WORKING PAPERS

N° 1535

May 2024

"Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information"

Gerard Maideu-Morera

Toulouse School of Economics

Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information*

Gerard Maideu-Morera †

May 2024

Abstract

I study a dynamic cash flow diversion model between a risk neutral lender and a risk averse entrepreneur who has persistent private information about the firm's productivity. In the optimal contract, the firm's size is always distorted downwards and its distortions inherit the autoregressive properties of the type process. The entrepreneur's compensation is smoothed and decoupled from the firm size dynamics. These results contrast those of equivalent models with risk neutrality. I use numerical simulations to study a quasi-implementation with simpler contracts, which highlights that this class of models is unable to generate realistic firm size and equity share dynamics simultaneously.

Keywords: Firm dynamics, financing constraints, recursive contracts, persistent private information.

JEL codes: D82, G32, L14.

*I am indebted to Christian Hellwig for his advice and guidance throughout the project. I have benefited very much from advice and discussions from Charles Brendon, Fabrice Collard, Alessandro Pavan and, especially, Nicolas Werquin. I would also like to thank George-Marios Angeletos, Matteo Broso, Eugenia Gonzalez-Aguado, Johannes Hörner, David Martimort, B. Ravikumar, Jean-Charles Rochet, Andreas Schaab, Stéphane Villeneuve and seminar participants at the TSE Macro Workshop, EEA 2022, NSEF PhD Workshop 2022, EWMES 2022, ESEM 2023, Northwestern Macro Lunch Seminar and Midwest Macro Fall 2023.

†Toulouse School of Economics. Email: gerard.maideumorera@tse-fr.eu

1 Introduction

Financing constraints slow down growth over the lifecycle of a firm. Dynamic contracting models have proved useful in understanding the underlying agency frictions that generate financing constraints and can account for several stylized facts on the lifecycle of firms. The canonical setting in this literature is the dynamic contracting model with cash flow diversion: at each period, an entrepreneur needs funds from a lender to operate a project, but only the entrepreneur observes the project's returns (i.e. cash flows) and can secretly divert them for consumption. These contracting problems have typically been analyzed assuming a risk-neutral entrepreneur and i.i.d productivity shocks (Clementi and Hopenhayn (2006), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007a), Biais *et al.* (2007)).[1] Under these assumptions, in the optimal contract, the firm size (i.e. working capital invested) drifts upwards, and the entrepreneur is compensated once the undistorted first best size is reached. Notably, the optimal contract generates a one-to-one link between firm size and compensation dynamics, and implementations typically link the entrepreneur's promised compensation to her equity share in the firm (Clementi and Hopenhayn (2006)).

In this paper, I revisit the predictions of the optimal contracting solution of dynamic cash flow diversion models. I depart from the previous literature by allowing the entrepreneur to be risk averse and to have persistent private information about the firm's productivity.[2] Risk aversion creates a consumption smoothing motive that makes it costly to backload the entrepreneur's compensation. Persistence allows the entrepreneur to have more information about the firm's future profitability, thus making her preferences for future contract arrangements depend on the current profitability.

I show that introducing risk aversion and persistent private information fundamentally changes the nature of the optimal contract and the resulting firm size and compensation dynamics. First, the firm size and compensation dynamics are decoupled. The entrepreneur is compensated for her current and past returns, but the firm size dynamics are driven by the costs of increasing the entrepreneur's ability to divert funds at a particular period or history. Hence, the dynamics of the two variables can be essentially characterized separately.

[1] A notable exception is Fu and Krishna (2019), who study a similar cash-flow diversion model with risk neutrality but with persistent shocks. However, as I will show, the role of persistence on firm dynamics depends on the entrepreneur's risk aversion.

[2] With risk neutrality, the contracting problem amounts to maximizing the value of the firm, which can be justified under complete markets. However, this assumption may be harder to justify in an entrepreneurship context as it involves significant non-diversifiable risks.

Second, the firm's size never converges to the first best and its distortions inherit the autoregressive properties of the type process. Their drift depends on the initial uncertainty about the entrepreneur's productivity. In particular, if the initial type is known – as assumed in the literature – the distortions drift upwards, so firm size tends to decrease over time. Third, the entrepreneur's compensation is smoothed intertemporally, but the variance of consumption grows over time (as in Thomas and Worrall (1990) and Atkeson and Lucas (1992)). Fourth, I show that implementing the optimal contract with risk aversion requires separately keeping track of the entrepreneur's wealth and equity share in the firm. Finally, I argue that cash flow diversion models always generate a tight link between the firm's size and the entrepreneur's equity share, posing a challenge for this class of models to generate realistic dynamics for the two variables simultaneously. In particular, they cannot account for the dilution of the entrepreneur's equity share as the firm grows.

I first derive a recursive characterization of the optimal contracting problem building on recent advances in dynamic mechanism design (Pavan *et al.* (2014), Kapička (2013), Farhi and Werning (2013)). With i.i.d shocks, the literature has characterized the optimal contract using the entrepreneur's promised continuation utility as a state variable. I can also derive a recursive representation of the problem with risk aversion and persistent private information by adding dynamic information rents as an extra state variable. These two state variables break the tight link between firm size and compensation dynamics. Moreover, risk aversion creates a consumption smoothing motive that qualitatively and quantitatively modifies the relation between promised utility and firm size. As a result, in the optimal contract, the promised utility drives the compensation dynamics, and the dynamic information rents drive the firm size dynamics with little interaction between the two.

I characterize the firm size dynamics with return-dependent investment wedges, which reduce capital below its first best level. Investment wedges are positive because more productive entrepreneurs have a relatively higher ability to divert funds as capital increases. The size of the investment wedges depends on the upper Pareto coefficient of the distribution of the marginal product of capital and a normalized shadow cost of information rents. With risk aversion and persistent private information, the lender reduces the cost of screening types at period t by promising to lower future expected information rents. However, this increases the shadow costs of information rents at $t+1$ and onwards.[3] Consequently, when the initial

[3] With persistent private information, more productive types at t prefer contracts with higher expected information rents at period $t+1$. Therefore, committing to lower information rents at $t+1$ reduces the cost of screening types at t. This is the same reason why the labor wedges tend to increase over time in dynamic

productivity of the entrepreneur is known, the investment wedges tend to increase over time and firm size tends to decrease because the firm starts operating without any promise to lower information rents. By contrast, with sufficiently high uncertainty about the initial type, the wedges decrease over time as the distortions from the initial screening problem gradually vanish. Finally, with i.i.d productivity shocks, there is no gain of promising to lower future information rents, so wedges and firm size are (approximately) stationary.

Next, I show that the entrepreneur's consumption process satisfies a Generalized Inverse Euler Equation (GIEE) similar to Hellwig (2021). With risk aversion, the lender smooths the entrepreneur's compensation intertemporally. After a history of high (low) productivity shocks, the entrepreneur is rewarded with high (low) consumption. Hence, the cross-sectional variance of the entrepreneur's consumption grows over time (Thomas and Worrall (1990), Atkeson and Lucas (1992)). However, the dispersion in compensation does not translate to firm size distortions. Numerically, I find that the investment wedges are essentially uncorrelated with compensation. For instance, after several periods, the firm's size can be distorted downwards, but the entrepreneur receives a high compensation.

To further understand the compensation dynamics, I use numerical simulations and analyze a (quasi-)implementation with simpler contracts. With i.i.d shocks, the following contract gets very close to the optimal allocation. The lender gives the entrepreneur a constant equity share in the firm's reported returns. Then, the entrepreneur can pledge her shares as collateral to borrow and smooth consumption given his implied wealth. Pledging shares is a common practice (Fabisik (2019)); this implementation shows how it can be rationalized as part of a nearly optimal contract. Moreover, the implementation is independent of dividend payout policies, which are typically used to implement the compensation with risk neutrality.

With persistent private information, we need an extra instrument to replicate the dynamic information rents. The entrepreneur's equity share is a natural candidate because it controls the sensitivity of the entrepreneur's compensation to productivity shocks. Intuitively, varying the equity share is informative for the lender because a more productive entrepreneur expects higher returns in the future and is less inclined to give up equity. Hence, the entrepreneur's equity share tracks the dynamics of the expected information rents. The equity share should be high when the expected information rents are high, and conversely. In particular, this implies that the equity share decreases over time if there is no uncertainty about the initial productivity.

Mirrlees models (see Farhi and Werning (2013) and Makris and Pavan (2020)).

The equity share dynamics help understand the distinct firm size dynamics with risk neutrality and risk aversion. Regardless of the entrepreneur's preferences, the lender provides more capital when the equity share is high because of the lower incentives to divert funds –implying that there is always a positive link between the entrepreneur's equity share and the firm's size. Hence, the opposite drift in the firm's size across different models can be understood from the opposite drift in the equity share. With risk neutrality, the equity share instead maps to the promised utility, which drifts upwards (Clementi and Hopenhayn (2006)). As a result, the equity share and the firm's size also drift upwards.

The equity share dynamics of the risk neutral model may be at odds with what we observe in the data. For example, in the venture capital industry, the founder's equity share gets diluted over the financing rounds as the firm grows (Sahlman (1990)). By introducing screening about the initial productivity, the model with risk aversion and persistence can generate more realistic firm dynamics, but this again implies an increasing equity share. Therefore, it appears to be challenging for this class of cash flow diversion models to break the embedded link between the firm's size and the entrepreneur's equity share and, consequently, generate realistic dynamics of these two variables simultaneously.

I explore three extensions to the main model: (i) limited commitment of the entrepreneur, which can generate dynamics where firm size increases over time as in Albuquerque and Hopenhayn (2004); (ii) a model where the entrepreneur can choose the fraction of funds invested and diverted, which delivers the same characterizations of the firm dynamics and the GIEE; and (iii) allowing the lender to terminate the contract, which may be optimal but does not affect the equations characterizing the optimal contract. Moreover, in a simplified version of the model, I show that if termination is optimal, termination probabilities increase with the persistence of the process. The intuition is similar to that of the equity share dynamics.

Related literature. This paper contributes to the dynamic financial contracting literature. Important early work on this class of models includes Clementi and Hopenhayn (2006), Albuquerque and Hopenhayn (2004), Biais *et al.* (2007), Biais *et al.* (2010), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007a), DeMarzo and Fishman (2007b) and DeMarzo *et al.* (2012).[4] In particular, I contribute to the literature by studying a workhorse dynamic cash flow diversion model with risk aversion and persistent private information.

[4]That firm size decreasing over time can be the outcome of an optimal contract has also been shown in Clementi *et al.* (2010). They study a dynamic moral hazard model where the firm's productivity distribution

Models with persistence have been recently analyzed in DeMarzo and Sannikov (2016), Fu and Krishna (2019) and Krasikov and Lamba (2021), but all these papers assume a risk neutral entrepreneur. To my knowledge, this is the first paper in the dynamic financial contracting literature with both persistent private information and risk aversion. I show that the entrepreneur's preferences determine the effects of persistent private information on the firm dynamics. In particular, the distortions to firm size inherit the autoregressive properties of the type process only with risk aversion. Fu and Krishna (2019) and Krasikov and Lamba (2021) show that distortions gradually vanish as with i.i.d types, but the speed of convergence to the first best decreases with the persistence of the process.

Khan *et al.* (2020) analyze the efficient allocation of capital across a continuum of risk-averse agents subject to i.i.d productivity shocks. They show that it is optimal to allocate more capital to agents with higher promised utility. Numerically, this link is weak in my model, so the dynamic information rents determine the allocation of capital. He (2012) and Di Tella and Sannikov (2021) also study dynamic contracting models with risk aversion. Both papers study a hidden savings problem, so the entrepreneur has persistent private information about his savings. I do not allow for hidden savings but allow for persistent private information about the firm's productivity.

Throughout the paper, I use tools and insights from the dynamic Mirrlees literature. I use the first-order approach (FOA) (Pavan *et al.* (2014)) and set up the principal's problem recursively using dynamic information rents as state variables as in Kapička (2013), Farhi and Werning (2013) or Golosov *et al.* (2016a). The FOA consists of solving a relaxed problem with the local incentive compatibility constraints and allows for solving the model with persistent private information.[5] Applying the FOA to a cash flow diversion model with risk aversion, an extra challenge emerges because the marginal information rents depend on consumption. So if the principal adjusts the consumption of some type θ, the slope of the profile of information rents for types $\theta' > \theta$ changes. This problem does not arise in commonly studied Dynamic Mirrlees problems with additively separable preferences between income and consumption, but it does when preferences are nonseparable. Following Hellwig (2021), I use incentive-adjusted probability measures to derive analytical characterizations of the optimal contract with risk aversion.[6] The incentive-adjusted measures reweight the

depends on the entrepreneur's costly effort exerted. In their model, the entrepreneur becomes wealthier over time, which lowers the effort and, consequently, firm size.

