A Theory of Dynamic Contracting with Financial Constraints*

Ilya Krasikov[†]

Rohit Lamba[‡]

December 2016

Abstract

We study a dynamic principal-agent model where the agent has access to a persistent private technology but is strapped for cash. Financial constraints are generated by the periodic interaction between incentives (private information) and feasibility (being strapped for cash). This interaction produces dynamic distortions that are a sum of two effects: backloading of incentives and illiquidity. Bad technology shocks increase distortions and monotonically push the agent further away from efficiency. An endogenous number of good shocks is required for the agent to become liquid, and eventually for the contract to become efficient. Efficiency is an absorbing state that is reached almost surely. The optimal allocation can be implemented through a mechanism which is precisely pinned down by a dynamic information operator. The shares of principal and agent in the net present value of economic surplus are endogenous to the evolution of technology shocks. Surplus itself is increasing in the share of the agent, and in his type contingent utility spread. By comparing the agent's utility with and without financial constraints, the model provides a foundation for the usefulness of limited liability in dynamic contracts.

0 Introduction

Financial constraints are ubiquitous, and economists have arguably converged to a view that their incorporation in our models is inevitable, even desirable. The emphasis has been pronounced post the Great Recession, so much so that a close cousin- financial frictions- is now colloquially understood to be an important causal element of inefficient allocation of capital at a micro level, and a key driver of crises episodes at the macro.

In an elegant note, Nobuhiro Kiyotaki advocates "a mechanism design approach to illustrate how different environments of private information and limited commitment generate different financial frictions."¹ In the spirit of the said agenda, this paper posits financial constraints as a

*We are indebted to Nageeb Ali for his advice and encouragement. Thanks also to Felix Bierbrauer, Ed Green, Nima Haghpahan, Vijay Krishna, Bruno Strulovici, and to the participants at the Pennsylvania Economic Theory Conference, Vanderbilt Mechanism Design Conference, and Northwestern University for their comments.

[†]Pennsylvania State University, izk113@psu.edu

[‡]Pennsylvania State University, rlamba@psu.edu.

¹Kiyotaki [2012].

product of the interaction between (i) "persistent agency frictions", and (ii) "limitations on the ability of agents to generate timed cash flows".

Agency frictions can be modeled as moral hazard or adverse selection. All dynamic financial contracting papers thus far have explored moral hazard (or cash flow diversion models), and almost all have focussed on iid technology. In contrast, this paper looks at a dynamic adverse selection (or screening) model with persistence in technology. It provides a tractable framework of endogenously generated financial constraints, and delivers clear insights on the short-run and long-run outcomes of contracting in such an environment.

Persistence in technology shocks lends an important empirical relevance to financial contracting. It is a natural assumption for durable technologies owned and managed by a firm over a period of time. For example, İmrohoroglu and Tüzel [2014] find the average persistence in total factor productivity of firms in Compustat data from 1962 to 2009 to be 0.7. Most dynamic financial contracting models thus far have restricted themselves to iid models of cash flow diversion.

Specifically, we study a dynamic contracting model with persistent asymmetric information and cash (or limited liability) constraints. A big firm (principal) repeatedly producing a final good contracts with a smaller firm (agent) that supplies an important input. The marginal cost of production for the smaller firm is its *private information*, and it is imperfectly correlated across time. Moreover, the small firm lacks capital- it cannot post a bond, collateralize its assets or forgo payments. In simple terms it is *strapped for cash*. Mathematically, this restriction demands positivity of per-period (or stage) utility of the agent.²

The big firm is tasked with designing a contract which sets supply of inputs by the small firm, and regular payments for its production. As motivated, financial constraints are produced by the simultaneous interaction of persistent private information and the cash-strapped constraint. Relaxing either of these leads to greater or complete efficiency, arguably limiting the empirical appeal of the model.

Ideally, the big firm would like the small firm to produce the efficient quantity in return for requisite payments. However, agency frictions prevent the efficient allocation from being implemented. The optimal quantity (or allocation) is determined by the trade-off between efficiency and rents:

$$\max_q [\text{surplus} - \text{information rent}] \Rightarrow \underbrace{d_q(\text{surplus})}_{\text{marginal benefit}} = \underbrace{d_q(\text{information rent})}_{\text{marginal cost}} \quad (\star)$$

Equation ($\star$)- pointwise optimization of quantities for each type of the agent- forms the backbone of our analysis. The left-hand side reports the derivative of surplus produced by a realized type. The small firm can have a low ("good shock") or high ("bad shock") marginal cost. Economically speaking it captures the marginal benefit associated with the optimal supply contract. The right hand side reports the derivative of information rent that has to be paid to the small firm to

²There is a large literature explaining inability of start-ups or small businesses to raise capital, which forms the core empirical motivation of our model. For example, Campello et al. [2010] conducted a survey of 1050 CFOs across the US, Europe and Asia, and found considerable impact of credit constraints on firm behavior in the aftermath of the Great Recession. Banerjee and Duflo [2014] exploit a change in policy by the Indian government to show that most firms in their study were credit constrained, and a relaxation of the same led to a spurt in growth.

Figure 1: The optimal contract

incentivize it to follow the recommendation of the big firm. It captures the marginal cost associated with incentive provision.³ Our paper pins down how optimal distortions determined by the equation evolve in a *dynamic contract* with *financial constraints*.

Structure of the Optimal Contract. The contract offered by the big firm is described by two vectors: allocation ($\mathbf{q}$) and expected utility ($\mathbf{U}$) of the small firm. The latter refers to the discounted sum of lifetime utility of the small firm at any point of the contract tree. It can be broken down into two components: current (or stage) and promised (or continuation) utility. Let h^t be a typical history of cost realizations. For example, $h^3 = (\theta_L, \theta_H, \theta_H)$ is a three period history where a low cost was followed by two consecutive high costs.⁴ A plausible contract identifies quantities and expected utilities that satisfy incentive compatibility and the cash-strapped constraint along every such history. An optimal contract maximizes the big firm's profit amongst the set of plausible ones. Moreover, by varying the ex ante bargaining power of the agent at the optimum we also map the Pareto frontier.

The evolution of $\mathbf{q}$ and $\mathbf{U}$ at the optimum is explained through two pictures. Figure 1a depicts a sequence of technology shocks. For $i = L, H$, let $q(\theta_i|h^t)$ and $U(\theta_i|h^t)$ be the allocation and expected utility for cost realizations (h^t, θ_i) . At this point, if the right hand side of equation ($\star$) is zero, then $q(\theta_i|h^t) = q^e(\theta_i)$, that is the efficient quantity is supplied. If it is positive, then $q(\theta_i|h^t) = q^e(\theta_i) - d(\theta_i|h^t)$ where d measures the history dependent optimal distortion.

As is standard, the low cost type always supplies the efficient quantity: $q(\theta_L|h^t) = q^e(\theta_L)$. On the other hand, each "bad shock" increases optimal distortions: $q(\theta_H|h^t, \theta_H) < q(\theta_H|h^t) < q^e(\theta_H)$. This is in striking contrast to dynamic mechanisms without financial constraints that em-

³This insight has been employed extensively in the literature on contract theory, see for example Stole [2001], and auctions (where it is called virtual valuation), see for example Myerson [1981].

⁴A low cost realization is always better for economic surplus than a high cost realization.

phasize progressively decreasing distortions along all histories (see Besanko [1985] and Battaglini [2005]).

Allowing for a long-term contract helps mitigate the problem of agency frictions- this mitigation is achieved by backloading incentives. Financial constraints, though, restrict the extent of backloading. Dynamic distortions in our framework are an additive sum of two effects: backloading of incentives and illiquidity due to financial constraints; the latter increases with each "bad shock", overturning the standard result.

Moreover, the realization of a "good shock" decreases the optimal distortion: $q(\theta_H|h^t) < q(\theta_H|h^t, \theta_L)$. An endogenous number of consecutive "good shocks", say $n(h^t)$, is required for the optimal distortion to reach zero. For every additional "bad shock", as distortions increase, the number increases: $n(h^t, \theta_H) \geq n(h^t)$. Once the optimal distortion reaches zero it stays zero, that is, efficiency is an absorbing state. In the long run, the optimal contract almost surely converges to the efficient allocation.

Each period the big firm provides the small firm with transfers and makes a long-term promise of expected utility: $U(\theta_i|h^t)$. With reference to Figure 1a, the expected utilities of both the low cost and high type go up after a "good shock" and go down after a "bad shock". That is, $(U(\theta_L|h^t, \theta_H), U(\theta_H|h^t, \theta_H)) \ll (U(\theta_L|h^t), U(\theta_H|h^t)) \ll (U(\theta_L|h, \theta_L), U(\theta_H|h^t, \theta_L))$.

The evolution of the optimal contract is divided into three regions- illiquidity, liquidity and efficiency- by two thresholds on the vector of expected utilities; see Figure 1b. A contract typically starts in the illiquid region- both the incentive and cash-strapped constraint bind to produce financial constraints that bite. A low cost type keeps the contract in illiquidity or can transition it to liquidity. A high cost type decreases the expected utility of the small firm which keeps it illiquid. After an endogenous number of low cost realizations, the expected utility of the small firm reaches a critical threshold at which the big firm agrees to lax the cash-strapped constraint and provide the small firm with some credit. This is called the liquid region.

Liquidity is not an absorbing state, a high cost realization can push the small firm back into illiquidity. The liquid region forms a penultimate zone towards efficiency. Once liquid, the realization of one more low cost pushes the expected utility of the small firm beyond the second threshold that propels the optimal contract into the absorbing state of efficiency.⁵

At a technical level, we use a mixture of sequential and recursive approaches to characterize the optimal contract. A novelty we bring to the table is the existence of a "shell", a subset of the recursive domain which houses the optimal constrained contract. To the best of our knowledge, this is a new feature of dynamic contracts. We show that as long as the optimal contract is inefficient, the expected utility of the agent must always lie in this shell. It allows us to show all the aforementioned monotonicity properties of the evolution of the optimal contract. It also clearly elucidates the optimal contract in the two limiting cases: perfectly persistent and iid types.

An Economic Implementation. In addition to providing a complete characterization of the

⁵In a cash flow diversion model, Fu and Krishna [2016] also establish liquidity as a penultimate zone towards efficiency. However, all the short-run properties of monotonicity in q , U and n are unique to our paper.

optimal contract, we propose an economically meaningful implementation. At the end of every period, the sum of the continuation utilities of the two firms constitutes the economic surplus. It is equal to the net present value of the "meta firm" borne out of their partnership. In keeping with the literature, we call the continuation utility of the big firm (principal) its profit and that of the small firm (agent) its promised utility. The two dimensional vector of profit and promised utility constitutes the *capital structure* of the "meta firm", which is endogenous to the realization of technology shocks.⁶

At any level of promised utility, the expected utility of the small firm is marked up on the realization of a "good shock" and marked down for a "bad shock". The difference between the two values is termed the utility spread. After a history h^t , the utility spread is simply $U(\theta_L|h^t) - U(\theta_H|h^t)$. It is a measure of the dynamic information rent paid to the low cost type over and above the high cost one. We can exactly pin down the utility spread from the so called dynamic envelope formula.

In order to fully characterize the capital structure we solve for the Pareto optimal contract- each point on the Pareto frontier corresponds to a specific capital structure. As the contract evolves, it endogenously chooses points according to the history of technology shocks. In the illiquid region, the cash-strapped constraint binds and the big firm only provides *working capital* to the small firm. Through a sequence of consecutive low cost realizations, the small firm has to earn its way into liquidity. In the liquid region, the big firm promises to *take-over* the small firm on the realization one more low cost type for a determinable strike price. Thereafter, the small firm operates in-house, producing the efficient quantity.

The value of the "meta firm" is increasing in the share of the agent. As the agent assumes a greater stake of the total surplus, incentives align with bargaining power reducing agency frictions and increasing the size of pie towards its efficient value. A less obvious observation is that along the optimal contract, the value of the "meta firm" is also increasing in the utility spread. Why does an increase in information rent increase the economic surplus? A "good shock" decreases the optimal distortion which in turn increases both the information rent and surplus. A "bad shock" on the other hand, increases optimal distortions which reduces the information rent of the agent and *downsizes* the meta firm by reducing the economic surplus.

The efficient contract represents a mature "meta firm" that has been able to overcome financial frictions- both incentive and cash-strapped constraints are slack. However, in contrast to the iid model the value of the mature firm is Markovian. The total economic surplus fluctuates between two values depending on the last period technology shock. Obviously the corresponding value for the "good shock" is higher than that for a bad one.

The paper provides a framework for the design of real contracts. In the face of uncertain technology and credit constraints, contracts should provide incentives through delayed payments. The terms should be flexible- they should sequentially deteriorate in the event of bad outcomes, and improve with good outcomes. After a certain number of "success", the "agent" should be deemed

⁶In the corporate finance view of our model, the Modigliani-Miller Theorem does not hold since capital structure matters for the value of the firm.

credit worthy, and if he continues to show results, the "principal" should take-over the technology and promote it in-house. Of course, our model is highly stylized and we only state this as a plausible motivation for some real contracts.

Foundation for Limited Liability. We also provide a conceptual foundation for a productive role of limited liability for small businesses. We ask- when should the positivity of stage utility of the small firm be interpreted as a limited liability constraint (as opposed to a cash constraint)? The answer depends on whether the agent does better in the benchmark model without financial constraints or in our model with financial constraints.

Consider the principal profit maximizing contract on the Pareto frontier, that is the big firm has all the bargaining power. The ex ante expected utility of the small firm from the contract is determined endogenously as part of the optimum. We show that for a large measure of parameters (and almost all economically relevant ones), the ex ante expected utility of the agent is higher in our model than in the benchmark. Thus, being "protected by limited liability" helps the agent in a utilitarian sense and is a meaningful interpretation.

Literature. This paper connects two related literatures- dynamic mechanism design and dynamic financial contracting. Our work can be seen as an attempt to bridge the two strands. Dynamic mechanism design has predominantly been focussed on (the challenging task of) modeling dynamic incentives. We arguably require more models in pushing the envelope on what Myerson calls feasibility constraints. Feasibility in our model of course manifests in the agent being strapped for cash. Incentives and feasibility interact every period, allowing us to uncover hitherto unexplored nature of dynamic distortions.

To dynamic financial contracting, we bring some newly developed techniques from the former literature. We pose familiar questions, but move them away from the moral hazard and cash flow diversion models to frameworks of screening and adverse selection. It allows us to illustrate the short-run properties of the optimal contract (through equation (★)), and hence the capital structure of the "meta firm". While the Kiyotaki proposal motivates our modeling choice, we stay fairly loyal to the seminal work by Clementi and Hopenhayn [2006] in our execution of it. Our paper can be viewed as the Markovian dynamic screening counterpart to their iid cash flow diversion model. In a complementary study, Fu and Krishna [2016] consider the Markovian version of the cash flow diversion model. We provide a detailed comparison to both papers in Section 6.

This marriage of ideas in the two literatures can potentially produce many other interesting applications. Two immediately come to mind- optimal taxation and double auctions. It might be reasonable to assume, for example in the presence of incomplete credit markets, that a government cannot prevent citizens from forgoing consumption beyond a certain limit in any given work period. A version of our cash-strapped constraint would then interact with the standard private information of labor productivity in the dynamic Mirrlees model of taxation. In economic transactions captured by repeated double auctions, it might be realistic to assume that neither party can commit to posting large upfront capital that will be honored over time. Again the cash-strapped

constraint will interact with the willingness to buy or sell. These are further discussed in the conclusion as potentials for future research.

1 Model

The key economic forces in dynamic contracting with persistent agency frictions can be formulated through various related models. We choose the repeated version of the marginal cost screening model, based on Laffont and Martimort [2002].

A big firm (principal) specializing in a final good requires a non-durable input that is produced by a smaller firm (agent) every period at a cost θq , where θ is the small firm's private information.⁷ The principal values the final good at $V(q)$, where $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfies the Inada conditions.⁸

The principal pays a price p to the agent for supplying her the intermediate good (or input), and the utility of both is linear in prices.⁹ Therefore, the per-period (or stage) utility for the principal and agent is given by $V(q) - p$ and $p - \theta q$, respectively. The contract lasts for T periods, where $T \leq \infty$. There is a common discount factor δ .

The marginal cost θ , oft referred to as the agent's type, can take on two values: $\Theta = \{\theta_L, \theta_H\}$, where $0 < \theta_L < \theta_H$. It is drawn from a prior $\mu = \{\mu_L, \mu_H\}$, and then evolves according to a Markov process: $f(\theta_L|\theta_i) = \alpha_i$, and $f(\theta_H|\theta_i) = 1 - \alpha_i$ for $i = H, L$. Distributions have full support: $\mu \gg 0$ and $f \gg 0$. For simplicity of exposition, from hereon we will assume a symmetric and persistent Markov process: $\alpha_L = 1 - \alpha_H = \alpha \geq \frac{1}{2}$. In the appendix, we consider the more general asymmetric case.

At the start of every period, the agent privately learns his marginal cost, θ . Given the Markovian nature of shocks, θ is also informative about future types. Both the principal and the agent can commit to a dynamic contract. We solve for the entire Pareto frontier by introducing as a parameter the agent's minimum ex ante share of the total economic surplus, v_0 . The total set of parameters is thus given by: $\Gamma = \{V(\cdot), \Theta, \mu, f, \delta, v_0\}$.

Invoking the revelation principle, we can focus on the direct mechanism. The principal offers a menu of history dependent price-quantity pairs to the agent. Every period the agent reports his marginal type to the principal, which translates into a selection from the menu. The principal's objective is to maximize her expected profit subject to incentive and feasibility constraints, and minimum ex ante utility for the agent.

Formally the mechanism is: $m = \langle \mathbf{p}, \mathbf{q} \rangle = \left(p(\hat{\theta}_t | h^{t-1}), q(\hat{\theta}_t | h^{t-1}) \right)_{t=1}^T$, where h^{t-1} and $\hat{\theta}_t$ are, respectively, the history of reports up to $t - 1$ and current report at time t .¹⁰ The reported history h^t is recursively defined as $h^t = \{h^{t-1}, \hat{\theta}_t\}$ starting with $h^0 = \emptyset$. The set of possible histories at time t is denoted by H^t . As is standard, the contract is restricted to lie in l^∞ .

