compute the optimal $\mathbf{w}_g$ by solving

$$\begin{aligned} \Psi(v_g) &:= \max \delta Q(\mathbf{w}_g, g) \\ &\text{s.t.} \quad \mathbf{w}_g \in V, \quad v_g = \delta[(1 - p_g)w_{gb} + p_g w_{gg}] \end{aligned}$$

This allows us to put some structure on the evolution of contingent equity.

Proposition 5.2. The optimal contingent equities satisfy:

(a) $w_{gb} < v_g/\delta < w_{gg}$.
(b) $w_{bb} < v_b/\delta < w_{bg}$.
(c) $\mathbf{w}_g$ is increasing in $\mathbf{v}$ for each $s \in S$.

The first two parts of the proposition follow immediately from Proposition 5.1 and the promise keeping constraints. The most important observation here is that $\mathbf{w}_g$ is independent of v_b and s. This follows immediately from the fact that the constraint set $\Gamma(\mathbf{v}, s)$ is independent of s, which implies $\Psi(v_g)$ is independent of v_b and s. The supermodularity of $Q(\mathbf{w}_g, g)$ in $\mathbf{w}_g$ means that increasing v_g (pointwise) increases $\mathbf{w}_g$.

5.3. Optimal Financing

Even in the absence of private information, optimal capital advancement is stochastic. Indeed, in that case, $\bar{k}_s$ satisfies, $p_s R'(\bar{k}_s) = 1$, and because s follows a Markov process, $\bar{k}_s$ follows the same Markov process. The presence of private information, indeed, private information that follows a Markov process, means that capital advancement may well be inefficient. This is intuitive because the size of investment determines the rents for the agent, and if contingent equity is low, then investment (and hence rents) cannot be high.

We shall say that capital advancement is perpetually efficient, if, starting at some date, capital advancement is efficient at each subsequent date and in each subsequent state $s \in S$.

Proposition 5.3. Let $(\mathbf{v}^0, s)$ be such that the optimal financing is inefficient, ie, $k(\mathbf{v}^0, s) < \bar{k}_s$. Also, let $\mathbf{w}_i(\mathbf{v}^0, s)$ be the optimal levels of continuation contingent equity. For a given p_b, there exists a critical level $\varphi(p_b)$ such that:

(a) If $\Delta < \varphi(p_b)$ and $\mathbf{w}_i(\mathbf{v}^0, s) \notin E_i$, then $k(\mathbf{w}_i(\mathbf{v}^0, s), i) < \bar{k}_i$ for $i = b, g$.
(b) If $\Delta \geq \varphi(p_b)$ and $\mathbf{w}_g(\mathbf{v}^0, s) \notin E_g$, then $k(\mathbf{w}_g(\mathbf{v}^0, s), g) < \bar{k}_g$.
(c) There exists a neighbourhood of $\mathbf{v}^b$ such that if $\Delta \geq \varphi(p_b)$ and $\mathbf{w}_b(\mathbf{v}^0, s) \notin E_b$, then $k(\mathbf{w}_b(\mathbf{v}^0, s), b) < \bar{k}_b$ unless $\mathbf{v}$ is in said neighbourhood of $\mathbf{v}^b$, in which case $k(\mathbf{w}_b(\mathbf{v}^0, s), b) = \bar{k}_b$.

In the iid case, if capital advancement is efficient at any date, then it is optimal from every date thereafter. This need not hold in the Markovian case. To see this, consider the contingent equity level $(0, v_g)$, where $v_g > \bar{v}^* g_g > R(\bar{k}_s)$. Promise keeping [PK$_b$] implies $\mathbf{w}_b = \mathbf{0}$. Then, by [IC*], we see that $v_g > R(\bar{k}_s)$, so that the incentive constraint is slack. It follows immediately from [FOCk] that $k\big((0, v_g), s\big) = \bar{k}_s$. However, if the state in the current period is bad, then $\mathbf{w}_b = \mathbf{0}$ implies that capital advancement is then 0 in perpetuity. By letting $v_b \approx 0$, we see that a similar argument holds, although capital advancement following a bad state will be small and positive, and not zero.

Of course, it is reasonable to ask if contingent equity levels of the form $(0, v_g)$ considered above can ever arise along the optimal contract's path. We show in Proposition E.2 in the appendix that it cannot. The other property that is peculiar to the Markovian setting is that we can compare capital advancement levels as a function of the last period's state, or more precisely, the principal's belief about the last period's state.

Lemma 5.4. For any $\mathbf{v} < \bar{\mathbf{v}}^b$, $k(\mathbf{v}, g) \geq k(\mathbf{v}, b)$.

Thus, the lemma says that conditional on having contingent equity level $\mathbf{v} < \bar{\mathbf{v}}^b$, capital advancement is higher if the last period's state was g instead of b. Intuitively, this is because if the last period's state was g, then the probability of high output is greater in the current period because of our assumption that $p_g \geq p_b$. (This is the only effect because $\Gamma(\mathbf{v}, s)$ is independent of s.)

To recapitulate, the qualitative properties of the early stages of the optimal contract are exactly the same as those of the contract in the iid case. This is because, the short run properties arise exclusively from the back loading of rents, and as noted above, back loading stems two factors: (i) the agent is risk neutral, and (ii) by back loading and paying the agent in equity, the principal can provide stronger incentives because she can always adjust the equity position by dilution or concentration. Thus, increased equity ameliorates the liquidity constraints, an observation that is independent of the degree of persistence.

Of course, with persistence, one would suspect, and our numerical calculation show this to be true for some parameter values, that increased persistence leads to increased volatility in investment. This is natural because with a success in the current period, for instance, the probability of success in the next period cannot go down, but can be strictly higher. Indeed, with high persistence, because the states are highly correlated, investment is also highly correlated across periods. While these observations are intuitive, these properties are nevertheless difficult to establish analytically in our discrete-time framework.

In the next section, we describe in detail the long-run properties of the optimal contract.

6. Long-run Properties — Maturing of the Firm

Recall again the Principal's (recursive) problem given by the functional equation

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

where $(k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$. In the iid model of Clementi and Hopenhayn (2006), where ex ante promised utility is the state variable, the long-run properties of the contract are uncovered using the observation that the derivative of the value function is a martingale. This observation was first made (also in an iid setting) by Thomas and Worrall (1990).

As we are working with interim, contingent utilities, the relevant martingale is a little more subtle, though still intuitive. First, a definition. For a fixed $s \in S$, the *directional derivative* of Q at $\mathbf{v}$ in the direction $(1, 1)$ is $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) := \lim_{h \to 0} \big[Q(\mathbf{v} + (h, h), s) - Q(\mathbf{v}, s) \big] / h$. Theorem 1 ensures that Q is differentiable everywhere, which in turn implies that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) = \langle \mathsf{D} Q, (1, 1) \rangle = Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s)$. We can now state the main result of this section.

Theorem 2. *An optimal contract induces a process $\mathsf{D}_{(1,1)}\, Q = Q_b + Q_g$ that is a non-negative martingale. The martingale $\mathsf{D}_{(1,1)}\, Q$ converges to 0 in finite time almost surely. Thus, the sets E_s represent threshold levels of equity in the sense that once in these sets, equity levels never leave these sets. In a maximal rent contract, contingent equity converges to $\bar{\mathbf{v}}^g$, and then cycles on the set $\{\bar{\mathbf{v}}^b, \bar{\mathbf{v}}^g\}$, with transitions according to the Markov process on S.*

Theorem 2 states that from any initial level of contingent equity $\mathbf{v}^0 \in V$ in state s, an optimal contract converges to contingent equity levels $\bar{\mathbf{v}}^g$ in finite time almost surely. This last part echoes the iid case, in that the only way to reach the efficient sets is by experiencing one final good production shock. In other words, if the agent has not yet achieved the threshold levels of contingent equity, then a bad shock will never place him in the efficient sets E_s.

Another implication of Theorem 2 and part (c) of Proposition 4.2 is that once the optimal contract promises contingent utilities in the sets E_s, the agent never leaves these sets. It is this property that justifies the nomenclature 'efficient sets'. Of course, since we are restricting attention to maximal rent contracts (see Proposition 5.1), any transition to the threshold sets means that contingent utility transitions initially to $\bar{\mathbf{v}}^g$, and then cycles on the set $\{\bar{\mathbf{v}}^g, \bar{\mathbf{v}}^b\}$ according to the Markov process on states S. This last part of Theorem 2 is a major difference between the iid case and the Markovian case with persistence. In the former, $\bar{\mathbf{v}}^b = \bar{\mathbf{v}}^g$, and so the cycling is trivial. However, in the Markovian case, the existence of non-trivial dynamics even after reaching the efficient sets leads to interesting conclusions and testable implications, which we discuss in Section 8 below.

To see why the process $\mathsf{D}_{(1,1)}\, Q$ is a non-negative martingale, recall first from Theorem 1 that even though $Q_g \geq 0$, for some $(\mathbf{v}, s)$, we have $Q_b(\mathbf{v}, s) < 0$. Nonetheless, Theorem 2 says that $Q_b + Q_g \geq 0$ for all $(\mathbf{v}, s)$. This is because if we start the optimal contract at a point $(\mathbf{v}^0, s)$ where $Q_b(\mathbf{v}^0, s) > 0$, then, along the path induced by an optimal contract, we will always have $Q_b(\cdot, s) \geq 0$. This implies $Q_b + Q_g$ is always strictly positive — recall part (c) of Theorem 1, which tells us that $Q_g \geq 0$ everywhere — until it takes the value 0, and this occurs if, and only if, $Q_b = Q_g = 0$, which happens precisely on the sets E_s.

The martingale property of $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s)$ is easy to see. From the envelope conditions (see Section 5), we see that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s) = \eta_b(\mathbf{v}, s) + \eta_g(\mathbf{v}, s)$. From the first order conditions, we obtain $(1 - p_s)\, \mathsf{D}_{(1,1)}\, Q(\mathbf{w}, b) + p_s\, \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) = \eta_b(\mathbf{v}, s) + \eta_g(\mathbf{v}, s)$, ie, $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}, s)$ is a martingale.

To understand why $\mathsf{D}_{(1,1)}\, Q = Q_b + Q_g$ *must* be a martingale, let us reason as Thomas and Worrall (1990) do. Consider an increase in $\mathbf{v}$ by $(\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. One way to accomplish this is by increasing $\mathbf{w}_i$ by $(\varepsilon/\delta, \varepsilon/\delta)$ for each $i = b, g$. It is easy to see that such a change to continuation contingent utilities is incentive compatible and also satisfies promise keeping, and the resulting change in the value of the firm from this increase in contingent utilities is precisely $(1-p_s)\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b, b) + p_s\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$. An envelope argument shows that this change in $\mathbf{w}_i$ is locally optimal, and so the change in firm value is $\mathsf{D}_{(1,1)}\,Q(\mathbf{v}, s) = Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s)$, ie, $\mathsf{D}_{(1,1)}\,Q$ is a martingale. Doob's Martingale Convergence Theorem 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, $\mathsf{D}_{(1,1)}\,Q$ converges to 0.

While such a conclusion is also drawn in the iid case, we hasten to point out an important difference. In the iid case, when using ex ante promised utility as a state variable, it is the derivative of the resulting value function that is a martingale. If the value function is strictly concave, there is a one-to-one relationship between the derivative and ex ante promised utility. (Even if strict concavity doesn't hold, there is nevertheless a tight relationship.) In the case of persistence, with interim equities as state variables, knowing $\mathsf{D}_{(1,1)}\,Q(\mathbf{v}, s) = c$ for some $c > 0$ does not pin down $\mathbf{v}$. 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, which essentially requires us to show that $\mathsf{D}_{(1,1)}\,Q(\mathbf{v}, s)$ cannot converge to a strictly positive value, rather more subtle.

Theorem 2 also states that convergence occurs in finite time almost surely. This last part is established as follows: Recall the set $A_{1,s}$, which is the one-step away set. If $(\mathbf{v}, s) \in A_{1,s}$, then a good outcome in the present period will place contingent equity at $\bar{v}^g$ in the following period. The proof shows that the set $A_{1,g}$ have a non-empty interior, and because almost every path converges to $\bar{\mathbf{v}}^g$, it cannot spend infinite amounts of time in an open subset of $A_{1,s}$, and hence in a neighbourhood of $\bar{\mathbf{v}}^g$. Thus, all convergence occurs in finite time almost surely.

7. Optimal Contract — The Mature Firm

A mature firm is one where the equity holders get dividend payments. The structure of the mature firm is stark in the iid case. The agent is the residual claimant of the firm, investment is constant over time, and the ratio of debt-to-equity is constant over time, which implies that the firm's β is also constant over time, which in turn implies that the firm's valuation is constant over time. These implications are seldom seen in practice — see Section 8 for some stylised facts about firms. As we shall see below, when output displays persistence, none of these conclusions hold. Moreover, the empirical implications of our model with persistence are in consonance with the data.

Our results will depend on the degree of persistence Δ. To capture the impact of persistence on the optimal contract for a mature firm, let us define the following sets:

$$\begin{aligned} B_+ &:= \{(p_b, p_g) : R(\bar{k}_b) < R(\bar{k}_g)\delta p_g/(1+\delta p_g)\} \\ B_= &:= \{(p_b, p_g) : R(\bar{k}_b) = R(\bar{k}_g)\delta p_g/(1+\delta p_g)\} \\ B_- &:= \{(p_b, p_g) : R(\bar{k}_b) > R(\bar{k}_g)\delta p_g/(1+\delta p_g)\} \end{aligned} \tag{7.1}$$

To understand the sets, let us fix p_g and suppose p_b is sufficiently small. Notice that $R(\bar{k}_g)\delta p_g/(1+\delta p_g)$ is independent of p_b, and that $R(\bar{k}_b)$ increases with p_b because $\bar{k}_b$ does. For sufficiently small p_b, we have $R(\bar{k}_b) < R(\bar{k}_g)\delta p_g/(1+\delta p_g)$. The set B_+ therefore delineates all probabilities (p_b, p_g) such that $p_g - p_b = \Delta$ is sufficiently large. Similarly, B_- denotes the set of all probabilities (p_b, p_g) such that $p_g - p_b$ is not too large. In particular, if $p_b = p_g$ and we are in the iid case, then $(p_b, p_g = p_b) \in B_-$.

Even for a mature firm, the promise keeping conditions [PK_b] and [PK_g] must hold. Proposition K.1 shows that for $s = b$, $(p_b, p_g) \in B_+$ if, and only if, [IC] holds with equality and [LL] holds with inequality, while $(p_b, p_g) \in B_-$ if, and only if, [IC] holds with inequality and [LL] holds with equality. Clearly, the set $B_=$ is the boundary between B_- and B_+, ie, is the intersection of their closures. Proposition K.1 shows that in case of failure, there are no transfers to or from the agent, and if the previous period had a success, then conditional on a good shock in the current period, the agent keeps some of the output and the incentive constraint [IC] holds as an equality.

Given Proposition K.1, which tells us precisely which constraints hold for various values of (p_b, p_g), we are now in a position to explicitly describe the threshold levels of equity $\bar{\mathbf{v}}^s$.

Lemma 7.1. The threshold levels of contingent equity are as follows:

(a) Suppose $(p_b, p_g) \in B_+$. Then,

$$\bar{\mathbf{v}}^b = (\bar{v}_b^b, \bar{v}_g^b) = \left(\frac{\delta p_b \bar{v}_g^b}{1 - \delta(1 - p_b)}, \frac{\delta p_g R(\bar{k}_g)}{(1 + \delta p_g)[1 - \delta p_g - \frac{\delta^2 p_b (1-p_g)}{1-\delta(1-p_b)}]} \right)$$
$$\bar{\mathbf{v}}^g = (\bar{v}_b^g, \bar{v}_g^g) = \left(\frac{\delta p_b \bar{v}_g^b}{1 - \delta(1 - p_b)}, \bar{v}_g^b + \frac{R(\bar{k}_g)}{1 + \delta p_g} \right)$$

(b) Suppose $(p_b, p_g) \in B_-$. Then,

$$\bar{\mathbf{v}}^b = (\bar{v}_b^b, \bar{v}_g^b) = \left(\frac{\delta p_b R(\bar{k}_b)}{(1 - \delta)(1 - \delta \Delta)}, \frac{1 - \delta(1 - p_b)}{(1 - \delta)(1 - \delta \Delta)} R(\bar{k}_b) \right)$$
$$\bar{\mathbf{v}}^g = (\bar{v}_b^g, \bar{v}_g^g) = \left(\frac{\delta p_b R(\bar{k}_b)}{(1 - \delta)(1 - \delta \Delta)}, R(\bar{k}_g) + \frac{\delta(p_g - \delta \Delta)}{(1 - \delta)(1 - \delta \Delta)} R(\bar{k}_b) \right)$$

The maximal rent contract for a mature firm is described next.

Theorem 3. *Suppose the firm is mature and consider a maximal rent contract, so that contingent equity levels are always $\bar{\mathbf{v}}^s$ if the previous period's shock was $s \in \{b, g\}$. The contract $(k, m_i, \mathbf{w}_i)$ takes the form:*

$$k(\bar{\mathbf{v}}^b, b) = \bar{k}_b, \quad k(\bar{\mathbf{v}}^g, g) = \bar{k}_g$$
$$\mathbf{w}_g(\bar{\mathbf{v}}^s, s) = \bar{\mathbf{v}}^g, \quad \mathbf{w}_b(\bar{\mathbf{v}}^s, s) = \bar{\mathbf{v}}^b$$
$$\bar{m}_g(\bar{\mathbf{v}}^b, b) = R(\bar{k}_b), \quad \bar{m}_g(\bar{\mathbf{v}}^g, g) = \frac{\delta p_g R(\bar{k}_g)}{1 + \delta p_g} \quad \textit{if } (p_b, p_g) \in B_+$$
$$\bar{m}_g(\bar{\mathbf{v}}^b, b) = \bar{m}_g(\bar{\mathbf{v}}^g, g) = \delta p_g \left[R(\bar{k}_g) - R(\bar{k}_b) \right] \quad \textit{if } (p_b, p_g) \in B_-$$

The most striking feature of the optimal contract is that even if the firm is mature, as long as $\Delta > 0$, the agent makes positive payments to the principal, thereby highlighting the sensitivity of the iid model. Theorem 3 says in particular that if the previous period had a good shock, then in the current period, conditional on another good shock, the agent keeps a positive fraction, but not all, of the output, and transfers some of the output to the principal. In other words, with positive persistence, the principal always gets some part of the output.

Indeed, if Δ is sufficiently large, then $(p_b, p_g) \in B_+$, and in this case, contingent on the previous period's shock being bad, the agent gives all of the output to the principal, while retaining only a positive (strictly less than one) fraction of the output only if the previous period's shock was good. This can be interpreted as a high-water mark contract whereby the agent is paid only if he reaches the previous best level of performance. To see the intuition behind this result, fix $p_g \in (0, 1)$, and suppose $p_b \approx 0$, which implies $\bar{k}_b \approx 0$. Then, if the previous period's shock was bad, [IC*] does not hold with equality because $\bar{v}_g^b - \bar{v}_g^b$ is bounded away from zero, while $R(\bar{k}_b) \approx 0$.

The fact that investment after a bad shock is extremely low gives the principal an added threat: After a good shock, investment will be high, so the agent's rent should be high. But if the agent lies, his utility in the next period will be very low because investment after a reported bad shock is low, which in turn reduces the agent's rent in the subsequent period. Thus, the principal can carve out a payment even after a good shock by (rationally) threatening to reduce investment, and thereby rent, if a bad shock is ever reported.

This intuition is for the case where Δ is sufficiently large relative to p_b. But the same tradeoffs are also present for lower levels of Δ. Indeed, as long as $\Delta > 0$, such a tradeoff is always present, which is why as long as there is the slightest bit of persistence, the principal always gets some payment.

It is clear that the discussion above relies on the presence of persistence. If, however, shocks are iid over time (so $p_b = p_g$), we get the following contingent equities:

$$\bar{\mathbf{v}}^{\text{iid}} = \left(\frac{\delta p}{1-\delta} R(\bar{k}), \frac{1-\delta+\delta p}{1-\delta} R(\bar{k}) \right)$$

which are constant over time and independent of the previous period. Notice that the *ex ante* expected utility from the contingent equity $\bar{\mathbf{v}}^{\text{iid}}$ is precisely $\bar{v}^{\text{iid}} = (1-p)\frac{\delta p}{1-\delta} R(\bar{k}) + p\frac{1-\delta+\delta p}{1-\delta} R(\bar{k}) = pR(\bar{k})/(1-\delta)$, just as in Clementi and Hopenhayn (2006). These generate transfers $\bar{m}_g^{\text{iid}} = \bar{m}_b^{\text{iid}} = 0$, as can easily be seen from the last displayed line in Theorem 3. In other words, with iid shocks, the agent is the residual claimant.

8. Empirical Implications

Given a level $\mathbf{v} = (v_b, v_g)$ of contingent equity, the firm's value is $Q(\mathbf{v}, s)$. To ensure that the agent's equity is a junior claim on cash flows, the agent is typically restricted to holding restricted stocks in the firm which he cannot freely sell, lest he destroy the incentive scheme, a practice documented by Murphy (1999). The debt the firm holds is $Q(\mathbf{v}, s) - (1 - p_s)v_b - p_s v_g$. This is a reasonable interpretation because debt is a claim on expected operating cash flows, while stocks (in the form of equity) pay dividends only when the firm reaches some threshold size or performance, consistent with the findings of Kaplan and Strömberg (2003). Moreover, our model predicts that young firms don't pay dividends, while mature and large firms do, consistent with the findings of Bulan, Subramanian and Tanlu (2007). Also, the optimal contract entails the principal either forgiving some debt in case of failure or rolling it over in case of success. However, notice that none of these findings depend on the degree of persistence. This suggests that these features of optimal contracts are driven entirely by risk neutrality and liquidity constraints, which is consistent with the finding of Kaplan and Strömberg (2004) that risk sharing is not an important consideration in venture capitalist (VC) contract design, at least relative to the agency problem.