[5]A priori, global incentive compatibility constraints may bind. Following the procedure in Kapička (2013) and Farhi and Werning (2013), I verify ex-post that this is not the case in all the numerical simulations.

[6]The idea of using incentive-adjusted measures was developed in Hellwig (2021) to study a dynamic

density of types such that the lender's evaluation of allocations accounts for the changes in information rents, and therefore, incentive compatibility is preserved. Moreover, the finding that firm size can drift downwards follows from the insight of the Dynamic Mirrlees literature that labor wedges tend to increase over time (Farhi and Werning (2013), Makris and Pavan (2020)).

Finally, this paper is also related to the literature on insurance with persistent private information (Williams (2011), Bloedel *et al.* (2023b) and Bloedel *et al.* (2023a)). With fixed capital, the cash flow diversion model studied in this paper is equivalent to the hidden endowment model used in this literature. Their focus is on the role of persistent private information for the long-run distribution of consumption and whether or not it features immiseration (Thomas and Worrall (1990) and Atkeson and Lucas (1992)). In the paper, I present some results and discussion on the long-run consumption dynamics. Nevertheless, numerical simulations show that in this model, immiseration is a very long-run phenomenon so that it may be irrelevant for the usual lifespan of a firm.

Outline. The rest of the paper is organized as follows. Section 2 describes the model, sets up the relaxed planning problem, and describes the first best allocation. Section 3 presents the main results on the firm size and consumption dynamics, and Section 4 illustrates them with numerical simulations. Section 5 studies the quasi-implementation. Section 6 discusses the differences in models with risk neutrality and risk aversion and their implications. Finally, Section 7 briefly summarizes the extensions to the main model, and Section 8 concludes.

2 Model

Time is discrete and indexed by $t = 0, 1, ..., \infty$. Every period, an entrepreneur (the agent, "he") needs funds k_t from a lender (the principal, "she") to operate a project. Both the entrepreneur and the lender are long-lived. At period t, the project generates returns equal to $f(k_t, \theta_t)$, where $\theta_t \in [\underline{\theta}, \overline{\theta}]$ is the entrepreneur's productivity type. The agent's type history is denoted by $\theta^t = \{\theta_0, ..., \theta_t\}$ and is the agent's private information. θ_t follows a first-order Markov process with conditional density $\varphi_t(\theta_t|\theta_{t-1})$ (and CDF $\Phi_t(\theta_t|\theta_{t-1})$), and the initial type θ_0 is drawn from the density $h(\theta_0)$ (and CDF $H(\theta_0)$).

Mirrlees taxation problem with non-separable preferences between consumption, leisure, and productivity.

The lender cannot observe the returns and instead relies on the entrepreneur's report. The entrepreneur can misreport and divert a fraction of the returns for his consumption. There is a deadweight loss $(1-\iota) \in [0,1)$ on diverted funds. That is, for every dollar of funds diverted, the entrepreneur only gets to consume a fraction ι. Capital k_t fully depreciates at the end of every period. After the entrepreneur reports returns $f(k_t, \widetilde{\theta}_t)$, the lender asks for a repayment $b_t(\widetilde{\theta}_t)$ and advances funds $k_{t+1}(\widetilde{\theta}_t)$ for the next period. The entrepreneur cannot privately save, so the entrepreneur's period t consumption if the true returns are $f(k_t, \theta_t)$ but he reports $f(k_t, \widetilde{\theta}_t)$ is

$$c_t = f(k_t, \theta_t) - (1-\iota)\left(f(k_t, \theta_t) - f(k_t, \widetilde{\theta}_t)\right) - b_t(\widetilde{\theta}_t). \tag{1}$$

In particular, if the entrepreneur does not misreport returns, he consumes $c_t = f(k_t, \theta_t) - b_t(\theta_t)$. As is common, I further assume that the agent cannot overreport his returns. That is, reports are restricted to $\widetilde{\theta}_t \leq \theta_t$. This assumption is motivated by the restriction that the entrepreneur cannot save outside the contract with the lender. The entrepreneur also has limited liability, so his consumption must always be non-negative $c_t \geq 0$. The entrepreneur is risk averse, derives utility $u(c_t)$ from consumption, and discounts the future with factor $\beta \in (0,1)$. Throughout the paper, I will use the following notation for the derivatives of the production function

$$f_k(k_t, \theta_t) \equiv \frac{\partial f(k_t, \theta_t)}{\partial k_t} \quad f_\theta(k_t, \theta_t) \equiv \frac{\partial f(k_t, \theta_t)}{\partial \theta_t} \quad f_{\theta k}(k_t, \theta_t) \equiv \frac{\partial^2 f(k_t, \theta_t)}{\partial \theta_t \partial k_t}.$$

Below, I summarize all the assumptions on the functions f and u and the productivity process.

Assumptions:

A1: The utility function satisfies $u'' < 0 < u'$, and the Inada conditions $\lim_{c \to 0} u'(c) = \infty$ and $\lim_{c \to \infty} u'(c) = 0$.

A2: The production function is twice differentiable and satisfies $f_{kk} < 0 < f_k$, $f_\theta > 0$, the Inada conditions $\lim_{k \to 0} f_k(k, \theta) = \infty$ and $\lim_{k \to \infty} f_k(k, \theta) = 0$, and $f_{\theta k} > 0$.

A3: The conditional density $\varphi_t(\theta_t|\theta^{t-1})$ is differentiable with respect to the second argument and persistent, i.e.

$$\mathcal{E}(\theta_t, \theta_{t-1}) \equiv \frac{\frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}}}{\varphi_t(\theta_t|\theta_{t-1})}$$

is non-decreasing in θ_t.

Assumption (A1) implies that the agent is risk averse and the optimal allocation is generally interior. Assumption (A2) states that there is decreasing marginal product of investment, higher types obtain higher returns and have a higher marginal product. This last assumption ($f_{\theta k} > 0$) is key as it will imply that higher capital increases information rents. Finally, assumption (A3) imposes that the type process has either positive persistence or is independent over time, in which case $\frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}} = 0$. The process is allowed to be time-dependent. Differentiability will be needed to use the envelope condition for the local incentive constraint. For future use, it is useful to define:

$$\rho_t(\theta^t) \equiv \frac{1-\Phi_t(\theta_t|\theta_{t-1})}{\varphi_t(\theta_t|\theta_{t-1})}\mathbb{E}\left[\mathcal{E}(\theta', \theta_{t-1})|\theta' \geq \theta_t, \theta_{t-1}\right] = \frac{\frac{\partial}{\partial \theta_{t-1}}\left(1-\Phi_t(\theta_t|\theta_{t-1})\right)}{\varphi_t(\theta_t|\theta_{t-1})}. \tag{2}$$

This is the impulse response of θ_t to θ_{t-1} as defined in Pavan, Segal and Toikka (2014). It is a measure of the persistence of the process. If the type process follows an AR(1) with autoregressive parameter ρ, then $\mathcal{E}(\theta_t, \theta_{t-1}) = -\rho \frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_t} / \varphi_t(\theta_t|\theta_{t-1})$ and $\rho_t(\theta^t) = \rho$.

2.1 Lender's problem

The lender is risk neutral and discounts the future with factor $q \in (0,1)$. At an ex-ante stage ($t = 0$) before the firm starts operating, the lender must screen over the entrepreneur's initial type (θ_0) with continuation contracts for periods $t = 1$ and onwards. In this initial screening stage, the entrepreneur does not consume and there is no production. By the revelation principle, it is without loss to focus on direct revelation mechanisms. At any history, the entrepreneur sends a report $r \in [\underline{\theta}, \theta_t]$ about θ_t to the lender. Define a reporting strategy by $\sigma = \{\sigma_t(\theta^t)\}$, it implies a history of reports $\sigma^t(\theta^t) = \{\sigma_1(\theta_0), ..., \sigma_t(\theta^t)\}$. Let $\Sigma = \{\sigma | \sigma_t(\theta^t) \leq \theta_t \quad \forall \theta^t \in [\underline{\theta}, \overline{\theta}]^t\}$ be the set of feasible reporting strategies. For $t \geq 1$, the entrepreneur's continuation utility with truth-telling can be written recursively as

$$w_t(\theta^t) = u(c(\theta^t)) + \beta \int w_{t+1}(\theta^t, \theta_{t+1}) \varphi_{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}, \tag{3}$$

where $c(\theta^t) = f(k_t(\theta^{t-1}), \theta_t) - b_t(\theta^t)$. Similarly, the continuation utility of type θ^t with reporting strategy σ is

$$w_t^\sigma(\theta^t) = u(c(\theta_t, \sigma^t(\theta^t))) + \beta \int w_{t+1}^\sigma(\theta^t, \theta_{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}, \tag{4}$$

where

$$c(\theta_t, \sigma^t(\theta^t)) = \iota f(k_t(\sigma^{t-1}(\theta^{t-1})), \theta_t) + (1-\iota)f(k_t(\sigma^{t-1}(\theta^{t-1})), \sigma_t(\theta_t)) - b_t(\sigma^t(\theta^t)). \tag{5}$$

At $t = 0$, the continuation utility with truth-telling writes

$$w_0(\theta_0) = \beta \int w_1(\theta^1)\varphi(\theta_1|\theta_0)d\theta_1, \tag{6}$$

and with reporting strategy σ

$$w_0^\sigma(\theta_0) = \beta \int w_1^\sigma(\theta^1)\varphi(\theta_1|\theta_0)d\theta_1. \tag{7}$$

Finally, at the start of the contract, the lender must deliver a minimum level of compensation to the entrepreneur equal to

$$(1-\kappa)w_0(\theta_0) + \kappa \int w_0(\theta_0')h(\theta_0')d\theta_0' \geq v_- \tag{8}$$

for all $\theta_0 \in [\underline{\theta}, \overline{\theta}]$ and where $\kappa \in \{0, 1\}$.[7] I set this up as Makris and Pavan (2020), which allows us to consider both ex-ante and ex-post participation constraints. If $\kappa = 1$, we have an ex-ante constraint that requires the principal to deliver expected utility v_- to the agent. Conversely, if $\kappa = 0$, we have an ex-post constraint so the agent's utility must be at least v_- for all θ_0 realizations.

The lender's problem consists of choosing an allocation $\{k_{t+1}(\theta^t), b_t(\theta^t)\}$ to minimize its expected discounted cost subject to the participation, incentive compatibility, and limited

[7] The constant v_- may correspond to the entrepreneur's outside option, or it can be pinned down by a break-even condition for the lender.

liability constraints:

$$K_0(v_-) \quad = \quad \min_{\{k_{t+1}(\theta^t), b_t(\theta^t)\}} \mathbb{E}\left[k_1(\theta_0) + \sum_{t=1}^{\infty} q^t \left(k_{t+1}(\theta^t) - b_t(\theta^t)\right)\right] \tag{9}$$

$$s.t \qquad (1-\kappa)w_0(\theta_0) + \kappa\mathbb{E}[w_0(\theta)] \geq v_- \quad \forall \theta_0 \in [\underline{\theta}, \overline{\theta}] \quad (PK)$$

$$w_t(\theta^t) \geq w_t^{\sigma}(\theta^t) \quad \forall \theta^t \in [\underline{\theta}, \overline{\theta}]^t \text{ and } \sigma \in \Sigma. \quad (IC)$$

$$c_t(\theta^t) \geq 0 \quad \forall \theta^t \in [\underline{\theta}, \overline{\theta}]^t \quad (LL)$$

2.1.1 Relaxed problem at $t \geq 1$

I start by deriving a – relaxed – recursive representation of the problem for periods $t \geq 1$, and then set up the time-0 screening problem in Section 2.1.2. With Markov shocks, it is sufficient to consider only the temporary incentive compatibility constraint (Fernandes and Phelan (2000), Kapička (2013))

$$w_t(\theta^t) = \max_{r \in [\underline{\theta}, \theta_t]} u(c(\theta_t, (\theta^{t-1}, r))) + \beta \int w_{t+1}(\theta^{t-1}, r, \theta_{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}, \tag{10}$$

where type's θ_t consumption if he reports r, $c(\theta_t, (\theta^{t-1}, r))$, is given by equation (5). This allows us to solve a recursive problem. Write entrepreneur's continuation utility under truth-telling as

$$w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \tag{11}$$

$$v_t(\theta^t) = \int w_{t+1}(\theta^{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}. \tag{12}$$

Following Kapička (2013), Farhi and Werning (2013) and Pavan, Segal and Toikka (2014), I use a first-order approach. That is, I solve a relaxed problem with the local IC constraint.[8] The envelope condition of the temporary IC (10) gives

$$\frac{\partial}{\partial \theta_t} w_t(\theta^t) = \underbrace{u'(c(\theta^t))\iota f_\theta(k_t(\theta^{t-1}), \theta_t)}_{\text{Static marginal info rent}} + \underbrace{\beta \Delta_t(\theta^t)}_{\text{Dynamic marginal info rent}} \tag{13}$$

$$\Delta_t(\theta^t) = \int w_{t+1}(\theta^{t+1})\frac{\partial \varphi_{t+1}(\theta_{t+1}|\theta_t)}{\partial \theta_t}d\theta_{t+1}. \tag{14}$$

[8] Following Kapička (2013) and Farhi and Werning (2013), I verify numerically that the global IC constraints do not bind. More details can be found in Section 4 and Appendix E.