Define the private history of the agent to be $h_A^t = \{h_A^{t-1}, \theta_t, \hat{\theta}_{t-1}\}$, starting from $h_A^0 = \{\theta_1\}$,

⁷We can introduce a fixed cost of production: $c(\theta, q) = \theta q + F$ without changing any of our results. For simplicity it is normalized to zero: $F = 0$.

⁸Technically: (i) $V'(q) > 0$, $V''(q) < 0$ for all $q \geq 0$, (ii) $V(0) = 0$, (iii) $\lim_{q \rightarrow 0} V'(q) = \infty$, $\lim_{q \rightarrow \infty} V'(q) = 0$.

⁹Throughout, the principal will be referred to as a 'she' and the agent as a 'he'.

¹⁰At the cost of minimal confusion, subscripts will be used interchangeably for time and L/H .

where $\hat{\theta}_t$ and θ_t are the reported and actual types, respectively. Fixing the set of parameters Γ , for a given direct mechanism m , we have a dynamic decision problem described by $\langle m, \Gamma \rangle$ in which the strategy for the agent, $(\sigma_t)_{t=1}^T$, is simply a function that maps his private history into an announcement every period: $h_A^t \mapsto \sigma_t(h_A^t) \in \Theta$.

Define the stage and expected utility of the agent (under truthful reporting) after any history of the contract tree to be

$$u(\theta_t | h^{t-1}) = p(\theta_t | h^{t-1}) - \theta_t q(\theta_t | h^{t-1})$$

$$U(\theta_t | h^{t-1}) = u(\theta_t | h^{t-1}) + \delta \mathbb{E} [U(\tilde{\theta}_{t+1} | h^{t-1}, \theta_t) | \theta_t]$$

It is straightforward to show that the contract space can then be expressed as $\langle \mathbf{u}, \mathbf{q} \rangle$ or $\langle \mathbf{U}, \mathbf{q} \rangle$. We shall use the three formulations interchangeably.

The constraints on the space of contracts can be divided into two categories- equilibrium and feasibility. In the mechanism design lexicon, we refer to the former as incentive compatibility. The contract $\langle \mathbf{U}, \mathbf{q} \rangle$ is said to be *incentive compatible* if truthful reporting is profitable for the agent. Using the one shot deviation principal, formally, $\forall h^{t-1} \in H^{t-1} \forall t$:

$$U(\theta_t | h^{t-1}) \geq p(\hat{\theta}_t | h^{t-1}) - \theta_t q(\hat{\theta}_t | h^{t-1}) + \delta \mathbb{E} [U(\tilde{\theta}_{t+1} | h^{t-1}, \hat{\theta}_t) | \theta_t]$$

for all $\theta_t, \hat{\theta}_t \in \Theta$. The Markovian assumption on stochastic evolution of types ensures that the agent wants to report truthfully even if he has lied in the past.

Two types of feasibility constraints are explored in the paper- individual rationality and strapped for cash. A contract is said to be *individually rational* if it offers each type of the agent a non-negative expected utility after every history. Formally:

$$U(\theta_t | h^{t-1}) \geq 0 \quad \forall \theta_t \in \Theta, h^{t-1} \in H^{t-1}, \forall t$$

And, the contract is said to be *cash-strapped* if it must provide each type of the agent a non-negative stage utility at every history. Formally:

$$u(\theta_t | h^{t-1}) \geq 0 \quad \forall \theta_t \in \Theta, h^{t-1} \in H^{t-1}, \forall t$$

The individual rationality constraint should be seen as a tempering of the commitment assumption on the side of the agent. It keeps the expected "lifetime" utility of the agent in every period above an exogenous threshold normalized to zero. However, it makes it possible that agent be asked to forgo payments or deposit large upfront capital with a promise of being compensated for it later. The cash-strapped constraint precludes contracts with such delayed promises.¹¹

Being cash-strapped is a physical or financial restriction on the magnitude of per period payments. We want the reader to view it primarily as a credit constraint. The small firm has to get

¹¹Note that if a contract is strapped for cash it necessarily satisfies individual rationality while the opposite is not necessarily true.

short term credit to produce the intermediate good and cannot pledge any collateral.^{12, 13}

This constraint has been invoked in various guises in the financial contracting literature, especially in models of moral hazard and cash flow diversion (see for example Clementi and Hopenhayn [2006], Biais et al. [2013], and Myerson [2012]). It is motivated therein primarily as a limited liability restriction. We provide a comparison between the two interpretations- limited liability and being strapped for cash- in Section 4.2.

2 Optimal contract

In order to appreciate the role of each of moving part in the model, and simplify the technical exposition that follows, we divide this section into four parts. First, we briefly exposit the optimal contract in the benchmark model with individual rationality. Since the agent has access to deep pockets in this framework, contrasting it with our results would help the reader understand the role played by unrestricted movement of transfers across time. Second, we explicitly solve the two-period model with financial constraints to communicate the key forces as simply as possible. Third, extending the sequential characterization, we point to an informative Lagrangian approach. And, finally we solve for the optimal contract using a recursive formulation of the problem.

Define $s(\theta, q) = V(q) - \theta q$ to be the static surplus, succinctly expressed as $s(\theta) = V(q(\theta)) - \theta q(\theta)$ for the direct mechanism. It is straightforward to note that the *efficient quantity* that maximizes the surplus is given by $V'(q^e(\theta)) = \theta$. Moreover, let $\bar{S} = \sum_{t=1}^T \delta^{t-1} \mathbb{E} [s(\tilde{\theta}_t)]$ be the (ex ante) expected surplus. The principal's problem, $(\mathcal{P}^*)$, can be stated as:

$$(\mathcal{P}^*) \quad \max_{\langle \mathbf{U}, \mathbf{q} \rangle} \quad \bar{S} - [\mu_L U(\theta_L) + \mu_H U(\theta_H)]$$

subject to $\mathbf{q} \geq 0$,

$$\mu_L U(\theta_L) + \mu_H U(\theta_H) \geq v_0, \text{ and} \quad (\text{PK})$$

$$IC_L(h^{t-1}), IC_H(h^{t-1}), C_L(h^{t-1}), C_H(h^{t-1}) \forall h^{t-1} \in H^{t-1} \forall t$$

where (PK) is the ex ante promise keeping constraint, and $IC_i(h^{t-1})$ and $C_i(h^{t-1})$ are the incentive and cash-strapped constraints, respectively, for type θ_i in period t after history h^{t-1} . Note that v_0 parameterizes the bargaining power of the agent, and maps the Pareto frontier. Since quantity is always non-negative at the optimum, we shall drop that constraint.

We consider a relaxed problem where we ignore $IC_H(h^{t-1})$ for all histories. A justification of this is provided in Section 2.6, including sufficient conditions for global optimality. The principal's

¹²An alternate way to express the cash-strapped constraint would be $p_t \geq C$ for all t , that is payment to the small firm or the agent has to be above a minimum constant amount every period. We allow the boundary to move with the supply contract, that is $p_t \geq \theta_t q_t$, both for realism and tractability. Think of a third party that disburses short term credit to the smaller firm that is able to verify its type and amount of production.

¹³Just as the incentive compatibility constraint, the cash-strapped constraint holds both "on" and "off" path. Even if the agent may have misreported in the past, the principal delivers a non-negative stage utility to him if he is truthful today.

relaxed problem, $(\mathcal{RP}^*)$, reads as follows:

$$(\mathcal{RP}^*) \quad R^*(v_0) = \max_{\langle \mathbf{U}, \mathbf{q} \rangle} \bar{S} - [\mu_L U(\theta_L) + \mu_H U(\theta_H)]$$

subject to (PK), and

$$IC_L(h^{t-1}), C_L(h^{t-1}), C_H(h^{t-1}) \forall h^{t-1} \in H^{t-1} \forall t$$

where $R^*(v_0)$ is the value of the objective- the principal's profit at the constrained optimum. Also, we shall denote the ex ante economic surplus generated by the optimal contract by $S^*(v_0)$.

The Myersonian quest herein is to write down an optimization problem equivalent to $(\mathcal{RP}^*)$ where a subset of binding incentive constraints is used to eliminate $\mathbf{U}$, and the objective and all remaining constraints are expressed only in terms of $\mathbf{q}$. Pointwise optimization of allocations along all histories then yields the efficient quantity for the low cost type: $q(\theta_L|h^{t-1}) = q^e(\theta_L)$, and for the high cost type:

$$\underbrace{\mathbb{P}(h^{t-1}, \theta_H) \left(V'(q(\theta_H|h^{t-1})) - \theta_H \right)}_{\text{marginal benefit}} = \underbrace{r(h^{t-1})}_{\text{marginal cost}} \quad (*)$$

where $r(h^{t-1})$ represents the marginal impact of history h^{t-1} on the information rent of the agent. It is a function of Γ , the parameters of the model. Equation (*) is the exact counterpart to equation (★) from the introduction. We shall use a mixture of sequential and recursive approaches to precisely characterize $r(h^{t-1})$ for all histories, which in turn pins down the evolution of optimal quantities and expected utilities.

2.1 Benchmark: dynamic model without financial constraints

Before presenting novel results with the cash-strapped constraint, we consider the model where the agent has access to deep pockets. In doing so we want to emphasize the role played by the free movement of transfers across time, in particular the ability of the agent to deposit upfront capital.

The problem looks exactly the same as $(\mathcal{P}^*)$, except that $C_i(h^{t-1})$ is replaced by $IR_i(h^{t-1})$ for $i = L, H$ and for all h^{t-1} . Dynamic contracting models of this form have been studied amongst others by Besanko [1985], Courty and Li [2000], Battaglini [2005] and Pavan et al. [2014]. In the context of our framework, the optimal allocation, $\mathbf{q}^\#$, is characterized by two facts.

Facts. Let $\theta_H^t \in H^t$ represent the history where each report until period t has been θ_H . The optimal allocation in the benchmark model:

1. becomes efficient forever as soon as the agent becomes a low cost type: $q^\#(\theta_L|h^{t-1}) = q^e(\theta_L) \forall h^{t-1}$, and $q^\#(\theta_H|h^{t-1}) = q^e(\theta_H) \forall h^{t-1} \neq \theta_H^{t-1}$
2. has decreasing distortions along the constant high cost history: $q^\#(\theta_H|\theta_H^{t-1}) = q^e(\theta_H) - d(\theta_H|\theta_H^{t-1})$, where $d(\theta_H|\theta_H^{t-1})$ is decreasing in t .

Battaglini [2005] terms these *generalized no distortion at the top* and *vanishing distortions at the bottom*, respectively. Drawing from equation (*), $r(h^{t-1}) = 0$ for all $h^{t-1} \neq \theta_H^{t-1}$. Once a "good shock" arrives, the marginal cost of incentive provision becomes zero. On the other hand, $d(\theta_H|\theta_H^{t-1}) \propto \frac{r(\theta_H^{t-1})}{\mathbb{P}(\theta_H^t)}$ which decreases over time. That is, marginal cost of incentive provision decreases along the "lowest history".¹⁴

Multiple prices implement the optimal allocation. Incentive and individual rationality constraints are satisfied by a continuum of values with the restriction that first period expected utilities- $U(\theta_L)$ and $U(\theta_H)$ - are uniquely pinned down. Without financial constraints, incentives and feasibility interact only in the first period- IR_H is the only individual rationality constraint that binds at the optimum. How much is this instrument of free movement of prices across time exploited by the principal?

2.2 Two period model

Consider problem $(\mathcal{RP}^*)$ for $T = 2$. All possible sequence of type realizations are depicted in Figure 2. In addition to (PK), we have to consider the following set of constraints:

$$IC_L, C_L, C_H, \text{ and } IC_L(\theta_i), C_L(\theta_i), C_H(\theta_i) \text{ for } i = L, H$$

It can be shown that $C_L(\theta_L)$ and $C_L(\theta_H)$ are implied by the other constraints. IC_L , $IC_L(\theta_H)$, C_H and $C_H(\theta_H)$ (shown in blue solid arrow and circle) in Figure 2, all bind at the optimum. Moreover, C_L (shown in a red square) can bind sometimes. Other constraints in the relaxed problem, $IC_L(\theta_L)$ and $C_H(\theta_L)$ (shown in a red dotted arrow and red diamond respectively) can be assumed to hold as an equality, they bind if C_L does.

Using the set of binding constraints, we want to express $\mu_L U(\theta_L) + \mu_H U(\theta_H)$ as a function of quantities. The key constraint here is the first period incentive constraint IC_L which binds at the optimum:

$$\begin{aligned} U(\theta_L) &= \Delta\theta q(\theta_H) + u(\theta_H) + \delta [\alpha u(\theta_L|\theta_H) + (1 - \alpha)u(\theta_H|\theta_H)] \\ &= U(\theta_H) + \Delta\theta q(\theta_H) + \delta(2\alpha - 1) [u(\theta_L|\theta_H) - u(\theta_H|\theta_H)] \end{aligned}$$

The term $(2\alpha - 1)$ is essentially the impact of misreport by agent on his expected utility in period 2. Using the second period binding incentive constraint $IC_L(\theta_H)$ further gives

$$U(\theta_L) = U(\theta_H) + \Delta\theta q(\theta_H) + \delta(2\alpha - 1)\Delta\theta q(\theta_H|\theta_H) \quad (1)$$

Rewriting it in the form of an "envelope formula":

$$\frac{U(\theta_L) - U(\theta_H)}{\Delta\theta} = q(\theta_H) + \delta \left(\frac{2\alpha - 1}{\alpha} \right) \alpha q(\theta_H|\theta_H) \quad (2)$$

Equation (2) is a mini version of a much more general formula elegantly derived for continuous

¹⁴Formal details are provided in the appendix in Section 8.1.

Figure 2: The binding constraints in the two period model

type spaces in Pavan et al. [2014].¹⁵ The term $\left(\frac{2\alpha-1}{\alpha}\right)$ has been referred to variedly in the literature as the informativeness measure, impulse response and dynamic distortion.

In the benchmark model, the binding IR_H constraint would deliver zero expected utility for the high cost type. However, in the presence of the cash-strapped constraint we have¹⁶

$$U(\theta_H) = \delta(1 - \alpha)\Delta\theta q(\theta_H|\theta_H) \quad (3)$$

Through equations (1) and (3), it is clear that being strapped for cash interacts with the incentive constraint to ensure information rents need to be paid to both the low and high cost types. Additively:

$$\begin{aligned} \mu_L U(\theta_L) + \mu_H U(\theta_H) &= \Delta\theta \mu_L q(\theta_H) + \delta \Delta\theta (\mu_L \alpha + \mu_H (1 - \alpha)) q(\theta_H|\theta_H) \\ &= \Delta\theta \mathbb{P}(\theta_1 = \theta_L) q(\theta_H) + \delta \Delta\theta \mathbb{P}(\theta_2 = \theta_L) q(\theta_H|\theta_H) \end{aligned} \quad (4)$$

where $\mathbb{P}(\theta_t = \theta_L)$ is the ex ante probability of being the low cost effective in period t . Define the threshold generated by equation (4) for the efficient quantity to be $\bar{v}$:

$$\bar{v} = \Delta\theta \sum_{t=1}^2 \delta^{t-1} \mathbb{P}(\theta_t = \theta_L) q^e(\theta_H|\theta_H^{t-1}) = \Delta\theta \sum_{t=1}^2 \delta^{t-1} \mathbb{P}(\theta_t = \theta_L) q^e(\theta_H) \quad (5)$$

and that generated by the optimal contract when we ignore (PK) to be $\underline{v}$.¹⁷

Finally, what about C_L ? It is easy to see that

$$\begin{aligned} u(\theta_L) &= U(\theta_L) - \delta [\alpha u(\theta_L|\theta_L) + (1 - \alpha) u(\theta_H|\theta_L)] \\ &= \Delta\theta q(\theta_H) + \delta \Delta\theta \alpha [q(\theta_H|\theta_H) - q(\theta_H|\theta_L)] \end{aligned} \quad (6)$$

¹⁵Battaglini and Lamba [2015] derive the same formula for discrete types. Eső and Szentes [2015] also have a clean derivation of the result for continuous types.

¹⁶Equation (3) is generated by the binding C_H , $C_H(\theta_H)$ and $IC_L(\theta_H)$ constraints: $U(\theta_H) = u(\theta_H) + \delta [(1 - \alpha)u(\theta_L|\theta_H) + \alpha u(\theta_H|\theta_H)] = \delta(1 - \alpha)\Delta\theta q(\theta_H|\theta_H)$.

¹⁷ $\underline{v}$ refers to the agent's expected utility on the principal profit maximizing point on the Pareto frontier. If $v_0 < \underline{v}$, then (PK) does not bind.

Therefore, C_L can be expressed as

$$C_L : q(\theta_H) + \delta\alpha q(\theta_H|\theta_H) \geq \delta\alpha q(\theta_H|\theta_L)$$

If ignored, parameters leading to large distortions along the high cost history would ensure that C_L is violated. So we introduce this as a constraint in terms of quantities, in place of $u(\theta_L) \geq 0$. We can now pin down the optimal allocation rule. The principal chooses $\mathbf{q}$ to maximize $\bar{S} - [\mu_L U(\theta_L) + \mu_H U(\theta_H)]$ subject to (PK) and C_L , where $\mu_L U(\theta_L) + \mu_H U(\theta_H)$ is given by equation (4). The precise closed form solution is provided in Lemma 3 in the appendix. Here we deliver the basic economic message.

Proposition 1. *The optimal contract, $\mathbf{q}^*$, with promised utility $v_0 \in [0, \bar{v})$, is characterized by the following allocation rule:*

1. $q^*(\theta_L|h) = q^e(\theta_L)$ for $h = \emptyset, \theta_L, \theta_H$.
2. $q(\theta_H|\theta_L) = q^e(\theta_H) - d(\theta_H|\theta_L)$, where $d(\theta_H|\theta_L) \geq 0$, and $d(\theta_H|\theta_L) > 0 \Leftrightarrow C_L$ binds.
3. $q^*(\theta_H|h) = q^e(\theta_H) - d(\theta_H|h)$ for $h = \emptyset, H$, where $d(\theta_H|\theta_H) > d(\theta_H) > 0$.