Our model predicts that for a young firm with sufficiently high levels of contingent equity and for a sufficiently high degree of persistence, investment after a bad shock can be efficient, but for a young firm, investment after a good shock must necessarily be inefficient. Moreover, investment is positively correlated over time, which is also true in the iid setting, but is exaggerated due to the persistence of the production shocks.

Recall that with persistence, there are three possible levels of output for a mature firm, namely 0, $R(\bar{k}_b)$, and $R(\bar{k}_g)$. A *high-water mark* compensation scheme entails the agent being paid only if the output is $R(\bar{k}_g)$. In the iid case, for a mature firm, investments are constant over time (ie, $R(\bar{k}_b) = R(\bar{k}_g)$), and the agent receives a dividend payment after a success is reported. (Thus, the agent is the residual claimant.) This allows us to interpret the compensation scheme for a mature firm as a high-water mark contract, because the agent is paid only after the firm produces the highest possible (efficient) level of output — the high-water mark — and other receives nothing. As mentioned above, an attractive feature of the high-water mark contract is that implicitly adjusts equity levels without having to do so explicitly. This suggests that studies such as Fama and French (2005) *underestimate* the issuance of equity, because these studies only look at explicit issuance, and ignore implicit issuance of equity.

However, as is clear, this result depends crucially on the fact that $R(\bar{k}_b) =$

$R(\bar{k}_g)$, ie, investment is independent of the past. With persistence, there are two levels of efficient investment (which depend on the previous period's output) and three levels of output, and as Theorem 3 shows, the high-water mark contract is not optimal for arbitrarily small (but strictly positive) levels of persistence. Instead, our model predicts that for small levels of persistence, even for a mature firm, the principal receives a small payment after a success, while the agent retains the residual cash flow. These payments diminish with persistence, and so with zero persistence, the principal does not receive any payment.

If persistence in the production shocks is sufficiently high, Theorem 3 says that the agent is paid only if the output is $R(\bar{k}_g)$, which can only occur after two successes in a row. Thus, dividend payments for a mature firm *do* resemble high-water mark compensation schemes. Notice that if persistence is sufficiently high, there is a significant difference between investment levels after successes and failures, namely $\bar{k}_g$ and $\bar{k}_b$ respectively. Therefore, we may interpret Theorem 3 as saying that the use of high-water mark compensation schemes should be seen when there is significant sensitivity of investment to past performance, ie, when investment is expected to be strongly correlated over time. This is consistent with Aragon and Qian (2010) who show that greater flow-performance sensitivity associated with hedge funds that use high-water mark compensation is 'driven by investors' response to superior past performance'. The important fact in our model is that the high-water mark compensation in our model appears as a consequence of persistence, and predicts that such a compensation scheme will only be seen when persistence is sufficiently high.

Theorem 3, which details the compensation scheme for a mature firm, asserts that mature firms issue equity, and moreover, use equity as compensation for managers. This is consistent with the findings of Fama and French (2005), who demonstrate that firms of all sizes and ages issue equity, often, and for a variety of reasons, compensation being one of them. Again, this is in contrast with the iid case, where a mature firm does not issue any more equity.

Our model also predicts that the valuation of mature firms can fluctuate over time. In particular, the debt-to-equity ratio, and hence the β for the firm also vary over time. This is in addition to investment in mature firms, which also varies over time. Moreover, our model predicts that equity issuance is procyclical, which comports with the findings of Covas and Den Haan (2011), who emphasise that it is important to exclude the largest of firms (approximately top 5%, but definitely the top 1%) when looking at the data.

Finally, our model also predicts that initial conditions — for instance, the state of the economy — matter insofar as they dictate whether the project will be funded by the principal, and if the project is funded, how much debt the firm will start with. Proposition 9.1 shows that if the principal believes that the production shock is likely to be good in the present period, then the initial level of debt is higher than if the production shock is likely to be bad. This is in line with the observations of Gompers et al. (2008), who argue that much of the volatility in initial VC funding of projects can be explained as responses to 'public market signals of investment opportunities'.

9. Discussion and Extensions

We now consider some of the salient features of the optimal contract as well as some extensions.

9.1. Initialisation

The initial vector of contingent equity $\hat{\mathbf{v}} = (\hat{v}_b, \hat{v}_g)$ is chosen by the principal. Clearly, the principal can choose this initial level so that firm value is maximised, ie, she can choose $\hat{\mathbf{v}}^s \in E_s$. However, this clearly gives too much rent to the agent. Therefore, assuming the principal's belief about the probability of success in the current period is p_s where $s \in \{b, g\}$, the principal chooses initial equity $\hat{\mathbf{v}}^s$ so as to maximise

$$\mathscr{D}(\hat{\mathbf{v}}^s, s) = Q(\hat{\mathbf{v}}^s, s) - (1 - p_s)\hat{v}_b^s + p_s \hat{v}_g^s$$

where $\mathscr{D}(\hat{\mathbf{v}}^s, s)$ is the *debt* that the firm holds. (As noted above, the value of the firm is simply the sum of the debt and the expected equity for the agent.) The next proposition shows that initial debt is higher if the principal believes there is a greater chance of success.

Proposition 9.1. Initial firm debt is greater if $s = g$, ie, $\mathscr{D}(\hat{\mathbf{v}}^g, g) > \mathscr{D}(\hat{\mathbf{v}}^b, b)$. In addition, initial investment is always inefficient for all $s \in \{b, g\}$.

The dependence of initial debt values of a firm on the surrounding business climate is a subject that has received much attention because of the volatility in venture capital investments — see Gompers et al. (2008). Proposition 9.1 also says that initial funding of firms results in investments that are necessarily suboptimal, also in line with observed funding behaviour.

9.2. Path Dependence

As in the iid case, the optimal contract exhibits strong path dependence in that the evolution of $\mathbf{v}$ depends on the sequence of shocks. However, there is a stronger form of path dependence in the present setting than in, say, the risk sharing model of Thomas and Worrall (1990) who show that with CARA utility, the optimal contract (in an iid setting) only depends on the fraction of good shocks rather than the specific order. In our setting, the order of shocks is crucial. This follows immediately from Theorem 2, which says that convergence to maturity occurs in finite time almost surely. Therefore, the agent would rather his good shocks come sooner rather than later.

It follows from Theorem 2 that it takes a good shock for a young firm to reach maturity, ie, for equity levels to reach the set E_g. This is also true in the iid case, as noted by Clementi and Hopenhayn (2006) and Krishna, Lopomo and Taylor (2013). However, if Δ is sufficiently large, and this is unique to the Markovian case, it necessarily takes two consecutive good production shocks for the firm to mature. This is formally stated in the next proposition.

Proposition 9.2. For fixed p_g, if Δ is sufficiently large, then $A_{1,b} = \varnothing$.

The proposition says that the one-step away set is empty if the last period's shock was bad. This makes intuitive sense because (i) it takes a good shock for the firm to mature, and (ii) if Δ is sufficiently large, then after a bad shock, the agent's contingent equity is sufficiently low that it cannot reach $\bar{\mathbf{v}}^g$ in one step after a good shock.

10. Conclusion

In this paper, we explore the question of how a firm is financed when the firm's output, modulo investment levels, is persistent over time. In particular, we consider a principal who provides an agent with funds to operate a firm. Even though shocks to revenue are persistent, we formulate the mechanism design as a recursive problem, and show that the firm's value is a solution to a recursive dynamic program where the agent's contingent equity (which is a vector) is a state variable, along with the beliefs about high revenue in the current period.

We show that in the optimal contract, when the firm is young, it faces financing constraints. Investment in the firm is suboptimal as long as the financing constraints bind. The incentive scheme involves the agent being compensated exclusively through adjustments to his contingent equity, and all payments are backloaded. There exist threshold levels of contingent equity, ie, minimal levels of equity such that if the agent reaches these levels of equity, the firm no longer faces financing constraints, and investment is forever hence optimal.

The long-run dynamics of the firm are captured by the observation that the directional derivative of the firm's value function in the direction $(1, 1)$ is a martingale. This allows us to show that with probability one, the agent's equity reaches the threshold levels. In other words, with probability one, the firm matures and is then forever free of financing constraints, and operates at its efficient level.

In contrast to the iid setting, with persistence, the mature firm's size varies over time, the agent is no longer the residual claimant, the agent is compensated through cash as well as adjustments to his contingent equity, and the firm's debt-to-equity ratio vary over time. Moreover, the structure of the optimal contract depends crucially on the degree of persistence. For instance, the model predicts that in a mature firm, the agent is compensated via a high-water mark contract only if persistence is sufficiently high. These predictions are in accord with stylised facts about mature firms.

There are many interesting extensions to our basic model. A first would be to embed our model in a general equilibrium setting, in order to understand how frictions in firm financing affect the macro-economy. Another interesting direction would be to allow for the possibility that capital does not depreciate completely between periods. This would correspond to the setting where the principal provides the agent with capital sporadically, but is nevertheless active in monitoring the agent's day-to-day performance. We leave these extensions to future work.

Appendices

A. Recursive Domain

We first present the proof of Proposition 4.1. 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 4.3) that $(k, m_i, \mathbf{w}_i)_{i=b,g}$ *implements* (v_b, v_g) if $(k, m_i, \mathbf{w}_i)$ satisfies the incentive compatibility, promise keeping, and limited liability constraints. Let $\mathscr{K}$ 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 $\Phi : \mathscr{K} \to \mathscr{K}$ as follows: for $C \in \mathscr{K}$, 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, Proposition 4.1 delineates such a set.

Proof of Proposition 4.1. 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

$$\begin{aligned} v_b &= -m_b + \delta\big[(1 - p_b)w_{bb} + p_b w_{bg}\big] \\ &\geq -(1 - p_b + p_b\alpha)m_b + \delta[(1 - p_b)w_{bb} + p_b\alpha w_{bb}] \\ &= (1 - p_b + p_b\alpha)(\delta w_{bb} - m_b) \end{aligned}$$

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

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

Plugging this into [IC*], we obtain

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

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

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

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

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

To see that $V := \{(v_b, v_g) \in \mathbb{R}^2_+ : v_g \geq v_b\}$ is a fixed point of Φ, we apply the operator Φ to V. Take any 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. Bellman Operator

In this section, we will define the Bellman operator corresponding to maximizing the firm's value and show that the resulting value function satisfies various properties. Let $C(V \times S)$ be the space of continuous functions on the domain $V \times S$ and let $\mathcal{F} := \{P \in C(V \times S) : 0 \leq P(\mathbf{v}, s) \leq \bar{Q}(s)\}$ be endowed with the 'sup' metric, where $\bar{Q}(s)$ is defined in equations [4.1] and [4.2]. It is easy to see that $\mathcal{F}$ so defined is a complete metric space. Define the Bellman operator $T : \mathcal{F} \to \mathcal{F}$ as:

$$\text{[P1]} \quad (TP)(\mathbf{v}, s) = \max_{k, m_i, \mathbf{w}_i} \big(-k + p_s[R(k) + \delta P(\mathbf{w}_g, g)] + (1 - p_s)\delta P(\mathbf{w}_b, b)\big)$$
$$\text{s.t.} \quad (k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$$

where $\Gamma(\mathbf{v}, s) = \{(k, m_i, \mathbf{w}_i) : (k, m_i, \mathbf{w}_i) \text{ implements } \mathbf{v} \text{ and } \mathbf{w}_i \in V\}$.

It is standard to show that $T : \mathcal{F} \to \mathcal{F}$ is well defined. In what follows, we shall show that T maps certain subsets of $\mathcal{F}$ to themselves. We begin by showing that T maps concave functions to concave functions. Let $\mathcal{F}_1 := \{P \in \mathcal{F} : P(\cdot, s) \text{ is concave for all } s \in S\}$.

Lemma B.1. If $P \in \mathcal{F}_1$, then $TP \in \mathcal{F}_1$.

Proof. Let $P \in \mathcal{F}_1$. For any $k \geq 0$, let $R(k) = r$, so that $c(r) = k$, where $c(r) := R^{-1}(r)$. Because R is increasing, c is well defined. The concavity of R implies that c is convex. Thus, we can let the choice variables be $(r, m_i, \mathbf{w}_i)$, so that the objective becomes

$$(TP)(\mathbf{v}, s) = \max_{r, m_i, \mathbf{w}_i} \big(-c(r) + p_s[r + \delta P(\mathbf{w}_g, g)] + (1 - p_s)\delta P(\mathbf{w}_b, b)\big)$$
$$\text{s.t.} \quad (r, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$$

It is easy to see that with this transformation, $\Gamma(\mathbf{v}, s)$ is the intersection of finitely many affine sets, and so is convex. Moreover, the objective, $-c(r) + p_s[r +$

$\delta P(\mathbf{w}_g, g)] + (1 - p_s)\delta P(\mathbf{w}_b, b)$ is concave in $(r, \mathbf{w}_i)$. Standard arguments now imply $TP(\mathbf{v}, s) \in \mathcal{F}_1$. □

Recall that the efficient sets E_s where the firm is unconstrained are defined as

$$E_b := \left\{ \mathbf{v} \geq \bar{\mathbf{v}}_b : v_g - v_b \geq \frac{R(\bar{k}_b)}{(1-\delta\Delta)} \right\}$$

$$E_g := \left\{ \mathbf{v} \geq \bar{\mathbf{v}}_g : v_g - v_b \geq R(\bar{k}_g) + \delta\Delta \max \left[\frac{\delta \bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta} \right] \right\}$$

where $\bar{\mathbf{v}}_b$ and $\bar{\mathbf{v}}_g$ are given in Lemma 7.1. Notice that $\bar{\mathbf{v}}_b$ and $\bar{\mathbf{v}}_g$ in Lemma 7.1 satisfy $\bar{v}_{sb} = \delta[p_b \bar{v}_{bg} + (1-p_b)\bar{v}_{bb}]$ and $\bar{v}_{sg} = R(\bar{k}_s) - \bar{m}_g(s) + \delta[p_g \bar{v}_{gg} + (1 - p_g)\bar{v}_{gb}]$. Limited liability [LL] implies that $R(\bar{k}_s) \geq \bar{m}_g(s)$, which yields (along with $\bar{\mathbf{v}}_s \in E_s$)

$$\bar{v}_{sb} = \delta[p_b \bar{v}_{bg} + (1-p_b)\bar{v}_{bb}] \tag{B.1}$$

$$\bar{v}_{sg} \geq \delta[p_g \bar{v}_{gg} + (1-p_g)\bar{v}_{gb}] \tag{B.2}$$

We now show that T maps the space of functions that achieve $\bar{Q}(s)$ on the sets E_s to itself. Let $\mathcal{F}_2 := \{P \in \mathcal{F} : P(\mathbf{v}, s) = \bar{Q}(s) \text{ for all } \mathbf{v} \in E_s,\ s \in S\}$.

Lemma B.2. If $P \in \mathcal{F}_2$, then $TP \in \mathcal{F}_2$.

Proof. Suppose $P \in \mathcal{F}_2$. We will show that for any $\mathbf{v} \in E_s$, there exists a policy such that $TP(\mathbf{v}, s) = \bar{Q}(s)$.

Let us first consider the case where $\mathbf{v} \in E_b$, so that $s = b$. For such a $\mathbf{v}$, consider the policy

- $k = \bar{k}_b$, $\mathbf{w}_g = \bar{\mathbf{v}}_g$
- $w_{bg} = \bar{v}_{bg}$, $w_{bb} = \bar{v}_{bg} - \min\{v_g - v_b, \bar{v}_{bg} - \bar{v}_{bb}\}$
- $m_g = R(\bar{k}_b) + \delta[p_g \bar{v}_{gg} + (1-p_g)\bar{v}_{gb}] - v_g$
- $m_b = \delta[p_b w_{bg} + (1-p_b)w_{bb}] - v_b$

We shall first show this policy satisfies all the constraints in [P1]. The constraints [PK_b] and [PK_g] follow from the definitions of m_b and m_g. Because $\mathbf{v} \in E_b$, we have

$$\begin{aligned} v_g - v_b &\geq R(\bar{k}_b) + \delta\Delta(v_g - v_b) \\ &\geq R(\bar{k}_b) + \delta\Delta \min\{v_g - v_b, \bar{v}_{bg} - \bar{v}_{bb}\} \\ &= R(\bar{k}_b) + \delta\Delta(w_{bg} - w_{bb}) \end{aligned}$$

The first line is from the definition of E_b while the second is just arithmetic. So the constructed policy satisfies [IC*]. By the definition of w_{bb}, we either have $w_{bb} = \bar{v}_{bb}$ or $w_{bb} = v_b - (v_g - \bar{v}_{bg}) \leq v_b$; the definition of w_{bb} also implies that in this latter case, $w_{bb} = v_b - (v_g - \bar{v}_{bg}) \geq v_b - (\bar{v}_{bg} - \bar{v}_{bb}) = \bar{v}_{bb}$.

In sum, we obtain $\bar{v}_{bb} \leq w_{bb} \leq v_b$, so from [B.1] it follows that

$$\begin{aligned} v_b \geq w_{bb} &\geq \delta[p_b \bar{v}_{bg} + (1 - p_b) w_{bb}] \\ &= \delta[p_b w_{bg} + (1 - p_b) w_{bb}] \end{aligned}$$

which means the constructed transfer $m_b \leq 0$, ie, [LL] for $s = b$ is satisfied. Moreover, the constructed transfer m_g satisfies

$$m_g \leq R(\bar{k}_b) + \delta[p_g \bar{v}_{gg} + (1 - p_g)\bar{v}_{gb}] - \bar{v}_{bg} \leq R(\bar{k}_b)$$

The first inequality is from $\bar{v}_{bg} \leq v_g$, and the second is by [B.2]. Moreover, $m_g \leq R(\bar{k}_b)$ implies [LL] is satisfied for $s = g$.

By definition, we have $w_{bg} - w_{bb} = \min\{v_g - v_b, \bar{v}_{bg} - \bar{v}_{bb}\} \geq \frac{R(\bar{k}_b)}{1-\delta\Delta}$, where the inequality is because $\mathbf{v}, \bar{\mathbf{v}}_b \in E_b$. Therefore, our choice of $\mathbf{w}_b \in E_b$, which in conjunction with the assumption that $P \in \mathcal{F}_2$ implies $P(\mathbf{w}_b, b) = P(\bar{\mathbf{v}}_b, b)$.

As all the constraints of [P1] are satisfied, we have

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

where we have used the definition of $\bar{Q}$ and the fact that $P(\mathbf{w}_b, b) = P(\bar{\mathbf{v}}_b, b)$ to obtain the last equality. Therefore $TP(\mathbf{v}, b) \in \mathcal{F}_2$.

Now consider the case where $\mathbf{v} \in E_g$ and $s = g$. For such a $\mathbf{v}$, consider the policy

- $k = \bar{k}_g$, $\mathbf{w}_g = \bar{\mathbf{v}}_g$
- $w_{bg} = \bar{v}_{bg}$, $w_{bb} = \bar{v}_{bg} - \max\left[\frac{\delta \bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right]$
- $m_g = R(\bar{k}_g) + \delta[p_g \bar{v}_{gg} + (1 - p_g)\bar{v}_{bg}] - v_g$
- $m_b = \delta[p_b w_{bg} + (1 - p_b) w_{bb}] - v_b$

The constraints [PK$_b$], [PK$_g$] are satisfied by construction. Because $\mathbf{v} \in E_g$, we find

$$\begin{aligned} v_g - v_b &\geq R(\bar{k}_g) + \delta\Delta \max\left[\frac{\delta\bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right] \\ &= R(\bar{k}_g) + \delta\Delta(w_{bg} - w_{bb}) \end{aligned}$$

which means [IC*] is satisfied. The constructed transfer m_g satisfies:

$$m_g \leq R(\bar{k}_g) + \delta[p_g\bar{v}_{gg} + (1-p_g)\bar{v}_{gb}] - \bar{v}_{gg} \leq R(\bar{k}_g)$$

where the first inequality is because $\bar{v}_{gg} \leq v_g$, and the second follows from [B.2]. So [LL] for $s = g$ is satisfied. The constructed transfer m_b satisfies

$$\begin{aligned} m_b &= \delta[p_b w_{bg} + (1-p_b)w_{bb}] - v_b \\ &= \delta\bar{v}_{bg} - \delta(1-p_b)\max\left[\frac{\delta\bar{v}_{bg} - v_b}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right] - v_b \\ &\leq 0 \end{aligned}$$

so that [LL] for $s = b$ is satisfied. Therefore, all the constraints of [P1] are satisfied, and the policy is feasible.

Notice also that

$$\begin{aligned} w_{bb} &\geq \bar{v}_{bg} - \max\left[\frac{\delta\bar{v}_{bg} - \bar{v}_{bb}}{\delta(1-p_b)}, \frac{R(\bar{k}_b)}{1-\delta\Delta}\right] \\ &\geq \bar{v}_{bg} - (\bar{v}_{bg} - \bar{v}_{bb}) = \bar{v}_{bb} \end{aligned}$$

The first line is from $\bar{v}_{bb} \leq v_b$, while the second line follows because $\frac{\delta\bar{v}_{bg}-\bar{v}_{bb}}{\delta(1-p_b)} = \bar{v}_{bg} - \bar{v}_{bb}$ (by [B.1]) and $\frac{R(\bar{k}_b)}{1-\delta\Delta} \leq \bar{v}_{bg} - \bar{v}_{bb}$ (from the definition of E_g). By construction, $w_{bg} - w_{bb} \geq \frac{R(\bar{k}_b)}{1-\delta\Delta}$, so it follows that $\mathbf{w}_b \in E_b$. The assumption $P \in \mathscr{F}_2$ then implies $P(\mathbf{w}_b, b) = P(\bar{\mathbf{v}}_b, b)$. Putting this all together, we obtain

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

where we have used the definition of $\bar{Q}(g)$ and the fact that $P(\mathbf{w}_s, s) = P(\bar{\mathbf{v}}_s, s)$ for $s = b, g$. Therefore $TP(\mathbf{v}, g) \in \mathscr{F}_2$, as desired. □

We now show that T preserves functions that are locally decreasing in v_b at (v, v) for $v > 0$. More precisely, let

$$\mathcal{F}_3 = \{P(\mathbf{v}, s) \in \mathcal{F} : P((v, v), s) \text{ is local minimum of } P((v_b, v), s)\ \forall v > 0\}$$

Lemma B.3. If $P(\mathbf{v}, s) \in \mathcal{F}_3$, then $TP(\mathbf{v}, s) \in \mathcal{F}_3$.