With persistent private information, the marginal information rents depend on two terms. The static component captures how much the agent can gain by marginally misreporting returns in the current period. The dynamic marginal information rent, which can be rewritten as $\Delta_t(\theta^t) = \mathbb{E}\left[\rho(\theta^{t+1})\frac{\partial w(\theta^{t+1})}{\partial \theta_{t+1}}|\theta^t\right]$, captures the rent that the agent obtains by having more information about future types than the principal. If types are i.i.d we have $\Delta_t(\theta^t) = 0$.

If the entrepreneur is risk averse, the static marginal information rent $(u'(c(\theta^t))\iota f_\theta(k_t(\theta^{t-1}), \theta_t))$ depends on the entrepreneur's consumption. Intuitively, if the entrepreneur's productivity increases by $d\theta_t$, he generates an extra return of $f_\theta(k_t(\theta^{t-1}), \theta_t)d\theta_t$. The entrepreneur can then decide to mimic the returns of the type right below him and divert the extra funds, he can then obtain $\iota f_\theta(k_t(\theta^{t-1}), \theta_t)d\theta_t$ extra consumption units. This extra information rent has to be transformed into utils by multiplying by $u'(c(\theta^t))$. The fact that information rents depend on the entrepreneur's consumption poses a challenge for characterizing the solution to this problem. If the principal increases the type's θ_t consumption, then this type's information marginal rent changes. But then the information rents of all types $\theta' > \theta_t$ must be adjusted non-linearly in order to preserve incentive compatibility. In Section 3.2, I will show how the incentive-adjusted probability measures developed in Hellwig (2021) can be used to take into account these changes in information rents.

The principal solves a dynamic programming problem where, within every period, there is an optimal control problem. I drop the limited liability constraints from the problem and verify ex-post that they do not bind.[9] For $t \geq 1$, the relaxed problem is

$$
\begin{aligned}
&K_t(v_{t-1}, \Delta_{t-1}, \theta_{t-1}, k_t) = \\
&\min_{\substack{\{k_{t+1}(\theta^t), b_t(\theta^t),\\ w_t(\theta^t), v_t(\theta^t), \Delta_t(\theta^t)\}}} \int \left[k_{t+1}(\theta^t) - b_t(\theta^t) + qK_{t+1}(v_t(\theta^t), \Delta_t(\theta^t), \theta_t, k_{t+1}(\theta^t))\right] \varphi_t(\theta_t|\theta_{t-1}) d\theta_t \\
&s.t \quad (PK) \quad w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \qquad [\varphi_t(\theta_t|\theta_{t-1})\xi_t(\theta^t)] \\
&\qquad\qquad v_{t-1} = \int w_t(\theta^t) \varphi_t(\theta_t|\theta_{t-1}) d\theta_t \qquad [\varphi_t(\theta_t|\theta_{t-1})\lambda_t] \\
&\qquad (IC) \quad \dot{w}_t(\theta^t) = u'(c(\theta^t)) \iota f_\theta(k_t, \theta_t) + \beta \Delta_t(\theta^t) \qquad [\mu_t(\theta^t)] \\
&\qquad\qquad \Delta_{t-1} = \int w_t(\theta^t) \frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}} d\theta_t \qquad [\varphi_t(\theta_t|\theta_{t-1})\gamma_t] \\
&(Feasibility) \quad c(\theta^t) = f(k_t, \theta_t) - b_t(\theta^t)
\end{aligned}
\tag{15}
$$

Note that I write the multipliers associated with each constraint inside square brackets. The Hamiltonian of this problem and the derivation of the optimality conditions can be found in Appendix B. To economize notation, I will write directly $u(\theta^t)$ and $f(\theta^t)$ instead of $u(c(\theta^t))$ and $f(k_t(\theta^{t-1}), \theta_t)$.

Along with the promised utility, v_{t-1}, the previous period's type, θ_{t-1}, and the funds advanced at $t-1$, k_t, the past dynamic information rents, Δ_{t-1}, become an extra state variable of the problem. The principal can lower current dynamic information rents by promising to reduce information rents in future periods. Intuitively, she does so by reducing the expected sensitivity of the entrepreneur's value to his productivity, i.e. by reducing his exposure to returns. Because the past promises must be satisfied, Δ_{t-1} has to be added as an extra state variable of the problem. Throughout the paper, I will refer to this state variable as the promised information rent. The co-state variable of the within period Hamiltonian is $\mu_t(\theta^t)$. This co-state variable will become key for the dynamics later. We will refer to it as the shadow cost of information rents, as it captures the principal's resource gain from reducing information rents.

[9] The allocation is generally interior because of the Inada condition on u. However, without extra assumptions on the utility function, understanding the behavior of the contract around the boundary $c_t = 0$ is more complex with persistent private information (see Bloedel *et al.* (2023b)). In any case, this is only a concern in the immiseration limit, which is not the focus of the paper, and numerically, I find that consumption is always strictly positive.

2.1.2 Time-0 problem

At $t = 0$, the principal screens over θ_0 with the continuation contracts for periods $t = 1$ and onwards. I also employ the FOA and solve the following relaxed problem[10]

$$K_0(v_-) = \min_{\substack{\{k_1(\theta_0), w_0(\theta_0), \\ v_0(\theta_0), \Delta_0(\theta_0)\}}} \int \left(k_1(\theta_0) + qK_1(v_0(\theta_0), \Delta_0(\theta_0), \theta_0, k_1(\theta_0))\right) h(\theta_0) d\theta_0$$

$$s.t \quad w_0(\theta_0) = \beta v_0(\theta_0)$$

$$v_- = (1-\kappa) w_0(\underline{\theta}) + \kappa \int w_0(\theta_0) h(\theta_0) d\theta_0$$

$$\dot{w}_0(\theta_0) = \beta \Delta_0(\theta_0)$$

When there is no uncertainty about the initial type (i.e. θ_0 is fixed), the problem can be directly solved by treating Δ_0 and k_1 as free variables and setting $v_- = v_0$: $K_0(v_-) = \min_{\Delta_0, k_1} k_1 + qK_1(v_-, \Delta_0, \theta_0, k_1)$.

2.2 First Best

To gain intuition on the model, it is useful to first look at the first best allocation, i.e. with no private information. The results are summarized in the following proposition.

Proposition 1. *In the First Best, at any history θ^t, there is*

1. *No diversion of funds*
$$f(k_t, \widetilde{\theta}_t) = f(k_t, \theta_t). \tag{16}$$

2. *No distortion of the firm's size*
$$\frac{1}{q} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1}) | \theta_t\right]. \tag{17}$$

3. *Full insurance and intertemporal consumption smoothing*
$$u'(c(\theta^t)) = \frac{\beta}{q} u'(c(\theta^{t+1})). \tag{18}$$

[10]Notice that in the ex-post participation constraint, I have used the fact that if $w_0(\underline{\theta}) = v_-$, the constraint must also hold for $\theta_0 > \underline{\theta}$ due to the incentive constraint.

By the revelation principle and because diverting funds is inefficient, there will also be no diversion of funds in the second-best. However, points 2. and 3. of the proposition do not hold in the second best allocation. In particular, firm size is distorted downwards, and the entrepreneur is exposed to risk. In the following section, we will study how, with private information, the firm size and compensation dynamics differ from the first best benchmark and the risk neutral and i.i.d cases.

3 Optimal allocation

In this section, I present the main results on the dynamics of the optimal allocation. I begin with a brief overview of the optimal contract in the benchmark with risk neutrality and discuss why risk aversion breaks the tight link between compensation and firm size dynamics (Section 3.1). As a result, the firm size and compensation dynamics can be characterized separately. I start with the firm size dynamics (Section (3.2)). First, I show that they are driven by the dynamics of the normalized shadow cost of information rents ($\widetilde{\mu}_t$). Second, I introduce the incentive-adjusted probability measures as in Hellwig (2021) to characterize $\widetilde{\mu}_t$ and its dynamics. Then, I turn to the compensation dynamics (Section (3.3)). I again use incentive-adjusted measures to characterize the entrepreneur's consumption process and discuss the implications.

3.1 Risk neutral benchmark and breaking the size-compensation link

To facilitate the comparison, I now simplify the model as in Clementi and Hopenhayn (2006), but I allow the entrepreneur to be risk averse. That is, I assume binary i.i.d shocks $\theta \in \{0, 1\}$, a return function of the form $\theta f(k)$ and that there is no deadweight loss on diverted funds. Dropping the time subscripts, the incentive constraint of the high type ($\theta = 1$) now writes

$$u(f(k) - b^H) + \beta v^H \geq u(f(k) - b^L) + \beta v^L,$$

or

$$\beta(v^H - v^L) \geq u(f(k) - b^L) - u(f(k) - b^H),$$

where superscripts H and L denote allocations for types $\theta = 1$ and $\theta = 0$, respectively. With risk-neutrality, Clementi and Hopenhayn (2006) show that the limited liability constraint binds for both types (i.e. $b^H = f(k)$ and $b^L = 0$) outside the region with no distortions. Intuitively, the principal wants as high-powered incentives as possible with risk neutrality, so the agent receives zero compensation until the first best is reached. In this case, the incentive constraint, which always binds, writes

$$u(f(k)) = \beta(v^H - v^L).$$

So, the current firm size uniquely pins down the required spread in continuation utilities. The cost of spreading continuation utilities increases with the concavity of the lender's value function. With risk neutrality, the value function is increasing and concave, and it becomes flat for a high enough v. As a result, firm size increases with v until it reaches the first best.

When the entrepreneur is risk averse, the consumption smoothing motive implies that it is never optimal (except, possibly, in the immiseration limit) to set $b^H = f(k)$ or $b^L = 0$. Hence, firm size does not directly pin down the required spread in continuation utilities because the incentive constraints also depend on the repayments b^H and b^L. Therefore, consumption smoothing breaks the tight link between the promised utility and firm size. It may still be the case that, in the optimal contract, capital increases with the promised utility as in Khan *et al.* (2020).[11] However, in the numerical simulations (Section 4), I find that firm size is approximately constant with i.i.d shocks. As I show in the following section, persistent private information generates time-varying dynamic information rents that drive the firm size dynamics with risk aversion.

3.2 Firm size dynamics

The firm dynamics implied by this cash flow diversion model with risk neutrality are well understood. On average, firm size tends to increase over time until it converges to the first best (Clementi and Hopenhayn (2006)). This is true regardless of whether the shocks are i.i.d or persistent (Fu and Krishna (2019)). The firm dynamics are remarkably different when we allow the entrepreneur to be risk averse. First, the firm's size is always below its first best

[11] Higher capital increases information rents and requires higher sensitivity of consumption. Khan *et al.* (2020) show that, with decreasing absolute risk aversion, this implies that it is optimal to allocate more capital to agents with higher promised utility.

level. With persistent private information, the firm size distortions inherit the autoregressive properties of the type process, and the drift depends on the uncertainty about the initial type. In particular, if θ_0 is fixed – as typically assumed in the literature – the firm's size tends to decrease over time.