It always profitable to supply the efficient quantity to the low cost type for the marginal cost of this provision is zero- there are no "upward" incentive constraints in the relaxed problem. Using the framework of equation (*), for the high cost type, the marginal cost of incentive provision (and hence dynamic distortion) is an additive sum of two economic forces:

$$r(h) = \underbrace{\text{backloading of incentives}(h)}_{r_1(h):\text{benchmark marginal cost}} + \underbrace{\text{financial constraints}(h)}_{r_2(h):\text{added marginal cost}}$$

where $h = \emptyset, \theta_L, \theta_H$. Recollect $d(\theta_H|h) \propto \frac{r(h)}{\mathbb{P}(h, \theta_H)}$. In the benchmark model, $r_2(h) = 0 \forall h$, and thus $r = r_1$. Since backloading of incentives is costless after a "good shock", $r_1(\theta_L) = 0$. With financial constraints, however, $r_2(\theta_L) > 0$ when C_L binds. Hence, $q^*(\theta_H|\theta_L) < q^e(\theta_H)$ and generalized no distortion at the top does not hold. Next, $r_2(\emptyset) < r_2(\theta_H)$, that is the impact of financial constraints strengthens for consecutive "bad shocks", culminating into $q^*(\theta_H|\theta_H) < q^*(\theta_H) < q^e(\theta_H)$. This overturns the vanishing distortions at the bottom result to increasing distortions at the bottom.¹⁸ Note that distortions along the "lowest" history increase irrespective of whether C_L binds.

When does C_L bind? Equation (6) clearly establishes that low values of $q(\theta_H)$ and $q(\theta_H|\theta_H)$ would violate C_L . To compensate, we must simultaneously distort $q(\theta_H|\theta_L)$ downwards, and $q(\theta_H)$ and $q(\theta_H|\theta_H)$ upwards, in proportion to the shadow price imposed by the constraint.¹⁹ Fig-

¹⁸We have $r_1(\emptyset) > r_2(\theta_H)$ and $r_2(\emptyset) < r_2(\theta_H)$. But, in aggregate $\frac{r(\emptyset)}{\mu_H} < \frac{r(\theta_H)}{\mu_H\alpha}$, which using equation (*) ensures that distortions increase at the "lowest" history.

¹⁹An equivalent way to think about the constraint optimization problem is as follows. The total rent paid to the low cost type appears in two constraints: $U(\theta_L) \geq \Delta\theta q(\theta_H) + \delta\alpha q(\theta_H|\theta_H)$, and $U(\theta_L) \geq \delta\alpha q(\theta_H|\theta_L)$. These are IC_L and C_L respectively. Employing a standard trick they can be written together as $U(\theta_L) = \gamma_1\Delta\theta [q(\theta_H) + \delta\alpha q(\theta_H|\theta_H)] + \gamma_2\Delta\theta\delta\alpha q(\theta_H|\theta_L)$ where $\gamma_1, \gamma_2 \geq 0$ and $\gamma_1 + \gamma_2 = 1$; and γ_i is the shadow price of the respective constraint. Thus the information rent (potentially) depends on all three quantities: $q(\theta_H)$, $q(\theta_H|\theta_H)$ and $q(\theta_H|\theta_L)$, and whenever $\gamma_2 > 0$, the last one is also distorted for the marginal cost of providing incentives at that node of the contract tree is positive. Introducing C_L as a separate constraint, as we did is equivalent.

Figure 3: $q^*(\theta_H|\theta_L)$ and $S^*(v_0)$ for the two period model

Figure 3a shows the parametric range for which C_L binds. In the $\mu_H \times \alpha$ rectangle, it plots $q^*(\theta_H|\theta_L)$ -shades represent numerical values as shown on the vertical key on the right. The lightest region is the efficient quantity, and darker the shade greater the optimal distortion. It is clear that low values of μ_H and high values of α correspond to the largest liquidity crunch.²⁰

What about the utility of the agent? It is obvious from equation (1) that the first period expected utility of the low cost type is higher than that of the high cost type. However, once a low or high cost is realized, we can say more about how the vector of utilities evolves.

Corollary 1. *In the optimal contract, $(u^*(\theta_L|\theta_L), u^*(\theta_H|\theta_L)) \geq (u^*(\theta_L|\theta_H), u^*(\theta_H|\theta_H))$.*

Evolution of the dynamic contract provides a clear ranking on the utility vector of the agent. For a "good shock" the next period's utility is larger for both types than that for a "bad shock". This observation furnishes the underpinning of the recursive approach we shall employ to solve the general model.

Finally, (PK) ensures that distortions are muted due to a greater share of the agent. A higher v_0 shifts bargaining power towards the possessor of private information, increasing the total size of the pie. This reduces optimal distortions and increases total surplus.

Formally, let $R^*(v_0)$ be principal's ex ante expected utility generated by the optimal allocation rule in Proposition 1. Total economic surplus is the sum of the utilities of the principal and agent: $S^*(v_0) = R^*(v_0) + \max\{\underline{v}, v_0\}$. The economic relationship between the principal and the agent creates a "meta firm". Each point $(R^*(v_0), \max\{\underline{v}, v_0\})$ on the Pareto frontier corresponds to a capital structure composed of their respective shares with total value $S^*(v_0)$. Figure 3b plots the value of the "meta firm" as a function of v_0 .²¹ As we increase the ex ante bargaining power of the agent, total economic surplus at first remains flat till $\underline{v}$, then rises to efficient value S^e at $\bar{v}$, and for $v_0 \geq \bar{v}$ it stays there. Hence the value of the "meta firm" is increasing in the ex ante share of the agent.

Corollary 2. *$S^*(v_0)$ is increasing in v_0 , and strictly so for $v_0 \in [\underline{v}, \bar{v}]$.*

²⁰Unless specified otherwise, throughout the paper: $V(q) = 10\sqrt{q}$, $\delta = 1$, $\theta_L = 3$, $\theta_H = 4$, $v_0 = 0$.

²¹The persistence parameter α is fixed at 0.75.

The two period model illustrates the key economic forces at play. It allows us to make educated guesses about what general results may look like. Are optimal distortions increasing in the number of consecutive "bad shocks"? Would these distortions be reduced by a "good shock"? Since a greater stake of the agent in total surplus is good for efficiency, and "good shocks" increase his utility, would a sequence of consecutive good shocks propel the contract into efficiency? If yes, what determines how many good shocks are needed? Is the relaxation of the cash-strapped constraint synonymous with efficiency? Would the principal provide credit to the agent in the long run? Is efficiency a certainty in the long run?

2.3 Sequential approach

A natural next step is to extend the two period analysis to the T period model, that is solve $(\mathcal{RP}^*)$ for a general time horizon. Using a subset of binding constraints, express $\mu_L U(\theta_L) + \mu_H U(\theta_H)$ and all remaining cash-strapped constraints in terms of quantities.

To this end, the first step is to inductively apply the binding constraints $IC_L(h^{t-1})$ and $C_L(h^{t-1})$ and generalize equation (4). Total expected utility for the agent is given by:

$$\mu_L U(\theta_L) + \mu_H U(\theta_H) = \Delta\theta \sum_{t=1}^T \delta^{t-1} \mathbb{P}(\theta_t = \theta_L) q(\theta_H | \theta_H^{t-1}) \quad (7)$$

Define the threshold generated by equation (7) for the efficient quantity to be $\bar{v}$:

$$\bar{v} = \Delta\theta \sum_{t=1}^T \delta^{t-1} \mathbb{P}(\theta_t = \theta_L) q^e(\theta_H) \quad (8)$$

and that generated by the optimal contract when we ignore (PK) to be $\underline{v}$. Moreover, for $T = \infty$, $C_L(h^{t-1})$ can be stated in terms of the quantity vector as:

$$C_L(h^{t-1}) : \sum_{s=0}^{\infty} \delta^s a_s q(\theta_H | h^{t-1}, \theta_H^s) \geq \sum_{s=1}^{\infty} \delta^s a_s q(\theta_H | h^{t-1}, \theta_L, \theta_H^{s-1})$$

where $a_s = \mathbb{P}(\theta_{t+s} = \theta_L | \theta_t = \theta_L) = \mathbb{P}(\theta_{s+1} = \theta_L | \theta_1 = \theta_L)$. Introducing Lagrange multipliers for these, we can pin down the optimal allocation rule.

Proposition 2. *Let $T = \infty$. The optimal contract (solution to $(\mathcal{RP}^*)$) is characterized by the following allocation rule:*

$$\begin{aligned} V'(q^*(\theta_L | h^{t-1})) &= \theta_L \quad \forall h^{t-1} \\ V'(q^*(\theta_H | \theta_H^{t-1})) &= \theta_H + \Delta\theta(1 - \lambda)Y_t - \Delta\theta \sum_{\tau=0}^{t-1} \eta(\theta_H^{t-\tau-1})X_{\tau} \quad \forall t \\ V'(q^*(\theta_H | h^{t-1}, \theta_L, \theta_H^{s-1})) &= \theta_H + \Delta\theta\eta(h^{t-1})X_s - \Delta\theta \sum_{\tau=0}^{s-1} \eta(h^{t-1}, \theta_L, \theta_H^{s-\tau-1})X_{\tau} \quad \forall s \end{aligned}$$

where $Y_s = \frac{\mathbb{P}(\theta_s = \theta_L)}{\mathbb{P}(\theta_H^s)}$, $X_s = \frac{a_s}{\mathbb{P}(\theta_H^s)}$, and λ and $\delta^{t-1}\eta(h^{t-1})\mathbb{P}(h^{t-1}, \theta_H)$ are respectively the Lagrange

Figure 4: Failure of the naive relaxed problem to satisfy cash-strapped constraint

*multipliers on (PK) and $C_L(h^{t-1})$, jointly determined through complementary slackness.*²²

One can immediately note that for positive values of η , distortions are pervasive. A binding $C_L(h^{t-1})$ leaves a legacy of distortions on all high cost quantities that follow- $q^*(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$. It is also important to note that distortions are a function of shadow prices as measured from the last time a low cost type was realized. Beyond these, it hard to drive home general arguments about the nature of dynamic distortions because η s are endogenous and jointly determined at the optimum.

To appreciate the role of financial constraints, it useful to consider a relaxed problem where we ignore $C_L(h^{t-1})$ for all histories, call it the naive relaxed problem ($n\mathcal{RP}$). The optimal allocation rule therein will obviously be the same as in Proposition 2, except that $\eta = 0$. Therefore, the solution will satisfy generalized no distortion at the top and for $v_0 < \bar{v}$, increasing distortions at the bottom.²³ Let $\mathbf{q}^n$ be the solution to ($n\mathcal{RP}$). We have the following result.

Corollary 3. *Let $T = \infty$. The solution to ($n\mathcal{RP}$) is not optimal: $\mathbf{q}^n \neq \mathbf{q}^*$.*

Dynamic distortions grow large enough that the principal wants even the low cost type to forgo payments in order to relax dynamic incentives- technically speaking $C_L(h^{t-1})$ is violated by $\mathbf{q}^n$. An intuitive explanation is provided in Figure 4. Distortions along the high cost history strictly increase over time, without bound. At some point, expected utility which is defined through binding incentives along the high cost history, will become very small. However, expected utility must also satisfy identities along the efficient history (shown in dashed red), which through incentive constraints ensure a reasonably large information rent. The two competing forces lead to a contradiction.²⁴

We turn to the recursive approach to reduce the curse of dimensionality and provide a precise characterization of the optimal contract.

²²At the cost of some extra notation $C_L(h^{t-1})$ and the proposition can be easily restated for a general $T \leq \infty$.

²³Distortion at the "lowest" history is increasing because $\frac{\mathbb{P}(\theta_{t+1}=\theta_H)}{\mathbb{P}(\theta_L^{t+1})} = \frac{\alpha\mathbb{P}(\theta_t=\theta_H)+(1-\alpha)\mathbb{P}(\theta_t=\theta_L)}{\alpha\mathbb{P}(\theta_L^t)} = \frac{\mathbb{P}(\theta_t=\theta_H)}{\mathbb{P}(\theta_L^t)} + \frac{1-\alpha}{\alpha} \frac{\mathbb{P}(\theta_t=\theta_L)}{\mathbb{P}(\theta_L^t)} > \frac{\mathbb{P}(\theta_t=\theta_H)}{\mathbb{P}(\theta_L^t)}$

²⁴Fixing the other parameters, this result will also hold for a finite but large enough T . As we already saw, even in the two period model, $C_L(\theta_L)$ can bind at the optimum.

2.4 Recursive formulation: a full characterization

In this section, we change the technical course of the economic problem by converting it to its recursive avatar. The recursive approach to dynamic contracting, understood at least since Green [1987], and Spear and Srivastava [1987], provides an able tool to characterize the optimal contract as a stationary solution of an optimal control problem with requisite state variables. As noted by Fernandes and Phelan [2000], with *Markovian* agency frictions, the recursive domain of "promised utility" is not one-dimensional. This *self generating set* depends on the cardinality of the type space.

Assume throughout that $T = \infty$. Let $S(h^{t-1})$ be the expected total surplus generated by the sequential contract from period t onwards:

$$S(h^{t-1}) = \sum_{s=0}^{\infty} \delta^s \mathbb{E} \left[s(\tilde{\theta}_{t+s}, q(\tilde{\theta}_{t+s} | \tilde{h}^{t+s-1})) \mid \tilde{h}^{t+s} \in H^{t+s} |_{h^{t-1}} \right]$$

where $H^t |_{h^\tau}$ for $\tau \leq t$ is set of all histories of length t whose first τ elements are h^τ .

Suppose that the agent reported (h^{t-1}, θ_j) truthfully. The principal is committed to deliver exactly w_i to the agent of type θ_i at this date. Then, for $w = (w_L, w_H)$, define $Q_j^*(w)$ for $j = L, H$ to be

$$\boxed{\star} \quad Q_j^*(w) = \max_{\langle U, q \rangle} S(h^{t-1}, \theta_j)$$

subject to $w_i = U(\theta_i | h^{t-1}, \theta_j)$ for $i = L, H$, and

$$IC_L(h^{t+s}), C_L(h^{t+s}), C_H(h^{t+s}) \forall h^{t+s} \in H^{t+s} |_{(h^{t-1}, \theta_j)}$$

Here $Q_j^*(w)$ is the maximal surplus (and hence maximal expected profit for the principal) generated by the optimal contract given that the previous period type was θ_j , and the agent has to be provided an expected utility vector exactly equal to w .

It is standard practice to show that conditional on w , the optimal value is independent of h^{t-1} , thus the sparse expression $Q_j^*(w)$. That is, all the history dependence is encoded in the two-dimensional expected utility and j . Let W be the largest set of w such that the constraints set above is non-empty. Again, this set does not depend on h^{t-1} .²⁵

The problem from $t = 2$ is *recursive* and it reads as follows:

$$(\mathcal{RF}) \quad Q_j^*(w) = \max_{\langle z_L, z_H, q \rangle} \alpha_j [s(\theta_L, q_L) + \delta Q_L^*(z_L)] + (1 - \alpha_j) [s(\theta_H, q_H) + \delta Q_H^*(z_H)]$$

subject to $\langle z_L, z_H, q \rangle \in W^2 \times \mathbb{R}_+^2$, and

$$w_L - w_H \geq \Delta \theta q_H + \delta(2\alpha - 1)(z_{HL} - z_{HH})$$

$$w_L \geq \delta[\alpha z_{LL} + (1 - \alpha)z_{LH}]$$

$$w_H \geq \delta[(1 - \alpha)z_{HL} + \alpha z_{HH}]$$

²⁵Given the time structure of the problem, and its recursive formulation it can also be shown that W is independent of j .

where $\alpha_L = 1 - \alpha_H = \alpha$, and by $(\mathcal{RF})$ we mean recursive formulation. Note that $q = (q_L, q_H)$ is the optimal allocation rule, $z_L = (z_{LL}, z_{LH})$ is the expected utility vector of the agent for the next period if his type today is θ_L , and $z_H = (z_{HL}, z_{HH})$ is the expected utility vector of the agent for the next period if his type today is θ_H . Given these choice variables, the first constraint is the incentive constraint for the low cost type, and next two are cash-strapped constraints for the low and high cost types, respectively.

At date $t = 1$ the problem is different for two reasons. First, the belief is equal to the prior, and second the contract has not been yet initialized.

$$(\mathcal{RF}_0) \quad R^*(v_0) = \max_{\langle \mathbf{w}, \mathbf{z}_L, \mathbf{z}_H, \mathbf{q} \rangle} \mu_L[s(\theta_L, q_L) + \delta Q_L^*(z_L)] + \mu_H[s(\theta_H, q_H) + \delta Q_H^*(z_H)] - \bar{U}$$

subject to $\langle \mathbf{w}, \mathbf{z}_L, \mathbf{z}_H, \mathbf{q} \rangle \in W^3 \times \mathbb{R}_+^2$, and

$$\begin{aligned} \bar{U} &= \mu_L w_L + \mu_H w_H \geq v_0 \\ w_L - w_H &\geq \Delta \theta q_H + \delta(2\alpha - 1)(z_{HL} - z_{HH}) \\ w_L &\geq \delta[\alpha z_{LL} + (1 - \alpha)z_{LH}] \\ w_H &\geq \delta[(1 - \alpha)z_{HL} + \alpha z_{HH}] \end{aligned}$$

where Q_L^* and Q_H^* are calculated in $(\mathcal{RF})$.

In what follows we present intuitive arguments that characterize the optimal contract; the full and mathematically precise details are provided in the appendix. Some of these details are worth mentioning in the main text and we do so concretely as we go along.

The recursive domain W , that is the set of all possible expected utilities that generate themselves in an incentive compatible and feasible manner, is precisely the positive orthant above the 45 degree line.²⁶ The existence and shape of the recursive domain can be established through iterative approximations. Figure 5 plots the recursive domain.