Proof. Suppose $P(\mathbf{v}, s) \in \mathcal{F}_3$ and pick any $v > 0$. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at $((v, v), s)$. From [IC*], we must have $k = 0$ and $w_{bg} = w_{bb}$. From [PK$_b$], we then obtain $w_{bg} = w_{bb} = (v + m_b)/\delta$. Let $v' = v - \frac{1-p_b}{1-p_g} R(\varepsilon)$ for any arbitrary small $\varepsilon > 0$, and $k' = R(\varepsilon)$, $w'_{bb} = w_{bb} - \frac{R(\varepsilon)}{\delta(1-p_g)}$.

It is easy to verify that $(k', m_b, m_g, (w'_{bb}, w_{bg}), \mathbf{w}_g) \in \Gamma((v', v), s)$, ie, it is a feasibly policy, so that we have

$$\begin{aligned} &TP((v', v), s) - TP((v, v), s) \\ &\quad \geq -\varepsilon + p_s R(\varepsilon) + (1 - p_s)\delta \left[P\left((w'_{bb}, w_{bg}), b\right) - P((w_{bb}, w_{bg}), b)\right] \\ &\quad \geq 0 \end{aligned}$$

The last inequality is implied by the fact that $P \in \mathcal{F}_3$ and the assumption that $R'(0) = \infty$, which implies $p_s R(\varepsilon) > \varepsilon$ for $\varepsilon > 0$ sufficiently small. (In particular, all we require is that ε be such that $p_s R(\varepsilon) \geq \varepsilon$), which defines a large interval containing the efficient levels of investment in both states.) This implies that $TP((v_b, v), s)$ is locally minimised at $v_b = v$, ie, $TP(\mathbf{v}, s) \in \mathcal{F}_3$. $\square$

We now show that T preserves functions that generate higher value in state g than in state b. In particular, let $\mathcal{F}_4 := \{P(\mathbf{v}, s) \in \mathcal{F} : P(\mathbf{v}, g) \geq P(\mathbf{v}, b)\}$.

Lemma B.4. If $P(\mathbf{v}, s) \in \mathcal{F}_4$, then $TP(\mathbf{v}, s) \in \mathcal{F}_4$.

Proof. Suppose $P(\mathbf{v}, s) \in \mathcal{F}_4$. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at $(\mathbf{v}, b)$. By [IC], we know

$$\begin{aligned} v_g &\geq R(k) - m_b + \delta[p_g w_{bg} + (1 - p_g) w_{bb}] \\ &\geq \delta[p_g w_{bg} + (1 - p_g) w_{bb}] \end{aligned}$$

The second inequality is from [LL] which requires $m_b \leq 0$. Let $m'_g = \delta[p_g w_{bg} + (1 - p_g) w_{bb}] - v_g + R(k)$ which implies $m'_g \leq R(k)$. Thus, the policy $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}_b)$

satisfies [PK_g] and [LL] for $s = g$. As the other constraints do not change, $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}_g) \in \Gamma(\mathbf{v}, b)$, which implies

$$\begin{aligned} TP(\mathbf{v}, b) &= -k + p_b R(k) + \delta p_b [P(\mathbf{w}_g, g) - P(\mathbf{w}_b, b)] + \delta P(\mathbf{w}_b, b) \\ &\geq -k + p_b R(k) + \delta p_b [P(\mathbf{w}_b, g) - P(\mathbf{w}_b, b)] + \delta P(\mathbf{w}_b, b) \end{aligned} \tag{B.3}$$

where the inequality reflects the choice of a (possibly) suboptimal policy. From [B.3] we then get $P(\mathbf{w}_g, g) - P(\mathbf{w}_b, b) \geq P(\mathbf{w}_b, g) - P(\mathbf{w}_b, b) \geq 0$ where the second inequality follows from the assumption that $P \in \mathcal{F}_4$. Moreover, because $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g) \in \Gamma(\mathbf{v}, g)$, we find

$$\begin{aligned} TP(\mathbf{v}, g) &\geq -k + p_g R(k) + \delta p_g [P(\mathbf{w}_g, g) - P(\mathbf{w}_b, b)] + \delta P(\mathbf{w}_b, b) \\ &\geq -k + p_b R(k) + \delta p_b [P(\mathbf{w}_g, g) - P(\mathbf{w}_b, b)] + \delta P(\mathbf{w}_b, b) \\ &= TP(\mathbf{v}, b) \end{aligned}$$

where the second line holds because $P(\mathbf{w}_g, g) - P(\mathbf{w}_b, b) \geq 0$ and $R(k) \geq 0$. Thus, $TP(\mathbf{v}, s) \in \mathcal{F}_4$ as claimed. □

Lemma B.5. Let $\mathcal{F}_5 = \{P(\mathbf{v}, s) \in \mathcal{F} : P(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq P(\mathbf{v}, s), \forall \varepsilon > 0\}$. Then $TP(\mathbf{v}, s) \in \mathcal{F}_5$. Moreover, $P(\mathbf{v}, s) \in \mathcal{F}_5$ implies there exists a policy with $m_b(\mathbf{v}, s) = 0$ that is optimal in [P1].

Proof. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at state $(\mathbf{v}, s)$. Since $(k, m_b - \varepsilon, m_g - \varepsilon, \mathbf{w}_b, \mathbf{w}_g) \in \Gamma(\mathbf{v} + (\varepsilon, \varepsilon), s)$ and m_b, m_g do not appear in the objective of [P1], we know $TP(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq TP(\mathbf{v}, s)$. (Notice that the proof does not require $P \in \mathcal{F}_5$. It only requires that $P \in \mathcal{F}$, which ensures the existence of a maximiser.)

To prove the second claim, let us suppose $m_b < 0$, and consider the policy $(k', m'_b, m'_g, \mathbf{w}'_b, \mathbf{w}'_g) = (k, 0, m_g, \mathbf{w}_b - (m_b, m_b)/\delta, \mathbf{w}_g)$. Obviously, $(k', m'_b, m'_g, \mathbf{w}'_b, \mathbf{w}'_g) \in \Gamma(\mathbf{v}, s)$. Since $P \in \mathcal{F}_5$, we know $P(\mathbf{w}'_b, b) \geq P(\mathbf{w}_b, b)$. Hence the new policy $(k', m'_b, m'_g, \mathbf{w}'_b, \mathbf{w}'_g)$ at least weakly increases the objective of [P1]. Thus, there exists a policy with $m_b = 0$ that is optimal, completing the proof. □

Because the optimal contract lies in the interior of the feasible set (in an appropriate sense), the continuous differentiability of TP and P follows from standard results as, for instance, in Stokey et al. (1989). To establish further properties of the value function, we will consider the optimality conditions for the problem [P1].

In what follows, $\hat{\eta}_b(\mathbf{v}, s)$ and $\hat{\eta}_g(\mathbf{v}, s)$ are the Lagrange multipliers for the promise keeping constraints [PK_b] and [PK_g] in [P1], $\hat{\lambda}(\mathbf{v}, s)$ is the Lagrange multiplier for the incentive constraint [IC], and $\hat{\mu}_b(\mathbf{v}, s)$ and $\hat{\mu}_g(\mathbf{v}, s)$ are the Lagrange multipliers for the limited liability constraints [LL] respectively. The first order conditions for [P1] are

$$\text{[BFOC}k\text{]} \qquad R'(k(\mathbf{v}, s)) = 1/[p_s - \hat{\eta}_g(\mathbf{v}, s) + \hat{\mu}_g(\mathbf{v}, s)]$$

$$\text{[BFOC}w_{bb}\text{]} \qquad (1 - p_s)P_b(\mathbf{w}_b(\mathbf{v}, s), b) = \hat{\eta}_b(\mathbf{v}, s)(1 - p_b) + \hat{\lambda}(\mathbf{v}, s)(1 - p_g)$$

$$\text{[BFOC}w_{bg}\text{]} \qquad (1 - p_s)P_g(\mathbf{w}_b(\mathbf{v}, s), b) = \hat{\eta}_b(\mathbf{v}, s)p_b + \hat{\lambda}(\mathbf{v}, s)p_g$$

$$\text{[BFOC}w_{gb}\text{]} \qquad p_s P_b(\mathbf{w}_g(\mathbf{v}, s), g) = \hat{\eta}_g(\mathbf{v}, s)(1 - p_g) - \hat{\lambda}(\mathbf{v}, s)(1 - p_g)$$

$$\text{[BFOC}w_{gg}\text{]} \qquad p_s P_g(\mathbf{w}_g(\mathbf{v}, s), g) = \hat{\eta}_g(\mathbf{v}, s)p_g - \hat{\lambda}(\mathbf{v}, s)p_g$$

The envelope conditions for [P1] are:

$$\text{[BEn}v_b\text{]} \qquad (TP)_b(\mathbf{v}, s) = \hat{\eta}_b(\mathbf{v}, s)$$

$$\text{[BEn}v_g\text{]} \qquad (TP)_g(\mathbf{v}, s) = \hat{\eta}_g(\mathbf{v}, s)$$

The following lemma establishes some properties of the Lagrange multipliers. Writing down the Lagrangean is straightforward and is hence omitted.

Lemma B.6. For any $(\mathbf{v}, s) \in V \times S$, the Lagrange multipliers in [P1] satisfy the following:

(a) The coefficient of m_b in the Lagrangean is $\hat{\eta}_b(\mathbf{v}, s) + \hat{\lambda}(\mathbf{v}, s) - \hat{\mu}_b(\mathbf{v}, s) \geq 0$,
(b) The complementary slackness condition $m_b(\mathbf{v}, s)[\hat{\eta}_b(\mathbf{v}, s) + \hat{\lambda}(\mathbf{v}, s) - \hat{\mu}_b(\mathbf{v}, s)] = 0$, and
(c) The coefficient of m_g in the Lagrangean is $\hat{\eta}_g(\mathbf{v}, s) - \hat{\lambda}(\mathbf{v}, s) - \hat{\mu}_g(\mathbf{v}, s) = 0$.

Proof. To see parts [a] and [b], notice first that m_b, m_g only appear in the Lagrangean for [P1] in a linear way. The term multiplying m_b in Lagrangean of [P1] is $\hat{\eta}_b + \hat{\lambda} - \hat{\mu}_b$. This term must be nonnegative. Otherwise, at $s = b$, maximizing the Lagrangean means that the optimal transfer is $m_b = -\infty$, since m_b is unbounded below. But this simply means the Lagrangean is unbounded above which is a contradiction since P is bounded above by $\bar{Q}$. Moreover, if $\hat{\eta}_b + \hat{\lambda} - \hat{\mu}_b > 0$ then maximizing the Lagrangean implies $m_b = 0$, because $m_b \leq 0$ by [LL]. Hence we always have $m_b(\hat{\eta}_b + \hat{\lambda} - \hat{\mu}_b) = 0$. These observations establish parts [a] and [b].

Now we show that part [c] holds. As m_g is unbounded below, using the observations in the paragraph above, we find that $\hat{\eta}_g - \hat{\lambda} - \hat{\mu}_g \geq 0$. Let

$(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at any $(\mathbf{v}, s)$. For any $\varepsilon > 0$, consider another policy: $(k', m_b', m_g', \mathbf{w}_b', \mathbf{w}_g') = \left(R^{-1}(R(k) + \varepsilon), m_b, m_g, \mathbf{w}_b, \mathbf{w}_g + \frac{\varepsilon}{\delta}(1, 1)\right)$. By definition,

$$
\begin{aligned}
(TP)_g(\mathbf{v}, s) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[TP((v_b, v_g + \varepsilon), s) - TP((v_b, v_g), s)] \\
&\geq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{ p_s R(k') - k' - p_s R(k) + k + p_s \delta \left[P\left(\mathbf{w}_g', g\right) - P(\mathbf{w}_g, g) \right] \right\} \\
&= p_s - \frac{1}{R'(k)} + p_s [P_b(\mathbf{w}_g, g) + P_g(\mathbf{w}_g, g)] \\
&= p_s - \frac{1}{R'(k)} + \hat{\eta}_g(\mathbf{v}, s) - \hat{\lambda}(\mathbf{v}, s) \qquad \text{[B.4]}
\end{aligned}
$$

The inequality in the second line is because $(k', m_b', m_g', \mathbf{w}_b', \mathbf{w}_g') \in \Gamma((v_b, v_g + \varepsilon), s)$. The equality in the third line is L'Hôpital's rule for the second term, while the third term is the directional derivative of $P(\mathbf{w}_g, g)$ in the direction $(1, 1)$. The last equality on the fourth line is obtained by adding [BFOCw_{gb}] and [BFOCw_{gg}]. Moreover,[BFOCk], [BEnv_g], and [B.4] together imply $\hat{\eta}_g(\mathbf{v}, s) - \hat{\lambda}(\mathbf{v}, s) - \hat{\mu}_g(\mathbf{v}, s) \leq 0$. In conjunction with our earlier observation that $\hat{\eta}_g - \hat{\lambda} - \hat{\mu}_g \geq 0$, this proves part [c]. □

Note that we can simplify [BFOCk] by using part [c] of Lemma B.6 as,

$$R'(k(\mathbf{v}, s)) = 1/[p_s - \hat{\lambda}(\mathbf{v}, s)] \qquad \text{[BFOC}k\text{]}$$

Let us say that a function $f : (0, 1) \to \mathbb{R}$ is *locally increasing* if for any $x \in (0, 1)$, there exists $\varepsilon > 0$ such that f is increasing on $[x, x + \varepsilon]$. (Here, increasing is taken to mean non-decreasing.) The following is a useful lemma that shows that continuous locally increasing functions on an interval are increasing.

Lemma B.7. Let $f : (0, 1) \to \mathbb{R}$ be continuous and locally increasing. Then, f is increasing, ie, for all $x, y \in (0, 1)$, $x \leq y$ implies $f(x) \leq f(y)$.

Proof. Let $x, y \in (0, 1)$ such that $x < y$. We want to show that $f(x) \leq f(y)$. Because f is continuous, f achieves a maximum, M, on $[x, y]$. Moreover, because f is continuous, the set $Z := \{z \in [x, y] : f(z) = M\}$ is closed. Let z^* be the supremum of Z, which is in Z because Z is closed. If $z^* = y$, we are done, because $f(x) \leq M$. If, however, $z^* < y$, then by the hypothesis that f is locally increasing, there exists $\varepsilon > 0$ (which depends on z^*) such that f is increasing on $[z^*, z^* + \varepsilon] \subset [x, y]$. This implies $f(z^* + \varepsilon) = M$, because M is the maximum value of f on $[x, y]$, which contradicts the definition of z^*, namely that z^* is the supremum of the set Z and that $z^* < y$. $\square$

The following lemma establishes that there exist optimal solutions to problem [P1] where the choice of $\mathbf{w}_g$ at $(\mathbf{v}, s)$ is independent of v_b.

Lemma B.8. Let $P \in \mathscr{F}$ and fix $(\mathbf{v}, s) \in V \times S$. Then, in the problem [P1], any $\mathbf{w}_g(\cdot, s)$ that is optimal at $\mathbf{v} \in [(v, 0), (v, v)]$ is also optimal at any other $\mathbf{v}' \in [(v, 0), (v, v)]$ for all $v \geq 0$. In this sense, every optimal solution $\mathbf{w}_g(\mathbf{v}, s)$ is independent of v_b.

Proof. Recall that problem [P1] entails solving

$$\text{[P1]} \qquad (TP)(\mathbf{v}, s) = \max_{k, m_i, \mathbf{w}_i} \big(-k + p_s[R(k) + \delta P(\mathbf{w}_g, g)] + (1 - p_s)\delta P(\mathbf{w}_b, b) \big)$$
$$\text{s.t.} \quad (k, m_i, \mathbf{w}_i) \in \Gamma(\mathbf{v}, s)$$

where $\Gamma(\mathbf{v}, s) = \{(k, m_i, \mathbf{w}_i) : (k, m_i, \mathbf{w}_i) \text{ implements } \mathbf{v} \text{ and } \mathbf{w}_i \in V\}$. To proceed with the proof, consider the following auxiliary problem

$$\text{[P2]} \qquad \max_{\mathbf{x} \geq 0} P(\mathbf{x}, g) \quad \text{s.t.} \quad y \geq \delta\big(p_s x_g + (1 - p_s) x_b\big)$$

Because P is continuous, and because the set of feasible $\mathbf{x}$ in problem [P2] is compact, it follows that there exists an optimal solution, that we denote by $\mathbf{x}^*(y, s)$.

Let $(k^*, m_i^*, \mathbf{w}_i^*) \in \Gamma(\mathbf{v}, s)$ be an optimal solution to [P1], and let $\mathbf{w}'_g := \mathbf{x}^*(v_g, s)$ be a solution to [P2] at $y = v_g$. We will show that $P(\mathbf{w}_g^*, g) = P(\mathbf{w}'_g, g)$. In particular, we will show that for a fixed $(\mathbf{v}, s) \in V \times S$, the following are true:

(a) If $(k^*, m_i^*, \mathbf{w}_i^*)$ is a solution to [P1], then $\mathbf{w}_g^*$ is a solution to [P2] at $y = v_g$.
(b) If $\mathbf{w}'_g$ is a solution [P2] at $y = v_g$, then there exists m'_g such that $(k^*, m_b^*, m'_g, \mathbf{w}_b^*, \mathbf{w}'_g)$ is feasible, ie, $(k^*, m_b^*, m'_g, \mathbf{w}_b^*, \mathbf{w}'_g) \in \Gamma(\mathbf{v}, s)$, and $(k^*, m_b^*, m'_g, \mathbf{w}_b^*, \mathbf{w}'_g)$ is a solution to [P1].

This will prove the lemma because every solution to [P2] is independent of v_b, so any $(k^*, m_i^*, \mathbf{w}_i^*)$ which is a solution to [P1] features $\mathbf{w}_g^*$ that is independent of v_b.

First, let $(k^*, m_i^*, \mathbf{w}_i^*)$ be a solution to [P1] and $\mathbf{w}'_g$ be a solution to [P2] at $y = v_g$. Notice that [PK_g] requires $v_g = R(k) - m_g + \delta\big(p_s w_{gg} + (1 - p_s) w_{gb}\big)$ while the limited liability constraint [LL] stipulates that $R(k) \geq m_g$. These two facts combine to give us $v_g \geq \delta\big(p_s w_{gg} + (1-p_s)w_{gb}\big)$, which implies $\mathbf{w}_g^*$ is feasible in the auxiliary problem [P2] at $y = v_g$. Optimality of $\mathbf{w}'_g$ at v_g in [P2] implies that $P(\mathbf{w}'_g, g) \geq P(\mathbf{w}_g^*, g)$.

Next, consider the policy $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}'_g)$ where $\mathbf{w}'_g = \mathbf{x}^*(v_g, s)$ as before, and $m'_g := R(k) + \delta\big(p_s w'_{gg} + (1-p_s)w'_{gb}\big) - v_g$. Notice that with this definition, $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}'_g) \in \Gamma(\mathbf{v}, s)$. But this feasible strategy cannot exhibit a higher value in the problem [P1] than the optimal strategy $(k^*, m_i^*, \mathbf{w}_i^*)$, which implies that $P(\mathbf{w}_g^*, g) \geq P(\mathbf{w}'_g, g)$.

This establishes that $P(\mathbf{w}_g^*, g) = P(\mathbf{w}'_g, g)$, that is, the optimal $\mathbf{w}_g(\mathbf{v}, s)$ in any solution to [P1] is also an optimal solution to [P2], thereby establishing (a). Moreover, $P(\mathbf{w}_g^*, g) = P(\mathbf{w}'_g, g)$ implies that $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}'_g) \in \Gamma(\mathbf{v}, s)$ is a solution to [P1], which establishes (b). This concludes the proof. □

We now establish that the operator T preserves supermodularity in $\mathbf{v}$ for all s. Let $\mathcal{F}_6 = \{P(\mathbf{v}, s) \in \mathcal{F}_1 \cap \mathcal{F}_5 : P_g(\mathbf{v}, s) \text{ is increasing in } v_b\}$.

Lemma B.9. If $P(\mathbf{v}, s) \in \mathcal{F}_6$, then $TP(\mathbf{v}, s) \in \mathcal{F}_6$.

Proof. Suppose $P(\mathbf{v}, s) \in \mathcal{F}_6$. Then $m_b(\mathbf{v}, s) = 0$ is optimal by Lemma B.5. From [IC*] we find that $\delta\Delta(w_{bg} - w_{bb}) \leq v_g - v_b - R(k)$. Using this in [PK$_b$] at any $(\mathbf{v}, s)$, we obtain

$$\text{[B.5]} \qquad w_{bb}(\mathbf{v}, s) \geq \frac{p_b R(k(\mathbf{v}, s)) + p_g v_b - p_b v_g}{\delta\Delta}$$

$$\text{[B.6]} \qquad w_{bg}(\mathbf{v}, s) \leq \frac{1-p_b}{\delta\Delta}\left[v_g - \frac{1-p_g}{1-p_b}v_b - R(k(\mathbf{v}, s))\right]$$

where the second inequality follows from the first and [PK$_b$]. Notice that [IC*] holds as equality if, and only if, both [B.5] and [B.6] hold as equalities.

Let $\varepsilon > 0$ and define $\mathbf{v}' = \mathbf{v} + (\varepsilon, 0)$. We will show in two cases below that for $\varepsilon > 0$ sufficiently small, $(TP)_g(\mathbf{v}', s) \geq (TP)_g(\mathbf{v}, s)$.