Following the Public Finance tradition, it is helpful to describe the optimal allocation in terms of implicit wedges, i.e. distortions in the second best allocation relative to the first best. I define the lending wedges as distortion to the cost of capital faced by the lender

$$\frac{1}{q(1-\widetilde{\tau}^k(\theta^t))} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})|\theta_t\right]. \tag{19}$$

Besides the direct effect of the productivity process $\{\theta^t\}$, now the dynamics of the firm's size ($k_{t+1}(\theta^t)$) also depend on the dynamics of the lending wedge. Therefore, to characterize the firm size dynamics in the second best, it is sufficient to focus on the dynamics of the lending wedge. The following proposition shows that this wedge can be characterized with return-dependent investment wedges.

Proposition 2. *At any history θ^t, the lending wedge $\widetilde{\tau}^k(\theta^t)$ satisfies*

$$\widetilde{\tau}^k(\theta^t) = \frac{\mathbb{E}[f_k(\theta^{t+1})\tau^k(\theta^{t+1})|\theta^t]}{\mathbb{E}[f_k(\theta^{t+1})|\theta^t]} \tag{20}$$

where the return-dependent investment wedges satisfy

$$\tau^k(\theta^{t+1}) \equiv \iota\Psi^{f_k}(\theta^{t+1}) \times \widetilde{\mu}_{t+1}(\theta^{t+1}) \geq 0$$

where

$$\Psi^{f_k}(\theta^{t+1}) = \frac{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}{f_k(\theta^{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)} f_{\theta,k}(\theta^{t+1}) \geq 0.$$

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \frac{\mu_{t+1}(\theta^{t+1})}{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)} u'(\theta^{t+1}) \geq 0$$

Because $\tau^k(\theta^{t+1}) \geq 0$ we have $\widetilde{\tau}^k(\theta^t) \geq 0$, so the lending wedge lowers capital below its first best level, i.e. $k^{SB}_{t+1}(\theta^t) \leq k^{FB}_{t+1}(\theta^t)$. From now on, I will focus on the return-dependent investment wedges $\tau^k(\theta^{t+1})$. The first term of $\tau^k(\theta^{t+1})$ equals the upper Pareto coefficient of the distribution of the marginal product of capital, $\Psi^{f_k}(\theta^{t+1})$, times the ability to consume diverted funds, ι.[12] Intuitively, because $f_{\theta k} > 0$, increasing capital increases the returns

[12] The upper Pareto coefficient of the distribution of the marginal product is defined as $\Psi^{f_k}(\theta^{t+1}) \equiv$

of higher types relatively more. Therefore, their ability to divert funds increases, i.e. the higher types' information rents (in consumption units) increase by more, which is costly for the lender. Hence, $\Psi^{f_k}(\theta^{t+1})$ measures the total increase in information rents above θ_{t+1}. Finally, the cost of increasing the information rents is proportional to the normalized shadow cost $\widetilde{\mu}(\theta^{t+1})$. This term increases when the lender wants (or has promised) to lower information rents. So when $\widetilde{\mu}(\theta^{t+1})$ is high, increasing information rents is more costly.

For log-additive production functions, i.e. $f(k_{t+1}, \theta_t) = g(\theta_{t+1})\tilde{f}(k_{t+1})$, $\Psi^{f_k}(\theta^{t+1})$ is only a function of θ_{t+1}. Consequently, this term usually only depends on the (exogenous) type process, and is stationary as long as the process is also stationary. Therefore, the normalized shadow cost of information rents ($\widetilde{\mu}(\theta^{t+1})$) typically drives all the wedge dynamics.

It has been shown in dynamic screening models that with risk aversion and persistent private information, these shadow costs (and so wedges) are persistent, and - if the initial type is fixed- they tend to increase over time (Farhi and Werning (2013), Makris and Pavan (2020)). In what follows, we first characterize the shadow costs at periods $t \geq 2$ and $t = 1$ and then discuss how the dynamics depend on the time-0 uncertainty.

Characterization of $\widetilde{\mu}$. As discussed, the main challenge for characterizing the optimal allocation in this problem is that the static marginal information rents, $u'(c(\theta^t))\iota f_\theta(k_t(\theta^{t-1}), \theta_t)$, depend on consumption. This implies that any perturbation of the allocation of some type θ_t – either through a change in capital or consumption – induces a non-uniform change in the information rents of types $\theta_t' > \theta_t$. To understand this, fix a history θ^{t-1}, and consider a perturbation in the consumption of type θ_t that changes its utility by $\Delta u(\theta^t) > 0$. This changes its marginal information rent by $\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}\Delta u(\theta^t)$. For types $\theta_t' > \theta_t$, due to the change in the slope of the profile of information rents, incentive compatibility requires a change in utility equal to:

$$\Delta u(\theta^{t-1}, \theta_t') = \exp\left(\int_{\theta_t}^{\theta_t'} \frac{u''(\theta^{t-1}, \theta'')\iota f_\theta(\theta^{t-1}, \theta'')}{u'(\theta^{t-1}, \theta'')} d\theta''\right) \Delta u(\theta^t).$$

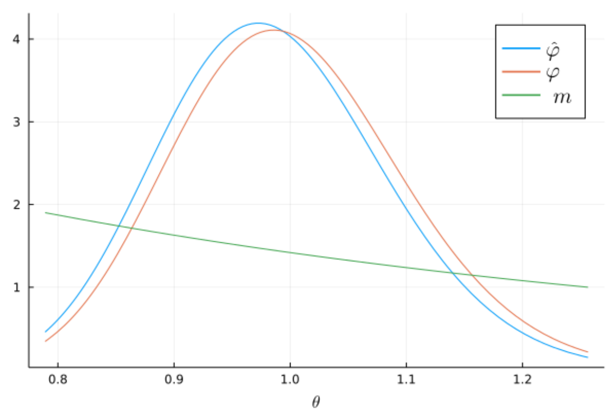
This non-separability between consumption and information rents is also a feature of dynamic Mirrlees taxation problems with arbitrary non-separable preferences between consumption,

$\frac{1-G^{f_k}(f_k(\theta^{t+1})|\theta^t)}{f_k(\theta^{t+1})g^{f_k}(f_k(\theta^{t+1})|\theta^t)}$, where g^{f_k} (G^{f_k}) are the density (CDF) of the distribution of the marginal product of capital. Then, using $G^{f_k}(f_k(\theta^{t+1})|\theta^t) = \Phi(\theta_t|\theta_{t-1})$ and $g^{f_k}(f_k(\theta^{t+1})|\theta^t) = (1 - \varphi(\theta_t|\theta_{t-1}))f_{\theta,k}(\theta^{t+1})$ we get the equation in the proposition.

income, and productivity $U(c, y, \theta)$ (see Hellwig (2021)). Hellwig (2021) shows that accounting for this non-separability amounts to evaluating the changes in utility of each type θ_t according to $m(\theta^t) \equiv e^{-\int_{\theta_t}^{\overline{\theta}} \frac{u''(\theta^{t-1},\theta')\iota f_\theta(\theta^{t-1},\theta')}{u'(\theta^{t-1},\theta')} d\theta'}$ (and the changes in consumption according to $M(\theta^t) \equiv \frac{1}{u'(\theta^t)} m(\theta^t)$). Crucially, the factor $m(\theta^t)$ can be interpreted as a reweighting of the type distribution. Accordingly, we define the incentive-adjusted probability measures as

$$\hat{\varphi}_t(\theta_t|\theta^{t-1}) \equiv \frac{\varphi_t(\theta_t|\theta_{t-1})m(\theta^t)}{\mathbb{E}[m(\theta^t)|\theta_{t-1}]}. \tag{21}$$

Figure 1: Incentive-adjusted probability measure



Note: The plot is computed with the same calibration as the main simulations in Section 4 for i.i.d types. Observe that m is monotonically decreasing, and the incentive-adjusted measure $\hat{\varphi}$ puts more weight on the lower type realizations.

Therefore, the new measure $\hat{\varphi}$ reweights the density of types such that these perturbations preserve incentive compatibility. Because $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} < 0$, the function $m(\theta^t)$ is decreasing in θ_t. So, $\Phi_t(\cdot|\theta_{t-1})$ first-order stochastically dominates $\hat{\Phi}_t(\cdot|\theta^{t-1})$. That is, incentive compatibility requires evaluating allocations as if the principal puts more weight on lower types, see Figure 1. Intuitively, because lower types have a higher marginal utility, their information rents are more sensitive to changes in consumption. Therefore, the incentive-adjusted measure that guarantees incentive compatibility has to put more weight on lower types. The following proposition uses the incentive-adjusted measure to characterize $\widetilde{\mu}_t$ for periods $t \geq 2$.

Proposition 3. *(Hellwig (2021)) The normalized shadow cost of information rents* $\widetilde{\mu}_{t+1}(\theta^{t+1})$

satisfies, for $t \geq 1$,

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \hat{MB}(\theta^{t+1}) + \hat{\rho}(\theta^{t+1})\frac{\beta}{q}\frac{1-\Phi_t(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi_t(\theta_t|\theta_{t-1})}\widetilde{\mu}_t(\theta^t), \tag{22}$$

with

$$\hat{MB}(\theta^{t+1}) = \frac{\mathbb{E}\left(m(\theta^t,\theta')|\theta' \geq \theta_{t+1},\theta_t\right)}{M(\theta^{t+1})}\left\{\hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta',\theta^t\right)} \mid \theta' \geq \theta_{t+1},\theta^t\right] - \hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta^{t+1}\right)}|\theta^t\right]\right\} \geq 0 \tag{23}$$

$$\hat{\rho}(\theta^{t+1}) \equiv \frac{\mathbb{E}\left(m(\theta^t,\theta')|\theta' \geq \theta_{t+1},\theta_t\right)}{M(\theta^{t+1})}\left\{\hat{\mathbb{E}}\left[\mathcal{E}(\theta',\theta^t) \mid \theta' \geq \theta_{t+1},\theta^t\right] - \hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta^t)|\theta^t\right]\right\} \geq 0. \tag{24}$$

Note that the operator $\hat{\mathbb{E}}$ denotes expectations under the measure $\hat{\varphi}$. The proposition shows that the normalized shadow cost of information rents is a function of two terms. The first, $\hat{MB}(\theta^{t+1})$, is the current marginal benefit of redistributing consumption from types $\theta' > \theta_{t+1}$ to $\theta'' < \theta_{t+1}$ while preserving the promise-keeping constraint.[13] The incentive-adjusted measure accounts for the non-uniform utility adjustments resulting from the changes in marginal information rents. Because lower types have higher marginal utility, this perturbation carries a cost reduction for the lender, implying that this term is always (weakly) positive. The second is a backward-looking term that accounts for how changes in information rents at $t+1$ affect information rents at t. Intuitively, it measures the pass-through of information rents from periods $t+1$ to t (captured by $\hat{\rho}(\theta^{t+1})$), times the cost of the resulting change in information rents at t. Hence, these shadow costs inherit the autoregressive properties of the type process.

Proposition 3 shows that the shadow costs (and wedges) inherit the autoregressive properties of the type process. The time-0 screening problem determines the starting value of this process, i.e. $\widetilde{\mu}_1(\theta^1)$. The following proposition shows that, at $t = 1$, a similar backward-looking term accounts for the promises to lower information rents in the time-0 screening problem.

[13] The literature on dynamic mechanism design with risk aversion typically analyses hidden effort models with separable preferences of the form $U(\theta, e, c) = u(c) - \psi(e, \theta)$, where e is the (unobservable) agent's effort. This includes dynamic taxation models with separable preferences, but also models of managerial compensation, among others. With these preferences, the static marginal information rents $\psi_\theta(e, \theta)$ are independent of consumption. So, if the principal increases the utility of type θ_t, it is sufficient to increase utility uniformly to all types $\theta'_t > \theta_t$ to preserve incentive compatibility, so consumption has to be redistributed in proportion to $\frac{1}{u'(\theta')}$. As a result, in these settings, one can derive the same characterization but under the original measure φ and with $m(\theta^t) = 1$ (Makris and Pavan (2020), Brendon (2013), Hellwig (2021)).

Proposition 4. *At $t = 1$, the normalized shadow cost of information rents satisfies*

$$\widetilde{\mu}_1(\theta^1) = \hat{MB}(\theta^1) + \hat{\rho}(\theta^1) MB_0(\theta^0) \tag{25}$$

where $\hat{MB}(\theta^1)$ and $\hat{\rho}(\theta^1)$ are given by equations (23) and (24), respectively, and $MB_0(\theta^0)$ solves

$$MB_0(\theta^0) = \frac{1 - H(\theta_0)}{h(\theta_0)} \left\{ \mathbb{E}_h \left[\lambda_1(\theta_0') | \theta_0' > \theta_0 \right] - \kappa \mathbb{E}_h \left[\lambda_1(\theta_0) \right] \right\}, \tag{26}$$

where

$$\lambda_1(\theta_0) = \hat{\mathbb{E}} \left[\frac{1}{u'(\theta^1)} | \theta_0 \right] + MB_0(\theta^0) \hat{\mathbb{E}} \left[\mathcal{E}(\theta_1, \theta_0) | \theta_0 \right]. \tag{27}$$

In particular, if the initial type θ_0 is fixed, $MB_0(\theta^0) = 0$.