Next, let E be the largest subset of W such that the constraints set in $(\mathcal{RF})$ and $(\mathcal{RF}_0)$ is non-empty when $q_i = q^e(\theta_i)$ for $i = L, H$. We term this the efficient set. E too is self-generating, hence an absorbing subset. Once the optimal contract becomes efficient, it stays efficient. Figure 5 plots the efficient set E - it is characterized by its lowest point $\underline{w}^e$ and two rays.²⁷ Intuitively, one can note that $\bar{v} = \mu_L \underline{w}_L^e + \mu_H \underline{w}_H^e$.

With some work, we can also show that the Bellman operator has a unique, continuous bounded fixed point Q^* which is concave, supermodular and continuously differentiable. Importantly the value functions in the sequential and recursive problems coincide.

Lemma 1. *Value functions in $\boxed{\star}$ and $(\mathcal{RF})$, and in $(\mathcal{RP}^*)$ and $(\mathcal{RF}_0)$ coincide.*

²⁶Self generation refers to the idea of identifying the largest possible set such that given an expected utility vector w in that set, there exists some feasible policy choice q, z_L, z_H such that z_L, z_H also lie in the set.

²⁷Just like W , E too is independent of j . This is in contrast to the cash flow diversion models, and leads to greater tractability of our framework.

Figure 5: Recursive domain W and the efficient set E .

Shape of the optimal contract

Let $(1 - \alpha_j)\beta$, $\alpha_j\rho_L$, and $(1 - \alpha_j)\rho_H$ be the Lagrange multipliers for the constraints in $(\mathcal{RF})$ in the order they appear. The optimal contract is characterized by the first-order and envelope conditions. These are provided in the appendix. Here we geometrically explain the structure of expected utilities that arise as part of the optimal contract.

First, there exists a threshold w_L^{liq} on the expected utility of the low cost type, above which the contract becomes liquid: that is $\rho_L = 0$ for $w_L \geq w_L^{liq}$. Critically, this threshold lies below the efficient level: $w_L^{liq} < \underline{w}_L^e$, see Figure 6a. Also, we show that for any w such that $w_L \geq w_L^{liq}$, $z_L \in E$ —once in the liquid region, an additional realization of the low cost type necessarily pushes the optimal contract into the efficient region. Therefore, $\rho_L = 0$ forms a penultimate zone towards efficiency.

Next, we draw the two level curves that enclose the optimal contract in the inefficient region. To understand their geometry, think of simple price theory. Cull the following sub-problem from $(\mathcal{RF})$:

$$\max_{z_L} Q_L^*(z_L) \quad s.t. \quad w_L = \delta [\alpha z_{LL} + (1 - \alpha)z_{LH}]$$

η_L is locus of points where "marginal rate of substitution = relative prices" for this optimization problem. That is,

$$\eta_L(w) = 0 \Leftrightarrow \frac{D_L Q_L^*(w)}{\alpha} = \frac{D_H Q_L^*(w)}{1 - \alpha}$$

where D_i is the directional derivative of Q_j^* for $i = L, H$. Figure 6a plots $\eta_L = 0$. We show that it is an increasing curve that joins the origin and $\underline{w}^e$. Similarly, we cull the following problem from $(\mathcal{RF})$:

$$\max_{z_H} Q_H^*(z_H) \quad s.t. \quad w_H = \delta [(1 - \alpha)z_{HL} + \alpha z_{HH}]$$

Figure 6: Constructing the space wherein the optimal contract lies

to generate a locus η_H that satisfies:

$$\eta_H(w) = 0 \Leftrightarrow \frac{D_H Q_H^*(w)}{1 - \alpha} = \frac{D_L Q_H^*(w)}{\alpha}$$

Figure 6a plots $\eta_H = 0$. We show that it is an increasing curve that joins the origin and w^e . Moreover, it lies above $\eta_L = 0$. The interior of the two curves, which we call the shell (B), delineates the region where the incentive constraint binds, $\beta > 0$, and the optimal constrained contract lies on or within it.

Figure 6b gives an example of the evolution of expected utility. Starting at w , it moves to z_L on the realization of a low cost type, and to z_H on high cost type. In fact the example is chosen so that on the realization of two consecutive low cost types, the contract becomes liquid (at z'_L), and therefore a third "good shock" carries it over to the efficient region, E . It is entirely possible that standing at z'_L , a "bad shock" would have carried the contract back into illiquidity.

Finally, as Figure 6b depicts, the realization of a low cost realizations always choose the expected utility vector in the northeast direction *on* the locus $\eta_L = 0$, that is z_L and z'_L lie *on* the curve. Whereas, realization of a high cost type chooses a point in the southwest direction in the *interior* of the shell.

2.5 The main result

Collecting all the key insights from the sequential and the recursive approach, we enlist a complete characterization of the optimal contract. For consistency, we state all the results in the lexicon of the sequential model. For any history h^{t-1} , we say that allocation is efficient if $q(\theta_i|h^{t-1}) = q^e(\theta_i)$, and the contract is liquid if $C_L(h^{t-1})$ does not bind. The lowest point of efficient set of expected utilities, E , is w^e . "Good" and "bad" shock refer to low and high cost realizations, respectively.

We assume that the prior is first-order stochastically ranked around its Markov evolution: $1 - \alpha \leq \mu_L \leq \alpha$. This is strictly more general than assuming a "seed type", that is $\mu_L = \alpha$ or $\mu_L = 1 - \alpha$ which is the standard in the recursive contracting literature.

Theorem 1. *Let $T = \infty$. The optimal contract, $\langle \mathbf{U}^*, \mathbf{q}^* \rangle$, (solution to $(\mathcal{RP}^*)$) is characterized by the following properties.*

A Optimal distortions:

1. *Optimal contract is downward distorted: $\forall h^{t-1} \in H^{t-1}, \forall t, q^*(\theta_L|h^{t-1}) = q^e(\theta_L)$, and $q^*(\theta_H|h^{t-1}) = q^e(\theta_H) - d(\theta_H|h^{t-1})$, where $d(\theta_H|h^{t-1}) \geq 0$.*
2. *Distortions are strictly increasing: for $q^*(\theta_H|h^{t-1}) < q^e(\theta_H)$, $d(\theta_H|h^{t-1}, \theta_H^s)$ is strictly increasing in s .*
3. *Distortions are muted after a "good shock": for $q^*(\theta_H|h^{t-1}) < q^e(\theta_H)$, $d(\theta_H|h^{t-1}, \theta_L) < d(\theta_H|h^{t-1})$.*

B Expected utility:

4. *Expected utility strictly increases (decreases) with a low (high) cost type: for $q^*(\theta_H|h^{t-1}) < q^e(\theta_H)$, we have $(U^*(\theta_L|h^{t-1}, \theta_H), U^*(\theta_H|h^{t-1}, \theta_H)) \ll (U^*(\theta_L|h^{t-1}), U^*(\theta_H|h^{t-1})) \ll (U^*(\theta_L|h^{t-1}, \theta_L), U^*(\theta_H|h^{t-1}, \theta_L))$.*

C Liquidity:

5. *The contract becomes liquid above a fixed threshold which is below the efficient level: $\exists \underline{w}_L^{liq} < \underline{w}_L^e$ such that for $U^*(\theta_L|h^{t-1}) \geq \underline{w}_L^{liq}$, $C_L(h^{t-1})$ is slack.*

D Efficiency:

6. *Efficiency is an absorbing state: $d(\theta_H|h^{t-1}) = 0 \Rightarrow d(\theta_H|h^{t+s-1}) = 0$, and $(U^*(\theta_L|h^{t-1}), U^*(\theta_H|h^{t-1})) \in E \Rightarrow (U^*(\theta_L|h^{t+s-1}), U^*(\theta_H|h^{t+s-1})) \in E \forall h^{t+s-1} \in H^{t+s-1}|_{h^{t-1}}$.*
7. *An endogenous and monotonic number of "good shocks" are required for efficiency: $\forall h^{t-1} \in H^{t-1}, \forall t, \exists n^*(h^{t-1}) \in \mathcal{N}$ such that $d(\theta_H|h^{t-1}, \theta_L^{n^*}) = 0$, and $n^*(h^{t-1}, \theta_H) \geq n^*(h^{t-1})$.*
8. *Efficiency is achieved through a penultimate slacking of the cash-strapped constraint: $C_L(h^{t-1})$ binds $\Leftrightarrow n^*(h^{t-1}) \geq 2$, and $C_L(h^{t-1})$ is slack $\Leftrightarrow d(\theta_H|h^{t-1}, \theta_L) = 0$.*

E Long run:

9. *Efficiency is a certainty: $\forall h^{t-1} \in H^{t-1}, d(\theta_H|h^{t+s-1}) \rightarrow_s 0$, and $(U^*(\theta_L|h^{t+s-1}), U^*(\theta_H|h^{t+s-1})) \rightarrow_s w \in E$ almost surely.*

Part A of the theorem characterizes the optimal allocation rule through dynamic distortions produced by the periodic interaction between incentives and feasibility. The low cost type supplies the efficient quantity, and the high cost is distorted downwards. As in the two period model, using

the framework of equation (*), the evolution of $r(h^{t-1})$ determines the optimal distortions d , which drives the main theorem.

After any sequence of types, if the contract is inefficient then every further high cost realization strictly increases the dynamic distortions, thereby strictly decreasing the optimal quantity. In Figure 1a, quantity along the history (h^{t-1}, θ_H^2) is less than quantity along (h^{t-1}, θ_H) . This is in stark contrast to the standard results in dynamic mechanism design (without financial constraints) that emphasize decreasing distortions over time. Moreover, a "good shock" reduces the optimal distortions- quantity along $(h^{t-1}, \theta_L, \theta_H)$ is greater than that along (h^{t-1}, θ_H) . These rankings of optimal distortions form the bedrock of our analysis.

Part B tracks the optimal path of expected utility. For any history of types, for the inefficient contract, expected utility strictly increases along both dimensions after a "good shock" and reduces after a "bad shock". In terms of Figure 1a, vector of squares is larger than circles which is larger diamonds. And, in the two dimensional domain of expected utility a "good shock" takes it in the northeast direction and "bad shock" in the southwest; always within the shell we constructed in Section 2.4.

Part C characterizes liquidity. The interval $[w_L^{liq}, \underline{w}_L^e)$ for $U^*(\theta_L|h^{t-1})$ witnesses slacking of $C_L(h^{t-1})$. The region can only be attained through a "good shock" (given the contract does not start in this region). It is important to note that liquidity is not an absorbing state, and it is not synonymous with efficiency. Even in the liquid region, a "bad shock" can revert the contract back into the illiquid region, $[0, w_L^{liq})$.

The path to efficiency is completed in part D. First, efficiency is an absorbing state (point 6). Once $q(\theta_H|h^{t-1}) = q^e(\theta_H)$, the contract is efficient thereafter: $q(\theta_H|h^{t-1}, h^s) = q^e(\theta_H)$. That is, once the distortion for the high cost type reduces to zero, it stays zero. This result has remnants of most dynamic contracting models with moral hazard and cash flow diversion (see for example Biais et al. [2013]). To the best of our knowledge, this is its first appearance in a dynamic screening or adverse selection framework with strong feasibility constraints.

Next the efficient region, E , can be attained only through an endogenous number of "good shocks". This number depends on the path of reported types. In particular, on a high cost realization, the number of types required to reach the efficient region increases (point 7). This is in sharp contrast to the benchmark model where the contract becomes efficient on the realization of one low cost type. Figure 7 plots $n^*(w_L)$ - the number of consecutive low cost realizations required to reach the efficient region as a function of w_L which encodes all the history dependence required for n^* . As w_L decreases, n^* can become quite large. Monotonicity in the endogenous number of shocks is an intuitive but formally novel addition to dynamic contracting.

In addition, for most parametric settings (where expected utility starts in the region $[0, w_L^{liq})$), the efficient region is achieved through a penultimate liquid region- $[w_L^{liq}, \underline{w}_L^e)$ (point 8). Once the cash-strapped constraint is slack, efficiency is attained through one more low cost type. Complementarily, Fu and Krishna [2016] show that this result also holds in the Markov extension of the cash flow diversion model of Clementi and Hopenhayn [2006].

Finally in part E, we close the theorem with a fairly intuitive result on convergence of the

Figure 7: Number of consecutive low cost types required to reach E

optimal contract to the efficient region. At any point on the contract tree (and hence any level of expected utility), the contract will converge to the efficient region almost surely. It employs the martingale converges theorem. It is reminiscent of the immiseration result in a model of insurance from stochastic income in Thomas and Worrall [1990], long-term efficiency result in investment with agency frictions in Clementi and Hopenhayn [2006], and retirement of the agent in a compensation model of Sannikov [2008].

Summarizing the key point, $r(h^t)$ in equation (*) positive if and only if (i) $h^t = \theta_H^t$, or (ii) $h^t = (h^{t-1}, \theta_L, \theta_H^{t-s})$ for $1 \leq s \leq t-1$ and $C_L(h^{s-1})$ binds, or (iii) $\theta_t = \theta_L$, $r(h^{t-1}) > 0$ and $C_L(h^{t-1})$ is slack. Moreover, $\frac{r(h^{t-1}, \theta_H^s)}{\mathbb{P}(h^{t-1}, \theta_H^s)}$ is increasing in s , that is distortions are increasing in the number of consecutive "bad shocks", and $\frac{r(h^{t-1}, \theta_H)}{\mathbb{P}(h^{t-1}, \theta_H)} > \frac{r(h^{t-1}, \theta_L, \theta_H)}{\mathbb{P}(h^{t-1}, \theta_L, \theta_H)}$, that is distortions reduce after a "good shock". Once distortions reach zero, they stay zero.

Note that the assumption $1 - \alpha \leq \mu_L \leq \alpha$ is made to ensure that the optimal contract starts in shell defined in Section 2.4. If these conditions are not satisfied then the optimal contract enters the shell the moment it gets a low cost realization. All our points still hold expect for the "lowest history" of continued high cost realizations.²⁸

2.6 Sufficiency conditions and ensuring global optimality

Recollect that $\Gamma = \{V(\cdot), \Theta, \mu, \alpha, \delta, v_0\}$ is the entire set of parameters. We say that the first-order approach is valid if the solution to $(\mathcal{RP}^*)$ defined in section 2.3 is incentive compatible, that is the high cost type or "upward" incentive constraints do not bind at the optimum.

When is the first-order approach valid? In the two period model discussed in section 2.2 the "upward" incentive constraint, IC_H , never binds. It is possible, however as we argue largely implausible, that for a long enough time horizon and large enough discount factor the "upward" incentive constraint may bind. In a nutshell, the measure of parameters for which we need to add the "upward" incentive constraint to the relaxed problem after some history is very small and therefore the economic message delivered by our solution worth consideration.²⁹

²⁸Even if this assumption was violated, our results will continue to hold almost surely.

²⁹Note that so far in the literature, papers have altogether avoided this discussion by simply assuming the "upward"

Figure 8: Numerical examples depicting the shell and the region where $\mathcal{RP}^*$ is valid

First, we document that irrespective of the other parameters of the model, as $\alpha \rightarrow 1/2$ and as $\alpha \rightarrow 1$, the first-order optimal contract is indeed optimal. Thus, in the neighborhood of both iid types and perfect persistence, "upward" incentive constraints can be safely ignored.

Proposition 3. *For any $\Gamma \setminus \{\alpha\}$, the first-order approach is valid as $\alpha \rightarrow 1/2$ and $\alpha \rightarrow 1$.*

Second, we enlist sufficiency conditions that ensure that the first-order optimal contract is globally optimal. These are presented in the appendix (Section 8.6). The primary motivation behind them is the following. After any history h^{t-1} , using the set of binding constraints in $(\mathcal{RP}^*)$, the "upward" incentive constraint and the cash-strapped constraint can respectively be expressed as:

$$IC_H(h^{t-1}) : \quad q^e(\theta_L) + \sum_{s=1}^{\infty} \delta^s (2\alpha - 1)^s q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1}) \geq \sum_{s=0}^{\infty} \delta^s (2\alpha - 1)^s q(\theta_H|h^{t-1}, \theta_H^s)$$

$$C_L(h^{t-1}) : \quad \sum_{s=0}^{\infty} \delta^s a_s q(\theta_H|h^{t-1}, \theta_H^s) \geq \sum_{s=1}^{\infty} \delta^s a_s q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$$

When $C_L(h^{t-1})$ is slack, $q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$ are efficient for all $s \geq 1$, therefore, $IC_H(h^{t-1})$ necessarily holds. When does $C_L(h^{t-1})$ bind? It binds when quantities on the left hand side of $C_L(h^{t-1})$, that is $q(\theta_H|h^{t-1}, \theta_H^s)$ for $s \geq 0$, are highly distorted owing to the interaction of binding incentive and cash-strapped constraints in previous periods. But, it is precisely when these quantities are highly distorted that it is easy for $IC_H(h^{t-1})$ to be satisfied for they appear on the right hand side of the constraint. Combining the efficient and inefficient regions, the measure of parameters for which the "upward" incentive constraint may bind after some history is quite small.

Third, we have numerically calculated the optimal contract for a large range of parameters to show that the first-order approach is indeed valid. The code for these numerical simulations has been made available online to test any combination of parameter values.³⁰ Two such examples are

incentive constraint does not bind, or ruling it out as a technological assumption. See for example, Clementi and Hopenhayn [2006] and Fu and Krishna [2016].

³⁰The code can be found at www.rohitlamba.com/research. We have used the parametric setting: $V(q) = 10\sqrt{q}$,

presented in Figure 8. The shaded region is the recursive domain for the inefficient contract (easy to see that the efficient contract is first-order optimal). The darkly shaded region is the set of expected utility vectors for which the "upward" constraint is slack at the optimum. The shell, wherein the optimal contract resides, lies within the darker shaded area. Hence the first-order approach is valid.

3 Dynamics of payments

Now that the optimal allocation rule for the Pareto problem is completely characterized, a next obvious question to ask is- how do we implement it? What are the set of prices and expected utility vectors that are incentive compatible with the optimal allocation?

There are three salient features of an incentive compatible payment schedule that implements the optimal allocation. First, as long as the optimal contract is liquid, delayed payments are optimal. Second, at any given history, promised utility is marked up after a "good shock" and marked down after a "bad shock", both in proportion to the history dependent information rent. And, third, the (history dependent) total economic surplus is increasing in both the share of the agent and his utility spread.