Case 1: [IC] holds as a strict inequality at $(\mathbf{v}, s)$. Complementary slackness implies $\hat{\lambda}(\mathbf{v}, s) = 0$. For an $\varepsilon > 0$ sufficiently small, continuity of the optimal policy (for the Lagrangean) implies that $\hat{\lambda} = 0$ on the interval $[\mathbf{v}, \mathbf{v}']$.[17] Moreover, from Lemma B.8, any $\mathbf{w}_g$ that is part of an optimal solution at any point on

(17) Here, $[\mathbf{v}, \mathbf{v}']$ denotes the closed interval connecting the vectors $\mathbf{v}$ and $\mathbf{v}'$, ie, $[\mathbf{v}, \mathbf{v}'] := \{t\mathbf{v} + (1-t)\mathbf{v}' : t \in [0, 1]\}$. Similarly, $[\mathbf{v}, \mathbf{v}')$ denotes the corresponding open interval.

the interval $[\mathbf{v}, \mathbf{v}']$ is also part of an optimal solution at every other point on the interval. That is, on this interval, the left hand side of [BFOCw_{gg}] does not vary with v_b. This implies, from [BFOCw_{gg}], that $\hat{\eta}_g(\mathbf{v}, s) = \hat{\eta}_g(\mathbf{v}', s)$. Then [BEn$v_g$] implies $(TP)_g(\mathbf{v}', s) = (TP)_g(\mathbf{v}, s)$.

Case 2: [IC] holds as equality at $(\mathbf{v}, s)$. We shall first show that $\hat{\lambda}(\mathbf{v}', s) \geq \hat{\lambda}(\mathbf{v}, s)$. If $\hat{\lambda}(\mathbf{v}, s) = 0$, this follows immediately because $\hat{\lambda}(\mathbf{v}', s) \geq 0$. So, suppose that $\hat{\lambda}(\mathbf{v}, s) > 0$ and $\hat{\lambda}(\mathbf{v}', s) < \hat{\lambda}(\mathbf{v}, s)$. By [BFOC$k$], we know $k(\mathbf{v}', s) > k(\mathbf{v}, s)$. From [B.5] at $(\mathbf{v}, s)$ and $(\mathbf{v}', s)$, we obtain

$$\begin{aligned} w_{bb}(\mathbf{v}', s) &\geq \frac{p_b R(k(\mathbf{v}', s)) + p_g v_b' - p_b v_g}{\delta \Delta} \\ &> \frac{p_b R(k(\mathbf{v}, s)) + p_g v_b - p_b v_g}{\delta \Delta} = w_{bb}(\mathbf{v}, s) \end{aligned}$$

where the equality is because [IC] holds as an equality at $(\mathbf{v}, s)$ by assumption. From [B.6] at $(\mathbf{v}, s)$ and $(\mathbf{v}', s)$, we obtain

$$\begin{aligned} w_{bg}(\mathbf{v}', s) &\leq \frac{1 - p_b}{\delta \Delta} \left[v_g - \frac{1 - p_g}{1 - p_b} v_b' - R(k(\mathbf{v}', s)) \right] \\ &< \frac{1 - p_b}{\delta \Delta} \left[v_g - \frac{1 - p_g}{1 - p_b} v_b - R(k(\mathbf{v}, s)) \right] = w_{bg}(\mathbf{v}, s) \end{aligned}$$

where, once again, the equality is because [IC*] holds as equality at $(\mathbf{v}, s)$. Because $P_g(\mathbf{v}, s)$ is increasing v_b (recall that $P \in \mathscr{F}_6$) and because P_g is decreasing in v_g (recall that by $P \in \mathscr{F}_6 \subset \mathscr{F}_1$ is concave in $\mathbf{v}$), we have $P_g(\mathbf{w}_b(\mathbf{v}', s), b) \geq P_g(\mathbf{w}_b(\mathbf{v}, s), b)$. This implies

$$\begin{aligned} \hat{\eta}_b(\mathbf{v}', s) &= (1 - p_s) P_g(\mathbf{w}_b(\mathbf{v}', s), b) - p_g \hat{\lambda}(\mathbf{v}', s) \\ &> (1 - p_s) P_g(\mathbf{w}_b(\mathbf{v}, s), b) - p_g \hat{\lambda}(\mathbf{v}, s) \\ &= \hat{\eta}_b(\mathbf{v}, s) \end{aligned}$$

where the first and last equalities obtain from [BFOCw_{bg}], while the strict inequality follows because, as noted above, $P \in \mathscr{F}_6$ and because $\hat{\lambda}(\mathbf{v}', s) < \hat{\lambda}(\mathbf{v}, s)$ (by assumption).

The envelope condition [BEnv_b] now implies $(TP)_b(\mathbf{v}', s) > (TP)_b(\mathbf{v}, s)$. However, from Lemma B.1, $TP(\mathbf{v}, s) \in \mathscr{F}_1$ and therefore concave in $\mathbf{v}$, which requires that $(TP)_b(\mathbf{v}', s) \leq (TP)_b(\mathbf{v}, s)$, a contradiction. Hence $\hat{\lambda}(\mathbf{v}', s) \geq \hat{\lambda}(\mathbf{v}, s)$.

By Lemma [B.8], the optimal policy $\mathbf{w}_g$ is constant on the interval $[\mathbf{v}, \mathbf{v}']$. Therefore, the left hand side of [BFOCw_{gg}] is constant on this interval, which implies $\hat{\eta}_g - \hat{\lambda}$ is constant on this interval. Having established that $\hat{\lambda}(\mathbf{v}', s) \geq \hat{\lambda}(\mathbf{v}, s)$, this implies $\hat{\eta}_g(\mathbf{v}', s) \geq \hat{\eta}_g(\mathbf{v}, s)$, which further implies $(TP)_g(\mathbf{v}', s) \geq (TP)_g(\mathbf{v}, s)$ by [BEnv$_g$]. Given the results in the two cases considered above, $(TP)_g(\mathbf{v}, s)$ is locally increasing in v_b. By Lemma B.7, $(TP)_g(\mathbf{v}, s)$ is therefore increasing in v_b, and so $TP(\mathbf{v}, s) \in \mathcal{F}_6$. $\square$

Theorem 4. *The unique fixed point of T, which we call Q, lies in $\bigcap_{i=1}^{6} \mathcal{F}_i$. Therefore, Q satisfies the following:*

(a) *$Q(\mathbf{v}, s)$ is concave in $\mathbf{v}$,*
(b) *$Q(\mathbf{v}, s) = \bar{Q}(s)$ for any $\mathbf{v} \in E_s$,*
(c) *$Q(\mathbf{v}, s)$ is decreasing in v_b at $((v, v), s)$ for any $v > 0$,*
(d) *$Q(\mathbf{v}, g) \geq Q(\mathbf{v}, b)$,*
(e) *$Q(\mathbf{v} + (\varepsilon, \varepsilon), s) \geq Q(\mathbf{v}, s)$, and*
(f) *$Q_g(\mathbf{v}, s)$ is increasing in v_b.*

Proof. It is easy to see that T is monotone (whereby $P_1 \leq P_2$ implies $TP_1 \leq TP_2$) and satisfies discounting (wherein $T(P + a) = TP + \delta a$), which implies T is a contraction mapping on $\mathcal{F}$ and hence has a unique fixed point in $\mathcal{F}$. We have established (in Lemmas B.1 through B.9) that if $P \in \bigcap_{i=1}^{6} \mathcal{F}_i$, then $TP \in \bigcap_{i=1}^{6} \mathcal{F}_i$. This implies the unique fixed point of T also lies in $\bigcap_{i=1}^{6} \mathcal{F}_i$. $\square$

C. Proofs from Section 4

In this section, we show various properties regarding value functions and unconstrained sets.

Proof of Theorem 1. (a) Because the only feasible policy at state $(\mathbf{0}, s)$ is

$$(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g) = (0, 0, 0, \mathbf{0}, \mathbf{0})$$

we must have $Q(\mathbf{0}, s) = 0$. We first show that $Q_g(\mathbf{0}, s) = \infty$, and $\mathsf{D}_{(1,1)}\, Q(\mathbf{0}, s) =$

∞. Then we use these two facts to show $Q_b((0,v),s) = \infty$. Note that

$$
\begin{aligned}
Q_g(\mathbf{0},s) &= \lim_{\varepsilon\to 0} \frac{Q((0,\varepsilon),g)}{\varepsilon} \\
&\geq \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[-R^{-1}(\varepsilon) + p_s\varepsilon + \delta p_s Q\left((0,\frac{\varepsilon}{\delta p_g}),g\right)\right] \\
&= p_s + \frac{p_s}{p_g} Q_g(\mathbf{0},g)
\end{aligned} \tag{C.1}
$$

The inequality is because

$$
(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g) = \left(R^{-1}(\varepsilon), 0, \varepsilon, \mathbf{0}, \left(0, \frac{\varepsilon}{\delta p_g}\right)\right) \in \Gamma((0,\varepsilon),s)
$$

Because $p_g > 0$, $Q_g(\mathbf{0},g)$ has to be ∞ for [C.1] to hold when $s = g$. Then $Q_g(\mathbf{0},b) = \infty$ is implied by [C.1] when $s = b$.
Next,

$$
\begin{aligned}
\mathsf{D}_{(1,1)}\, Q(\mathbf{0},s) &= \lim_{\varepsilon\to 0} \frac{\mathsf{D}_{(1,1)}\, Q((\varepsilon,\varepsilon),s)}{\varepsilon} \\
&\geq \lim_{\varepsilon\to 0} \frac{\delta}{\varepsilon}\left[p_s Q\left((0,\frac{\varepsilon}{\delta p_g}),g\right) + (1-p_s) Q\left((\frac{\varepsilon}{\delta},\frac{\varepsilon}{\delta}),b\right)\right] \\
&= \frac{p_s}{p_g} Q_g(\mathbf{0},g) + (1-p_s)\, \mathsf{D}_{(1,1)}\, Q(\mathbf{0},b)
\end{aligned} \tag{C.2}
$$

The inequality is because

$$
(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g) = \left(0,0,0,\left(\frac{\varepsilon}{\delta},\frac{\varepsilon}{\delta}\right),\left(0,\frac{\varepsilon}{\delta p_g}\right)\right) \in \Gamma((\varepsilon,\varepsilon),b)
$$

Because $Q_g(\mathbf{0},g) = \infty$, [C.2] at $s = b$ implies $\mathsf{D}_{(1,1)}\, Q(\mathbf{0},b) = \infty$. Then [C.2] at $s = g$ implies $\mathsf{D}_{(1,1)}\, Q(\mathbf{0},g) = \infty$.
Next, let $(k, 0, m_g, \mathbf{0}, \mathbf{w}_g)$ be the optimal policy at $((0,v),s)$ for any $v > 0$. The optimal $m_b = w_{bi} = 0$ is implied by [PK_b] at $v_b = 0$. Then we have

$$
\left(k, 0, m_g, \left(\frac{\varepsilon}{\delta},\frac{\varepsilon}{\delta}\right), \mathbf{w}_g\right) \in \Gamma((\varepsilon,v),s)
$$

which implies

$$
\begin{aligned}
Q_b((0,v),s) &= \lim_{\varepsilon\to 0} \frac{Q((\varepsilon,v),s) - Q((0,v),s)}{\varepsilon} \\
&\geq \lim_{\varepsilon\to 0} \frac{(1-p_s)\delta}{\varepsilon}\left[Q\left(\left(\frac{\varepsilon}{\delta},\frac{\varepsilon}{\delta}\right),b\right) - Q(\mathbf{0},b)\right] \\
&= (1-p_s)\, \mathsf{D}_{(1,1)}\, Q(\mathbf{0},b) = \infty
\end{aligned}
$$

Next, let's define $Q_b((v,v),s)$ as the left derivative since only $v_g \geq v_b$ is feasible. Using the same argument as in Lemma B.3, for arbitrary small $\varepsilon > 0$ and $v' = v - \frac{1-p_b}{1-p_g}R(\varepsilon)$, we get

$$
\begin{aligned}
&Q((v',v),b) - Q((v,v),b) \\
\text{[C.3]} \quad &\geq -\varepsilon + p_b R(\varepsilon) + (1-p_b)\delta\left[Q\left((w_{bg},w),b\right) - Q((w,w),b)\right] \geq 0
\end{aligned}
$$

where $w = \frac{v+m_b}{\delta}$, and $w_{bg} = w - \frac{R(\varepsilon)}{\delta(1-p_g)}$. Divide both sides of [C.3] by $\frac{1-p_b}{1-p_g}R(\varepsilon)$, and take the limit as ε converges to zero, we obtain:

$$
\text{[C.4]} \qquad -Q_b((v,v),b) \geq \frac{p_b(1-p_g)}{1-p_b} - Q_b((w,w),b)
$$

Because $Q_b((w,w),s) \leq 0$ by part (c) of Theorem 4, we know $-Q_b((v,v),b) \geq \frac{p_b(1-p_g)}{1-p_b}$. The same argument shows that $-Q_b((w,w),b) \geq \frac{p_b(1-p_g)}{1-p_b}$. So it must be that $-Q_b((v,v),b) \geq \frac{2p_b(1-p_g)}{1-p_b}$. Repeating this procedure, we get the result $-Q_b((v,v),b) \geq \frac{np_b(1-p_g)}{1-p_b}$ for any $n \in \mathbb{N}$. Hence, we must have $Q_b((v,v),b) = -\infty$. Now let $(k, m_b, m_g, (w,w), w_g)$ be the optimal policy at $((v,v),s), v > 0$, where $w = \frac{v+m_b}{\delta}$. And let $w'_{bb} = w - \varepsilon, m'_b = m_b - (1-p_b)\delta\varepsilon$, and $m'_g = m_g - \varepsilon$. Then we have

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

. which implies

$$
\begin{aligned}
Q_g((v,v),s) &\geq \lim_{\varepsilon\to 0} \frac{\delta(1-p_s)}{\varepsilon}\left[Q\left((w-\varepsilon,w),b\right) - Q((w,w),b)\right] \\
&= -\delta(1-p_s)Q_b((w,w),b) = \infty
\end{aligned}
$$

(b) Because it is always feasible to advance zero capital at all time, make repayments $m_g = -v_g, m_b = -v_b$ in the first period and no repayment in all subsequent periods, $Q(\mathbf{v},s) \geq 0$. Moreover, the surplus $Q(\mathbf{v},s)$ is uniformly bounded above by the efficient surplus $\bar{Q}(g)$.

(c) 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 $Q_g(\mathbf{v},s) \geq 0$.

(d) Shown in part (d) of Theorem 4.
(e) Recall that the proof of Lemma B.4 (needs reference) shows that any function in $\mathscr{F}_4$ satisfies this property. Because $Q \in \mathscr{F}_4$, so the result holds.
(f) Shown in part (f) of Theorem 4.

□

Lemma C.1. The efficient supluses 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 [??] and [??] that jointly determine $\bar{Q}(s)$. □

To properly define the threshold contingent utilities, we use 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 C.2. For each $v_b \geq 0$, there exists $f_s(v_b) > v_b$ such that $Q_g(\mathbf{v}, s) = 0$ if $v_g \geq f_s(v_b)$, and $Q_g(\mathbf{v}, s) > 0$ if $v_b \leq v_g < f_s(v_b)$. Moreover, $f_s(v_b)$ is increasing in v_b.

Proof. First, we show that for any $v_b \geq 0$, there exists some $v_g > v_b$ such that $Q_g(\mathbf{v}, s) = 0$. Note that part (b) of Theorem 4 shows that there exists some $\hat{\mathbf{v}}_s$ with $\hat{v}_{sg} > \hat{v}_{sb}$ such that $Q(\hat{\mathbf{v}}_s, s) = \bar{Q}(s)$ in a small neighborhood around $\hat{\mathbf{v}}_s$, implying $Q_g(\hat{\mathbf{v}}_s, s) = 0$. For any $v_b \leq \hat{v}_{sb}$, the supermodularity of Q implies that $0 \leq Q_g((v_b, \hat{v}_{sg}), s) \leq Q_g(\hat{\mathbf{v}}_s, s) = 0$. For any $v_b > \hat{v}_{sb}$, we consider $v_g = v_b + \hat{v}_{sg} - \hat{v}_{sb}$. Part (b) of Theorem 4 shows that $Q(\mathbf{v}, s) = \bar{Q}(s)$ in a small neighborhood around $\mathbf{v}$, implying $Q_g(\mathbf{v}, s) = 0$.

Next, we fix any $v_b \geq 0$ and define $f_s(v_b) := \min\{v_g : Q_g(\mathbf{v}, s) = 0\}$. Because $Q_g((v_b, v_b), s) = \infty$, we must have $f_s(v_b) > v_b$. By this definition, $Q_g(\mathbf{v}, s) > 0$ if $v_b \leq v_g < f_s(v_b)$. Moreover, the concavity of Q implies that $0 \leq Q_g(\mathbf{v}, s) \leq Q_g[(v_b, f_s(v_b)), s] = 0$ if $v_g \geq f_s(v_b)$.

Now we show that $f_s(.)$ is increasing. Take any v_b, v_b' such that $0 \leq v_b < v_b'$. Supermodularity of Q implies $0 \leq Q_g[(v_b, f(v_b')), s] \leq Q_g[(v_b', f(v_b')), s] = 0$. So we have $Q_g[(v_b, f_s(v_b')), s] = 0$, which further implies that $f_s(v_b) \leq f_s(v_b')$ by the definition of $f_s(.)$. $\square$

Lemma C.3. For any $v_b \geq 0$, we have $Q_b[(v_b, f_s(v_b)), s] \geq 0$. Moreover, there exists some $\hat{\mathbf{v}}_s$ in the interior of V such that $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] = 0$.

Proof. Note that the definition of $f_s(.)$ means $Q_g[(v_b, f_s(v_b), s] = 0$. So we have $Q_b[(v_b, f_s(v_b)), s] = Q_b[(v_b, f_s(v_b), s] + Q_g[(v_b, f_s(v_b), s] \geq 0$ for any $v_b \geq 0$. The inequality is implied by part (e) of Theorem 4. Moreover, part (b) of Theorem 4 implies that there exists some $\hat{\mathbf{v}}_s$ in the interior of V such that $Q_b(\hat{\mathbf{v}}_s, s) = Q_g(\hat{\mathbf{v}}_s, s) = 0$. By the definition of $f_s(.)$, we know $f_s(\hat{v}_{sb}) \leq \hat{v}_{sg}$. Supermodularity of Q then implies $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] \leq Q_b(\hat{\mathbf{v}}_s, s) = 0$. So it has to be that $Q_b[(\hat{v}_{sb}, f_s(\hat{v}_{sb})), s] = 0$. $\square$

We are now ready to define the threshold contingent utilities and the unconstrained sets. Let $\bar{v}_{sb} = \min\{v_b : Q_b((v_b, f_s(v_b)), s) = 0\}$ be the threshold of continuation utility contingent on bad shock. From Theorem 1, $Q_b((0, f(0)), s) = \infty$. So by definition $\bar{v}_{sb} > 0$. Let $\bar{v}_{sg} = f_s(\bar{v}_{sb})$ be the threshold continuation utility contingent on good shock. The unconstrained sets of contingent utilities are defined as $E_s := \{\mathbf{v} \in V : v_b \geq \bar{v}_{sb}, v_g \geq f_s(v_b)\}$.

Proof of Proposition 4.2. (a) Take any $\mathbf{v} \in E_s$. The definition of E_s simply requires $v_b \geq \bar{v}_{sb}$. Because $f_s(.)$ is increasing, we have $v_g \geq f_s(v_b) \geq f_s(\bar{v}_{sb}) = \bar{v}_{sg}$.

(b) We first show that $Q_b((\bar{v}_{sb}, v_g), s) = 0$ if $v_g \geq \bar{v}_{sg}$. For each $v_b \leq \bar{v}_{sb}$, because $f_s(.)$ is increasing, we know $f_s(v_b) \leq f_s(\bar{v}_{sb}) = \bar{v}_{sg}$. This means $Q_g(\mathbf{v}, s) = 0$ when $v_b \leq \bar{v}_{sb}$ and $v_g \geq \bar{v}_{sg}$. Take any $\hat{v}_g > \bar{v}_{sg}$ and any small $\varepsilon > 0$. We have $Q((\bar{v}_{sb}, \hat{v}_g), s) = Q((\bar{v}_{sb}, \bar{v}_{sg}), s)$ and $Q((\bar{v}_{sb} - \varepsilon, \hat{v}_g), s) = Q((\bar{v}_{sb} - \varepsilon, \bar{v}_{sg}), s)$, which implies

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

Now take any $\mathbf{v} \in E_s$. Concavity of Q implies $Q_b(\mathbf{v}, s) \leq Q_b((\bar{v}_{sb}, v_g), s) = 0$. The definition of $f_s(.)$ and E_s simply implies that $Q_g(\mathbf{v}, s) = 0$. Moreover, because $Q_b(\mathbf{v}, s) + Q_g(\mathbf{v}, s) \geq 0$ by part (e) of Theorem 4, $Q_g(\mathbf{v}, s) = 0$ further implies $Q_b(\mathbf{v}, s) \geq 0$. Therefore, it has to be $Q_b(\mathbf{v}, s) = 0$.

(c) Take any $\mathbf{v} \in E_s$. Since Q is concave, $Q_b(\mathbf{v}, s) = Q_g(\mathbf{v}, s) = 0$ implies $Q(\mathbf{v}, s)$ achieves its maximum $\bar{Q}(s)$. Suppose $k(\mathbf{v}, s) < \bar{k}_s$. By the definition of $Q(\mathbf{v}, s)$ and $\bar{Q}(s)$, we have:

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

which is a contradiction. So $k(\mathbf{v}, s) = \bar{k}_s$.