When $\kappa = 1$, the backward-looking term in $\widetilde{\mu}_1$ measures the marginal benefit of redistributing expected continuation utilities $v_0(\theta_0)$ from types $\theta' > \theta_0$ to $\theta'' < \theta_0$. With the ex-post constraint ($\kappa = 0$), this is simply the marginal benefit of lowering the continuation utilities of types $\theta' > \theta_0$. In both cases, this term is zero when θ_0 is fixed, so there is no backward-looking term in $\widetilde{\mu}_1$. More generally, as I find numerically, this term should typically increase with the variance of θ_0. Moreover, because $\mathbb{E}_h \left[\lambda_1(\theta_0) \right] > 0$, this marginal benefit will usually be much larger with the ex-post constraint, especially if the constraint is tight so that the multipliers $\lambda_1(\theta_0)$ are large.[14]

Dynamics of $\widetilde{\mu}$. Proposition 3 shows that the shadow costs $\widetilde{\mu}_{t+1}$ are persistent. Iterating backward on equation (22) and using (25) we get

$$\begin{aligned} \widetilde{\mu}_{t+1}(\theta^{t+1}) = & \sum_{\tau=0}^{t} \left(\frac{\beta}{q} \right)^{\tau} \prod_{s=0}^{\tau-1} \left(\hat{\rho}_{t+1-s}(\theta^{t+1-s}) \frac{1 - \Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s}) \varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})} \right) \hat{MB}(\theta^{t+1-\tau}) \\ & + \left(\frac{\beta}{q} \right)^{t} \prod_{s=0}^{t-1} \left(\hat{\rho}_{t+1-s}(\theta^{t+1-s}) \frac{1 - \Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s}) \varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})} \right) \hat{\rho}(\theta^1) MB_0(\theta^0) \end{aligned} \tag{28}$$

The formula shows that the shadow costs of information rents are a function of current and past marginal benefits of redistribution $\{\hat{MB}(\theta^{t+1-\tau})\}$ and $MB_0(\theta^0)$. In particular, because the passthrough terms $\hat{\rho}_{t+1-s}(\theta^{t+1-s}) \frac{1-\Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s})\varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}$ are always positive, the shadow

[14] Notice also that $MB_0(\theta_0)$ is typically inverse U-shaped in θ_0 with $\kappa = 1$ but decreasing with $\kappa = 0$. Thus, the firm size distortions will be the largest for the intermediate (lowest) types with ex-ante (ex-post) constraints.

costs are always increasing in the past marginal benefits. Moreover, the drift will depend on whether the time-0 marginal benefits $MB_0(\theta^0)$ are larger than the marginal benefits in subsequent periods $\{\hat{MB}(\theta^{t+1-\tau})\}$.

Consider first the case where θ_0 is fixed so that $MB_0(\theta_0) = 0$. Then, $\widetilde{\mu}_{t+1}(\theta^{t+1})$ and $\tau^k(\theta^{t+1})$ will tend to grow with the distance from the starting period. The intuition is the following. With persistent private information, different types θ_t have different preferences for period $t+1$ contracts. In particular, higher types value relatively less contracts with low information rents at $t+1$, as they know they are expected to be more productive then and so collect higher information rents. The principal can use this to lower the resource cost of screening types at every period. More concretely, if the principal promises to lower the future information rents to type (θ^{t-1}, θ') (i.e lowers $\Delta_t(\theta^{t-1}, \theta')$) this relaxes the incentive constraints of types (θ^{t-1}, θ') with $\theta'' > \theta'$. Because every period the principal can gain by promising to lower future information rents, the shadow costs $\widetilde{\mu}_t$ will tend to increase over time. More concretely, the gain of relaxing type (θ^{t-1}, θ')'s incentives constraint equals the marginal benefit of redistributing consumption around him. Hence, the increases in $\widetilde{\mu}_{t+1}$ are proportional to the intertemporal passthroughs of information rents times the marginal benefits $\{\hat{MB}(\theta^{t+1-\tau})\}$. However, as will be shown in the numerical simulations, the wedges may, over time, converge to a stationary distribution. I use this intuition to explain why the lender may want to use equity purchases in the implementation (see Section 5.2).[15] It is important to stress that, in this case, for every type θ_t firm size $(k_{t+1}(\theta_t))$ is never larger than in the initial period. The reason is that the principal initializes the contract by setting Δ_0 freely. So Δ_0 is set to not have any "promises" to lower dynamic information rents. Consequently, for every $\theta_t \in [\underline{\theta}, \overline{\theta}]$, the wedges will not be smaller than in the initial period.

In the time-0 problem, by the same logic, the lender also lowers dynamic information rents $(\Delta_0(\theta_0))$ to reduce the cost of screening over θ_0. Hence, with initial uncertainty, the shadow costs at $t = 1$ can already be high, which can make wedges decrease over time. Consider the extreme case where all the uncertainty about the entrepreneur's productivity is realized in the initial period, i.e. $\theta_t = \theta_0$ for all $t \geq 1$. Then, the marginal benefits of redistribution would be zero for all periods following $t = 0$, i.e. $\hat{MB}(\theta^t) = 0$ for all $t \geq 1$. If the passthrough term

[15] Alternatively, imagine the principal increases consumption of all types $(\theta^{t-1}, \widetilde{\theta}_t)$ with $\widetilde{\theta}_t > \theta_t$. To preserve incentive compatibility, the principal needs to adjust the information rent of all types $(\theta^{t-2}, \theta^{'}_{t-1})$ with $\theta^{'}_{t-1} > \theta_{t-1}$. Because if types are persistent (i.e $\rho_t(\theta^t) > 0$), types $\theta^{'}_{t-1} > \theta_{t-1}$ have a higher probability of being type $\widetilde{\theta}_t$ at period t. This adjustment has to be done for all types $(\theta^{\tau-1}, \theta^{'}_{\tau-1})$ with $\theta^{'}_{\tau-1} > \theta_{\tau-1}$ at all periods $\tau < t$. Therefore, these costs will tend to increase over time if types are persistent. For a clearer and more detailed intuition on this, see Makris and Pavan (2020).

times $\frac{\beta}{q}$ is smaller than one – as will typically be the case with a mean-reverting process– the effect of the time-0 screening problem on the shadow costs will gradually vanish and the investment wedges go to zero. More generally, the drift in the wedges will depend on the relative magnitudes of $MB_0(\theta^0)$ and $\{\hat{MB}(\theta^t)\}$. With sufficiently high variance in θ_0, $MB_0(\theta^0)$ can be high enough such that the shadow cost and investment wedges decrease over time. As I will show numerically, if the initial variance is not as high, the investment wedges can increase during a few initial periods and then gradually decrease.

Moreover, both risk aversion and persistence are necessary to have these investment wedge dynamics. If the agent is risk neutral we have $\hat{MB}(\theta^{t+1}) = 0$, which implies $\widetilde{\mu}_{t+1}(\theta^{t+1}) = 0$. If the type process is not persistent we have $\hat{\rho}_t(\theta^t) = \rho_t(\theta^t) = 0$ and

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \hat{MB}(\theta^{t+1}) \tag{29}$$

so past marginal benefits of lowering information rents do not affect the current shadow costs. However, it is still the case that wedges are always positive and so firm size is below the first best. As I will show in the numerical simulations, wedges are approximately stationary with i.i.d types.

The persistence of the wedges can be amplified or dampened with the incentive-adjusted measure relative to the impulse responses under the original type measure $\rho_t(\theta^t)$. The (unnormalized) persistence is

$$\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) \gtreqless \rho_t(\theta^t)$$

if $\rho_t(\theta, \theta^{t-1})\frac{u''(\theta,\theta^{t-1})f_\theta(\theta,\theta^{t-1})}{u'(\theta,\theta^{t-1})}$ is increasing/constant/decreasing in θ (see proposition 3 in Hellwig (2021)). This condition depends, in particular, on the properties of the utility function. For instance, assume that the type process is (log) AR(1) with autoregressive parameter ρ (i.e $\frac{\partial\varphi_t(\theta_t|\theta_{t-1})}{\partial\theta_{t-1}} = -\rho\frac{\partial\varphi_t(\theta_t|\theta_{t-1})}{\partial\theta_t}$ and $\rho_t(\theta^t) = \rho$) and that the production function is linear in the type (i.e $f_{\theta\theta} = 0$). Then, the persistence of the wedges is amplified, i.e. $\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) > \rho$, if the utility features decreasing absolute risk aversion (DARA), but $\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) = \rho$ with CARA utility.[16]

In the data, we consistently observe a strong lifecycle component in firm dynamics (Evans (1987)). Young firms are usually small and face strong financing constraints. Over time,

[16]Note CRRA utility functions belong to the DARA class.

the firm size tends to increase, and financing constraints are relaxed. A cash flow diversion model with a risk neutral agent and limited liability (Clementi and Hopenhayn (2006), Fu and Krishna (2019)) can qualitatively replicate the dynamics observed in the data. However, this is no longer the case once we introduce risk aversion and persistent private information in the benchmark without time-0 uncertainty. The opposite dynamics emerge: the firm size tends to decrease over time, and it never reaches the first best. In Section 6, I discuss in more detail why models with risk neutrality generate different firm dynamics using intuitions from the implementation. Introducing uncertainty about the starting productivity can solve this and generate dynamics where firm size increases over time.

3.3 Compensation dynamics

We now turn to the compensation dynamics. As in all dynamic insurance models, at the optimum, the principal equalizes the cost of increasing the agent's utility at periods t and $t+1$ in an incentive-compatible manner, i.e.

$$\lambda_{t+1}(\theta^t) = \frac{\beta}{q}\xi(\theta^t), \tag{30}$$

where $\lambda_{t+1}(\theta^t)$ is the multiplier on the period $t+1$ promise keeping constraint and $\xi(\theta^t)$ is the multiplier on the type's θ^t period t continuation utility constraint. Again, when the principal promises to increase utilities at period $t+1$, this changes all marginal information rents (as they depend on consumption). So, utility has to be distributed non-uniformly to preserve incentive compatibility. Hellwig (2021) shows how incentive-adjusted measures can be used to derive a Generalized Inverse Euler Equation (GIEE). I derive a similar characterization in this model.

Proposition 5. *In the optimal allocation, the following Generalized Inverse Euler Equation holds at any history θ^t*

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{1}{u'(\theta^t)}(1+s(\theta^t)) \tag{31}$$

where

$$s(\theta^t) = \left(\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} - \hat{\mathbb{E}}\left[\rho_{t+1}(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right]\right)\frac{f_k(\theta^t)}{f_{k\theta}(\theta^t)}\tau^k(\theta^t). \tag{32}$$

The GIEE provides an intuitive representation that clarifies what effects drive consumption dynamics and allows to perform direct comparative statics. As in the standard Inverse Euler Equation, the costs of increasing utility at period t and $t+1$ are proportional to $\frac{1}{u'(\theta^t)}$ and $\frac{1}{u'(\theta^{t+1})}$.[17] However, expectations are taken with respect to the incentive-adjusted probability measure because utility at $t+1$ has to be redistributed non-uniformly to preserve incentive compatibility. Moreover, an extra wedge emerges that captures how savings decisions affect marginal information rents at periods t and $t+1$. Changes in marginal information rents at $t+1$ lower information rents in period t at rate $\rho_{t+1}(\theta^{t+1})$. Therefore, the size and sign of the savings wedge depend on the persistence of the process. Intuitively, when the persistence is higher, increasing consumption at $t+1$ lowers the cost of incentive provision at t by more, and so the principal wants relatively higher savings. In general, if persistence (i.e. $\rho_{t+1}(\theta^{t+1})$) is not too high, we will have $s(\theta^t) < 0$. The savings wedge is then scaled by the cost of information rents at period t.