Define the promised utility of the agent to be:

$$v^*(\theta_j|h^{t-1}) = \frac{1}{\delta} [U^*(\theta_j|h^{t-1}) - u^*(\theta_j|h^{t-1})]$$

That is, given $U^*(\theta_j|h^{t-1})$, $u^*(\theta_j|h^{t-1})$ and $v^*(\theta_j|h^{t-1})$ represent how expected utility is split between current utility and promised utility.

Next, define the dynamic information operator as follows:

$$I(h^{t-1}) = \Delta\theta \sum_{s=1}^{\infty} \delta^{s-1} (2\alpha - 1)^{s-1} q^*(\theta_H|h^{t-1}, \theta_H^{s-1})$$

starting initially at $I = \Delta\theta \sum_{t=1}^{\infty} \delta^{t-1} (2\alpha - 1)^{t-1} q^*(\theta_H|\theta_H^{t-1})$. The dynamic information operator is essentially the utility spread offered to the agent- under binding $IC_L(h^{t-1})$ it measures the difference in expected utility between the low cost and high cost types.

The dynamics of payments are as follows. Fix the optimal allocation rule and initial promised utility v_0 .³¹ Solving the promised utility identity and the "envelope formula" together:

$$\mu_L U^*(\theta_L) + \mu_H U^*(\theta_H) = v_0$$

$$U^*(\theta_L) = U^*(\theta_H) + I$$

gives

$$U^*(\theta_L) = v_0 + \mu_H I \quad \text{and} \quad U^*(\theta_H) = v_0 - \mu_L I \tag{9}$$

$\delta = 0.8, \theta_L = 3, \theta_H = 4, v_0 = 0$.

³¹In case (PK) is not binding, replace v_0 with $\underline{v}$ in equation (9).

A realization of low cost type marks up the agent's expected utility, and a realization of a high type marks it down both in proportion to the dynamic information rent operator I .

Now, $U^*(\theta_i) = u^*(\theta_i) + \delta v^*(\theta_i)$. Choosing $u(\theta_i)$ automatically determines $v(\theta_i)$. Proceeding inductively, we have:

$$\alpha_j U^*(\theta_L|h^{t-1}, \theta_j) + (1 - \alpha_j) U^*(\theta_H|h^{t-1}, \theta_j) = v^*(\theta_j|h^{t-1})$$

$$U^*(\theta_L|h^{t-1}, \theta_j) = U^*(\theta_H|h^{t-1}, \theta_j) + I(h^{t-1}, \theta_j)$$

Solving the two equation gives us

$$\begin{aligned} U^*(\theta_L|h^{t-1}, \theta_j) &= v^*(\theta_j|h^{t-1}) + (1 - \alpha_j)I(h^{t-1}, \theta_j) \\ U^*(\theta_H|h^{t-1}, \theta_j) &= v^*(\theta_j|h^{t-1}) - \alpha_j I(h^{t-1}, \theta_j) \end{aligned} \tag{10}$$

Starting from promised utility v_0 and choosing per period transfers optimally, equations (9) and (10) inductively define future expected and promised utilities.

We choose a simple implementation rule- always give the agent zero stage utility till it enters the liquid region. And, in the liquid (and efficient) region provide the low cost type with positive utility according to inductively binding incentive compatibility constraints. We define the mechanism formally in the next proposition. For $j = L, H$, let $v_j^e = \alpha_j \underline{w}_L^e + (1 - \alpha_j) \underline{w}_H^e$ be the efficient promised utility offered to the agent at the lowest point of the efficiency set.

Proposition 4. *Suppose $v_0 \leq \bar{v}$. Given optimal allocation rule $\mathbf{q}^*$, expected utility vectors $\mathbf{U}^*$ defined by equations (9) and (10) implement it, where*

$$\delta v^*(\theta_H|h^{t-1}) = U^*(\theta_H|h^{t-1}) \quad \text{and} \quad \delta v^*(\theta_L|h^{t-1}) = \min \{U^*(\theta_L|h^{t-1}), v_L^e\}$$

Suppose $v_0 > \bar{v}$. Then $\mathbf{q}^ = \mathbf{q}^e$. A modified rule where the principal makes an initial transfer of $\eta = v_0 - \bar{v}$ to the agent, and then follows equations (9) and (10) as described above, but starting at $\bar{v} = v_0 - \eta$ (instead of v_0), implements $\mathbf{q}^*$.*

Proposition 4 precisely characterizes the optimal mechanism $\langle \mathbf{U}^*, \mathbf{q}^* \rangle$. But, what about the equivalent representation $\langle \mathbf{u}^*, \mathbf{q}^* \rangle$, that is, what about per-period transfers? Given $\mathbf{U}^*$, we can define the following:

$$u^*(\theta_H|h^{t-1}) = 0 \quad \forall h^{t-1} \quad \forall t \tag{11}$$

$$u^*(\theta_L|h^{t-1}) = \max \{U^*(\theta_L|h^{t-1}) - \delta v_L^e, 0\} \quad \forall h^{t-1} \quad \forall t \tag{12}$$

Equations (11) and (12) pin down one possible way of implementing the optimal allocation. An intuitive way to think about our mechanism is the following. In the illiquid region, the principal only compensates the agent with working capital. Note that the optimal mechanism is unique in the illiquid region. In liquid region the principal loosens his purse for the first time, with a push towards efficiency in the event of another "good shock". Once the contract becomes efficient, and hence the information rent of the agent is maximal and stationary, the big firm (principal) can

simply the take-over the small firm (agent) by allowing it to operate "in-house". The price of the take-over is precisely the promised utility of the agent at the time it becomes efficient, viz. v_L^e . After the take-over the big firm simply forwards the working capital (sans the information rent) every period. When $v_0 > \bar{v}$, the contract is efficient: $q^* = q^e$. Therefore, the take-over can happen at the inception.

Finally, we want to document correlation between the value of the "meta firm", and (i) promised utility of the agent and (ii) utility spread. The utility spread, that is the difference in expected utility for the low and high cost types, measures the information rent that the principal must pay to the agent for truthful reporting. Let $S^*(h^{t-1})$ the optimal value of the economic surplus defined in Section 2.4. In the next result we establish this economic surplus (which is the value of the "meta firm") is positively correlated with promised utility and utility spread.

Proposition 5. *For the optimal contract, total economic surplus is positively correlated with promised utility and utility spread:*

$$S^*(h^{t-1}) \propto v^*(h^{t-1}) \quad \text{and} \quad S^*(h^{t-1}) \propto I(h^{t-1})$$

In the appendix we show that the sequential approach can be made recursive with two new state variables: promised utility and utility spread. And, at the optimum, the derivative of the total surplus with respect to each is increasing, proving Proposition 5.

An economically sensible way of interpreting the result is that the value of the "meta firm" is increasing in the share of the agent and his information rent. The former is simply statement of aligning incentives with bargaining power. As the agent assumes a greater stake of the total surplus the effect of agency frictions is reduced which increases the size of pie towards its efficient value. The latter though may not be immediately obvious.

An increase in the utility spread is synonymous with greater information rents. Why should it increase the value of the economic surplus?³² The optimal utility spread increases on the realization of a "good shock". This decreases the optimal distortion which in turn increases the surplus. This process continues till the contract becomes efficient. Upon reaching efficiency, the both the utility spread and value of the "meta firm" become constant for each type. It should be noted that while the monotonic relationship of economic surplus with promised utility is a global one, that with utility spread is only valid along the optimum.

4 Discussion

4.1 Comparative statics

What is the role of persistence in the optimal contract? Without financial constraints, we know that persistence of private information is broadly bad for ex post efficiency- higher the persistence, greater the level of asymmetric information, and hence higher the dynamic distortions. Its role in

³²Note that from the second period onwards, since the principal has to deliver a specific utility to the agent, maximizing surplus is synonymous with maximizing profit.

Figure 9: Optimal quantity and surplus as a function of persistence for the two-period model

incentive provision with financial constraints is less well understood. As stated before, dynamic distortions are a sum of the standard incentive distortions due to maximal backloading, and an added component due to financial constraints. While the former continues to be increasing in persistence, the latter may not. Perhaps surprisingly, we show that the interaction of private information with the cash-strapped constraint every period produces distortions that can decrease with persistence. Figure 9a plots $q(\theta_H|\theta_H)$ for the two-period model.

What about the ex ante value of total surplus? Predictably, it depends on the prior. The optimal value of ex ante economic surplus is increasing in persistence for high values of μ_L , and it is decreasing in persistence for low values of μ_L . Figure 9b plots this relationship for our running example.

In Figure 7, we plotted the number of "good shocks" required to reach the efficient region of the optimal contract as a function of expected utility. Plotting the graph for different values of α shows that n^* is increasing in α —for a fixed value of w_L , number of consecutive "good shocks" required to attain efficiency increases in persistence.

Finally, it is also informative to note that as the model converges to iid, $\alpha \rightarrow 0.5$, the shell constructed in Section 2.4 as a subset of the recursive domain converges to a single line. This can already be noted in Figure 8b where $\alpha = 0.6$. On the other hand as $\alpha \rightarrow 1$, $w_H^e \rightarrow 0$, that is, the high cost component of the recursive domain shrinks to zero. It is immediately observable that a larger set of expected utilities can be rationalized by the flexibility provided by persistence, which translates into a richer composition of the capital structure of the "meta firm".

4.2 Limited liability versus strapped for cash

Should the positivity of stage utility be seen as a limited liability or cash-strapped constraint? The answer can be found in a comparison to the benchmark model. It is clear that both the profit of the principal and the total value of economic surplus is higher in the benchmark model than in the our model, cash-strapped is a stronger restriction than individual rationality. The ambiguity lies in the ex ante value of the agent's information rent or utility.

We argue that in the presence of binding (PK) constraint, the question should be settled in

Figure 10: $\mathbb{1}_{\{\underline{v}^* \geq \underline{v}^b\}}$ as a function of μ_H and α , shaded area is zero.

favor of the cash-strapped constraint. If the ex ante utility of the agent with individual rationality and positive stage utility is fixed at the same value v_0 , it moot to argue that in the latter he is "protected by limited liability." The right economic interpretation for us then is one of a credit or cash constraint- the agent cannot easily borrow money, which if possible would increase the total surplus and the principal's profit.

Suppose (PK) does not exist (or does not bind). Then we ask- in the principal profit maximizing contract, when is the agent (ex ante) better off? The ex ante expected utility of the agent in the two models is given by

$$\underline{v}^b = \mu_L U^b(\theta_L) + \mu_H \cdot 0 = \mu_L \Delta \theta \sum_{t=1}^T \delta^{t-1} (2\alpha - 1)^{t-1} q^\#(\theta_H | \theta_H^{t-1})$$

$$\underline{v}^* = \mu_L U^*(\theta_L) + \mu_H U^*(\mu_H) = \Delta \theta \sum_{t=1}^T \mathbb{P}(\theta_t = \theta_L) q^*(\theta_H | \theta_H^{t-1})$$

where both quantity vectors evaluated at $\lambda = 0$, and where b stands for benchmark model. A careful look at the two formulas would reveal that there is no obvious mathematical way of ranking $\underline{v}^b$ and $\underline{v}^*$.

We numerically calculate the values for two models explored above: $T = 2$ and $T = \infty$. In Figure 10 we present the results for the simpler model: $T = 2$. In the $\mu_H \times \alpha$ space, we have plotted $\mathbb{1}_{\{\underline{v}^* \geq \underline{v}^b\}}$. The shaded area represents a negative value of $\underline{v}^* - \underline{v}^b$, and white delineates when it is positive. Note that both figures are drawn for very small values of δ . When δ is moderately large, the space is completely white, that is, the ex ante expected utility of the agent in our model is higher than that in the benchmark model. And, this basic insight carries over for various parameter values and the infinite horizon framework.³³

³³We present the two period example here, because the recursive infinite horizon model involves an eight dimensional policy choice: $\langle \mathbf{w}, \mathbf{z}_L, \mathbf{z}_H, \mathbf{q} \rangle$, which would take the reader a long time to replicate. However, we do provide the code for completeness.

Conceptually, in the absence of (PK), we argue for the admittance of this numerical analysis as evidence of limited liability as a rightful interpretation of constraining the stage utility to be positive. When the agent is reasonably patient, he strictly gains more from signing a contract in which he is "protected" by a rule of law that he cannot be forced to pay large amounts of cash or liquidate high valued assets at any stage of the contract.

5 Extensions

5.1 Asymmetric Markov evolution

So far we have assumed a symmetric Markov matrix. We extend our results to the general model: $f(\theta_L|\theta_i) = \alpha_L$. We require some structure on the parameters to be able to precisely pin down the ranking of optimal distortions across various histories. These are specified in the appendix. All proofs are in fact presented for this more general model.

The conclusion remains the same: the interaction of incentives and feasibility every period produces markedly different dynamic distortions than we find in the standard model of dynamic mechanism design. Optimal distortions, weighted by asymmetries in Markov evolution, are increasing for a sequence of high cost realizations, and decrease for a low cost. Expected utility increases after a low cost, and decreases after a high cost realization. Liquidity forms a penultimate zone towards efficiency. And, efficiency is an absorbing state that is reached almost surely.

5.2 Generalized iid

Another interesting model, not completely captured by our specification is the general iid model: $\alpha_L = 1 - \alpha_H$. This is equivalent to $\alpha = 1/2$ in the symmetric Markov model. Extending our results to this framework is straightforward. The shell, depicted in Figure 6, in which inefficient region of the optimal contract lies, collapses to a line, that is $\text{int}(B) = \emptyset$. The recursive framework is presented in the appendix in Section 8.8. All key thresholds and monotonicity properties hold as a special case of the Markov model.

6 Related literature

This paper sits at the intersection of two literatures- dynamic mechanism design and dynamic financial contracting. We take up each in turn.

The last decade has seen a burgeoning literature in mechanism design where agents meet or transact in the market repeatedly. A key challenge therein is the modeling of dynamic incentives. This literature has been pioneered by papers such as Courty and Li [2000], Battaglini [2005] and Eső and Szentes [2007] with a focus on optimal contracts, and Athey and Segal [2007], Athey and Segal [2013] and Bergemann and Välimäki [2010] that focus on implementing the efficient allocation. Pavan et al. [2014] provide a unified treatment of dynamic incentives through what has

come to be referred as the dynamic envelope formula.³⁴

Most, if not all, of this literature does not ask the question of feasibility or the extent of commitment beyond the usual individual rationality constraint. In particular it allows agents to forgo payments or post large bonds, thereby alleviating dynamic incentives through maximal backloading of information rents. In a critique of these models, Eső and Szentes [2015] point out that these modelling choices lead to a kind of irrelevance of dynamics. Since incentives and feasibility interact only at the start of the contract, distortions are akin to the static model and are only augmented by the marginal "innovation" to agent types.³⁵

What would happen if this unbridled backloading instrument is not available? How does incentives interact with a stronger notion of feasibility? Our paper places a standard mechanism design problem in a dynamic framework where the agent is constrained on how much credit he can borrow- producing markedly different time structure of information rents, and hence distortions.

Krähmer and Strausz [2015] consider a (two-period) sequential screening model with ex post participation constraints. This comes close to our framework of demanding stronger notions of feasibility. They show that with these additional constraints the optimal contract is static and does not illicit the agent's information sequentially. Our model is multi-period, the optimal contract is not "static", and efficiency (and overcoming of private information) is attained eventually.

In a similar vein, Ashlagi et al. [2016] consider a framework where a monopolist wants to sell k goods in k periods, valuations are iid over time, and the mechanism must satisfy ex post individual rationality. They provide an implementation of the optimal mechanism through delayed payments where all the utility is paid in the last period. We look at a different model, solve for the optimal allocation rule, and provide a plausible implementation.

There are other papers that seek to temper the flexibility allowed by the linearity of transfers in dynamic models. Guo and Hörner [2015] analyze a dynamic screening model without transfers (see Lipnowski and Ramos [2015] for a related model without commitment). Amador et al. [2006], and Halac and Yared [2014] study models of delegation. Thomas and Worrall [1990], Garrett and Pavan [2015], Luz [2015], and Arve and Martimort [2016] consider dynamic models of private information where the agent is risk averse. Our paper presents a tractable model of restricting the feasible set of transfers which in turn interacts with incentives to produce distinctly different allocation rules.

The second strand of literature that our paper connects to is dynamic financial contracting (see excellent surveys by Biais et al. [2013], and Sannikov [2013]). Agency frictions here are motivated in two forms: moral hazard and cash flow diversion. Our paper is the most closely related to Clementi and Hopenhayn [2006] and Fu and Krishna [2016]. Both study the problem of cash flow diversion by the agent in a repeated setting, the only difference being that the former looks at an iid technology and the latter a Markovian one.

A direct and simple way to map their framework into ours would be to change the time structure of the supply of inputs and payments. At the start of every period the agent commits to a

³⁴See Bergemann and Said [2011], Vohra [2012], and Bergemann and Pavan [2015] for excellent surveys.

³⁵Their argument works only when the first-order approach is valid. When global incentive constraint binds, the full dynamic nature of the evolution of the types has to be internalized. See Battaglini and Lamba [2015].

production plan after which his cost type is realized. The type is reported, agreed upon input quantity is supplied, and the agent is compensated for by the principal. The interpretation here is that agent does not know whether his cost would be low or high when he makes the production decision. Despite being a low cost type, he can misreport to be a high cost type, supply some portion of the produced quantity and sell the rest in the black market- a diversion of the economic surplus.