(d) Take any $\mathbf{v} \in V \setminus E_s$. There are two possible cases. First, if $v_g < f_s(v_b)$, then by the definition of f_s, $Q_g(\mathbf{v}, s) > 0$. Second, if $v_b < \bar{v}_{sb}$ and $v_g \geq f_s(v_b)$, then the supermodularity of Q implies $Q_b(\mathbf{v}, s) \geq Q_b[(v_b, f_s(v_b)), s] > 0$. The strict inequality is from the definition of $\bar{v}_{sb}$. So we must have either $Q_g(\mathbf{v}, s) > 0$ or $Q_b(\mathbf{v}, s) > 0$, which implies $Q(\mathbf{v}, s) < \bar{Q}(s)$.

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

$$
w_{bg} - w_{bb} < \frac{\delta\bar{v}_{bg} - v_b}{\delta(1 - p_b)} = \frac{\delta\bar{v}_{bg} + m_b - \delta[p_b w_{bg} + (1 - p_b)w_{bb}]}{\delta(1 - p_b)} \tag{C.5}
$$

The equality is from [PK$_b$]. Rearranging [C.5] 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 $v_g - v_b \geq R(\bar{k}_s) + \delta\Delta\frac{\delta\bar{v}_{bg} - v_b}{\delta(1-p_b)}$.
Combing the above results, 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 $v \in E_s$.

□

D. Auxiliary Problem

To proceed the proof in Section 5 and beyond, it is convenient to define an auxiliary problem:

[P2] $$\Psi(z,s) = \max_{y \geq x \geq 0} \delta Q((x,y),s)$$
$$\text{s.t.} \quad z \geq \delta(p_s y + (1-p_s)x)$$

where $z \geq 0$ and $s = \{b,g\}$. To ease notation, let $x^*(z,s), y^*(z,s)$ be the optimal solution of problem [P2]. Also let $\bar{z}_b = \bar{v}_b^b, \bar{z}_g = \delta[p_g \bar{v}_g^g + (1-p_g)\bar{v}_b^g]$. In this part of the Appendix, we will show some useful properties of function $\Psi(z,s)$.

Lemma D.1. Function $\Psi(z,s)$ defined in [P2] satisfies:

(a) $\Psi(z,s)$ is increasing and concave in z;
(b) $\Psi(z,s)$ is strictly increasing in z and the constraint of [P2] binds when $z \in [0,\bar{z}_s)$;
(c) $\Psi_z(z,s) = \mathsf{D}_{(1,1)}\, Q((x^*(z,s), y^*(z,s)),s)$.

Proof. (a) Since rasing z always relaxes the constraint in [P2], we have $\Psi_z(z,s) \geq 0$. The objective of [P2] is concave by Theorem 1, and its constraint is convex. So $\Psi(z,s)$ is concave in z.

(b) Suppose there exists some $0 \leq \tilde{z}_s < \bar{z}_s$ such that $\Psi_z(\tilde{z}_s,s) = 0$. Concavity of Ψ implies that $\Psi(\tilde{z}_s,s) = \Psi(\bar{z}_s,s) = \delta\bar{Q}(s)$. The last equality is because $x = \bar{v}_b^s, y = \bar{v}_g^s$ are feasible at $z = \bar{z}_s$, and $\bar{Q}(s)$ is the upper bound of $Q((x,y),s)$. From Proposition 4.2, we know $x^*(\tilde{z}_s,s) \geq \bar{v}_b, y^*(\tilde{z}_s,s) \geq \bar{v}_g^s$. However, this implies $\tilde{z}_s \geq \bar{z}_s$, a contradiction. Therefore, $\Psi(z,s)$ is strictly increasing in z when $z < \bar{z}_s$. Let γ_s be the Lagrange multiplier of [P2]. The envelope condition is: $\Psi_z(z,s) = \gamma_s$. So when $z < \bar{z}_s$, we know $\gamma_s > 0$ and therefore the constraint of [P2] binds by complementary slackness.

(c) The first order conditions for [P2] are: $Q_b((x^*(z,s), y^*(z,s)),s) = (1-p_s)\gamma_s$ and$Q_g((x^*(z,s), y^*(z,s)),s) = p_s\gamma_s$. The envelope condition for problem [P2] implies $\Psi_z(z,s) = \gamma_s = \mathsf{D}_{(1,1)}\, Q((x^*(z,s), y^*(z,s)),s)$.

□

Lemma D.2. The optimal policy $\mathbf{w}_b(\mathbf{v}, s), \mathbf{w}_g(\mathbf{v}, s)$ satisfies:

(a) $\mathbf{w}_g(\mathbf{v}, s)$ is the solution of [P2] at (v_g, g), hence only a function of v_g;
(b) If $\lambda(\mathbf{v}, s) = 0$, then $\mathbf{w}_b(\mathbf{v}, s)$ is the solution of [P2] at (v_b, b), hence only a function of v_b.

Proof. (a) Note that w_{gb}, w_{gg} only appear in the constraint [PK_g] in problem [VF], if we consider the incentive compatibility constraint as [IC*]. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at $(\mathbf{v}, s)$. Suppose $\mathbf{w}_g$ is not the solution of [P2] at (v_g, g). Then consider $\mathbf{w}'_g = (x^*(v_g, g), y^*(v_g, g))$ and $m'_g = R(k) + \delta(p_g y^*(v_g, g) + (1 - p_g)x^*(v_g, g) - v_g \leq R(k)$. We have

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

The new policy strictly increases the objective of [VF], a contradiction.
(b) If $\lambda(\mathbf{v}, s) = 0$, then the same argument as in (i) shows that $\mathbf{w}_b(\mathbf{v}, s)$ must be the solution to [P2] at (v_b, b).

□

The first order conditions and envelope conditions of problem [VF] are:

[FOCk]	$R'(k(\mathbf{v}, s)) = 1/[p_s - \eta_g(\mathbf{v}, s) + \mu_g(\mathbf{v}, s)]$
[FOCw_{bb}]	$(1 - p_s)Q_b(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s)(1 - p_b) + \lambda(\mathbf{v}, s)(1 - p_g)$
[FOCw_{bg}]	$(1 - p_s)Q_g(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s)p_b + \lambda(\mathbf{v}, s)p_g$
[FOCw_{gb}]	$p_s Q_b(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s)(1 - p_g) - \lambda(\mathbf{v}, s)(1 - p_g)$
[FOCw_{gg}]	$p_s Q_g(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s)p_g - \lambda(\mathbf{v}, s)p_g$

[Envv_b]	$Q_b(\mathbf{v}, s) = \eta_b(\mathbf{v}, s)$
[Envv_g]	$Q_g(\mathbf{v}, s) = \eta_g(\mathbf{v}, s)$

E. Strict Concavity

In this section we show the strict concavity of $Q(\mathbf{v}, s)$ and Ψ on the set $H := \{(\mathbf{v}, s) \in V \times S : Q_b(\mathbf{v}, s) > 0, Q_g(\mathbf{v}, s) > 0\}$.

Given history h^{t-1}, let $k^t(h^{t-1}; \mathbf{v}, s)$ be the optimal capital advancement, and $\mathbf{w}_i^t(h^{t-1}; \mathbf{v}, s)$ be the optimal contingent utilities generated from the policy functions starting at state $(\mathbf{v}, s)$.

Lemma E.1. For $(\tilde{\mathbf{v}}, s), (\hat{\mathbf{v}}, s) \in V \times S$ with $\tilde{\mathbf{v}} \neq \hat{\mathbf{v}}$, and $\theta \in (0,1)$, if $k^t(h^{t-1}; \hat{\mathbf{v}}, s) \neq k^t(h^{t-1}; \tilde{\mathbf{v}}, s)$ for some history h^{t-1}, then $Q(\theta\hat{\mathbf{v}} + (1-\theta)\tilde{\mathbf{v}}, s) > \theta Q(\hat{\mathbf{v}}, s) + (1-\theta)Q(\tilde{\mathbf{v}}, s)$.

Proof. Define the average policies as:

$$\bar{k}^t(h^{t-1}) = \theta\hat{k}^t(h^{t-1}; \hat{\mathbf{v}}, s) + (1-\theta)\tilde{k}^t(h^{t-1}; \tilde{\mathbf{v}}, s)$$
$$\bar{\mathbf{w}}_i^t(h^{t-1}) = \theta\hat{\mathbf{w}}_i^t(h^{t-1}; \hat{\mathbf{v}}, s) + (1-\theta)\tilde{\mathbf{w}}_i^t(h^{t-1}; \tilde{\mathbf{v}}, s)$$

We change the control variable k to be $R = R(k)$. So $k = R^{-1}(R) = C(R)$. Then $C(.)$ is stictly convex and all the constraints are linear in R. Other control variables and the program will not change since R and k are one-to-one. Let us consider iterating T times of the Bellman operator starting from states $(\hat{v}, s)$ and $(\tilde{v}, s)$.

$$\begin{aligned} Q(\hat{v}, s) = &\sum_{t=0}^{T} \delta^t \mathbb{E}_0[-C(\hat{R}^t(h^{t-1})) + p_{s^{t-1}}\hat{R}^t(h^{t-1})] \\ &+ \delta^{T+1}\mathbb{E}_0[p_{s^{T-1}}Q(\hat{\mathbf{w}}_g^T(h^{T-1}), g) + (1-p_{s^{T-1}})Q(\hat{\mathbf{w}}_b^T(h^{T-1}), b)] \end{aligned}$$

$$\begin{aligned} Q(\tilde{v}, s) = &\sum_{t=0}^{T} \delta^t \mathbb{E}_0[-C(\tilde{R}^t(h^{t-1})) + p_{s^{t-1}}\tilde{R}^t(h^{t-1})] \\ &+ \delta^{T+1}\mathbb{E}_0[p_{s^{T-1}}Q(\tilde{\mathbf{w}}_g^T(h^{T-1}), g) + (1-p_{s^{T-1}})Q(\tilde{\mathbf{w}}_b^T(h^{T-1}), b)] \end{aligned}$$

Averaging for large enough T, we obtain

$$\begin{aligned} &\theta Q(\hat{\mathbf{v}}, s) + (1-\theta)Q(\tilde{\mathbf{v}}, s) \\ &= \sum_{t=0}^{T} \delta^t \mathbb{E}_0[-\theta C(\hat{R}^t(h^{t-1})) - (1-\theta)C(\tilde{R}^t(h^{t-1})) + p_{s^{t-1}}\bar{R}^t(h^{t-1})] \\ &\quad + \theta\delta^{T+1}\mathbb{E}_0[p_{s^{T-1}}Q(\hat{\mathbf{w}}_g^T(h^{T-1}), g) + (1-p_{s^{T-1}})Q(\hat{\mathbf{w}}_b^T(h^{T-1}), b)] \\ &\quad + (1-\theta)\delta^{T+1}\mathbb{E}_0[p_{s^{T-1}}Q(\tilde{\mathbf{w}}_g^T(h^{T-1}), g) + (1-p_{s^{T-1}})Q(\tilde{\mathbf{w}}_b^T(h^{T-1}), b)] \\ &< \sum_{t=0}^{T} \delta^t \mathbb{E}_0[-C(\bar{R}^t(h^{t-1})) + p_{s^{t-1}}\bar{R}^t(h^{t-1})] \\ &\quad + \delta^{T+1}\mathbb{E}_0[p_{s^{T-1}}Q(\bar{\mathbf{w}}_g^T(h^{T-1}), g) + (1-p_{s^{T-1}})Q(\bar{\mathbf{w}}_b^T(h^{T-1}), b)] \\ &\le Q(\theta\hat{\mathbf{v}} + (1-\theta)\tilde{\mathbf{v}}, s) \end{aligned}$$

The strict equality follows from the facts that $-C(.)$ is strictly concave and that the assumption that $\hat{R}^t(h^{t-1}) \neq \tilde{R}^t(h^{t-1})$ for some history h^{t-1}. The weak equality follows from the fact that the average plan $(\bar{R}^t, \bar{\mathbf{w}}_i^t)$ satisfy the constraints of the Bellman equation at every step of the iteration. $\square$

We restrict attention to the subset of domain $H := \{(\mathbf{v}, s) \in V \times S : Q_b(\mathbf{v}, s) > 0, Q_g(\mathbf{v}, s) > 0\}$, because the states induced by the optimal contract always locate in set H before reaching their thresholds.

Proposition E.2. Let $(\mathbf{v}^{(0)}, s_{-1})$ be the initial equity holdings of the agent, and let $(\mathbf{v}^{(n)}, s_{n-1})$ denote the evolution of $(\mathbf{v}, s)$ induced by the optimal contract. If $(\mathbf{v}^{(0)}, s_{-1}) \in H$ and if $(\mathbf{v}^{(n)}, s_{n-1}) \notin E_s$, then $(\mathbf{v}^{(n)}, s_{n-1}) \in H$.

Proof. We show this statement by induction. First, optimality at the initialization of the optimal contract implies $Q_b(\mathbf{v}^0, s^{-1}) = 1 - p_s > 0$ and $Q_b(\mathbf{v}^0, s^{-1}) = p_s > 0$, where $s^{-1} = s$. So $(\mathbf{v}^0, s^{-1}) \in H$. Suppose that $(\mathbf{v}^t, s^{t-1}) \in H$ for $t \geq 1$. If a bad shock occurs at time t, then from [FOCw_{bb}] and [FOCw_{bg}] we know $Q_b(\mathbf{v}^{t+1}, b) > Q(\mathbf{v}^t, s^{t-1})(1 - p_b)/(1 - p_s) > 0$, and $Q_g(\mathbf{v}^{t+1}, b) > Q(\mathbf{v}^t, s^{t-1})p_b/(1 - p_s) > 0$, where $s^{t-1} = s$. So $(\mathbf{v}^{t+1}, b) \in H$. If a good shock occurs at time t and the efficient set E_g is not achieved, then from [FOCw_{gb}], [FOCw_{gg}] and [G.2], we know $Q_b(\mathbf{v}^{t+1}, g) = (1 - p_g)\mu_g(\mathbf{v}^t, s^{t-1})/p_s \geq 0$ and $Q_g(\mathbf{v}^{t+1}, g) = p_g\mu_g(\mathbf{v}^t, s^{t-1})/p_s \geq 0$, where $s^{t-1} = s$. And since the efficient sets are not reached, we also know at least one of $Q_b(\mathbf{v}^t, s^{t-1})$, $Q_g(\mathbf{v}^t, s^{t-1})$ is positive. If $Q_b(\mathbf{v}^{t+1}, g) > 0$, then we must have $\mu_g(\mathbf{v}^t, s^{t-1}) > 0$, which further implies $Q_g(\mathbf{v}^{t+1}, g) > 0$. The same argument shows that if $Q_g(\mathbf{v}^{t+1}, g) > 0$ then $Q_b(\mathbf{v}^{t+1}, g) > 0$. Hence, we must have $Q_b(\mathbf{v}^{t+1}, g), Q_b(\mathbf{v}^{t+1}, g) > 0$, which means $(\mathbf{v}^{t+1}, g) \in H$. Therefore, $(\mathbf{v}^t, s) \notin E_s$ implies $(\mathbf{v}^t, s) \in H$. $\square$

Lemma E.3. For any $(\mathbf{v}, s) \in H$, $Q(\mathbf{v}, s)$ is strictly concave in both v_b and v_g.

Proof. Take $(\tilde{\mathbf{v}}, s), (\hat{\mathbf{v}}, s) \in H$ with $\tilde{v}_g < \hat{v}_g$ and $\tilde{v}_b = \hat{v}_b$. Then $Q(\hat{\mathbf{v}}, s) > Q(\tilde{\mathbf{v}}, s)$. This further means that the optimal capital advancement $k^t(h^{t-1}; \hat{\mathbf{v}}, s) \neq k^t(h^{t-1}; \tilde{\mathbf{v}}, s)$ for some history h^{t-1}. Otherwise, the surplus will be the same starting at $(\tilde{\mathbf{v}}, s)$ and $(\hat{\mathbf{v}}, s)$. Then Lemma E.1 implis $Q(\mathbf{v}, s)$ is strictly concave in v_g for $(\mathbf{v}, s) \in H$. The same argument shows $Q(\mathbf{v}, s)$ is strictly concave in v_b for for $(\mathbf{v}, s) \in H$. $\square$

Lemma E.4. For any $(\mathbf{v}, s) \in A_{1,s}$, investment $k(\mathbf{v}, s)$ is strictly increasing in v_g.

Proof. For any $(\mathbf{v}, s) \in A_{1,s}$, left hand side of [FOC$w_{gg}$] is zero. Then $la_g(\mathbf{v}, s)$ is strictly decreasing in v_g, since by Lemma E.3 $\eta_g(\mathbf{v}, s)$ is strictly decreasing in v_g. From [FOCk], $k(\mathbf{v}, s)$ is strictly increasing in v_g. $\square$

Proposition E.5. For $(\tilde{\mathbf{v}}, s), (\hat{\mathbf{v}}, s) \in H$ with $\tilde{\mathbf{v}} \neq \hat{\mathbf{v}}$ and $\theta \in (0, 1)$, $Q(\theta\hat{\mathbf{v}} + (1 - \theta)\tilde{\mathbf{v}}, s) > \theta Q(\hat{\mathbf{v}}, s) + (1 - \theta)Q(\tilde{\mathbf{v}}, s)$.

Proof. By Lemma E.1, it suffices to show that $k^t(h^{t-1}; \hat{\mathbf{v}}, s) \neq k^t(h^{t-1}; (\tilde{\mathbf{v}}, s))$ for some history h^{t-1}. Suppose it is not true. The surplus and capital advancement must be the same after any history starting at the two initial states. This means either $\tilde{v}_b > \hat{v}_b$ and $\tilde{v}_g < \hat{v}_g$, or $\tilde{v}_b < \hat{v}_b$ and $\tilde{v}_g > \hat{v}_g$. Otherwise, $Q(\hat{\mathbf{v}}, s) \neq Q(\tilde{\mathbf{v}}, s)$ and the conclusion is obtained. Without loss of generality, we assume $\tilde{v}_b > \hat{v}_b, \tilde{v}_g < \hat{v}_g$. Moreover, at history $h^0 = \{s, g\}$, we know $Q(\mathbf{w}_g^0(h^0; \hat{\mathbf{v}}, s), g) = Q(\mathbf{w}_g^0(h^0; \tilde{\mathbf{v}}, s), g)$, which implies $\Psi(\hat{v}_g, g) = \Psi(\tilde{v}_g, g)$ by Lemma D.2. If $(\tilde{\mathbf{v}}, s) \notin A_{1,s}$, then Lemma D.1 implies that we must have $\hat{v}_g = \tilde{v}_g$ since $\Psi(z, g)$ is strictly increasing in $z \in [0, \delta[p_g\bar{v}_g^g + (1 - p_g)\bar{v}_b])$, a contradiction. Consider the case of $(\tilde{\mathbf{v}}, s) \in A_{1,s}$. Then we know $(\hat{\mathbf{v}}, s) \in A_{1,s}$. Moreover, $k(\hat{\mathbf{v}}, s) > k((\hat{v}_b, \tilde{v}_g), s) \geq k(\tilde{\mathbf{v}}, s)$. The first inequality is by Lemma E.4, and the second inequality is by the fact $k(\mathbf{v}, s)$ is deceasing in v_b which is implied by the supermodularity of Q in Theorem 4. This forms a contradiction. $\square$

Lemma E.6. For any $z \in [0, \bar{z}_s)$, $\Psi(z, s)$ is strictly concave in z, where $\bar{z}_b = \bar{v}_b^b$ and $\bar{z}_g = \delta[p_g\bar{v}_g^g + (1 - p_g)\bar{v}_b^g]$

Proof. For any $z < \bar{z}_s$, $((x^*(z, s), y^*(z, s)), s) \in H$. And from Lemma D.1, when $z < \bar{z}_s$, $\delta[p_s y^*(z, s) + (1 - p_s)x^*(z, s)] = z$. So $(x^*(\hat{z}, s), y^*(\hat{z}, s)) \neq (x^*(\tilde{z}, s), y^*(\tilde{z}, s))$ for $\hat{z}, \tilde{z} < \bar{z}_s$ and $\hat{z} \neq \tilde{z}$. Then Proposition E.5 implies $\theta\Psi(\hat{z}, s) + (1 - \theta)\Psi(\tilde{z}, s) < \Psi(\theta\hat{z} + (1 - \theta)\tilde{z}, s)$. Therefore, $\Psi(z, s)$ is strictly concave when $z < z_s$. $\square$

F. Properties of the Directional Derivative

In this section, we show the directional derivative of the firm's value function is a nonnegative martingale and how it evolves in the optimal contract. We also show the directional derivative must split at the good state.

Lemma F.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. 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

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

where $\mathbf{w}_i = \mathbf{w}_i(\mathbf{v}, s)$. Moreover, by result (f) of Theorem 4, $\mathsf{D}_{(1,1)}\,Q(\mathbf{v}, s) = \lim_{\varepsilon\to 0}[Q(\mathbf{v} + (\varepsilon, \varepsilon), s) - Q(\mathbf{v}, s)] \geq 0$. So the process $\mathsf{D}_{(1,1)}\,Q$ is a nonnegative martingale. □

Using the martingale relation and first order conditions we can characterize the evolution of directional derivative martingale $\mathsf{D}_{(1,1)}\,Q$ on any optimal path in the following Lemma.

Lemma F.2. On any path induced by the optimal contract, the martingale $\mathsf{D}_{(1,1)}\,Q$ evolves according to:

$$\mathsf{D}_{(1,1)}\,Q(\hat{\mathbf{w}}_g, g) = \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g) - \frac{1}{p_g}\lambda(\mathbf{w}_g, g) \tag{F.2}$$

$$\mathsf{D}_{(1,1)}\,Q(\hat{\mathbf{w}}_b, b) = \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g) + \frac{1}{1 - p_g}\lambda(\mathbf{w}_g, g) \tag{F.3}$$

$$\mathsf{D}_{(1,1)}\,Q(\tilde{\mathbf{w}}_b, b) = \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b, b) + \frac{(1 - p_s)\lambda(\mathbf{w}_b, b) - \Delta\lambda(\mathbf{v}, s)}{(1 - p_b)(1 - p_s)} \tag{F.4}$$

$$\mathsf{D}_{(1,1)}\,Q(\tilde{\mathbf{w}}_g, g) = \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b, b) - \frac{(1 - p_s)\lambda(\mathbf{w}_b, b) - \Delta\lambda(\mathbf{v}, s)}{p_b(1 - p_s)} \tag{F.5}$$

where $\mathbf{w}_i = \mathbf{w}_i(\mathbf{v}, s)$, $\hat{\mathbf{w}}_i = \mathbf{w}_i(\mathbf{w}_g, g)$, $\tilde{\mathbf{w}}_i = \mathbf{w}_i(\mathbf{w}_b, b)$.