The savings wedge takes a particularly simple form with CARA utility $u(c) = -e^{-\sigma c}$ with $\sigma > 0$. Assume also an autoregressive process $\rho_t(\theta^t) = \rho$ and $f(k,\theta) = \theta \tilde{f}(k)$, then

$$\begin{aligned} s(\theta^t) &= -\sigma\iota\theta_t \times \tau^k(\theta^t) \times \left(\tilde{f}(k_t) - \rho\tilde{f}(k_{t+1}(\theta^t))\right) \\ &= -\sigma\iota \times \tau^k(\theta^t) \times \left(f(k_t,\theta_t) - \mathbb{E}\left[f(k_{t+1}(\theta^t),\theta_{t+1})|\theta_t\right]\right) \end{aligned}$$

Because $-\sigma\iota\theta_t \times \tau^k(\theta^t) \leq 0$, we have $s(\theta^t) \leq 0$ if $\rho \leq \frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$. Moreover, savings are, on the margin, more discouraged when the agent is more risk-averse (higher σ), the costs of diverting funds are small (high ι), and the costs of incentive provision are high (high $\tau^k(\theta^t)$). With fixed capital ($k_t = k$) and $\iota = 1$, this model nests a hidden endowment model.[18] In this case, $s(\theta^t) = 0$ if $\rho = 1$ and we can use the following result.

Proposition 6. *Assume* $\frac{q}{\beta} \leq 1$ *and* $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$*, if* $s(\theta^t) \geq 0$ *marginal utility follows a super-martingale*

$$u'(\theta^t) \geq \mathbb{E}\left[u'(\theta^{t+1})|\theta_t\right].$$ [19]

[17] As in the characterization of $\widetilde{\mu}$, in hidden effort models with separable preferences it is sufficient to increase utility uniformly across all realizations of θ_{t+1} to preserve incentive compatibility. So, equation (30) leads to the well know Inverse Euler Equation $\frac{1}{u'(c(\theta^t))} = \frac{\beta}{q}\mathbb{E}\left[\frac{1}{u'(c(\theta^{t+1}))}|\theta_t\right]$. One cannot derive this tight characterization in all other settings studied in the literature: this includes models with taste shocks (as in Atkeson and Lucas (1992)), hidden endowment (as in Thomas and Worrall (1990)), Mirrlees with non-separable preferences, and also this model. For this reason, results on the agent's consumption process are usually derived from the principal's marginal cost martingale (see Golosov *et al.* (2016b)).

[18] With CARA utility, it is also equivalent to a taste shocks model as in Atkeson and Lucas (1992).

[19] With i.i.d shocks, it is easy to verify that the inequality $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$ holds. However, with

Moreover, if $s(\theta^t) \geq 0$ *for all* θ^t, $u' \to 0$ *almost surely.*

The proposition shows that the marginal utility dynamics are preserved under the original measure when $s(\theta^t) \geq 0$. Therefore, in a hidden endowment model with a unit root process ($\rho = 1$) there is no immiseration (Thomas and Worrall (1990) and Atkeson and Lucas (1992)), and the contract sends the agent to bliss, which is consistent with the results in Bloedel *et al.* (2023b) and Bloedel *et al.* (2023a).[20] When $s(\theta^t) < 0$, we do not have direct implications for the dynamics under the original measure. The numerical simulations indicate, as expected, that consumption converges to zero, and so there is immiseration. However, the convergence is very slow, so these results may be irrelevant for the usual lifespan of a firm.

Compared to a hidden endowment model, time-varying capital generates an extra motive to increase the variance in compensation over time. For the parametric specification above, as long as $\frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$ is decreasing in θ_t, given some high enough ρ, there can exist a $\widetilde{\theta}_t$ such that $s(\theta^t) < 0$ if $\theta_t \leq \widetilde{\theta}_t$ and $s(\theta^t) \geq 0$ otherwise.[21] So, savings are on the margin more discouraged for lower types. Intuitively, because $f_{\theta k} > 0$, higher capital increases information rents. If lower types will have less capital at $t+1$, their incentive constraints will be less tight. Hence, the benefit of increasing consumption at $t+1$ to lower information rents is smaller for lower types.

In sum, the lender minimizes the cost of compensating the agent across periods in an incentive-compatible manner. For this reason, it is optimal to smooth the entrepreneur's compensation over time. Moreover, because the entrepreneur always needs to be compensated for reporting high returns, the cross-sectional variance of consumption grows over time.

---

persistent private information, the principal can provide incentive by lowering the dynamic information rents, and, a priori, this inequality may not hold (see Bloedel *et al.* (2023b)). Hence, I include it as an assumption in the proposition and verify that it holds in all numerical simulations.

[20] Bloedel *et al.* (2023b) and Bloedel *et al.* (2023a) have corrected the findings in Williams (2011) and shown (with more general utility functions and processes) that there is immiseration whenever there is some mean-reversion in the type process.

[21] The condition that $\frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$ is decreasing in θ_t would not be satisfied if, for some types $\theta'_t > \theta''_t$, the effect of higher wedges at $t+1$ for type θ'_t is stronger than from the higher expected productivity. The numerical simulations verify that $k_{t+1}(\theta^t)$ is indeed increasing in θ_t, see Figure 7 in Appendix A.

4 Numerical simulations

In this section, I numerically solve and simulate the model. This will help us better understand the results in the previous section and allow us to quantify the effect of persistent private information on firm size and compensation dynamics. The numerical simulations will also be used to guide the implementation in Section 5.

I assume the agent has CRRA utility

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

and the production function is given by

$$f(k, \theta) = z\theta k^{\alpha},$$

where $\alpha \in (0, 1)$ and z is a positive constant used to scale up the problem. The agent's productivity follows a geometric AR(1) process

$$\theta_t = \theta_{t-1}^{\rho} \varepsilon_t,$$

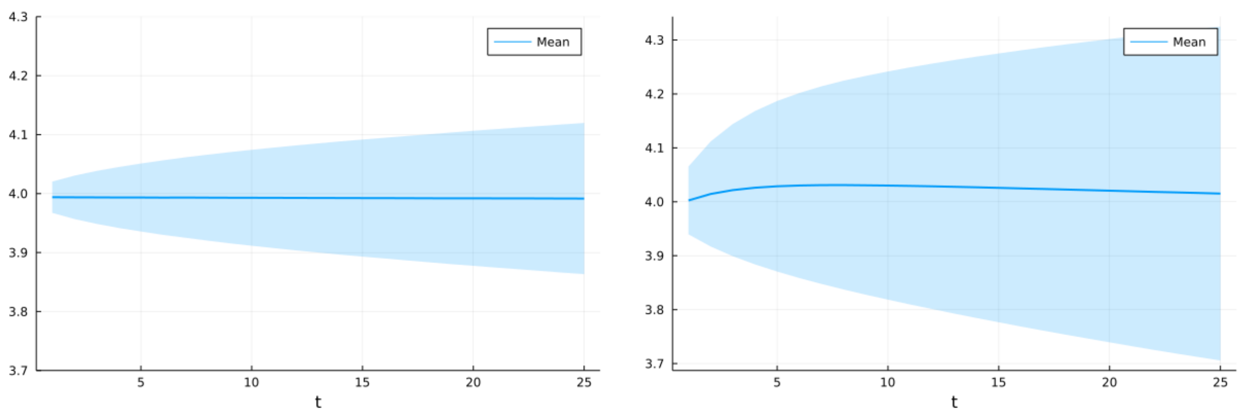
where $\log(\varepsilon_t) \sim N(\mu, \sigma_\varepsilon^2)$. I set $\alpha = 3/4$, $1 - \iota = 0.05$, $\sigma = 2$ and assume the lender and the entrepreneur have the same discount rate $\beta = q = 0.95$. For the productivity process, I set $\mu = 1$ and $\sigma_\varepsilon^2 = 0.01$. The comparative statics of this section focus on the effect of the persistence ρ. The model is solved with $\rho = 0$ (i.i.d types) and $\rho = 0.7$. I also solve the model with different parametrizations of the utility function (log utility ($\sigma = 1$) and CARA), qualitatively, the results are the same (see Appendix A). Details on the solution method, algorithm and the procedure to check global incentive compatibility can be found in Appendix E. After solving for the value functions (K, v and Δ), the policy functions (c_t, λ_{t+1}, γ_{t+1} and k_{t+1}), and the costate (μ_t), I run a Monte Carlo simulation with 10^6 draws over 25 periods each. I start with the benchmark case where θ_0 is known, and at the end of the section, introduce the screening over θ_0.

Consumption dynamics. Figure 2 illustrates the evolution of the mean and standard deviation of consumption along the cross-section over time with $\rho = 0$ and $\rho = 0.7$. As expected, the variance of consumption is permanently increasing in both cases. With i.i.d types, average consumption is approximately constant. With persistence, there is also a slight increase in average consumption in the initial periods. Since the savings wedge $s(\theta^t)$ is proportional to the investment wedge $\tau^k(\theta^t)$, this is consistent with the initial increase in the investment wedge that we will observe (see Figure 4). Moreover, because the agent is risk averse, the average marginal utility tends to increase over time.

To visualize the immiseration dynamics, in Figure 6 in Appendix A I plot the median and quantiles of the distribution of consumption over a long time horizon. The median consumption monotonically decreases, indicating that consumption will converge to its lower bound. However, the decrease is very slow, so it may be irrelevant for the usual lifespan of a firm.

Figure 2: Consumption dynamics

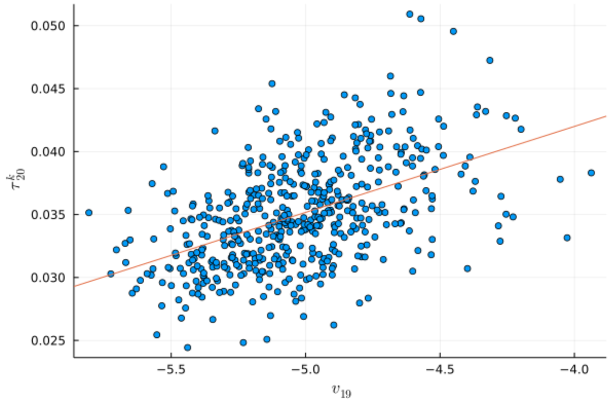
(a) i.i.d ($\rho = 0$) (b) Persistence ($\rho = 0.7$)



Note: For both panels, I initialize the simulation by setting $\lambda_0 = 15.9$. The initial types, θ_1, are drawn from the ergodic distribution of each process. For each period, the blue line is the mean consumption along the cross-section, and the shaded blue area is one standard deviation.

Separating compensation and firm size distortions. The separation between compensation and firm size (or wedge) dynamics can be illustrated very clearly with the numerical simulations. Figure 3 shows the relation between the promised utility and the investment wedge at age 20. There appears to be some positive association between the two variables, but they are not linked one to one. We can observe that there is some probability that at age 20, the entrepreneur receives a high compensation (high v_t) but that the firm is financially constrained (high τ^k). The converse is also possible: the compensation is low, but the financing constraints are also low. As discussed, this is not the case in a model with risk neutrality (Clementi and Hopenhayn (2006)), where the promised utility is linked one-to-one with the distortions to firm size.

Figure 3: Investment wedge and promised utility at $t = 20$ ($\rho = 0.7$)



Note: Each dot is a random realization of the investment wedge and promised utility at period 20. The red line is a linear regression line on the 500 draws ploted.

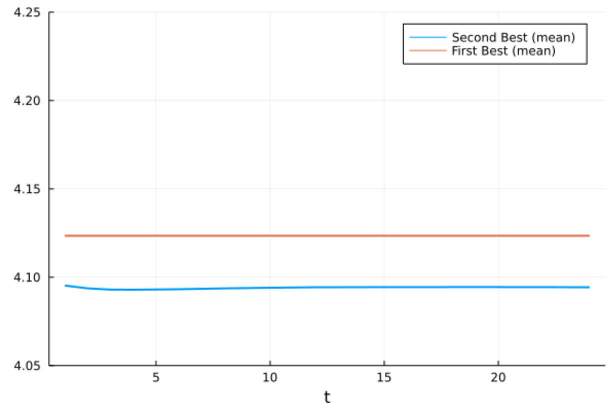
Firm and wedge dynamics without time-0 screening. Figure 4 plots the firm size and investment wedge dynamics. In both cases, the firm size closely follows the dynamics of the investment wedge. With i.i.d shocks, the wedges are approximately stationary, so firm size is constant (Panels 4a and 4c).[22] Hence, the firm size dynamics are essentially independent of the compensation dynamics. Intuitively, the lender compensates the entrepreneur by permanently increasing his consumption, not by lending more capital to the firm. Moreover, the wedges are small, so firm size is also very close to – but always below – the first best level.