The space of contracts or the mechanism now is defined by $m = \left(p(\theta_t|h^{t-1}), q(h^{t-1}) \right)_{t=1}^T$. The efficient contract is of course Markov:

$$V'(q^e(h^{t-1}, \theta_j)) = \alpha_j \theta_L + (1 - \alpha_j) \theta_H$$

Stage utility in this case has a different form: $u(\theta_t|h^{t-1}) = p(\theta_t|h^{t-1}) - \theta_t q(h^{t-1})$. The incentive constraint is given by:

$$U(\theta_t|h^{t-1}) \geq p(\hat{\theta}_t|h^{t-1}) - \theta_t q(h^{t-1}) + \delta \mathbb{E} \left[U(\tilde{\theta}_{t+1}|h^{t-1}, \hat{\theta}_t) \mid \theta_t \right]$$

and cash-strapped constraint simply by $p(\theta_t|h^{t-1}) - \theta_t q(h^{t-1}) \geq 0$. An immediate consequence of the time structure is that any finite horizon version of the modified model has a trivial solution. In the last period the principal does not have enough instruments to screen types, so pools both at best contract for the high cost type. By backward induction the same is true for all periods. In contrast, we provide an economically meaningful solution for the two period model.

Drawing on Clementi and Hopenhayn [2006], we can show that in the modified model, points 6 and 9 of Theorem 1 will continue to hold for iid types. Fu and Krishna [2016] show that points 5 and 8 also hold in the Markov extension of Clementi and Hopenhayn [2006]. Neither paper though can explicitly characterize the optimal distortions, that is, points 1, 2 and 3 in Theorem 1 are unique to our paper. As a consequence, points 4 and 7 too are novel.

The literature on dynamic financial contracting is also seeped in plausible implementations of the optimal allocation rules. DeMarzo and Fishman [2007] and Biais et al. [2007] are leading references. We show that such implementations can have natural interpretation in corresponding screening and adverse selection models through the dynamic information operator.

7 Conclusion

This paper motivates the study of financial constraints in dynamic contracting through the interaction between persistent asymmetric information and cash or liquidity constraints. The agent has access to a viable technology marred by agency frictions, and is strapped for cash. The paper situates itself firmly in between the literatures on dynamic mechanism design and dynamic financial contracting.

The key challenge is to characterize the evolution of informational distortions. We show that dynamic distortions increase with every consecutive "bad shock" and decrease with a "good shock". An endogenous number of "good shocks" are required for the contract to become liquid, and

eventually efficient, the latter being an absorbing state. We present a simple implementation of the optimal allocation in terms of an intuitive dynamic information operator. The share of the principal and agent in the total economic surplus evolve endogenously. The value of the "meta firm" is increasing in the share of the agent, and along the optimal contract also in his utility spread. We also provide a foundation for limited liability as a useful restriction for the agent in dynamic contracts.

An important missing link from our analysis, present in Clementi and Hopenhayn [2006], is liquidation or scrapping of the contract between the principal and agent. There is a sequence of histories for which the value of the "meta firm" could become very low, and perhaps the fire sale value of the assets at that point is greater than expected value of continuation. The evolution of liquidation decisions in dynamic financial contracting with persistent agency frictions is a worthwhile question for future research.

Theoretically speaking, the paper is limited to the two-types model because it is difficult to pin down the optimal allocation rule with more than two Markov types (see Battaglini and Lamba [2015]). Global incentive constraints are generically binding for high persistence, even for the benchmark model. Adding stronger feasibility will potentially complicate the analysis considerably. Looking for approximate optimal and easily characterizable contracts is a promising approach going forward.

Finally, the ideas developed in the paper can potentially hold promise for other economically meaningful questions such as optimal taxation and double auctions. In optimal taxation, the agent is the citizen with a private labor productivity. The principal is the government seeking to maximizing a utilitarian or Pareto-weighted welfare function. It can be shown (see for example Farhi and Werning [2013]) that the dual of this problem can be written down as follows.

$$\text{Principal: } y - c, \quad \text{Agent: } c - \theta h(y)$$

where y is the agent's income/endowment, c his consumption plan, and h is a concave production function. The principal seeks to maximize his objective (that is edowment net of consumption) subject to incentive, promise keeping, and a "consumption-strapped" constraint. Here, financial constraints have a meaningful interpretation- the agent cannot be forced, either by law or due to incomplete credit markets to forgo significant amounts of consumption in any given period.³⁶ Our model exactly fits this setup, and it would be interesting to explore the economic message of optimal taxation with a limited liability or consumption-strapped constraint.

In repeated transaction in financial markets, a double auction with liquidity constraints seems like a reasonable baseline model which could potentially have attractive empirical properties. The problem can be written down as follows.

$$\text{Seller: } p_S - \theta^S x, \quad \text{Buyer: } \theta^B x - p_B$$

³⁶A typical Mirrlees taxation model would have risk averse agents and individual rationality constraints. We have agents with quasilinear preferences, but the consumption-strapped constraint ensures that the agent is infinitely risk-averse at a utility below zero.

where x is the probability of trade, p the transfer, and θ^S and θ^B are the private measures of respective willingness to trade. So far the literature on repeated bilateral trade has only considered individual rationality constraint allowing for strong forms of commitment which may not be realistic in certain situations (see Athey and Miller [2007], Skrzypacz and Toikka [2015] and Lamba [2016]). The evolution of dynamic distortions in repeated bilateral trade with financial constraints is an interesting topic for future work.

8 Appendix

We divide the appendix into six sections- the proofs for two period model, the sequential approach period, followed by the recursive approach, then the main theorem, sufficiency conditions, and finally dynamics of payments. Throughout we will invoke the general model where $f(\theta_L|\theta_i) = \alpha_i$. We shall assume the following, their role aptly explained by their title:

(A1) Persistence: $\alpha_L, 1 - \alpha_H \geq 1/2$.

(A2) Limited asymmetry: $1 - \alpha_L \geq \alpha_H \geq (1 - \alpha_L)\alpha_L$.

(A3) Ranking of prior: $\alpha_H \leq \mu_L \leq \alpha_L$.

(A1) is assumed throughout. (A2) is used in constructing the shell in the recursive approach. (A3) ensures that the optimal contract starts in the shell.

8.1 Benchmark: dynamic model without financial constraints

Consider a relaxed problem where the principal chooses to maximize $\bar{S} - [\mu_L U(\theta_L) + \mu_H U(\theta_H)]$ subject to $IC_L(h^{t-1})$ and $IR_H(h^{t-1}) \forall h^{t-1}, \forall t$. All constraints can be assumed to hold as an equality, and it can be easily shown that the solution to the relaxed problem is globally optimal.

Inductively applying the binding $IC_L(\theta_H^{t-1})$ gives us

$$U(\theta_L) = U(\theta_H) + \Delta\theta \sum_{t=1}^T \delta^{t-1} (\alpha_L - \alpha_H)^{t-1} q(\theta_H|\theta_H^{t-1})$$

In addition, IR_H binds, which gives us $U(\theta_H) = 0$. Substituting back into the objective function, we get that

$$V'(q^\#(\theta_L|h^{t-1})) = \theta_L \forall h^{t-1}, \forall t, \text{ and } V'(q^\#(\theta_H|h^{t-1})) = \theta_H \forall h^{t-1} \neq \theta_H^{t-1}, \forall t$$

$$V'(q^\#(\theta_H|\theta_H^{t-1})) = \theta_H + (1 - \lambda) \frac{\mu_L}{\mu_H} \left(\frac{\alpha_L - \alpha_H}{1 - \alpha_H} \right)^{t-1} \Delta\theta$$

where λ is the Lagrange multiplier on (PK).

8.2 Two period model

We first establish the set of binding constraints for the relaxed problem. For the mechanism $\langle \mathbf{U}, \mathbf{q} \rangle$, the constraints can be written as:

$$\begin{aligned} IC_L : \quad & U(\theta_L) \geq U(\theta_H) + \Delta\theta q(\theta_H) + \delta(\alpha_L - \alpha_H) [u(\theta_L|\theta_H) - u(\theta_H|\theta_H)] \\ C_L : \quad & U(\theta_L) \geq \delta [(\alpha_L u(\theta_L|\theta_L) + (1 - \alpha_L)u(\theta_H|\theta_L))] \\ C_H : \quad & U(\theta_H) \geq \delta [(\alpha_H u(\theta_L|\theta_H) + (1 - \alpha_H)u(\theta_H|\theta_H))] \\ IC_L(\theta_i) : \quad & u(\theta_L|\theta_i) \geq \Delta\theta q(\theta_H|\theta_i) + u(\theta_H|\theta_i) \quad \text{for } i = L, H \\ C_H(\theta_i) : \quad & u(\theta_H|\theta_i) \geq 0 \quad \text{for } i = L, H \end{aligned}$$

Lemma 2. For $(\mathcal{RP})$: IC_L , $IC_L(\theta_H)$, C_H , $C_H(\theta_H)$ bind at the optimum, and $IC_L(\theta_L)$, $C_H(\theta_L)$ can be assumed to hold as equalities.

Proof. See online appendix. □

Next, we provide closed form solutions for the optimal allocation rule.

Lemma 3. The optimal contract with promised utility $v_0 \in [0, S^e]$ is characterized by the following allocation rule:

$$\begin{aligned} V'(q^*(\theta_L)) &= \theta_L, \quad V'(q^*(\theta_H)) = \theta_H + (1 - \lambda) \frac{\mathbb{P}(\theta_1 = \theta_L)}{\mathbb{P}(\theta_H^1)} \Delta\theta - \frac{\gamma}{\mu_H} \Delta\theta \\ V'(q^*(\theta_L|\theta_L)) &= \theta_L, \quad V'(q^*(\theta_H|\theta_L)) = \theta_H + \frac{1}{\mu_L} \frac{\gamma\alpha}{1 - \alpha} \Delta\theta \\ V'(q^*(\theta_L|\theta_H)) &= \theta_L, \quad V'(q^*(\theta_H|\theta_H)) = \theta_H + (1 - \lambda) \frac{\mathbb{P}(\theta_2 = \theta_L)}{\mathbb{P}(\theta_H^2)} \Delta\theta - \frac{\gamma}{\mu_H} \Delta\theta \end{aligned}$$

where γ and λ are jointly determined as follows: $\gamma \geq 0$ is given by the solution to the equation

$$u^*(\theta_L) = \Delta\theta q^*(\theta_H) + \delta\Delta\theta\alpha [q^*(\theta_H|\theta_H) - q^*(\theta_H|\theta_L)] \geq 0$$

through complimentary slackness, and

$$\lambda \begin{cases} = 0 & \text{if } v_0 \leq \underline{v} \\ \in (0, 1) \text{ and is strictly increasing in } v_0 & \text{if } \underline{v} < v_0 < \bar{v} \\ = 1 & \text{if } v_0 \geq \bar{v} \end{cases}$$

Proof. Using the binding constraints, we can write down the total information rent in equation (4):

$$\mu_L U(\theta_L) + \mu_H U(\theta_H) = \Delta\theta \sum_{t=1}^2 \delta^{t-1} \mathbb{P}(\theta_t = \theta_L) q(\theta_H|\theta_H^{t-1})$$

Substituting this, and the equation $u^*(\theta_L) \geq 0$ with Lagrange multiplier γ and taking derivatives with respect to $\mathbf{q}$ gives us the solution. Since the objective is concave in quantities, first-order conditions are indeed sufficient.

Next, we argue that λ is strictly increasing in v_0 , for $v_0 \in [\underline{v}, \bar{v}]$. Now, λ and γ are respectively the Lagrange multipliers on the following constraints are jointly determined:

$$(\lambda) : \quad \frac{v_0}{\Delta\theta} = \sum_{t=1}^2 \mathbb{P}(\theta_t = \theta_L) q^*(\theta_H | \theta_H^{t-1})$$

$$(\gamma) : \quad q^*(\theta_H) + \delta\alpha (q^*(\theta_H | \theta_H) - q^*(\theta_H | \theta_L)) = 0$$

It is clear that $q^*(\theta_H)$ and $q^*(\theta_H | \theta_H)$ are increasing functions of v_0 . Then, it is easy to see that the optimal value of the objective is strictly concave in v_0 , for $v_0 \in [\underline{v}, \bar{v}]$. The subdifferential of the optimal value of the objective at v_0 is $-\Lambda(v_0)$ where $\Lambda(v_0)$ consists of λ which supports the solution at v_0 . By strict concavity, Λ is a well-defined correspondence which is strictly decreasing in the strong set order.

It remains only to show λ can not exceed 1. We argue that $q^*(\theta_H)$, $q(\theta_H | \theta_L)$ and $q(\theta_H | \theta_H)$ can be only distorted downward from $q^e(\theta_H)$, then the claim follows. Indeed, these quantities appear only on the right side of the constraints in $(\mathcal{RP}^*)$. Hence, for example, if $q^*(\theta_H) > q^e(\theta_H)$, then this quantity can be decreased to $q^e(\theta_H)$ satisfying the constraints and strictly improving the objective. $\square$

Proposition 1 follows from Lemma 3 given that $\frac{\mathbb{P}(\theta_1=\theta_L)}{\mathbb{P}(\theta_H^1)} < \frac{\mathbb{P}(\theta_2=\theta_L)}{\mathbb{P}(\theta_H^2)}$ from (A2) and (A3). Next we show Corollaries 1 and 2.

Proof of Corollary 1. We have $u^*(\theta_H | \theta_H) = 0$ and $u^*(\theta_H | \theta_L) \geq 0$. Moreover,

$$u^*(\theta_L | \theta_L) \geq \Delta\theta q^*(\theta_H | \theta_L) + u^*(\theta_H | \theta_L) \quad \text{and} \quad u^*(\theta_L | \theta_H) = \Delta\theta q^*(\theta_H | \theta_H) + u^*(\theta_H | \theta_H)$$

If C_L is slack, then $q^*(\theta_H | \theta_L) = q^e(\theta_L) > q^*(\theta_H | \theta_H)$. If it binds, then

$$\delta\alpha (q^*(\theta_H | \theta_L) - q^*(\theta_H | \theta_H)) = q^*(\theta_H) > 0$$

Either way, we get $u^*(\theta_H | \theta_L) > u^*(\theta_H | \theta_H)$. $\square$

Proof of Corollary 2 . Recall that $S^*(v_0) = R^*(v_0) + \max\{\underline{v}, v_0\}$. The claim follows from the arguments in Lemma 3, because S^* is subdifferentiable at $v_0 \in [\underline{v}, \bar{v}]$. Its subdifferential at any such v_0 is $1 - \Lambda(v_0) \geq 0$, which is strictly increasing in the strong set order. Finally, notice that S^* is flat for $v_0 \notin [\underline{v}, \bar{v}]$. So, it is non-decreasing in the whole domain. $\square$

8.3 Sequential approach

The set of constraints in $(\mathcal{RP}^*)$ can be enlisted as follows.

$$\begin{aligned}
IC_L(h^{t-1}) : \quad & U(\theta_L|h^{t-1}) \geq U(\theta_H|h^{t-1}) + \Delta\theta q(\theta_H|h^{t-1}) \\
& \quad + \delta(\alpha_L - \alpha_H) [U(\theta_L|h^{t-1}, \theta_H) - U(\theta_H|h^{t-1}, \theta_H)] \\
C_L(h^{t-1}) : \quad & U(\theta_L|h^{t-1}) \geq \delta [(\alpha_L U(\theta_L|h^{t-1}, \theta_L) + (1 - \alpha_L)U(\theta_H|h^{t-1}, \theta_L)] \\
C_H(h^{t-1}) : \quad & U(\theta_H|h^{t-1}) \geq \delta [(\alpha_H U(\theta_L|h^{t-1}, \theta_H) + (1 - \alpha_H)U(\theta_H|h^{t-1}, \theta_H)]
\end{aligned}$$

We now provide the proofs of Proposition 2 and Corollary 3.

Proof of Proposition 2. Consider any history h^{t-1} . First, $C_H(h^{t-1})$ could be assumed to hold as an equality, because this relaxes the constraints. Second, $IC_L(h^{t-1})$ binds which is proven in the recursive approach in Lemma 22 and in the main theorem.

Proving Lemmata 10, 11 and 12, we argue that the optimal contract is interior and could be characterized using the Lagrangian method with some multipliers in l^1 . So, let $\delta^{t-1}\eta(h^{t-1})\mathbb{P}(h^{t-1}, \theta_H)$ be the multiplier for $C_H(h^{t-1})$, then the result readily follows. $\square$

Proof of Corollary 3. If we show that the optimal distortions in $(n\mathcal{RP})$ along the high cost history grows without bound, we are done. Now,

$$\frac{\mathbb{P}(\theta_{t+1} = \theta_L)}{\mathbb{P}(\theta_H^{t+1})} = \frac{\alpha_H + [\mu_L(1 - \alpha_L) - \mu_H\alpha_H](\alpha_L - \alpha_H)^t}{\mu_H(1 - \alpha_L + \alpha_H)(1 - \alpha_H)^t}$$

This expression clearly converges to ∞ . $\square$

8.4 Recursive approach

In this section, we convert $(\mathcal{RP}^*)$ into its recursive avatar. The recursive formulations have been defined as $(\mathcal{RF})$ and $(\mathcal{RF}_0)$ in the main text in section 2.4. Here we establish the validity of the recursive approach and the shape of the optimal contract. First, $(\mathcal{RF})$ can be restated for the general Markovian framework as follows:

$$(\mathcal{RF}) \quad Q_j^*(w) = \max_{\langle \mathbf{z}_L, \mathbf{z}_H, \mathbf{q} \rangle} \alpha_j [s(\theta_L, q_L) + \delta Q_L^*(z_L)] + (1 - \alpha_j) [s(\theta_H, q_H) + \delta Q_H^*(z_H)]$$

subject to $\langle \mathbf{z}_L, \mathbf{z}_H, \mathbf{q} \rangle \in W^2 \times \mathbb{R}_+^2$, and

$$\begin{aligned}
w_L - w_H &\geq \Delta\theta q_H + \delta(\alpha_L - \alpha_H)(z_{HL} - z_{HH}) \\
w_L &\geq \delta[\alpha_L z_{LL} + (1 - \alpha_L)z_{LH}] \\
w_H &\geq \delta[\alpha_H z_{HL} + (1 - \alpha_H)z_{HH}]
\end{aligned}$$

$(\mathcal{RF}_0)$ can similarly be rewritten for the general model.

Recursive domain

In the main text we introduced the recursive formulation, though we were sloppy in describing its domain W . It follows from the cash-strapped constraint that $W \subseteq \mathbb{R}_+^2$. In addition, it is easy to see that $\{w \in \mathbb{R}_+^2 : w_L \geq w_H\} \subseteq W$. We prove the converse inclusion by iterative approximations of W .