Proof. From [FOCw_{gg}] and [Env$_g$]:

$$p_s[\eta_g(\mathbf{w}_g, g) - \lambda(\mathbf{w}_g, g)] = p_g[\eta_g(\mathbf{v}, s) - \lambda(\mathbf{v}, s)] - p_s\lambda(\mathbf{w}_g, g) \tag{F.6}$$

Add [FOCw_{gb}] and [FOCw_{gg}] at time $t - 1$ and time t respectively to get:

$$p_s\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g) = \eta_g(\mathbf{v}, s) - \lambda(\mathbf{v}, s) \tag{F.7}$$

$$p_g\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_g, g) = \eta_g(\mathbf{w}_g, g) - \lambda(\mathbf{w}_g, g) \tag{F.8}$$

Then combine [F.6], [F.7], [F.8] and rearrange to get [F.2]. Next, from [FOCw_{bb}]:

$$\begin{aligned}&(1 - p_s)[\eta_b(\mathbf{w}_g, b) + \lambda(\mathbf{w}_b, b)]\\ &\quad = (1 - p_b)[\eta_b(\mathbf{v}, s) + \lambda(\mathbf{v}, s)] - \Delta\lambda(\mathbf{v}, s) + (1 - p_s)\lambda(\mathbf{w}_b, b)\end{aligned} \tag{F.9}$$

Add [FOCw_{bb}] and [FOCw_{bg}] respectively to get:

$$\text{[F.10]} \qquad (1-p_s)\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b, b) = \eta_b(\mathbf{v}, s) + \lambda(\mathbf{v}, s)$$

$$\text{[F.11]} \qquad (1-p_b)\,\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_b, b) = \eta_b(\mathbf{w}_b, b) + \lambda(\mathbf{w}_b, b)$$

Then combine [F.9],[F.10], [F.11] and rearrange to get [F.4]. Combe the martingale equation [F.1] at state $(\mathbf{w}_g, g)$ and equation F.2 to obtain [F.3]. Similarly, combine the martingale equation [F.1] at state $(\mathbf{w}_b, b)$ and equation F.4 to obtain [F.5] □

Let $\mathbf{w}_g = \mathbf{w}_g(\mathbf{v}, s)$ for any $(\mathbf{v}, s) \in V \times S$, and $\mathbf{w}'_g = \mathbf{w}_g(\mathbf{w}_g, g), \mathbf{w}'_b = \mathbf{w}_b(\mathbf{w}_g, g)$. Using the strict concavity of the auxiliary problem Ψ, we show in the following result that the directional derivative must split (goes down after a good shock and goes up after a bad shock) if last period had a good shock, i.e. $(\mathbf{w}_g, g)$ is the current period state.

Lemma F.3. At state $(\mathbf{w}_g, g)$, the directional derivative goes down after a good shock and goes up after a bad shock, i.e. $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_g, g) < \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$ and $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_b, b) > \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$, if $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g) > 0$.

Proof. If $(\mathbf{w}_g, g) \notin A_{1,g}$, then by Lemma D.2, $\mathbf{w}_g$ is the solution to [P2] at (v_g, g), and $\mathbf{w}'_g$ is the solution to [P2] at (w_{gg}, g). By Lemma D.2, $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g) = \Psi_z(v_g, g)$, and $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_{gg}, g) = \Psi_z(w_{gg}, g)$. From [PK$_g$] at $(\mathbf{v}, s)$, we know $v_g \le \delta[p_g w_{gg} + (1-p_g) w_{gb}] \le \delta w_{gg} < w_{gg}$. So $v_g < w_{gg} < \delta[p_g \bar{v}_g^g + (1-p_g)\bar{v}_b]$. The second inequality is by the assumption that $(\mathbf{w}_g, g) \notin A_{1,g}$. By Lemma E.6, $\Psi(z, g)$ is strictly concave in this region. So $\Psi_z(w_{gg}, g) < \Psi_z(v_g, g)$. Therefore, $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_g, g) < \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$. If $(\mathbf{w}_g, g) \in A_{1,g}$, then from [FOC$w_{gg}$], we know $\lambda(\mathbf{w}_g, g) > 0$. And equation [F.2] simply means $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_g, g) < \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$. By the martingale equation [F.1], $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_g, g) < \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$ implies $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}'_b, b) > \mathsf{D}_{(1,1)}\,Q(\mathbf{w}_g, g)$. □

G. Proofs from Section 5

In this section we show the optimal repayments, one-step set, and various properties regarding investments.

The condition derived in Lemma B.6 regarding Lagrange multipliers in problem [P1] also holds for Lagrange multipliers in problem [VF]. This is because function P in $(P1)$ satisfies all the properties of function $Q(\mathbf{v}, s)$ in [VF]. In particular, we know

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

[G.2] $\eta_g(\mathbf{v}, s) - \lambda(\mathbf{v}, s) - \mu_g(\mathbf{v}, s) = 0$

Moreover, we can use [G.2] to rewrite [FOCk] as:

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

Proof of Proposition 5.1. (a) If $\mu_b(\mathbf{v}, s) > 0$, then complementary slackness implies $m_b(\mathbf{v}, s) = 0$. If $\mu_b(\mathbf{v}, s) = 0$, then $\eta_b(\mathbf{v}, s) + \lambda(\mathbf{v}, s) - \mu_b(\mathbf{v}, s) > 0$, because $\eta_b(\mathbf{v}, s) > 0$ when $\mathbf{v} < \bar{\mathbf{v}}^s$ and $\lambda(\mathbf{v}, s) \geq 0$. Then [G.1] implies $m_b(\mathbf{v}, s) = 0$.

(b) Given $\mathbf{v} < \bar{\mathbf{v}}^s$, $(\mathbf{v}, s) \notin A_{1,s}$ implies $m_g(\mathbf{v}, s) = R(k(\mathbf{v}, s))$
Since $\mathbf{v} < \bar{\mathbf{v}}^s$, $(\mathbf{v}, s) \notin A_{1,s}$, we know $v_g < \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b]$. By Lemma D.2, $\mathbf{w}_g(\mathbf{v}, s)$ is the solution of [P2] at (v_g, g). And by lemma D.1, the constraint of [P2] at (v_g, g) must bind when $v_g < \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b]$. This means $\delta[p_g w_{gg}(\mathbf{v}, s) + (1 - p_g) w_{gb}(\mathbf{v}, s)] = v_g$. Therefore [PK$_g$] implies $m_g(\mathbf{v}, s) = R(k(\mathbf{v}, s))$.

(c) Suppose $(\mathbf{v}, s) \in A_{1,s}$. Given the optimal policy $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ at $(\mathbf{v}, s)$, let $\mathbf{w}'_g = \bar{\mathbf{v}}^g, m'_g = R(k(\mathbf{v}, s)) + \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b^g] - v_g$. Then $m'_g \leq R(k)$ because $v_g \geq \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b^g]$ by the assumption $(\mathbf{v}, s) \in A_{1,s}$. So $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}'_g) \in \Gamma(\mathbf{v}, s)$. Moreover, because $Q(\mathbf{w}'_g, g) = \bar{Q}(\mathbf{v}, s)$, changing the policy to $(k, m_b, m'_g, \mathbf{w}_b, \mathbf{w}'_g)$ at least weakly increases the objective of [VF]. So $\mathbf{w}(\mathbf{v}, g) = \bar{\mathbf{v}}^g \in E_g$.
Suppose $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. That is the contingent utilities reach the efficient region after a good shock. By Proposition 4.2, $w_{gg}(\mathbf{v}, s) \geq \bar{v}_g^g$ and $w_{gb}(\mathbf{v}, s) \geq \bar{v}_b$. Then by [PK$_g$] at $(\mathbf{v}, s)$, we know $v_g \geq \delta[p_g w_{gg}(\mathbf{v}, s) + (1 - p_g) w_{gb}(\mathbf{v}, s)] \geq \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b]$. Therefore, $(\mathbf{v}, s) \in A_{1,s}$.

(d) When $\mathbf{v} \in A_{1,s}$ and $v_g > \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b]$, part (c) implies $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. So in the maximum rent contrat, $\mathbf{w}_g(\mathbf{v}, s) = \bar{\mathbf{v}}^g$. Then [PK$_g$] implies $R(k(\mathbf{v}, s)) - m_g(\mathbf{v}, s) = v_g - \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b] > 0$.

□

We will show in the following proof that : (a) for any $\mathbf{v} < \bar{v}^b, k(\mathbf{v}, g) \geq k(\mathbf{v}, b)$; (b) $k(\mathbf{v}, s)$ is decreasing in v_b for any $(\mathbf{v}, s) \in V \times S$; (c) $k(\mathbf{v}, s)$ is increasing in v_g for any $(\mathbf{v}, s) \in V \times S$.

Proof of Lemma 5.4. (a) Let $\hat{k}, \hat{\mathbf{w}}_i$ and $k, \mathbf{w}_i$ be the optimal policies at states $(\mathbf{v}, b)$ and $(\mathbf{v}, g)$ respectively. Suppose $\hat{k} > k$. Since $\Gamma(\mathbf{v}, g) = \Gamma(\mathbf{v}, b)$, optimality at $(\mathbf{v}, g)$ implies

$$\begin{aligned} & -k + p_g[R(k) + \delta Q(\mathbf{w}_g, g)] + (1 - p_g)\delta Q(\mathbf{w}_b, b) \\ \text{[G.3]} \qquad & \geq -\hat{k} + p_g[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1 - p_g)\delta Q(\hat{\mathbf{w}}_b, b) \end{aligned}$$

Since $\mathbf{w}_g = \hat{\mathbf{w}}_g$ by Lemma [D.2], [G.3] then implies $Q(\mathbf{w}_b, b) > Q(\hat{\mathbf{w}}_b, b)$. Moreover, optimality at $(\mathbf{v}, b)$ implies

$$\begin{aligned} & -\hat{k} + p_b[R(\hat{k}) + \delta Q(\hat{\mathbf{w}}_g, g)] + (1 - p_b)\delta Q(\hat{\mathbf{w}}_b, b) \\ \text{[G.4]} \qquad & \geq -k + p_b[R(k) + \delta Q(\mathbf{w}_g, g)] + (1 - p_b)\delta Q(\mathbf{w}_b, b) \end{aligned}$$

Add [G.3], [G.4] and rearrange to get:

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

which is a contradiction with $Q(\mathbf{w}_b, b) > Q(\hat{\mathbf{w}}_b, b)$.

(b) Take any $(\mathbf{v}, s), (\mathbf{v}', s) \in V \times S$ with $v'_g = v_g, v'_b > v_b$. By Lemma D.2 we know $\mathbf{w}(\mathbf{v}, s) = \mathbf{w}(\mathbf{v}', s)$. Moreover, $Q_g(\mathbf{v}', s) > Q_g(\mathbf{v}, s)$ by result (g) of Theorem 4. Then from [FOCw_{gg}], we must have $\lambda(\mathbf{v}', s) > \lambda(\mathbf{v}, s)$. Hence, [FOC$k$] implies $k(\mathbf{v}', s) < k(\mathbf{v}, s)$.

(c) Take any $(\mathbf{v}, s), (\mathbf{v}', s) \in V \times S$ with $v'_b = v_b, v'_g > v_g$. Suppose that $k(\mathbf{v}, s) > k(\mathbf{v}', s)$. From [FOC$k$], $\lambda(\mathbf{v}', s) > \lambda(\mathbf{v}, s)$. From [IC*] and [PK$_b$] ($m_b(\mathbf{v}, s) = m(\mathbf{v}', s) = 0$ are optimal), we obtain

$$\begin{aligned} w_{bb}(\mathbf{v}', s) &= \frac{p_b R(k(\mathbf{v}', s)) + p_g v_b - p_b v'_g}{\delta \Delta} \\ &< \frac{p_b R(k(\mathbf{v}, s)) + p_g v_b - p_b v_g}{\delta \Delta} \leq w_{bb}(\mathbf{v}, s) \end{aligned}$$

and

$$\begin{aligned} w_{bg}(\mathbf{v}', s) &= \frac{1 - p_b}{\delta \Delta}\left[v'_g - \frac{1 - p_g}{1 - p_b} v_b - R(k(\mathbf{v}', s))\right] \\ &> \frac{1 - p_b}{\delta \Delta}\left[v_g - \frac{1 - p_g}{1 - p_b} v_b - R(k(\mathbf{v}, s))\right] \geq w_{bg}(\mathbf{v}, s) \end{aligned}$$

The equalities are because $\lambda(\mathbf{v}', s) > 0$. Then we know $Q_g(\mathbf{w}_b(\mathbf{v}, s), b) \geq Q_g(\mathbf{w}_b(\mathbf{v}', s), b)$ because Q_g is increasing in the first coordinate and decreasing in the second coordinate by Theorem 4. Moreover, $\eta_b(\mathbf{v}, s) = Q_b(\mathbf{v}, s) \leq Q_b(\mathbf{v}', s) = \eta_b(\mathbf{v}', s)$ by Theorem 4. From [FOCw_{bg}], we will have $\lambda(\mathbf{v}', s) \leq \lambda(\mathbf{v}, s)$, a contradiction. Therefore, $k(\mathbf{v}, s) \leq k(\mathbf{v}', s)$

□

To proceed the proof of Proposition 5.3, we first show that given a certain value of v_b, if v_g is sufficiently large, then efficient investment will be achieved. Moreover, for a certain value of v_b, there exists a threshold value of v_g such that investment is efficient if v_g is above the threshold and inefficient if below the threshold.

Lemma G.1. For any state $(\mathbf{v}, s) \in V \times S$ that sastisfies $v_g \geq R(\bar{k}_s) + \frac{p_g v_b}{p_b}$, we have $\lambda(\mathbf{v}, s) = 0, k(\mathbf{v}, s) = \bar{k}_s$.

Proof. By [PK$_b$], $\delta[p_b(w_{bg} - w_{bb}) + w_{bb}] \leq v_b + m_b \leq v_b$. So we have $w_{bg} - w_{bb} \leq \frac{v_b}{\delta p_b}$. Then the right hand side of [IC*] is smaller than $R(\bar{k}_s) + \frac{\Delta v_b}{p_b}$. Since $v_g \geq R(\bar{k}_s) + \frac{p_g v_b}{p_b}$, we know $v_g - v_b \geq R(\bar{k}_s) + \frac{\Delta v_b}{p_b}$. This means for all feasible policies at $(\mathbf{v}, s)$, [IC*] will not bind. Therefore, we must have $\lambda(\mathbf{v}, s) = 0, k(\mathbf{v}, s) = \bar{k}_s$. $\square$

Lemma G.2. For any $v_b \geq 0$, there exists $h_s(v_b) > v_b$ such that $v_g \geq h_s(v_b)$ implies $\lambda(\mathbf{v}, s) = 0$, and $v_b \leq v_g < h_s(v_b)$ implies $\lambda(\mathbf{v}, s) > 0$. Moreover, $h_s(v_b)$ is increasing and satisfies $h(0) = R(\bar{k}_s), h(\bar{v}_b^s) \leq \bar{v}_g^s$;

Proof. Take any $v_b > 0$. If v_g is sufficiently close to v_b, then investment $k(\mathbf{v}, s)$ will be sufficiently close to zero by [IC*]. We know $\lambda(\mathbf{v}, s) > 0$ by [FOCk]. If v_g is sufficiently large, by Lemma G.1, we know $\lambda(\mathbf{v}, s) = 0$. Hence, for a certain value of v_b we can define the smallest value of v_g such that $\lambda(\mathbf{v}, s) = 0$ as $h_s(v_b) = \inf\{v_g \geq v_b : \lambda(\mathbf{v}, s) = 0\}$. By Proposition 5.x and [FOCk] that $\lambda(\mathbf{v}, s)$ is increasing in v_b and decreasing in v_g. So we have $\lambda(\mathbf{v}, s) = 0$ if $v_g \geq h_s(v_b)$.

Then, $h(0) = R(\bar{k}_s)$, since the only feasible contingent utility vector $\mathbf{w}_b$ is $\mathbf{0}$ at $(0, R(\bar{k}_s))$. And since $\lambda(\bar{\mathbf{v}}^s, s) = 0$, we know $h(\bar{v}_b) \leq \bar{v}_g^s$. Now we show $h_s(v_b)$ is increasing. Take any v_b, v_b' with $v_b' > v_b$. We know $0 < \lambda((v_b, h_s(v_b) - \varepsilon), s) \leq \lambda((v_b', h_s(v_b) - \varepsilon), s)$ for any small $\varepsilon > 0$. The first inequality is by the definition of $h_s(v_b)$, and the second is by $\lambda(\mathbf{v}, s)$ is increasing in v_b. By definition, $0 = \lambda((v_b', h(v_b')), s) < \lambda((v_b', h_s(v_b) - \varepsilon), s)$. So $h(v_b') \geq h_s(v_b) - \varepsilon$, which implies $h(v_b') \geq h_s(v_b)$. $\square$

Lemma G.3. Take $(\mathbf{v}, s), (\hat{\mathbf{v}}, s) \in V \times S$ such that $\mathbf{v} = (v_b, h_s(v_b)), \hat{\mathbf{v}} = (v_b, \hat{v}_g)$ with $\hat{v}_g < h_s(v_b)$. If $\lambda(\mathbf{w}_b(\mathbf{v}, s), b) > 0$, then $\lambda(\mathbf{w}_b(\hat{\mathbf{v}}, s), b) > 0$.

Proof. Note that $Q(\mathbf{w}_b(\mathbf{v}, s), b) \geq Q(\mathbf{w}_b(\hat{\mathbf{v}}, s), b)$, since $\mathbf{w}_b(\mathbf{v}, s), \mathbf{w}_b(\hat{\mathbf{v}}, s)$ are both feasible in problem [P2] at (v_b, b) and $\mathbf{w}_b(\mathbf{v}, s)$ is the maximizer (because by construction $\lambda(\mathbf{v}, s) = 0$). Suppose $w_{bg}(\hat{\mathbf{v}}, s) - w_{bb}(\hat{\mathbf{v}}, s) > w_{bg}(\mathbf{v}, s) - w_{bb}(\mathbf{v}, s)$. Then there exists $k > k(\hat{\mathbf{v}}, s)$ such that

$$\begin{aligned} \hat{v}_g - v_b &\geq R(k(\hat{\mathbf{v}}, s)) + \delta\Delta[w_{bg}(\hat{\mathbf{v}}, s) - w_{bb}(\hat{\mathbf{v}}, s)] \\ &= R(k) + \delta\Delta[w_{bg}(\mathbf{v}, s) - w_{bb}(\mathbf{v}, s)] \end{aligned} \tag{G.5}$$

Let $m_b = \delta[p_b w_{bg}(\mathbf{v}, s) + (1 - p_s)w_{bb}(\mathbf{v}, s)] - v_b$. Then [G.5] means

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

so that

$$\begin{aligned} &-k(\hat{\mathbf{v}}, s) + p_s R(k(\hat{\mathbf{v}}, s)) + (1 - p_s)\delta Q(\mathbf{w}_b(\hat{\mathbf{v}}, s), b) \\ &\geq -k + p_s R(k) + (1 - p_s)\delta Q(\mathbf{w}_b(\mathbf{v}, s), b) \end{aligned}$$

which further implies $Q(\mathbf{w}_b(\mathbf{v}, s), b) < Q(\mathbf{w}_b(\hat{\mathbf{v}}, s), b)$, a contradiction. So $w_{bg}(\hat{\mathbf{v}}, s) - w_{bb}(\hat{\mathbf{v}}, s) \leq w_{bg}(\mathbf{v}, s) - w_{bb}(\mathbf{v}, s)$. Since $m_b(\mathbf{v}, s) = m_b(\mathbf{v}', s) = 0$ are optimal and from [PK_b] at both states $(\mathbf{v}, s)$ and $(\mathbf{v}', s)$, we know $w_{bg}(\hat{\mathbf{v}}, s) \leq w_{bg}(\mathbf{v}, s)$ and $w_{bb}(\hat{\mathbf{v}}, s) \geq w_{bb}(\mathbf{v}, s)$. So we have,

$$w_{bg}(\hat{\mathbf{v}}, s) \leq w_{bg}(\mathbf{v}, s) < h_b(w_{bb}(\mathbf{v}, s)) \leq h_b(w_{bb}(\hat{\mathbf{v}}, s))$$

The second inequality is by assumption $\lambda(\mathbf{w}_b(\mathbf{v}, s), b) > 0$. The last inequality is by the fact that h_s is increasing from Lemma G.2. Therefore, $w_{bg}(\hat{\mathbf{v}}, s) < h_b(w_{bb}(\hat{\mathbf{v}}, s))$ means $\lambda(\mathbf{w}_b(\hat{\mathbf{v}}, s), b) > 0$. □

We will show in the following proof that:

(a) $\lambda(\mathbf{w}_g(\mathbf{v}, s), g) > 0$ if $\mathbf{w}_g(\mathbf{v}, s) \notin E_g$;
(b) $\lambda(\mathbf{w}_b(\mathbf{v}, s), b) > 0$, for any $(\mathbf{v}, s) \in V \times S$ with $v_b < \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$;
(c) If $(p_b, p_g) \in \{\boldsymbol{p} : p_b \geq \phi(p_g)\}$, then $\lambda(\mathbf{w}_b(\mathbf{v}, s), b) > 0$ if $\mathbf{w}_b(\mathbf{v}, s) \notin E_b$.