For the persistent case, at the first best, the variation in firm size is driven only by differences in expected returns. Moreover, because the type process is mean-reverting, firm size is stationary. At the second best, on average, the wedges tend to increase over time and firm size tends to decrease (Panels 4b and 4d). However, the wedges do not increase indefinitely. Over time, they converge to a stationary distribution, and so does firm size. With log utility (lower risk aversion), the wedges and the decrease in firm size are smaller (see Figure 12 in Appendix A). Overall, the decrease in firm size will be larger the higher the risk aversion and persistence.

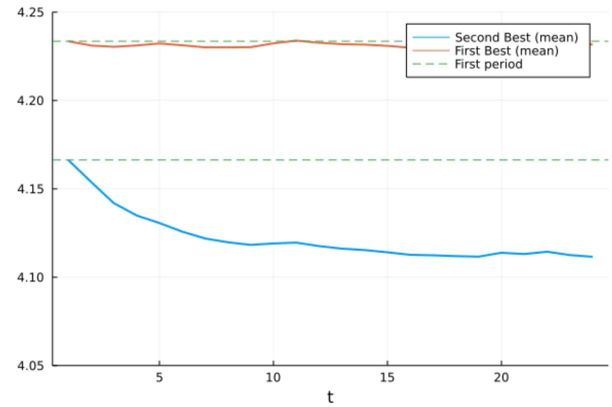
[22]This is also the case with the other parametrizations of the utility function.

Figure 4: Firm size and investment wedge dynamics (θ_0 fixed)

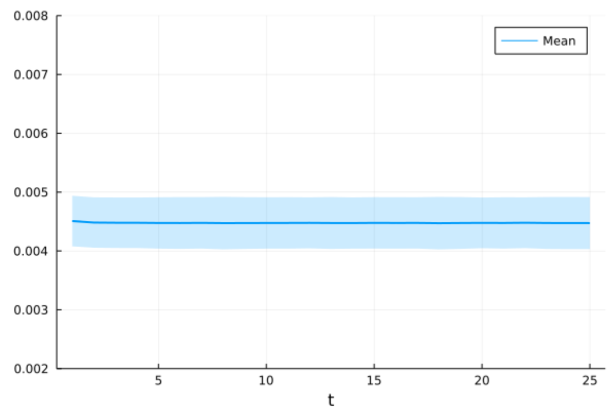
(a) Size (k) i.i.d ($\rho = 0$)



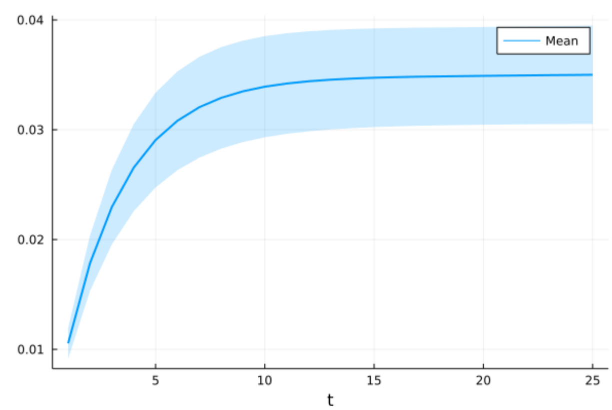
(b) Size (k) with persistence ($\rho = 0.7$)



(c) Investment wedge (τ^k) i.i.d ($\rho = 0$)



(d) Investment wedge (τ^k) with persistence ($\rho = 0.7$)

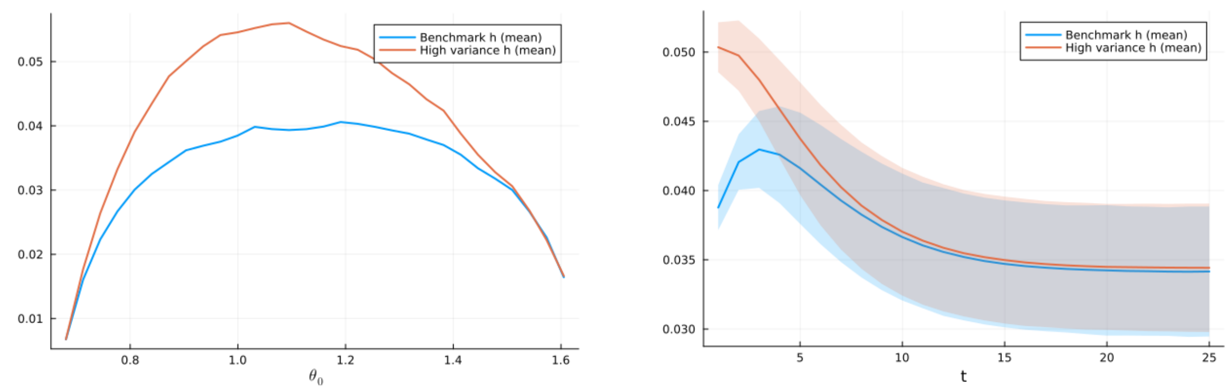


Note: Panel (a): The red line is the size at the first best (constant). The blue line is the average size at the second best; it is the same for almost all realizations as expected wedges are approximately constant.
Panel (b): The red (blue) line is the average size in the first (second) best. The dashed green lines are the average size in the first period.
Panels (c) and (d): The blue lines are the average investment wedges, and the shaded blue areas are one standard deviation.

Wedge dynamics with time-0 screening. I now introduce the time-0 screening problem. For this, I assume an ex-ante participation constraint and consider two parametrizations of h. In the benchmark, I set it equal to the ergodic distribution of θ_t but double the standard deviation, and I increase the standard deviation by five times in the second one. Panel 5a in Figure 5 plots the average investment wedge over the initial type θ_0. As usual, the average wedges are inverse U-shaped in θ_0 due to the no distortions at the top and bottom in the time-0 screening problem. Moreover, as expected, the average wedges are larger in the calibration with higher variance. Panel 5b plots the investment wedge dynamics starting with the θ_0 with the lowest promised information rent (i.e. lowest $\Delta_0(\theta_0)$). In the high variance calibration, we see now that the investment wedge gradually decreases over time. Interestingly, with a lower variance, the wedges initially increase for a few periods and then gradually decrease. Consistent with equation (28), in the first periods, the marginal benefits of redistribution from $t \geq 1$ onwards $(\hat{MB}(\theta^1), \hat{MB}(\theta^2)...)$ are summed with the initial marginal benefit $MB_0(\theta^0)$, making the initial wedges increase. Over time, the distortions from the time-0 screening start to vanish because the passthrough term is below one, and so wedges start to decrease. Accordingly, the wedges appear to converge to the same stationary distribution with both calibrations. As discussed, the wedges will typically be much larger with an ex-post participation constraint, so they could decrease over time even with a much smaller variance. Moreover, in that case, the average τ_1^K would be decreasing in θ_0 instead of inverse U-shaped.

Figure 5: Investment wedge with time-0 screening

(a) Average τ_1^K (b) Wedge dynamics (highest starting distortion)



Note: Panel (a): On a grid for θ_0, I compute the policy functions $(\lambda_1(\theta_0), \gamma_1(\theta_0), k_1(\theta_0))$ from the time-0 problem. Then, I use these policies as starting values to compute the average τ_1^K with a Montecarlo simulation.

Panel (b): For both paths, I start with the θ_0 with the lowest $\gamma_1(\theta_0)$. The shaded areas are one standard deviation.

5 Quasi-implementation

The optimal contract studied thus far may a priori be complex, which limits the insights we can derive from the problem. Therefore, it is helpful to study implementations of the optimal contract. This will also provide intuition on what drives the different firm size dynamics in models with risk neutrality and risk aversion. A full implementation of the optimal contract is challenging and left for future work. Instead, I analyze two simpler problems. In this section, I use numerical simulations to study a quasi-implementation with simpler contracts that approximate the optimal allocation. Then, in Appendix D, I study a full implementation in a simplified two-period version of the model.

The two implementations are also different in the sense that they are based on two different margins of distortion for the lender. In particular, we consider two ways in which the principal can lower dynamic information rents to better screen types. In the first one, the lender dynamically distorts the entrepreneur's compensation. Given the distortions to compensation, she may then optimally adjust the firm size (k_{t+1}).[23] In the second, the lender directly distorts firm size by discouraging the entrepreneur from reinvesting profits.

The approach to deriving the quasi-implementation with the numerical simulations will be the following. First, I use regressions with the model simulated data to better understand the compensation dynamics. Then, I propose a simple contract and use the simulated data and regression estimates to calibrate the parameters of the contract. Finally, I solve the entrepreneur's problem under the simple contract and compare the induced consumption dynamics with the optimal contract. With i.i.d types, firm size is constant, so I also fix capital to be constant in the implementation. For simplicity, I will also keep capital fixed for the persistent case. Therefore, we will focus solely on the compensation dynamics.

5.1 i.i.d types

I use the simulated data from Section 4 to run regressions of consumption on returns and promised utility. The regression results are in Table 4 in Appendix A; we make three observations:

1. Returns at any period $t-k$ have the same effect as returns at t on consumption at t (column 2). Relatedly, consumption follows a random walk (column 5). This suggests that compensation is perfectly smoothed across periods.

2. The sensitivity of compensation to returns does not depend on the current promised utility. Note the interaction $returns_t \times v_{t-1}$ is close to 0 in column 3.

[23] In this sense, the cash flow diversion model is equivalent to a hidden endowment model (as in Thomas and Worrall (1990)) where the principal has some control over the agent's income process.

3. The sensitivity of compensation to returns is close to linear. See the linear relation between consumption and returns in Figure 9 in Appendix A or note that $returns_t^2$ is close to 0 in column 4.

Points 2. and 3. suggest that a constant equity share can be a good approximation. If the promised utility (v_{t-1}) were related to the equity share, we would observe that it affects the sensitivity of consumption to returns, even if the entrepreneur is smoothing consumption intertemporally. Point 1. indicates that in the implementation, the entrepreneur's implicit wealth can be used to perfectly smooth consumption intertemporally. As is known, the promised utility can be naturally mapped to the agent's wealth (Atkeson and Lucas (1992)). Let W_t denote the agent's wealth and χ the (inside) equity share, i.e. the portion of the returns accruing to the entrepreneur. Let $\overline{f}(k_t) = \mathbb{E}\left[f(k_t, \theta_t)\right]$ denote the expected returns if capital is k_t. I fix capital to the optimum in the second best k_{SB}. The entrepreneur also receives initial cash W_0.[24] Therefore, at period 1, the entrepreneur's wealth is

$$W_1 = W_0 + \frac{\chi \overline{f}(k_{SB})}{1-q}.$$

At every period, after returns are realized, if the entrepreneur does not misreport, his wealth changes by $\chi\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right)$. So, the law of motion of the entrepreneur's wealth follows

$$c_t + W_{t+1} = \frac{1}{q} W_t + \chi\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right) \equiv C(W_t, \theta_t). \tag{33}$$

Given the entrepreneur's wealth, savings can be chosen to smooth consumption. Therefore, this contract is equivalent to allowing the entrepreneur to pledge his shares as collateral and borrow to consume. This practice is prevalent; Fabisik (2019) reports that between 2007 and 2016, 7.6% of CEOs of US public companies had pledged shares. Moreover, she estimates that 90.5% of CEOs use it to obtain liquidity while maintaining ownership. This motive is consistent with this implementation. Pledging shares aligns the entrepreneur's consumption with the firm's value but without having to sell shares, which is costly as it reduces the entrepreneur's incentives. Moreover, the implementation is independent of dividend payout policies. Notice that it is equivalent if the extra returns $\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right)$ are paid as dividends or are kept as savings inside the firm, and the entrepreneur and the firm face the same interest rate $\frac{1}{q} - 1$.

[24] This is just a free variable used to match the chosen initial promised utility in the second best, so we may also have $W_0 < 0$ if the entrepreneur initially transfers funds to the lender.