Lemma 4. $W = \{w \in \mathbb{R}_+^2 : w_L \geq w_H\}$.

Proof. See online appendix. $\square$

Properties: existence and equivalence

Lemma 5. *There exists the unique continuous bounded function satisfying the Bellman equation in the problem $(\mathcal{RF})$.*

Proof. The result follows from Exercise 9.7 of Stokey et al. [1989] $\square$

Let $(\bar{z}_L(w), \bar{z}_H(w), \bar{q}(w))$ be the set of maximizers in the problem $(\mathcal{RF})$ given w and θ_j . Notice that this set is independent of θ_j . The policy correspondence is a correspondence which maps w into $(\bar{z}_L(w), \bar{z}_H(w), \bar{q}(w))$.

Lemma 6. *The policy correspondence is non-empty, compact-valued and upper hemicontinuous.*

Proof. The result follows from Exercise 9.7 of Stokey et al. [1989] $\square$

Say that a contract is generated from the policy correspondence with $(U(\theta_L), U(\theta_H))$ if

$$\begin{aligned} U(\theta_j|h^{t-1}, \theta_i) &\in \bar{z}_{ij}[U(\theta_L|h^{t-1}), U(\theta_H|h^{t-1})] \\ q(\theta_i|h^{t-1}) &\in \bar{q}_i[U(\theta_L|h^{t-1}), U(\theta_H|h^{t-1})] \end{aligned} \quad (13)$$

Lemma 7. *A contract is generated from the policy correspondence with $(U(\theta_L), U(\theta_H))$ if and only if it solves the problem $\star$ with $w = (U(\theta_L), U(\theta_H))$.*

Proof. The result follows from Exercises 9.4, 9.5 of Stokey et al. [1989] and the Lemmata 5 and 6. $\square$

Finally, a proof of Lemma 1 follows from Lemma 7

Properties: efficiency

Let E be the largest subset of W such that the constraints set in the problem $(\mathcal{RF})$ and $(\mathcal{RF}_0)$ is non-empty when $q_i = q^e(\theta_i)$ for $i = L, H$. If the principal is committed to deliver $w \in E$, then it is possible to achieve the maximal surplus which is given by

$$Q_j^e(w) = \max_{\langle z_L, z_H \rangle} \alpha_j[s(\theta_L, q^e(\theta_L)) + \delta Q_L^e(z_L)] + (1 - \alpha_j)[s(\theta_H, q^e(\theta_H)) + \delta Q_H^e(z_H)]$$

Notice E does not depend on h^{t-1} . Importantly, this set is independent of θ_j , because a last period shock does not matter for this period efficient quantity.

Let $\kappa = \frac{\Delta q^e(\theta_H)}{1-\delta(\alpha_L-\alpha_H)}$, $(1-\delta)\underline{w}_H^e = \delta\alpha_H\kappa$, and $\underline{w}_L^e = \underline{w}_H^e + \kappa$. It is easy to see that $\{w \in W : w_H \geq \underline{w}_H^e \text{ and } w_L \geq w_H + \kappa\} \subseteq E$. To prove the converse, we use the same approach as in the section on the recursive domain.

Lemma 8. $E = \{w \in W : w_H \geq \underline{w}_H^e \text{ and } w_L \geq w_H + \kappa\}$.

Proof. See online appendix. □

Properties: shape

Lemma 9. Q_j^* is concave.

Proof. See online appendix. □

Lemma 10. Q_j^* is supermodular.

Proof. See online appendix. □

Properties: differentiability

In this section, we study differentiability of the value functions in $\boxed{\star}$ and $(\mathcal{RF})$. Unfortunately, the standard argument of Benveniste and Scheinkman [1979] is not applicable in our context, because it might not be possible to change q_H keeping z_L, z_H constant. Moreover, the other known result of Rincón-Zapatero and Santos [2009] also does not have a bite.

We approach the differentiability through the uniqueness of Lagrange multipliers by applying Theorem 2 of Morand and Reffett [2015]. Further details of this are provided in the online appendix.

Consider the problem $\boxed{\star}$ and notice that the constraints set in this problem could be described by a linear operator from l^∞ to itself. We shall call this operator the constraint map.

Lemma 11. Let $w \in \text{int}(W)$, then there exists a feasible point such that the constraint map is uniformly bounded away from 0.

Proof. See online appendix. □

Now, we establish differentiability of the value function. It is clear that for the concave problem differentiability is equivalent to the uniqueness of the Lagrange multipliers.

Lemma 12. Consider $\langle U^*, q^* \rangle$ which solves the problem $\boxed{\star}$ with $w \in \text{int}(W)$. Then, $q^*(\theta_H|h^{t+s}) > 0$ and $(U^*(\theta_L|h^{t+s}), U^*(\theta_H|h^{t+s})) \in \text{int}(W)$ for all $h^{t+s} \in H^{t+s}|_{(h^{t-1}, \theta_j)}$.

Proof. See online appendix. □

Lemma 13. Q_j^* is continuously differentiable on $\text{int}(W)$.

Proof. See online appendix. □

Lemma 14. $\lim_{w_L \rightarrow w_H} D_L Q_j^*(w) = \infty \forall w_H$, and $\lim_{w_H \rightarrow 0} D_H Q_j^*(w) = \infty \forall w_L \neq 0$.

Proof. See online appendix. $\square$

Properties: optimality

Let $(1 - \alpha_j)\beta$, $\alpha_j \rho_L$ and $(1 - \alpha_j)\rho_H$ be Lagrange multipliers for the respective constraints in $(\mathcal{RF})$. Let $w \in \text{int}(W)$. Since the optimum is interior by the Lemma 12, it is characterized by the following first-order conditions

$$\begin{aligned} D_L Q_L^*(z_L) &= \alpha_L \rho_L \\ D_H Q_L^*(z_L) &= (1 - \alpha_L) \rho_L \\ D_L Q_H^*(z_H) &= \alpha_H \rho_H + (\alpha_L - \alpha_H) \beta \\ D_H Q_H^*(z_H) &= (1 - \alpha_H) \rho_H - (\alpha_L - \alpha_H) \beta \\ D_q s(\theta_H, q_H) &= \Delta \theta \beta \end{aligned} \tag{14}$$

In addition, the following envelope conditions are satisfied

$$\begin{aligned} D_L Q_L^*(w) &= \alpha_L \rho_L + (1 - \alpha_L) \beta \\ D_H Q_L^*(w) &= (1 - \alpha_L) (\rho_H - \beta) \\ D_L Q_H^*(w) &= \alpha_H \rho_L + (1 - \alpha_H) \beta \\ D_H Q_H^*(w) &= (1 - \alpha_H) (\rho_H - \beta) \end{aligned} \tag{15}$$

At date $t = 1$, the problem is different. Let λ be a multiplier for the first constraint in $(\mathcal{RF}_0)$, and $\mu_H \beta$, $\mu_L \rho_L$ and $\mu_H \rho_H$ be the other multipliers. The first-order conditions with respect to z_L , z_H and q_H are the same as in (14). The extra first-order conditions are

$$\begin{aligned} \mu_L \rho_L + \mu_H \beta &= \mu_L \lambda \\ \mu_H (\rho_H - \beta) &= \mu_H \lambda \end{aligned} \tag{16}$$

Auxiliary results

Now, we prove a sequence of small results that lead to the proof of the main theorem. Note at the outset that $w_L^{liq} = \delta[\alpha_L \underline{w}_L^e + (1 - \alpha_L) \underline{w}_H^e]$.

Lemma 15. Let $w \in \text{int}(W)$, then $D_L Q_j^*(w) = D_H Q_j^*(w) = 0$ only if $w \in E$.

Proof. From (15), $D_L Q_j^*(w) = D_H Q_j^*(w) = 0$ holds if and only if $\rho_L = \rho_H = \beta = 0$ and it implies $D_L Q_L^*(v_L) = D_H Q_L^*(v_L) = D_L Q_H^*(v_H) = D_H Q_H^*(v_H) = 0$ with $q_H = q^e(\theta_H)$. Iterative application of the previous argument yields $w \in E$. $\square$

Next, define $H = \{w \in \text{int}(W) : DQ_L^*(w) \gg 0 \text{ and } DQ_H^*(w) \gg 0\}$; then by Lemma 14 $H \neq \emptyset$.

Lemma 16. $H \subseteq (0, \underline{w}_L^e) \times (0, \underline{w}_H^e)$.

Proof. See online appendix. $\square$

Lemma 17. *Let $w \in H$, then $z_H \in H$ and $z_L \in H \cup E$. Moreover, $z_{LL} > w_L$ when $w \notin E$, and $z_L \in E$ if and only if $w_L \geq w_L^{liq}$.*

Proof. From (15), $\rho_H > \beta \geq 0$ implying that $DQ_H^*(z_H) \gg 0$ by (14). Next, from (14), either $DQ_L^*(z_L) = 0$ or $DQ_L^*(z_L) \gg 0$ and $DQ_L^*(z_L) = 0$ if and only if $\rho_L = 0$, which proves the first part. $\delta \in (0, 1)$ and $z_L \in W \Rightarrow z_{LL} > w_L$. Finally, $w_L < w_L^{liq}$ if and only if $z_L \notin E$. Therefore, the last claim follows from Lemma 15. $\square$

Lemma 18. *Let $w \in H$, then β is non-increasing in w_L and non-decreasing in w_H*

Proof. See online appendix. $\square$

Lemma 19. *Q_j^* is strictly concave in each variable on H .*

Proof. See online appendix. $\square$

Define $\eta_j(w) = (1-\alpha_j)D_L Q_j^*(w) - \alpha_j D_H Q_j^*(w)$. Then $\eta_j(w) = \alpha_j(1-\alpha_j)(\rho_L - \rho_H) + (1-\alpha_j)\beta$, $\eta_H(z_H) = (\alpha_L - \alpha_H)\beta$ and

$$\frac{\eta_H(w)}{\alpha_H(1-\alpha_H)} = \frac{\eta_L(w)}{\alpha_L(1-\alpha_L)} + \frac{(\alpha_L - \alpha_H)\beta}{\alpha_L \alpha_H} \quad (17)$$

Lemma 20. *For any $w_H \in (0, \underline{w}_H^e)$, there exists unique $w_L^i(w_H)$ such that $\eta_i[w_L^i(w_H), w_H] = 0$ and w_L^i is strictly increasing in w_H .*

Proof. See online appendix. $\square$

Lemma 21. $\lim_{w_H \rightarrow 0} w_L^i(w_H) = 0$ and $\lim_{w_H \rightarrow \underline{w}_H^e} w_L^i(w_H) = \underline{w}_L^e$

Proof. See online appendix. $\square$

Define $B = \{w \in H : \eta_L(w) \leq 0 \leq \eta_H(w)\}$. In the main next, we call B the shell.

Lemma 22. *If $\alpha_L \neq \alpha_H$, then $B \neq \emptyset$ and $\beta > 0$ when $w \in B$.*

Proof. Consider $w \in H$ with $\eta_H(w) = 0$ and $w_L \geq w_L^{liq}$. Since $\rho_L = 0$, $\beta > 0$ implies that $\eta_L(w) < 0$. Let w^0 be such that $\eta_H(w^0) = 0$ with $w_L^0 = w_L^{liq}$. Then, there exists w^1 such that $\eta_L(w^1) = 0$ and $w_H^1 = w_H^0$, $w_L^1 < w_L^0$. From Lemma 18, $\beta(w_L^1, w_H) \geq \beta(w_L, w_H)$ for any $w_L > w_L^1 > w_H > 0$. In particular $\beta(w^1) \geq \beta(w^0) > 0$. Notice that $\eta_H(w^1) > 0$ by (17). Thus, there exists w^2 such that $\eta_H(w^2) = 0$ and $w_H^2 < w_H^1$, $w_L^2 = w_L^1$. By strict concavity on H (see Lemma 19) $\eta_L(w^2) < 0$, implying that $\beta(w^2) > 0$ by (17). Iterative application of the previous argument gives the result. $\square$

Lemma 23. *Let $\alpha_L \neq \alpha_H$ and $w \in H$ with $w_H \geq \delta[\alpha_H w_L + (1 - \alpha_H)w_H]$, then $z_{HL} > w_L$.*

Proof. Suppose instead $z_{HL} \leq w_L$. Then $z_{HH} \geq w_H$ implying that $D_H Q_H^*(z_H) \leq D_H Q_H^*(w)$ by concavity and supermodularity. But, (14), (15) and Lemma 22 yield $D_H Q_H^*(z_H) > D_H Q_H^*(w)$, a contradiction. $\square$

Lemma 24. *Let $\alpha_L \neq \alpha_H$ and $w \in H$ with $w_H \leq \delta[\alpha_H w_L + (1 - \alpha_H)w_H]$, then $z_{HH} < w_H$.*

Proof. Similar to the proof of Lemma 23. $\square$

Lemma 25. *Suppose $\alpha_L \neq \alpha_H$ and $\alpha_H \geq \alpha_L(1 - \alpha_L)$, then $B \subseteq \{w \in H : w_H < \delta[\alpha_H w_L + (1 - \alpha_H)w_H]\}$. In addition, $z_H \neq w$, and z_H and w are ordered.*

Proof. Consider w such that $\eta_L(w) \leq 0$, then by (17) $\alpha_L(\rho_L - \rho_H) + \beta \leq 0$. The assumption implies that $D_L Q_H^*(z_H) - D_L Q_H^*(w) = \alpha_H(\rho_H - \rho_L) - (1 - \alpha_L)\beta \geq 0$, hence $z_H \neq w$, and z_H and w are ordered by strict concavity and supermodularity. Then, Lemmata 23 and 24 imply that $\{w \in H : \eta_L(w) \leq 0\}$ and $\{w \in H : w_H = \delta[\alpha_H w_L + (1 - \alpha_H)w_H]\}$ do not intersect. $\square$

Lemma 26. *Let $\alpha_L \neq \alpha_H$ and consider $w \neq w' \in H$ such that $\eta_L(w) = \eta_L(w') = 0$, then $\eta_H(w) \neq \eta_H(w')$.*

Proof. Suppose that $\eta_H(w) = \eta_H(w')$, then $\beta(w) = \beta(w') = \rho_L(w) - \rho_H(w) = \rho_L(w') - \rho_H(w') = 0$, which is a contradiction to Lemma 22. $\square$

Lemma 27. *Suppose $\alpha_L \neq \alpha_H$, $\alpha_H \geq (1 - \alpha_L)\alpha_L$ and $\alpha_H \leq 1 - \alpha_L$. Then, $z_H \in B$ when $w \in B$.*

Proof. First, assume that $\eta_L(w) = 0$. By (17), and the second assumption, $\eta_H(w) = \frac{1 - \alpha_H}{\alpha_L} \eta_H(z_H) \geq \eta_H(z_H)$ when $\eta_H(z_H) \geq 0$, which is always satisfied. By Lemma 26, the level curves of η_H crosses at most once in H . Then, Lemmata 23 and 25 imply the result. The general case is implied by monotonicity of β and the previous result. $\square$

Lemma 28. *At date $t = 1$, $w \in B$ if $\alpha_H \leq \mu_L \leq \alpha_L$.*

Proof. (16) gives $\mu_L(\rho_L - \rho_H) + \beta = 0$. Then, using (17), $\alpha_L(\rho_L - \rho_H) + \beta \leq 0 \leq \alpha_H(\rho_L - \rho_H) + \beta$ if and only if $\alpha_H \leq \mu_L \leq \alpha_L$. $\square$

8.5 Main result

Proof of Theorem 1. 1. It follows from the first-order conditions in (14: $D_q s(\theta_H, q_H) = \Delta\theta\beta \geq 0$.

2. By Lemmata 26 and 27, the optimal contract lies in the shell with $\eta_L(w) \leq 0 \leq \eta_H(w)$. In addition, $\eta_H(w) = \frac{\alpha_H(1 - \alpha_H)}{\alpha_L(1 - \alpha_L)} \eta_L(w) + \frac{\alpha_H(1 - \alpha_H)}{\alpha_L \alpha_H} \eta_H(z_H)$ by equation (17) and $\eta_H(z_H) = (\alpha_L - \alpha_H)\beta$. Using $\eta_L(w) \leq 0$, $\eta_H(w) \leq \frac{1 - \alpha_H}{\alpha_L} \eta_H(z_H)$. Now, iterate forward on θ_H and use $\eta_H(z_H) = (\alpha_L - \alpha_H)\beta$ to get $\beta(w) \leq \frac{1 - \alpha_H}{\alpha_L} \beta(z_H)$. Finally, $D_q s(\theta_H, q_H) = \Delta\theta\beta$ implies the result, because $1 - \alpha_H = \alpha_L$ for the symmetric case. For the asymmetric case, quantities are strictly decreasing with the appropriate weights.

3. Lemmata 17 and 26 taken together imply that β is monotone along the one-dimensional curve w_L^L . Then, by Equation (14), $D_q s(\theta_H, q_H) = \Delta\theta\beta$ which proves the claim.
4. The first-order and envelope conditions given in (14) and (15) give $D_L Q_L(z_L) - D_L Q_L(w) = -(1 - \alpha_L)\beta$ and $D_H Q_L(z_L) - D_H Q_L(w) = (1 - \alpha_L)(\rho_L - \rho_H + \beta)$. By Lemmata 26 and 27, the optimal contract lies in the shell with $\eta_L(w) \leq 0 \leq \eta_H(w)$ and $\beta > 0$ in the shell by Lemma 22. Thus two expression are strictly negative. This implies that $z_L \neq w$ and they are ordered. Since, $z_{LL} > w_L$ when $w \notin E$ by Lemma 17, $z_{LH} > w_H$.
Next, $z_H \neq w$ and they are ordered when $w \ni nE$ by Lemma 25 and the shell lies in the region $\{w \in H : w_H < \delta[\alpha_H w_L + (1 - \alpha_H)w_H]\}$. Since the optimal contract lies in the shell (see reference above), Lemma 24 says that $z_{HH} < w_H$ given the position of the shell. So, it must be that $z_{HL} < w_L$.
5. This claim follows from Lemma 17.
6. By construction in Lemma 8.
7. It is implied by Lemmata 8 and 17.
8. It is implied by Lemmata 8 and 17.
9. By Lemma 28, the contract starts at $w \in \text{int}(W)$. Let $D^* = (D_L + D_H)$, by (14) and (15), the stochastic process $D^*Q_j^*$ is a non-negative martingale, namely

$$\begin{aligned} D^*Q_L^*(w) &= \alpha_L D^*Q_L^*(z_L) + (1 - \alpha_L) D^*Q_H^*(z_H) \geq 0 \\ D^*Q_H^*(w) &= \alpha_H D^*Q_L^*(z_L) + (1 - \alpha_H) D^*Q_H^*(z_H) \geq 0 \end{aligned}$$

So, the Martingale convergence theorem implies that $D^*Q_j^*$ converges almost surely. Therefore, the Lagrange multipliers are uniquely pinned by the limits through (15). Lemma 16 and equation (16) imply that w converges to a point in E , hence $q^*(\theta_i|h^{t-1})$ converges to $q^e(\theta_i)$ almost surely.