(a) means after a good shock investment is inefficient. (b) means if v_b is not sufficiently close to $\bar{v}_b^s$, then investment is inefficient after a bad shock. (c) means if persistence is low, then we always have investment is inefficient after a bad shock.

Proof of Proposition 5.3. (a) If $\mathbf{w}_g(\mathbf{v}, s) \in A_{1,g}$, then the left hand side of [$\mathrm{FOC}w_{gg}$] at $(\mathbf{w}_g(\mathbf{v}, s), g)$ is zero, and hence from its right hand side, $\lambda(\mathbf{w}_g(\mathbf{v}, s), g) = \eta_g(\mathbf{w}_g(\mathbf{v}, s), g) > 0$. Consider $\mathbf{w}_g(\mathbf{v}, s) \notin A_{1,g}$. Let $\mathbf{w}'_g = \mathbf{w}_g(\mathbf{w}_g(\mathbf{v}, s), g)$. Lemma F.3 shows that $\mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g(\mathbf{v}, s), g) > \mathsf{D}_{(1,1)}\, Q(\mathbf{w}'_g, g)$. And by [F.2], we must have $\lambda(\mathbf{w}_g(\mathbf{v}, s), g) > 0$.

(b) Note it suffices to show $\lambda(\mathbf{w}_b(\hat{\mathbf{v}},s),b)>0$ if $\hat{v}_b<\frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$, $\hat{v}_g=h_s(\hat{v}_b)$. If it is ture, then: (1) $\lambda((\mathbf{w}_b(\mathbf{v}',s),b)>0$ for any $(\mathbf{v}',s)$ with $v_b'=v_b, v_g'<\hat{v}_g$, by Lemma G.3; (2) $\lambda((\mathbf{w}_b(\mathbf{v}',s),b)=\lambda((\mathbf{w}_b(\mathbf{v},s),b)>0$ for any $(\mathbf{v}',s)$ with $v_b'=v_b, v_g'>\hat{v}_g$, because $\mathbf{w}_b(\mathbf{v}',s)=\mathbf{w}_b(\mathbf{v},s)$ by Lemma D.2.

Suppose $\lambda(\mathbf{w}_b(\hat{\mathbf{v}},s),b)=0$. And by construction $\lambda(\hat{v},s)=0$. Then [F.4] implies that $\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b',b)=\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b(\hat{\mathbf{v}},s),b)$, where $\mathbf{w}_b'=\mathbf{w}_b(\mathbf{w}_b(\hat{\mathbf{v}},s),b)$. By Lemma D.1, $\Psi_z(\hat{v}_b,b)=\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b(\hat{\mathbf{v}},s),b)$, and $\Psi_z(w_{bb}(\hat{\mathbf{v}},s),b)=\mathsf{D}_{(1,1)}\,Q(\mathbf{w}_b',b)$. So $\Psi_z(\hat{v}_b,b)=\Psi_z(w_{bb}(\hat{\mathbf{v}},s),b)$. The strict concavity of $\Psi(z,b)$ in z, by Lemma E.6, implies $w_{bb}(\hat{\mathbf{v}},s)=\hat{v}_b$. Then we know $\mathbf{w}_b'=\mathbf{w}_b(\hat{\mathbf{v}},s)$, since by Lemma D.2 $\mathbf{w}_b'$ is only function of $w_{bb}(\hat{\mathbf{v}},s)$, and $\mathbf{w}_b(\hat{\mathbf{v}},s)$ is only a function of $\hat{v}_b$. From [PK$_b$] and [IC*] at state $(\mathbf{w}_b(\hat{\mathbf{v}},s),b)$:

[G.6] $$w_{bb}(\hat{\mathbf{v}},s)\geq\delta[p_b w_{bg}(\hat{\mathbf{v}},s)+(1-p_b)w_{bb}(\hat{\mathbf{v}},s)]$$

[G.7] $$w_{bg}(\hat{\mathbf{v}},s)-w_{bb}(\hat{\mathbf{v}},s)\geq R(\bar{k}_b)+\delta\Delta[w_{bg}(\hat{\mathbf{v}},s)-w_{bb}(\hat{\mathbf{v}},s)]$$

[G.6] and [G.7] together imply $w_{bb}(\hat{\mathbf{v}},s)\geq\frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$, and $w_{bg}(\hat{\mathbf{v}},s)\geq\frac{1-\delta(1-p_b)}{(1-\delta)(1-\delta\Delta)}R(\bar{k}_b)$. Then from [PK$_b$] at $(\hat{\mathbf{v}},s)$, we know

$$\hat{v}_b\geq\delta[p_b w_{bg}(\hat{\mathbf{v}},s)+(1-p_b)w_{bb}(\hat{\mathbf{v}},s)]\geq\frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$$

This is a contradiction with the assumption that $\hat{v}_b<\frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$.

(c) If $(p_b,p_g)\in\{\mathbf{p}:p_b\geq\phi(p_g)\}$, then $\bar{v}_b=\frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$. And $\mathbf{w}_b(\mathbf{v},s)\notin E_b$ means $v_b<\bar{v}_b$. So from part (b), $\lambda(\mathbf{w}_b(\mathbf{v},s),b)>0$.

□

H. Proof of Theorem 2

Recall from section E of the Appendix that the set H is defined as $H:=\{(\mathbf{v},s)\in V\times S : Q_b(\mathbf{v},s)>0, Q_g(\mathbf{v},s)>0\}$. Define the process $\{(\mathbf{v}^{(t)},s_{t-1})\}_{t=0}^{\infty}$ to be the states induced by the optimal contract starting at some $(\mathbf{v}^{(0)},s_{-1})\in V\times S$. To establish the convergence results in Theorem 2, we first show some useful properties regarding H. Denote the closure of H as $\mathrm{cl}(H)$.

Lemma H.1. The set H has the following properties:

(a) $(\mathbf{v},s)\in H$ implies $v_b<\bar{v}_b^s$ and $v_g<\bar{v}_g^s$.
(b) For any $(\mathbf{v},s)\in\mathrm{cl}(H)$, $\mathsf{D}_{(1,1)}\,Q(\mathbf{v},s)=0$ implies $\mathbf{v}=\bar{\mathbf{v}}^s$.

(c) For any $(\mathbf{v}, s) \in V \times S$, either $\mathbf{w}_g(\mathbf{v}, s) \in H$ or $\mathbf{w}_g(\mathbf{v}, s) \in E_g$.

Proof. (a) Take any $(\mathbf{v}, s) \in H$. By the definition of H, $Q_b(\mathbf{v}, s) > 0$ and $Q_g(\mathbf{v}, s) > 0$. Suppose $v_b \geq \bar{v}_b^s$. Then from part (e) of Theorem 4, we know $Q_b(\mathbf{v}, s) \leq 0$, a contradiction. Suppose $v_g \geq \bar{v}_g^s$ and $v_b < \bar{v}_b^s$. From part (b) of Proposition 4.2, we know $Q_g((\bar{v}_b^s, v_g), s) = 0$ because $(\bar{v}_b^s, v_g) \in E_s$. Supermodularity of Q then implies $0 \leq Q_g(\mathbf{v}, s) \leq Q_g((\bar{v}_b^s, v_g), s) = 0$, a contradiction. So we must have $v_b < \bar{v}_b^s$ and $v_g < \bar{v}_g^s$.

(b) Let $A := \{(\mathbf{v}, s) \in V \times S : \mathbf{v} \leq \bar{\mathbf{v}}^s\}$. Part (a) implies $H \subset A$. Since A is a closed set, we know $\mathrm{cl}(H) \in A$. Take any $(\hat{\mathbf{v}}, s) \in \mathrm{cl}(H)$ such that $\mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, s) = 0$. The assumption $(\hat{\mathbf{v}}, s) \in \mathrm{cl}(H)$ implies $\hat{\mathbf{v}} \leq \bar{\mathbf{v}}^s$ and $Q_b(\hat{\mathbf{v}}, s) \geq 0$, $Q_g(\hat{\mathbf{v}}, s) \geq 0$. Then $\mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}, s) = 0$ implies $Q_b(\hat{\mathbf{v}}, s) = 0$, $Q_g(\hat{\mathbf{v}}, s) = 0$. From Proposition 4.2 we know $\hat{\mathbf{v}} \in E_s$, and hence, $\hat{\mathbf{v}} \geq \bar{\mathbf{v}}^s$. So we must have $\hat{\mathbf{v}} = \bar{\mathbf{v}}^s$.

(c) Take any $(\mathbf{v}, s) \in V \times S$. From the first order conditions [FOCw_{gb}] and [FOCw_{gg}], we know $p_g Q_b(\mathbf{w}_g(\mathbf{v}, s), g) = (1 - p_g) Q_g(\mathbf{w}_g(\mathbf{v}, s), g)$. So we must either have $Q_b(\mathbf{w}_g(\mathbf{v}, s), g) > 0$ and $Q_g(\mathbf{w}_g(\mathbf{v}, s), g) > 0$ or have $Q_b(\mathbf{w}_g(\mathbf{v}, s), g) = 0$ and $Q_g(\mathbf{w}_g(\mathbf{v}, s), g) = 0$. The former case means that $\mathbf{w}_g(\mathbf{v}, s) \in H$ and the latter case means that $\mathbf{w}_g(\mathbf{v}, s) \in E_g$ by part (b) of Proposition 4.2. □

In the high peresistence 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 infinitely often.

Lemma H.2. The sets $\{s_t = \alpha \quad \text{i.o.}\}$, where $\alpha \in S$, have full measure. Similarly, the sets $\{(s_{t-1}, s_t) = (\alpha, \beta) \quad \text{i.o.}\}$, where $\alpha, \beta \in S$, have full measure.

Intuitively, the lemma says that bad or good shocks occur infinitely often with probability one. Moreover, consecutive ‘bad-bad’, ‘bad-good’, ’good-bad’, and ‘good-good’ shocks also occur infinitely often with probability one.

Proof of Theorem 2. (a) We show that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges to 0 almost surely.

By Lemma F.1, the process $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ is a nonnegative martingale. So Doob’s Martingale Convergence Theorem ensures that $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges almost surely to a non-negative and integrable random variable.

Consider a path with the property that $\lim_{t\to\infty} \mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1}) = a > 0$ and good shock occurs infinitely many times. There exists a large T such that $\mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1}) > 0$ for $t \geq T$.

Now let us consider the subsequence that has only good shocks. Note that when $t \geq T$, this subsequence must stay in the set H. This is because by part (c) of Lemma H.1, after good shocks if directional derivative is strictly positive then the state induced by the optimal contract has to be in H. Moreover, by part (a) of Lemma H.1 we know this subsequence with only good shocks is bounded and must have a converging subsequence $\{(\mathbf{v}^{(\tau_t)}, g)\}_{t=0}^{\infty}$ with limit $(\hat{\mathbf{v}}, g)$.

By construction we have $s_{\tau_t - 1} = g$. And it is either the case that $s_{\tau_t} = b$ infinitely often (i.e. bad shock occurs infinitely often after the chosen good shocks), or the case that $s_{\tau_t} = g$ infinitely often. Suppose $s_{\tau_t} = b$ for infinitly many t. And let $\{(\mathbf{v}^{(\tau'_t)}, g)\}_{t=0}^{\infty}$ be a subsequence of $\{(\mathbf{v}^{(\tau_t)}, g)\}_{t=0}^{\infty}$ that has $s_{\tau'_t - 1} = g$ and $s_{\tau'_t} = b$. By construction, $\mathbf{v}^{(\tau'_t+1)} = \mathbf{w}_b(\mathbf{v}^{(\tau'_t)}, g)$. And hence, $\lim_{t\to\infty} \mathbf{v}^{\tau'_t+1} = \mathbf{w}_b(\hat{\mathbf{v}}, g)$ because $\mathbf{w}_b$ is continuous. And by the continuity of $\mathsf{D}_{(1,1)} Q$, we get $\lim_{t\to\infty} \mathsf{D}_{(1,1)} Q(\mathbf{v}^{(\tau'_t)}, g) = \mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) = a$, and $\lim_{t\to\infty} \mathsf{D}_{(1,1)} Q(\mathbf{v}^{(\tau'_t+1)}, b) = \mathsf{D}_{(1,1)} Q(\mathbf{w}_b(\hat{\mathbf{v}}, g), b) = a$. However, from Lemma F.3 the directional derivative martingale must strictly increase after a bad shock if $\mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) > 0$, which means $\mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) < \mathsf{D}_{(1,1)} Q(\mathbf{w}_b(\hat{\mathbf{v}}, g), b)$. This forms a contradiction.

In the case that $s_{\tau_t} = g$ for infinitely many t, we can use the same argument to show that $\mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) = \mathsf{D}_{(1,1)} Q(\mathbf{w}_g(\hat{\mathbf{v}}, g), g) = a$. It also contradicts Lemma F.3 because $\mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) > \mathsf{D}_{(1,1)} Q(\mathbf{w}_g(\hat{\mathbf{v}}, g), g)$ if $\mathsf{D}_{(1,1)} Q(\hat{\mathbf{v}}, g) > 0$.

Therefore, we must have $\lim_{t\to\infty} \mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$. By Lemma H.2, paths with only finitely many good shocks have measure zero, so it must be that $\lim_{t\to\infty} \mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ almost surely.

(b) We show that once contingent equities enter E_s, they will never leave these sets.

Take any $(\mathbf{v}, s) \in E_s$. Suppose that $\mathbf{w}_g(\mathbf{v}, s) \notin E_g$. By part (c) of Proposition 4.2, $Q(\mathbf{w}_g(\mathbf{v}, s), g) < \bar{Q}(g)$. This implies $Q(\mathbf{v}, s) < \bar{Q}(s)$, by the objective function defining $Q(\mathbf{v}, s)$. And by part (a) of Proposition 4.2, we must have $(\mathbf{v}, s) \notin E_s$, which is a contradiction. Therefore, $\mathbf{w}_g(\mathbf{v}, s) \in E_g$. The same argument shows we must also have $\mathbf{w}_b(\mathbf{v}, s) \in E_b$.

(c) We show that $\mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1})$ converges to 0 in finite time almost surely.

Consider a path with the property that $\mathsf{D}_{(1,1)} Q(\mathbf{v}^{(t)}, s_{t-1}) > 0$ for any finite t and good-good shock occurs infinitely many times. Take a subsequence

$\{(\mathbf{v}^{(\gamma_t)}, s_{\gamma_t - 1})\}_{t=0}^{\infty}$ that has only good-good shocks, i.e. $s^{\gamma_t - 1} = s^{\gamma_t} = g$. Note that the sequence $\{(\mathbf{v}^{(\gamma_t)}, s_{\gamma_t - 1})\}_{t=0}^{\infty}$ is in set H. This is because by part (c) of Lemma H.1, after good shocks if directional derivative is strictly positive then the state induced by the optimal contract has to be in H.

Since $(\bar{\mathbf{v}}^s, s)$ are the only points in the closure of H that have zero directional derivative by Lemma H.1, $\lim_{t\to\infty} \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\gamma_t)}, g) = 0$ implies $\lim_{t\to\infty} \mathbf{v}^{(\gamma_t)} = \bar{\mathbf{v}}^g$. This means there exists a large T and sufficiently small $\varepsilon > 0$ such that $\bar{v}_g^g - v_g^{(\gamma_t)} < \varepsilon$ for $t \geq T$. So $(\mathbf{v}^{(\gamma_T)}, s_{\gamma_T - 1}) \in A_1$. And since $s_T = g$ by construction, we must have $\mathbf{v}^{(\gamma_T + 1)} \in E_g$ and $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(\gamma_T+1)}, g) = 0$, a contradiction. Since the path with only finitely many good-good shocks has measure zero by Lemma H.2, the argument above shows $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ for some finite t almost surely. From part (b), we know if $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(T)}, s_{T-1}) = 0$, then $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0$ for any $t \geq T$. Therefore, $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges to 0 in finite time almost surely.

(d) We show in a maximum rent contract that contingent equities reach $\bar{\mathbf{v}}^g$ in finite time almost surely and cycle between $\bar{\mathbf{v}}^b, \bar{\mathbf{v}}^g$.

Consider a path in the maximum rent contract along which $\mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1})$ converges to zero in finite time. This means on the path that we consider, there exists $T = \min\{t : \mathsf{D}_{(1,1)}\, Q(\mathbf{v}^{(t)}, s_{t-1}) = 0\}$. From part (b) of Proposition 4.2, we know that $\mathbf{v}^{(T)} \in E_s$ where $s = s_{T-1}$. And since the contingent equities reach efficient sets only after good shocks, we must have $s_{T-1} = g$ and $\mathbf{v}^{(T)} = \bar{\mathbf{v}}^g$ in the maximum contract. Moreover, part (c) shows that the path that we are considering has measure one. Therefore, in the maximum rent contract, the contingent equities must reach $\bar{\mathbf{v}}^g$ in finite time almost surely. Since part (b) shows that $(\mathbf{v}, s)$ never leave the efficient sets once reach there, we know $\mathbf{v}^{(t)} = \bar{\mathbf{v}}^g$ if $s_{t-1} = g$ and $\mathbf{v}^{(t)} = \bar{\mathbf{v}}^b$ if $s_{t-1} = b$ for $t \geq T$. $\square$

I. Proofs from Section 7

In this section, we show the level of threshold equities and repayments of the mature firm.

For a given p_g, we can define a cutoff level $\phi(p_g)$ that satisfies: $\phi(p_g) R'(k^\phi) = 1$ and $R(k^\phi) = \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g}$. With this cutoff, we can partition the parameter space as: $B_- = \{\boldsymbol{p} : \phi(p_g) < p_b < p_g\}$, $B_= = \{\boldsymbol{p} : p_b = \phi(p_g)\}$, and $B_+ = \{\boldsymbol{p} : p_b < \phi(p_g)\}$. The parameter space with $p_b \geq \phi(p_g)$ characterizes the case of low persistence. And the parameter space with $p_b \leq \phi(p_g)$ characterizes the case of high persistence. The first result shows that the high and low persistence cases are separated by an increasing boundary $\phi(p_g)$.

Lemma I.1. (a) $0 < \phi(p_g) < p_g$, and $\phi(p_g)$ is increasing in p_g;
(b) $\boldsymbol{p} \in B_-$ if, and only if $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$;
(c) $\boldsymbol{p} \in B_=$ if, and only if $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] = R(\bar{k}_b)$;
(d) $\boldsymbol{p} \in B_+$ if, and only if $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] > R(\bar{k}_b)$.

Proof. (a) Because $\frac{\delta p_g}{1+\delta p_g} < 1$, by the definition of $\phi(.)$, $R(k^\phi) < R(\bar{k}_g)$, and hence $k^\phi < \bar{k}_g$. Concavity of R then implies $R'(k^\phi) > R'(\bar{k}_g)$. Moreover, by the definition of $\phi(.)$, we have $\phi(p_g)R'(k^\phi) = p_g R'(\bar{k}_g) = 1$. So $R'(k^\phi) > R'(\bar{k}_g)$ implies $\phi(p_g) < p_g$.
$R(k^\phi) = \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g}$ is increasing p_g. So k^ϕ is increasing in p_g. Hence, $\phi(p_g) = 1/R'(k^\phi)$ is increasing in p_g.

(b) By the definition of $\phi(.)$, we have $\phi(p_g)R'(k^\phi) = p_b R'(\bar{k}_b) = 1$. Then $\boldsymbol{p} \in B_-$ implies $R'(k^\phi) > R'(\bar{k}_b)$. Concavity of R implies $k^\phi < \bar{k}_b$. Moreover,by the definition of $\phi(.)$, we know $\frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} = R(k^\phi) < R(\bar{k}_b)$. Rearrange to obtain that $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$. If we know $\delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] < R(\bar{k}_b)$, then $R(\bar{k}_b) > \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} = R(k^\phi)$ which implies $\bar{k}_b > k^\phi$. By concavity of R, we know $R'(k^\phi) > R'(\bar{k}_b)$. So by definition of $\phi(.)$, we have $\phi(p_g) < p_b$, meaning $\boldsymbol{p} \in B_-$.

(c) Similar argument as in (b) shows the result.

(d) Similar argument as in (b) shows the result.

$\square$

Lemma I.2. At $(\bar{\mathbf{v}}^s, s)$, either [IC*] or [LL] for $s = g$ or both must hold as equality.

Proof. Suppose not. Then [IC] and [LL] for g both hold as inequality at $(\bar{\mathbf{v}}^s, s)$. Since $Q_b(\bar{\mathbf{v}}, s) = Q_g(\bar{\mathbf{v}}, s) = 0$, we know $(\bar{\mathbf{v}}^s, s) \in cl(H)$[18]. By continuity of policies, there exists $(\hat{\mathbf{v}}, s) \in H$ such that [IC] and (LL) for g both hold as inequality. Complementary slackness then implies $\mu_g(\hat{\mathbf{v}}, s) = 0$ and $\lambda(\hat{\mathbf{v}}, s) = 0$. And from [G.2], we know $\eta_g(\hat{\mathbf{v}}, s) = Q_g(\hat{\mathbf{v}}, s) = 0$, contradicted with $(\hat{\mathbf{v}}, s) \in H$. $\square$

Lemma I.3. (a) $\bar{m}_b^s = 0$; (b) $\bar{v}_b^g = \bar{v}_b^b$, $\bar{v}_g^b < \bar{v}_g^g$.

(18) Recall set H is defined in Section 4 of the Appendix as $\{(\mathbf{v}, s) \in V \times S : Q_b(\mathbf{v}, s) > 0, Q_g(\mathbf{v}, s) > 0\}$.

Proof. (a) From Proposition [5.1], $m_b(\mathbf{v}, s) = 0$. So by continuity $\bar{m}_b^s = m_b(\bar{\mathbf{v}}^s, s) = 0$.