□

8.6 Sufficiency conditions and global optimality

Here we study the conditions under which the solution to the relaxed problem $(\mathcal{RP}^*)$ is optimal, that is we document the validity of the first-order approach. After history h^{t-1} , the "upward" incentive constraint $IC_H(h^{t-1})$ can be expressed as

$$q^e(\theta_L) + \sum_{s=1}^{\infty} \delta^s (\alpha_L - \alpha_H)^s q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1}) \geq \sum_{s=0}^{\infty} \delta^s (\alpha_L - \alpha_H)^s q(\theta_H|h^{t-1}, \theta_H^s)$$

Moreover, the cash-strapped constraint, $C_L(h^{t-1})$ can be written as:

$$\sum_{s=0}^{\infty} \delta^s a_s q(\theta_H|h^{t-1}, \theta_H^s) \geq \sum_{s=1}^{\infty} \delta^s a_s q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$$

where $a_s = \mathbb{P}(\theta_{t+s} = \theta_L | \theta_t = \theta_L) = \frac{1}{1-\alpha_L+\alpha_H} \{\alpha_H + (1-\alpha_L)(\alpha_L - \alpha_H)^s\}$.

It is clear that $IC_H(h^{t-1})$ trivially holds if $C_L(h^{t-1})$ does not bind, because then $q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1}) = q^e(\theta_H)$ for all $s \geq 1$. In general, it is hard to check $IC_H(h^{t-1})$ as $q(\theta_H|h^{t-1}, \theta_H^s)$ might end up being greater than $q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$ for some $s \geq 1$.

Note that $IC_H(h^{t-1})$ trivially holds when $\alpha_L = \alpha_H$, and for $\alpha_H = 0$, $C_L(h^{t-1})$ implies $IC_H(h^{t-1})$; proving **Proposition 3** as a special case. Next, we state some more general sufficient conditions for $IC_H(h^{t-1})$ to valid.

Lemma 29. *The first-order approach is valid if either of the following condition holds.*

$$(\mathcal{S}_1) : \quad \alpha_H q^e(\theta_H) \left(\frac{\delta}{1-\delta} - \frac{\delta(\alpha_L - \alpha_H)}{1-\delta(\alpha_L - \alpha_H)} \right) \leq (1 - \alpha_L + \alpha_H) q^e(\theta_L)$$

$$(\mathcal{S}_2) : \quad q^e(\theta_H) \frac{1}{1-\delta(\alpha_L - \alpha_H)} \leq q^e(\theta_L)$$

Proof. To see $(\mathcal{S}_1)$, note that $a_s \propto \alpha_H + (1-\alpha_L)(\alpha_L - \alpha_H)^s = (1-\alpha_L+\alpha_H)(\alpha_L - \alpha_H)^s + \alpha_H[1 - (\alpha_L - \alpha_H)^s]$ and quantities are always distorted downward. Use $C_L(h^{t-1})$, which binds, and plug into $IC_H(h^{t-1})$. Finally, bound $q(\theta_H|h^{t-1}, \theta_L, \theta_H^{s-1})$ by $q^e(\theta_H)$ and $q(\theta_H|h^{t-1}, \theta_H^s)$ by 0, because $1 - (\alpha_L - \alpha_H)^s \geq 0$.

$(\mathcal{S}_2)$ is derived only from $IC_H(h^{t-1})$. These two conditions cover the cases when α_H is sufficiently small, provided α_L is not too big, and α_L, α_H are sufficiently close to each other and $\alpha_H - \alpha_L$ is close to one. $\square$

8.7 Dynamics of payments

The proof of **Proposition 4** is straightforward and follows from induction. To study the relationship between total economic surplus, and promised utility and utility spread, we introduce a new recursive representation. Define promised utility utility spread respectively by

$$v(\theta_j|h^{t-1}) = \alpha_j U(\theta_L|h^{t-1}, \theta_j) + (1 - \alpha_j) U(\theta_H|h^{t-1}, \theta_j) \quad (18)$$

$$U^s(\theta_j|h^{t-1}) = U(\theta_L|h^{t-1}, \theta_j) - U(\theta_H|h^{t-1}, \theta_j) \quad (19)$$

Our goal is to show how that expected surplus increases with expected utility and the utility spread.

Define $w^e = v(\theta_j|h^{t-1})$ and $w^s = U^e(\theta_j|h^{t-1})$ to be the expected utility and utility spread, respectively, that the principal is committed to deliver to the agent after history (h^{t-1}, θ_j) . Then,

★ can be restated in terms of the new state variables as:

$$\begin{aligned}
\tilde{Q}_j(w) &= \max_{(v_L, v_H, q)} \alpha_j[s(\theta_L, q_L) + \delta \tilde{Q}_L(z_L)] + (1 - \alpha_j)[s(\theta_H, q_H) + \delta \tilde{Q}_H(z_H)] \\
\text{s.t. } (z_L, z_H, q) &\in W_L^2 \times W_H^2 \times \mathbb{R}_+^2 \\
w^s &\geq \Delta q_H + \delta(\alpha_L - \alpha_H)z_H^s \\
w^e + (1 - \alpha_j)w^s &\geq \delta z_L^e \\
w^e - \alpha_j w^s &\geq \delta z_H^e
\end{aligned} \tag{20}$$

This problem is recursive, and analogous to $(\mathcal{RF})$. However, there are some key differences. The recursive domain W_j now depends on θ_j . Indeed, using transformations (18) and (19), and the previous recursive domain, one could show that $W_j = \{w \in \mathbb{R}_+^2 : w^e \geq \alpha_j w^s\}$. In addition, $\tilde{Q}_j$ is well-defined, bounded, continuously differentiable and concave as it is obtained from Q_j by the linear transformation of variables

$$\tilde{Q}_j(w) = Q_j[w^e + (1 - \alpha_j)w^s, w^e - \alpha_j w^s] \tag{21}$$

Let $(1 - \alpha_j)\beta$, $\alpha_j\rho_L$ and $(1 - \alpha_j)\rho_H$ be the Lagrange multipliers. Then the first-order and envelope conditions are given by:

$$\begin{aligned}
D_e \tilde{Q}_L(v_L) &= \rho_L \\
D_s \tilde{Q}_L(v_L) &= 0 \\
D_e \tilde{Q}_H(v_H) &= \rho_H \\
D_s \tilde{Q}_H(v_H) &= (\alpha_L - \alpha_H)\beta \\
D_q s(q_H, \theta_H) &= \Delta\beta
\end{aligned} \tag{22}$$

$$\begin{aligned}
D_e \tilde{Q}_L(w) &= \alpha_L \rho_L + (1 - \alpha_L)\rho_H \\
D_s \tilde{Q}_L(w) &= \alpha_L(1 - \alpha_L)(\rho_L - \rho_H) + (1 - \alpha_L)\beta \\
D_e \tilde{Q}_H(w) &= \alpha_H \rho_L + (1 - \alpha_H)\rho_H \\
D_s \tilde{Q}_H(w) &= \alpha_H(1 - \alpha_H)(\rho_L - \rho_H) + (1 - \alpha_H)\beta
\end{aligned} \tag{23}$$

In principle, the multipliers and the solution depend on θ_j , but we omit this dependence to ease notation. It follows from (22) and (23) that the expected surplus is non-decreasing in the expected utility (globally) and it is non-decreasing in the utility spread (for the optimal contract): $D\tilde{Q}_j(w) \geq 0$. Finally, it is straightforward to show that the ex ante value of economic surplus is increasing in the initial promised utility, v_0 . This proves **Proposition 5**.

8.8 General iid model

Suppose that $\alpha_L = 1 - \alpha_H = \mu_L$, then $\eta_L = \eta_H$ by (17) implying that the optimal contract lives on a one-dimensional curve. To characterize the optimal contract, it suffices to have only one state variable, namely expected promised utility. Notice that $Q_L^* = Q_H^*$, then for any $w \geq 0$ define Q^* by

$$Q^*(w) = \max_{z \in W} Q_j^*(z) \text{ s.t. } w = \mu_L z_L + \mu_H z_H \quad (24)$$

This definition is based on the problem $(\mathcal{RF})$, the problem $(\mathcal{RF}_0)$ is trivially modified. Importantly, that the value function in (24) solves the simpler Belman equation $(\mathcal{RF}')$, which could be used to characterize the optimal contract.

$$(\mathcal{RF}') \quad Q_j^*(w) = \max_{\langle z_L, z_H, q \rangle} \mu_L [s(\theta_L, q_L) + \delta Q_L^*(z_L)] + \mu_H [s(\theta_H, q_H) + \delta Q_H^*(z_H)]$$

subject to $\langle \mathbf{u}, \mathbf{z}, \mathbf{q} \rangle \in \mathbb{R}_+^6$, and

$$\begin{aligned} w &= \mu_L (u_L + \delta z_L) + \mu_H (u_H + \delta z_H) \\ u_L + \delta z_L &\geq \Delta q_H + u_H + \delta z_H \end{aligned}$$

The problem $(\mathcal{RF}')$ inherits many properties of the original problem and it has a simpler structure. In particular, Q^* is well-defined and unique in the space of continuous bounded functions. Let $Q^e = \mu_L Q_L^e + \mu_H Q_H^e$, then $Q^* \leq Q^e$ and $Q^* = Q^e$ if and only if $w \geq \bar{v} = \mu_L \underline{w}_L^e + \mu_H \underline{w}_H^e$. In addition, Q^* is continuously differentiable on $(0, +\infty)$ with a unbounded right derivative at 0 and strictly increasing, concave on $(0, \bar{v})$.

Consider $w \in (0, \bar{v})$. Given the shape of Q^* , it is easy to see that the constraints in the problem $(\mathcal{RF}')$ could be rewritten as $0 \leq \delta z_H = w - \mu_H \Delta q_H$ and $0 \leq \delta z_L \leq w + \mu_L \Delta q_H$. This implies that $0 < z_H < z_L \leq \bar{v}$ and there exists $v^{liq} \in (0, \bar{v})$ such that $z_L = \bar{v}$ if and only if $w \geq v^{liq}$. Finally, z_H is strictly increasing on $(0, \bar{v})$, z_L is also strictly increasing on $(0, v^{liq})$ and $0 < q_H < q^e(\theta_H)$ is strictly increasing on $(0, \bar{v})$.

References

- M. Amador, I. Werning, and G.-M. Angeletos. Commitment vs. flexibility. *Econometrica*, 74(2): 365–396, 2006.
- M. Arve and D. Martimort. Dynamic procurement under uncertainty: Optimal design and implications for incomplete contracts. *American Economic Review*, 106(11):3238–3274, 2016.
- I. Ashlagi, C. Daskalakis, and N. Haghpahan. Sequential mechanisms with ex-post participation guarantees. Stanford University and Massachusetts Institute of Technology and Pennsylvania State University, 2016.

- S. Athey and D. A. Miller. Efficiency in repeated trade with hidden valuations. *Theoretical Economics*, 2(3):299–354, 2007.
- S. Athey and I. Segal. Designing efficient mechanisms for dynamic bilateral trading games. *American Economic Review*, 97(2):131–136, 2007.
- S. Athey and I. Segal. An efficient dynamic mechanism. *Econometrica*, 81(6):2463–2485, 2013.
- A. V. Banerjee and E. Duflo. Do firms want to borrow more? testing credit constraints using a directed lending program. *The Review of Economic Studies*, 81(2):572–607, 2014.
- M. Battaglini. Long-term contracting with markovian consumers. *American Economic Review*, 95(3):637–658, 2005.
- M. Battaglini and R. Lamba. Optimal dynamic contracting: the first-order approach and beyond. Cornell University and Pennsylvania State University, 2015.
- L. M. Benveniste and J. A. Scheinkman. On the differentiability of the value function in dynamic models of economics. *Econometrica*, 47:727–732, 1979.
- D. Bergemann and A. Pavan. Introduction to symposium on dynamic contracts and mechanism design. *Journal of Economic Theory*, 159:679–701, 2015.
- D. Bergemann and M. Said. Dynamic auctions. *Wiley Encyclopedia of Operations Research and Management Science*, 2011.
- D. Bergemann and J. Välimäki. The dynamic pivot mechanism. *Econometrica*, 78(2):771–789, 2010.
- D. Besanko. Multi-period contracts between principal and agent with adverse selection. *Economics Letters*, 17(1-2):33–37, 1985.
- B. Biais, T. Mariotti, G. Plantin, and J.-C. Rochet. Dynamic security design: Convergence to continuous time and asset pricing implications. *Review of Economic Studies*, 74(2):345–390, 2007.
- B. Biais, T. Mariotti, and J.-C. Rochet. Dynamic financial contracting. In D. Acemoglu, M. Arelano, and E. Dekel, editors, *Advances in Economics and Econometrics: Tenth World Congress*, chapter 9, pages 125–171. Cambridge University Press, 2013.
- M. Campello, J. R. Graham, and C. R. Harvey. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics*, 97(3):470–487, 2010.
- G. L. Clementi and H. A. Hopenhayn. A theory of financing constraints and firm dynamics. *Quarterly Journal of Economics*, 121:229–265, 2006.
- P. Courty and H. Li. Sequential screening. *Review of Economic Studies*, 67(4):697–717, 2000.
- P. M. DeMarzo and M. J. Fishman. Optimal long-term financial contracting. *Review of Financial Studies*, 20(6):2079–2128, 2007.

- P. Eső and B. Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, 74(3):705–731, 2007.
- P. Eső and B. Szentes. Dynamic contracting: an irrelevance theorem. *Theoretical Economics*, 2015.
- E. Farhi and I. Werning. Insurance and taxation over the life cycle. *Review of Economic Studies*, 80(2):596–635, 2013.
- A. Fernandes and C. Phelan. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, 91(2):223–247, 2000.
- S. Fu and V. R. Krishna. Dynamic financial contracting with persistent private information. University of Rochester and Florida State University, 2016.
- D. Garrett and A. Pavan. Dynamic managerial compensation: A variational approach. *Journal of Economic Theory*, 159:775–818, 2015.
- E. J. Green. Lending and the smoothing of uninsurable income. In E. C. Prescott and N. Wallace, editors, *Contractual arrangements for intertemporal trade*, volume 1, chapter 1, pages 3–25. University of Minnesota Press, 1987.
- Y. Guo and J. Hörner. Dynamic mechanisms without money. Northwestern University and Yale University, 2015.
- M. Halac and P. Yared. Fiscal rules and discretion under persistent shocks. *Econometrica*, 82(5):1557–1614, 2014.
- A. İmrohoroglu and Ş. Tüzel. Firm-level productivity, risk, and return. *Management Science*, 60(8):2073–2090, 2014.
- N. Kiyotaki. A mechanism design approach to financial frictions. In F. Allen, M. Aoki, J.-P. Fitoussi, N. Kiyotaki, R. Gordon, and J. E. Stiglitz, editors, *The Global Macro Economy and Finance*, chapter 9, pages 177–187. Palgrave Macmillan UK, 2012.
- D. Krähmer and R. Strausz. Optimal sales contracts with withdrawal rights. *Review of Economic Studies*, 82(2):762–790, 2015.
- R. Lamba. Efficiency with(out) intermediation in repeated bilateral trade. Pennsylvania State University, 2016.
- E. Lipnowski and J. Ramos. Repeated delegation. University of Chicago and University of Southern California, 2015.
- V. F. Luz. Dynamic competitive insurance. University of British Columbia, 2015.
- O. Morand and K. Reffett. Lagrange multipliers in convex programs with applications to classical and nonoptimal stochastic one-sector growth models. University of Connecticut and Arizona State University, 2015.

- R. B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.
- R. B. Myerson. A model of moral-hazard credit cycles. *Journal of Political Economy*, 120(5):847–878, 2012.
- A. Pavan, I. Segal, and J. Toikka. Dynamic mechanism design: A myersonian approach. *Econometrica*, 82(2):601–653, 2014.
- J. P. Rincón-Zapatero and M. S. Santos. Differentiability of the value function without interiority assumptions. *Journal of Economic Theory*, 144(5):1948–1964, 2009.
- Y. Sannikov. A continuous-time version of the principal-agent problem. *Review of Economic Studies*, 75(3):957–984, 2008.
- Y. Sannikov. Dynamic security design and corporate financing. In G. M. Constantinides, M. Harris, and R. M. Stulz, editors, *Handbook of the Economics of Finance*, volume 2A, chapter 2, pages 71–122. Elsevier, 2013.
- A. Skrzypacz and J. Toikka. Mechanisms for repeated trade. *American Economic Journal: Microeconomics*, 7(4):252–293, 2015.
- S. E. Spear and S. Srivastava. On repeated moral hazard with discounting. *Review of Economic Studies*, 54(4):599–617, 1987.
- N. L. Stokey, R. E. Lucas Jr, and E. Prescott. *Recursive methods in economic dynamics*. Harvard University Press, 1989.
- L. Stole. Lectures on the theory of contracts and organizations. University of Chicago, 2001.
- J. Thomas and T. Worrall. Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory*, 51(2):367–390, 1990.
- R. V. Vohra. Dynamic mechanism design. *Surveys in Operations Research and Management Science*, 17(1):60–68, 2012.