(b) The right hand side of [PK_b] at $(\bar{\mathbf{v}}^s, s)$ is not contingent on s because $\bar{m}_b^s = 0$. So we must have $\bar{v}_b^b = \bar{v}_b^g$. Then we can rewrite [IC] at $(\bar{\mathbf{v}}^s, s)$ as

$$\bar{m}_g^s \leq \delta p_g(\bar{v}_g^g - \bar{v}_g^b) \tag{I.1}$$

Moreover, by [PK_g] for $(\bar{\mathbf{v}}^g, g)$ and $(\bar{\mathbf{v}}^g, b)$:

$$\bar{v}_g^g - \bar{v}_g^b = R(\bar{k}_g) - R(\bar{k}_b) - (\bar{m}_g^g - \bar{m}_g^b) \tag{I.2}$$

Suppose $\bar{v}_g^g \leq \bar{v}_g^b$. Then [I.1] implies $\bar{m}_g^g, \bar{m}_g^b \leq 0$. Also, [I.2] implies $\bar{m}_g^g - \bar{m}_g^b \geq R(\bar{k}_g) - R(\bar{k}_b) > 0$. So we must have $\bar{m}_g^b < \bar{m}_g^g \leq \delta p_g(\bar{v}_g^g - \bar{v}_g^b) \leq 0$. This means $\bar{m}_g^b < R(\bar{k}_b)$ and $\bar{m}_g^b < \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$, which contradicts with Lemma I.2.

□

Lemma I.4. In the maximum rent contract, the debt repayment satisfies:

(a) At state $(\bar{\mathbf{v}}^g, g)$: $\bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b) < R(\bar{k}_g)$;
(b) At state $(\bar{\mathbf{v}}^b, b)$: if $\boldsymbol{p} \in B_-$, then $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$;
(c) At state $(\bar{\mathbf{v}}^b, b)$: if $\boldsymbol{p} \in B_+$, then $\bar{m}_g^b = R(\bar{k}_b)$;
(d) At state $(\bar{\mathbf{v}}^b, b)$: if $\boldsymbol{p} \in B_=$, then $\bar{m}_g^b = R(\bar{k}_b) = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$.

Proof. (a) We will show that $\bar{m}_g^g < R(\bar{k}_g)$. Suppose $\bar{m}_g^g = R(\bar{k}_g)$. From [I.2], $\bar{m}_g^b - R(\bar{k}_b) = \bar{v}_g^g - \bar{v}_g^b > 0$. This means $\bar{m}_g^b > R(\bar{k}_b)$, which violates [LL] for b, a contradiction. Then $\bar{m}_g^g < R(\bar{k}_g)$ implies [IC] must hold as equality by Lemma I.1. Therefore, from [I.1], we know $\bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$.

(b) We will show that when $\boldsymbol{p} \in B_-$, we have $\bar{m}_g^b < R(\bar{k}_b)$. Suppose not. $\bar{m}_g^b = R(\bar{k}_b)$. By [I.2], we know $\bar{v}_g^g - \bar{v}_g^b = R(\bar{k}_g) - \bar{m}_g^g = R(\bar{k}_g) - \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. The last equality is from part (a) that $\bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. So $\delta p_g(\bar{v}_g^g - \bar{v}_g^b) = \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} < R(\bar{k}_b)$. The last inequality is from Lemma I.1 when $\boldsymbol{p} \in B_-$. So $\delta p_g(\bar{v}_g^g - \bar{v}_g^b) < R(\bar{k}_b) = \bar{m}_g^b$. However, comparing with [I.1], we know [IC] at $(\bar{\mathbf{v}}^b, b)$ is violated. Therefore, we must have $\bar{m}_g^b < R(\bar{k}_b)$. By Lemma I.2, [IC] must hold as equality. So $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$.

(c) We will show that when $\boldsymbol{p} \in B_+$, we have $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. Suppose not. $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. Then we know from part (a) that $\bar{m}_g^b = \bar{m}_g^g$. And from [I.2], $\bar{v}_g^g - \bar{v}_g^b = R(\bar{k}_g) - R(\bar{k}_b)$. So $\bar{m}_g^b = \delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] > R(\bar{k}_b)$. The last inequality is from Lemma I.1 when $\boldsymbol{p} \in B_+$. However, this means [LL] for

g at $(\bar{\mathbf{v}}^b, b)$ is violated. Therefore, we must have $\bar{m}_g^b < \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$, which means the [IC] holds as inequality. So [LL] for g must hold as equality by Lemma I.2, which means $\bar{m}_g^b = R(\bar{k}_b)$.

(d) When $\boldsymbol{p} \in B_=$, the same procedure as in part (b) and (c) shows that $\bar{m}_g^b = R(\bar{k}_b)$ and $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$ imply one another. Therefore, by Lemma I.2, they must both hold.

□

Proof of Theorem 3. When the firm is mature, the contingent equity levels in a maximum rent contract are $\mathbf{w}_g(\bar{\mathbf{v}}^s, s) = \bar{\mathbf{v}}^g$ if a good shock occurs, and $\mathbf{w}_b(\bar{\mathbf{v}}^s, s) = \bar{\mathbf{v}}^b$ if a bad shock occurs. From [PK$_b$] and [PK$_g$] at $(\bar{\mathbf{v}}^b, b)$ and $(\bar{\mathbf{v}}^g, g)$ respectively, we have

$$\bar{v}_b = \delta[p_g \bar{v}_g^b + (1 - p_g)\bar{v}_b] \tag{I.3}$$

$$\bar{v}_g^b = R(\bar{k}_b) - \bar{m}_g^b + \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b] \tag{I.4}$$

$$\bar{v}_g^g = R(\bar{k}_g) - \bar{m}_g^g + \delta[p_g \bar{v}_g^g + (1 - p_g)\bar{v}_b] \tag{I.5}$$

where $\bar{v}_b^g = \bar{v}_b^b = \bar{v}_b$, because $\bar{m}_b^s = 0$.

(a) When $p_b \geq \phi(p_g)$, or equivalently $\boldsymbol{p} \in B_-$ or $B_=$, Lemma I.4 implies:

$$\bar{m}_g^b = \bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b) \tag{I.6}$$

Comibine [I.3] to [I.5], and [I.6], we obtain the solution:

$$\bar{v}_b = \frac{\delta p_b R(\bar{k}_b)}{(1-\delta)(1-\delta\Delta)}$$

$$\bar{v}_g^b = \frac{1 - \delta(1 - p_b)}{(1-\delta)(1-\delta\Delta)} R(\bar{k}_b), \qquad \bar{v}_g^g = R(\bar{k}_g) + \frac{\delta(p_g - \delta\Delta)}{(1-\delta)(1-\delta\Delta)} R(\bar{k}_b)$$

(b) When $p_b \leq \phi(p_g)$, or equivalently $\boldsymbol{p} \in B_+$ or $B_=$, Lemma I.4 implies:

$$\bar{m}_g^b = R(\bar{k}_b),\ \bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b) \tag{I.7}$$

Comibine [I.3] to [I.5], and [I.7], we obtain the solution:

$$\bar{v}_b = \frac{\delta^2 p_b p_g R(\bar{k}_g)}{(1+\delta p_g)(1-\delta)(1-\delta\Delta)}$$

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

$$\bar{v}_g^g = R(\bar{k}_g) + \frac{\delta^2 p_g(p_g - \delta\Delta)}{(1+\delta p_g)(1-\delta)(1-\delta\Delta)} R(\bar{k}_g)$$

□

J. Proofs from Section 9

In this section, we show the initial debt value is higher at the good state.

At the beginning of the contracting relation, the principal chooses $\mathbf{v} \in V$ to maximize the expected debt value: $Q(\mathbf{v}, s) - p_s v_g - (1 - p_s)v_b$. Let $\hat{\mathbf{v}}^s$ be the optimal contingent utilities at initial stage, when the state is s. Let $\mathcal{D}(\hat{\mathbf{v}}^s, s) = Q(\hat{\mathbf{v}}^s, s) - p_s \hat{v}^s_g - (1 - p_s)\hat{v}^s_b$.

Proof of Proposition 9.1. Let $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g)$ be the optimal policy at state $(\hat{\mathbf{v}}^b, b)$. The optimal debt value at the bad state is:

$$\begin{aligned}
\mathcal{D}(\hat{\mathbf{v}}^b, b) = Q(\hat{\mathbf{v}}^b, b) &- p_b \hat{v}^b_g - (1 - p_b)\hat{v}^b_b \\
&= -k + p_b R(k) + \delta p_b [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] \\
&\quad + \delta Q(\mathbf{w}_b, b) - p_b(\hat{v}^b_g - \hat{v}^b_b) - \hat{v}^b_b \\
&= -k + p_b R(k) + \delta p_b [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - p_b(\hat{v}^b_g - \hat{v}^b_b) \\
&\quad + \delta [Q(\mathbf{w}_b, b) - p_b w_{bg} - (1 - p_b) w_{bb}] + m_b \\
&\leq -k + p_b R(k) + \delta p_b [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - p_b(\hat{v}^b_g - \hat{v}^b_b) \\
&\quad + \delta [Q(\mathbf{w}_b, b) - p_b w_{bg} - (1 - p_b) w_{bb}]
\end{aligned}$$

which implies

$$\begin{aligned}
&p_b \{R(k) + \delta [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (\hat{v}^b_g - \hat{v}^b_b)\} \\
&\qquad \geq \mathcal{D}(\hat{\mathbf{v}}^b, b) - \delta [Q(\mathbf{w}_b, b) - p_b w_{bg} - (1 - p_b) w_{bb}] + k
\end{aligned} \tag{J.1}$$

If $Q(\mathbf{w}_b, b) - p_b w_{bg} - (1 - p_b) w_{bb} \leq 0$, then the right hand side of [J.1] is positive. If $Q(\mathbf{w}_b, b) - p_b w_{bg} - (1 - p_b) w_{bb} > 0$, then because $\delta < 1$ and because $\hat{\mathbf{v}}^b$ maximizes debt value when $s = b$, we still have the right hand side of [J.1] is positive. Hence,

$$R(k) + \delta [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (\hat{v}^b_g - \hat{v}^b_b) > 0 \tag{J.2}$$

Then we know

$$\begin{aligned}
&\mathcal{D}(\hat{\mathbf{v}}^g, g) - \mathcal{D}(\hat{\mathbf{v}}^b, b) \\
&\quad \geq Q(\hat{\mathbf{v}}^b, g) - p_g \hat{v}^b_g - (1 - p_g)\hat{v}^b_b - [Q(\hat{\mathbf{v}}^b, b) - p_b \hat{v}^b_g - (1 - p_b)\hat{v}^b_b] \\
&\quad = Q(\hat{\mathbf{v}}^b, g) - Q(\hat{\mathbf{v}}^b, b) - \Delta(\hat{v}^b_g - \hat{v}^b_b) \\
&\quad \geq \Delta\{R(k) + \delta [Q(\mathbf{w}_g, g) - Q(\mathbf{w}_b, b)] - (\hat{v}^b_g - \hat{v}^b_b)\} > 0
\end{aligned}$$

The first inequality is from the fact that $(k, m_b, m_g, \mathbf{w}_b, \mathbf{w}_g) \in \Gamma(\hat{\mathbf{v}}^b, g)$, and the second inequality is from [J.2].

To see that investment is inefficient, notice that because $\hat{\mathbf{v}}^s$ maximizes debt value, we have

$$Q_b(\hat{\mathbf{v}}^s, s) = 1 - p_s, \; Q_g(\hat{\mathbf{v}}^s, s) = p_s \tag{J.3}$$

By [J.3], $1 = \mathsf{D}_{(1,1)}\, Q(\hat{\mathbf{v}}^s, s) = \Psi_z(\delta(p_s \hat{v}_g^s + (1 - p_s)\hat{v}_b^s), s)$. Suppose investment is efficient at the initial stage which implies $\lambda(\hat{\mathbf{v}}^g, g) = 0$ by [FOCk]. By adding up [FOCw_{gb}] and [FOCw_{gg}], we get $p_g\, \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) = Q_g(\hat{\mathbf{v}}^g, g) - \lambda(\hat{\mathbf{v}}^g, g) = Q_g(\hat{\mathbf{v}}^g, g) = p_g$, and hence $1 = \mathsf{D}_{(1,1)}\, Q(\mathbf{w}_g, g) = \Psi_z(\hat{v}_g^g, g)$. By the strict concavity of Ψ, we know $\hat{v}_g^g = \delta(p_g \hat{v}_g^g + (1 - p_g)\hat{v}_b^g) \geq \delta \hat{v}_g^g > \hat{v}_g^g$, a contradiction. □

K. Additional Statements

Proposition K.1. Suppose the firm is mature, and consider a maximal rent contract, so the equity level is $\bar{\mathbf{v}}^s$ for some $s \in \{b, g\}$. Then, the following hold:

(a) $\bar{m}_b(\bar{\mathbf{v}}^s, s) = 0$.
(b) If $s = g$, then [IC] holds with equality, and $\bar{m}_g\big((\bar{\mathbf{v}}^g, g)\big) < R(\bar{k}_g)$, ie, [LL] holds as an inequality.
(c) If $s = b$ and $(p_b, p_g) \in B_-$, then [IC] holds as an equality while limited liability [LL] is a strict inequality, ie, $\bar{m}_g\big((\bar{\mathbf{v}}^b, b)\big) < R(\bar{k}_b)$.
(d) If $s = b$ and $(p_b, p_g) \in B_+$, then [IC] holds as an inequality while limited liability [LL] is an equality, ie, $\bar{m}_g\big((\bar{\mathbf{v}}^b, b)\big) = R(\bar{k}_b)$.
(e) If $s = b$ and $(p_b, p_g) \in B_=$, then both [IC] and [LL] hold as an equality, so that $\bar{m}_g\big((\bar{\mathbf{v}}^b, b)\big) = R(\bar{k}_b)$.

Proof. (a) From Proposition [5.1], $m_b(\mathbf{v}, s) = 0$ on any path induced by the optimal contract. So by continuity $\bar{m}_b^s = m_b(\bar{\mathbf{v}}^s, s) = 0$.
(b) We will show that $\bar{m}_g^g < R(\bar{k}_g)$. Suppose $\bar{m}_g^g = R(\bar{k}_g)$. From [I.2], $\bar{m}_g^b - R(\bar{k}_b) = \bar{v}_g^g - \bar{v}_g^b > 0$. This means $\bar{m}_g^b > R(\bar{k}_b)$, which violates [LL] for g at $(\bar{\mathbf{v}}^b, b)$, a contradiction. Then $\bar{m}_g^g < R(\bar{k}_g)$ implies [IC] must hold as equality at $(\bar{\mathbf{v}}^g, g)$ by Lemma I.1.
(c) We will show that when $\boldsymbol{p} \in B_-$, we have $\bar{m}_g^b < R(\bar{k}_b)$. Suppose $\bar{m}_g^b = R(\bar{k}_b)$. By [I.2], we know $\bar{v}_g^g - \bar{v}_g^b = R(\bar{k}_g) - \bar{m}_g^g = R(\bar{k}_g) - \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. The last equality is from (b) that [IC] holds as equality at $(\bar{\mathbf{v}}^g, g)$ and hence

$\bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$ by [I.1]. So $\delta p_g(\bar{v}_g^g - \bar{v}_g^b) = \frac{\delta p_g R(\bar{k}_g)}{1+\delta p_g} < R(\bar{k}_b)$. The last inequality is from Lemma I.1 when $\boldsymbol{p} \in B_-$. So $\delta p_g(\bar{v}_g^g - \bar{v}_g^b) < R(\bar{k}_b) = \bar{m}_g^b$. However, comparing with [I.1], we know [IC] at $(\bar{\mathbf{v}}^b, b)$ is violated. Therefore, we must have $\bar{m}_g^b < R(\bar{k}_b)$. By Lemma I.2, [IC] must hold as equality at $(\bar{\mathbf{v}}^b, b)$.

(d) Suppose [IC] holds as equality at $(\bar{\mathbf{v}}^b, b)$ when $\boldsymbol{p} \in B_+$. Then [IC] holds as equality for both $(\bar{\mathbf{v}}^b, b), (\bar{\mathbf{v}}^g, g)$. From [I.1], we know $\bar{m}_g^b = \bar{m}_g^g = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$. And from [I.2], $\bar{v}_g^g - \bar{v}_g^b = R(\bar{k}_g) - R(\bar{k}_b)$. So $\bar{m}_g^b = \delta p_g[R(\bar{k}_g) - R(\bar{k}_b)] > R(\bar{k}_b)$. The last inequality is from Lemma I.1 when $\boldsymbol{p} \in B_+$. However, this means [LL] for g at $(\bar{\mathbf{v}}^b, b)$ is violated. Therefore, we must have [IC] holds as inequality, and hence [LL] for g holds as equality at $(\bar{\mathbf{v}}^b, b)$ by Lemma I.2.

(e) When $\boldsymbol{p} \in B_=$, the same procedure as in the previous two parts shows that $\bar{m}_g^b = R(\bar{k}_b)$ and $\bar{m}_g^b = \delta p_g(\bar{v}_g^g - \bar{v}_g^b)$ imply one another. Therefore, by Lemma I.2, both [IC] and [LL] for g hold as equality at $(\bar{\mathbf{v}}^b, b)$. □

References

Abreu, Dilip, David Pearce and Ennio Stacchetti. 1990. 'Toward a Theory of Discounted Repeated Games with Imperfect Monitoring'. *Econometrica* 58 (5): 1041–1063. (Cit. on pp. 17, 38).

Aragon, George O., and Jun Qian. 2010. *High-water Marks and Hedge Fund Compensation*. Tech. rep. (Cit. on p. 34).

Battaglini, Marco. 2005. 'Long-Term Contracting with Markovian Consumers'. *American Economic Review* 95 (3): 637–658. (Cit. on p. 9).

Biais, Bruno, Thomas Mariotti and Jean-Charles Rochet. 2013. 'Dynamic Financial Contracting'. In *Advances in Economics and Econometrics, Tenth World Congress, Volume 1, Economic Theory*, ed. by Daron Acemoglu, Manuel Arellano and Eddie Dekel, 125–171. Cambridge University Press. (Cit. on p. 3).

Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet. 2007. 'Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications'. *Review of Economic Studies* 74 (2): 345–390. (Cit. on pp. 3, 7, 8).

Bloedel, Alexander W., and R. Vijay Krishna. 2014. *Misery, Persistence, and Growth*. Tech. rep. (Cit. on p. 9).

Bolton, Patrick, and David S. Scharfstein. 1990. 'A Theory of Predation Based on Agency Problems in Financial Contracting'. *American Economic Review* 80 (1): 93–106. (Cit. on pp. 7, 13).

Bulan, Laarni T., Narayanan Subramanian and Lloyd Tanlu. 2007. 'On the Timing of Dividend Initiations'. *Financial Management* 36:31–65. (Cit. on p. 33).

Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. 'A Theory of Financing Constraints and Firm Dynamics'. *Quarterly Journal of Economics* 121:229–265. (Cit. on pp. 3, 4, 7, 8, 11, 24, 27, 32, 36).

Covas, Francisco, and Wouter J. Den Haan. 2011. 'The Cyclical Behavior of Debt and Equity Finance'. *American Economic Review* 101 (2): 877–899. (Cit. on p. 34).

DeMarzo, Peter, and Michael J. Fishman. 2007. 'Agency and Optimal Investment Dynamics'. *Review of Financial Studies* 20 (1): 151–188. (Cit. on pp. 3, 7, 8).

DeMarzo, Peter, and Yuliy Sannikov. 2006. 'Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model'. *Journal of Finance* 61 (6): 2681–2724. (Cit. on p. 8).

Fama, Eugene F., and Kenneth R. French. 2005. 'Financing Decisions: Who Issues Stock'. *Journal of Financial Economics* 76:549–582. (Cit. on pp. 5, 6, 33, 34).

Fernandes, Ana, and Christopher Phelan. 2000. 'A Recursive Formulation for Repeated Agency with History Dependence'. *Journal of Economic Theory* 91:223–247. (Cit. on pp. 6–9, 14, 17).

Gertler, Mark, and Simon Gilchrist. 1994. 'Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms'. *Quarterly Journal of Economics* CIX:309–340. (Cit. on p. 3).

Gomes, Joao F. 2001. 'Financing Investment'. *American Economic Review* 91 (5): 1263–1285. (Cit. on pp. 3, 4).

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. (Cit. on pp. 35, 36).

Guo, Yingni, and Johannes Hörner. 2014. *Dynamic Mechanisms without Money*. Tech. rep. (Cit. on p. 9).

Halac, Marina, and Pierre Yared. 2014. 'Fiscal Rules and Discretion under Persistent Shocks'. *Econometrica* 82 (5): 1557–1614. (Cit. on p. 9).

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

Kaplan, Steven N., and Per Strömberg. 2004. 'Characteristics, Contracts, and Actions: Evidence from Venture Capital Analyses'. *Journal of Finance* 59:2177–2210. (Cit. on p. 33).

– . 2003. 'Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts'. *Review of Economic Studies* 70 (2): 281–315. (Cit. on pp. 5, 33).

Krishna, R. Vijay, Giuseppe Lopomo and Curtis R. Taylor. 2013. 'Stairway to Heaven or Highway to Hell: Liquidity, Sweat Equity, and the Uncertain Path to Ownership'. *RAND Journal of Economics* 44 (1): 104–127. (Cit. on p. 36).

Murphy, Kevin J. 1999. 'Executive Compensation'. In *Handbook of Labor Economics, Volume 3*, ed. by Orley Ashenfelter and David Card. North-Holland, Amsterdam. (Cit. on p. 33).

Myers, Stewart C. 1984. 'The Capital Structure Puzzle'. *Journal of Finance* 39:575–592. (Cit. on p. 5).

Myers, Stewart C., and Nicholas S. Majluf. 1984. 'Corporate Financing and Investment Decisions when Firms Have Information That Investors Do Not Have'. *Journal of Financial Economics* 12:187–221. (Cit. on p. 5).

Thomas, Jonathan P., and Tim Worrall. 1990. 'Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem'. *Journal of Economic Theory* 51:367–390. (Cit. on pp. 8, 27, 28, 36).