The next step for the numerical implementation is to obtain a value for χ. I back out this value from the regressions on model simulated data. For an entrepreneur that does not misreport and is allowed to save by himself, to a first-order approximation, we have

$$\frac{dc_t}{df(k_t,\theta_t)} \approx (1-q)\chi.$$

So χ can be identified from the regressions as $\hat{\chi} = \frac{\beta_{returns}}{(1-q)} = \frac{0.0488}{0.05} \approx \iota$, where $\beta_{returns}$ is the regression coefficient on returns in column (1) of Table 4. So I set directly $\hat{\chi} = \iota$.[25] Then, given $\hat{\chi}$, the entrepreneur's recursive problem with wealth W_t and productivity θ_t is

$$\begin{aligned} \mathcal{W}(W_t,\theta_t) &= \max_{\widetilde{\theta}\leq\theta} u(\widetilde{c}_t) + \beta\mathcal{V}(W_{t+1}) \\ s.t \qquad W_{t+1} &= qC(W_t,\widetilde{\theta}_t) \\ c_t &= (1-q)C(W_t,\widetilde{\theta}_t) \\ \widetilde{c}_t &= c_t + \iota(f(k_{SB},\theta) - f(k_{SB},\widetilde{\theta})) \end{aligned} \tag{34}$$

where $\mathcal{V}(W_{t+1}) = \mathbb{E}\left[\mathcal{W}(W_{t+1},\theta_{t+1})\right]$, $C(W_t,\theta_t) = \frac{1}{q}W_t + \hat{\chi}\left(f(k_{SB},\widetilde{\theta}_t) - \overline{f}(k_{SB})\right)$ and W_0 is chosen such that $\mathcal{V}(W_1) = v_1$, i.e. the promised utility under the direct mechanism. Throughout the paper, I have assumed that the entrepreneur cannot secretly save. So in the implementation, there is a double deviation problem if the entrepreneur is allowed to save freely. That is, the entrepreneur deviates by misreporting funds and saving more. For this reason, I assume that the lender directly assigns a consumption/savings level given the entrepreneur's report and wealth $(W_t,\widetilde{\theta}_t)$.[26] Equivalently, we can imagine that the entrepreneur is penalized if the lender observes that his savings choices are not optimal given the reported type and wealth.

I solve numerically for the policy functions $\widetilde{\theta}(W_t,\theta_t)$ in the entrepreneur's problem (34). Then, I run the same Monte Carlo simulation as for the optimal allocation and compare the results.[27] Figure 10 in Appendix A shows that the consumption paths are very close

[25] It is a regular result in cash flow diversion models (especially in static versions) that the equity share is linked to the deadweight loss of diverting funds.

[26] I assign the consumption to be $c_t = (1-q)C(W_t,\widetilde{\theta}_t)$ because I observe that average consumption is approximately constant in the numerical simulations. But this is not the optimal savings level of the entrepreneur, as he would save more for precautionary motives. To relax this restriction, we could introduce an extra wedge (or tax) on the entrepreneur's returns on savings to exactly counteract the precautionary motive.

[27] To have accurate comparisons, in the Monte Carlo simulation, for each realization of the shock process

Table 1: Welfare comparisons i.i.d

	Total Welfare	Deadweight loss diversion of funds	Risk premium (relative to SB)
Optimal contract (SB)	-55.88	0	0
Quasi-Implementation ($\chi = 0.95$)	-56.11 (-0.4% loss)	5.7e-8	0.22

to the optimal allocation and that this contract induces minimal diversion of funds. Not surprisingly, this simple contract also reaps most of the benefits of the optimal allocation (see Table 1). Given a fixed initial promised utility (v_0), we can decompose the lender's loss from using the simple contract

$$K^I(v_0, \theta_0) - K^{SB}(v_0, \theta_0) =$$
$$\underbrace{(1-\iota)\mathbb{E}\left[\sum_{t=0}^{\infty} q^t \left(f(k_{SB}, \theta^t) - f(k_{SB}, \widetilde{\theta}_t(\theta^t))\right)|\theta_0\right]}_{\geq 0\text{,Deadweight loss diversion of funds}} + \underbrace{\mathbb{E}\left[\sum_{t=0}^{\infty} q^t \left(c^I(\theta^t) - c^{SB}(\theta^t)\right)|\theta_0\right]}_{\text{Risk premium,}>0\text{ if less risk in SB}}, \quad (35)$$

where the superscript I is used to denote allocations under the implementation. As shown in Table 1, most of the losses from the simple contract result from exposing the entrepreneur to more risk, but the differences are negligible. The implementation performs even better with log utility, see Figure 11 in Appendix A.

5.2 Persistent types

With persistent private information, the dynamic information rents (Δ_{t-1}) must be added as an extra state variable in the problem. Intuitively, this variable captures the expected sensitivity of the entrepreneur's utility to the productivity realization (i.e. his exposure), as equation (14) can be written as

$$\Delta_{t-1} = \mathbb{E}\left[\rho(\theta^t)\frac{\partial w(\theta^t)}{\partial \theta_t}|\theta_{t-1}\right]. \quad (36)$$

$\{\varepsilon_t\}_{t=1}^{25}$ I compute consumption for both the optimal allocation and the implementation. Then for each realization and period, I compute the distance and average across all draws. That is, I compute for every period $\bar{c}_t^{dist} = \sum_i \sqrt{\left(c_t^{SB}(\{\varepsilon_{i.\tau}\}_{\tau=1}^t) - c_t^I(\{\varepsilon_{i.\tau}\}_{\tau=1}^t)\right)^2}$, where c^{SB} is the consumption under the optimal allocation and c^I under the implementation.

We can also verify this in the regressions with model simulated data, where we obtain

$$c_t = -0.2^{***}\theta_t - 3.327^{***}\Delta_{t-1} + 1.011^{***}\theta_t \times \Delta_{t-1} + 0.652^{***}\theta_{t-1} + 0.382^{***}v_{t-1}.$$

The coefficient on the interaction term $\Delta_{t-1} \times \theta_t$ is positive. So, for a given level of persistence, when the lender has promised low information rents (i.e low Δ_{t-1}), the entrepreneur's exposure to returns decreases. In this implementation, the exposure of the entrepreneur is controlled by the equity share. Notice that for the i.i.d case (problem (34)), if it is optimal for the entrepreneur to not divert funds, we have

$$\frac{\partial \mathcal{W}(W_t, \theta_t)}{\partial \theta_t} = \chi \times u'(c_t) f_\theta(k_{SB}, \theta_t).$$

Thus, a full implementation of the optimal contract would need to allow for a time-varying equity share. In general, lowering the entrepreneur's equity is beneficial as it increases the entrepreneur's insurance against productivity shocks, but it also comes at the cost of increasing the incentives to misreport funds. If types are persistent, there is an extra gain of lowering the equity share at period $t+1$ because it helps screen types.

Why does buying equity help screen types? Imagine that, at period t, the lender offers to buy some equity from type (θ^{t-1}, θ'). Assume also that the lender offers to pay him a price $P_{\Delta\chi}((\theta^{t-1}, \theta'))$ such that he is indifferent between accepting the offer or rejecting it. If returns are persistent, types (θ^{t-1}, θ'') with $\theta'' > \theta'$ have higher expected returns at period $t+1$. So it is not attractive for them to sell equity at price $P_{\Delta\chi}((\theta^{t-1}, \theta'))$. That is, if the entrepreneur is not willing to sell equity at a fair price relative to the projected returns – which both parties agree upon given the current reported returns – it signals that he is misreporting funds. Therefore, the lender can use equity purchases, which inefficiently lower the equity share, to better screen types.

An implementation with a time-varying equity share is substantially more challenging.[28] However, I find that the contract with a constant equity share still delivers small welfare losses relative to the optimal contract.[29] Compared to the i.i.d case, we now only have to

[28] Now it is more challenging to infer the equity share from the regressions directly. Moreover, it may follow a complicated stochastic process. As it would be persistent but also because there is no distortion at the top ($\overline{\theta}$) and bottom ($\underline{\theta}$) in the promised information rents.

[29] I have experimented with contracts where the equity share is uniformly decreased over time for all types. The idea is that when the agent underreports at t, he experiences a capital loss but expects to recover it at $t+1$ with returns that are higher than expected. However, if his equity share is lower at $t+1$, he cannot fully recover the capital loss. However, I have not found any gains from these types of contracts.

Table 2: Welfare comparison with persistence

	Total Welfare	Deadweight loss diversion of funds	Risk premium (relative to SB)
Optimal contract (SB)	-58.49	0	0
Quasi-Implementation ($\chi = 0.95$)	-59.12 (-1.07% loss)	3e-3	0.572

make one modification. Notice that when the entrepreneur's period t returns increase, the net present value of the firm's future returns also increases. So, the firm's value increases and the entrepreneur experiences a capital gain. Define the value of the firm by

$$\overline{f}_{t+1}(k_{SB}, \theta_t) \equiv \mathbb{E}\left[\sum_{\tau=1}^{\infty} q^{\tau-1} f(k_{SB}, \theta_{t+\tau})|\theta_t\right].$$

Recall capital is fixed to the same level k_{SB} as in the i.i.d case. Then, the entrepreneur's cash on hand at period t if he reports type $\widetilde{\theta}_t$ and his past type report was $\widetilde{\theta}_{t-1}$ is

$$\begin{aligned} C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) &= \frac{1}{q}W_t + \chi\left(f(k_{SB}, \widetilde{\theta}_t) + q\overline{f}_{t+1}(k_{SB}, \widetilde{\theta}_t) - \overline{f}_t(k_{SB}, \widetilde{\theta}_{t-1})\right) \qquad (37) \\ &= \frac{1}{q}W_t + \chi\left(\underbrace{f(k_{SB}, \widetilde{\theta}_t) - \mathbb{E}\left[f(k_{SB}, \widetilde{\theta}_t)|\widetilde{\theta}_{t-1}\right]}_{\text{Innovation returns}} + \underbrace{q\left(\overline{f}_{t+1}(k_{SB}, \widetilde{\theta}_t) - \overline{f_{t+1}}(k_{SB}, \widetilde{\theta}_{t-1})\right)}_{\text{Capital gain}}\right). \end{aligned}$$

The entrepreneur's problem is the same as in (34) but with the cash on hand given by (37). Because the entrepreneur can borrow using his shares as collateral, the capital gains also increase the entrepreneur's consumption. Moreover, we also need to keep track of the past report $\widetilde{\theta}_{t-1}$ as an extra state variable because it affects the expected returns. Table 2 contains the welfare comparison and decompositions with the optimal contract, assuming that the initial type is fixed in both cases. The risk premium is higher than for the i.i.d case, but the welfare losses from the simple contract continue to be small.[30]

[30] In general, the optimal contract studied is not renegotiation-proof. When the principal lowers Δ_t at period t, this is inefficient from the period $t+1$ perspective. The optimal contracts are renegotiation-proof with i.i.d types because these effects are not present. I conjecture that this quasi-implementation with constant equity could be a good approximation to the optimal renegotiation-proof contract with persistent private information. The constant equity implies constant sensitivity as the optimal renegotiation proof contracts in Strulovici (2022). Moreover, as in Strulovici (2022), the sensitivity is increasing in the level of persistence because capital gains are larger. Because the constant equity contract delivers small welfare losses compared to the optimal, I further conjecture that, at least for the calibration used, the losses from restricting to renegotiation-proof contracts may be small.

Fu and Krishna (2019) study a similar model with persistent private information and a risk neutral entrepreneur. They show that as persistence increases, the sensitivity of the entrepreneur's compensation to returns also increases. In their implementation, this implies that the entrepreneur is compensated more with stock options and less with equity. A priori, a full implementation of the optimal contract could require using stock options. But only with equity the entrepreneur already has a higher exposure to returns. Because the entrepreneur experiences a capital gain and can borrow using his shares as collateral, his compensation increases by more than his equity share times the reported returns. Intuitively, simply accounting for capital gains may allow the implementation to approximate the optimal contract because it (approximately) generates the extra sensitivity in compensation required by the dynamic information rents. To understand this, notice that by expanding the dynamic information rent term (Equation (36)), we can write

$$\Delta_t = \iota \times \mathbb{E}\left[\sum_{j=1}^{\infty} I_t^{t+j}(\theta^{t+j})\beta^{j-1}u'(c(\theta^{t+j}))f_\theta(k_{t+j},\theta_{t+j})|\theta^t\right],$$

where $I_t^{t+j}(\theta^{t+j}) \equiv \rho(\theta^t) \times ... \times \rho(\theta^{t+j})$ are the impulse response functions as defined in Pavan *et al.* (2014). With $\chi = \iota$ and fixed capital, this is actually the capital gain of the entrepreneur from an increase in productivity $d\theta_t$ if the firm is priced with the entrepreneur's stochastic discount factor.

Finally, this implementation bears some resemblance with the "Dynamic Incentive Account" (DIA) implementation in Edmans *et al.* (2012) for a CEO compensation model. There, the DIA tracks the agent's NPV of future pay (i.e., his wealth) and invests a fraction in the company stock and a fraction in interest-bearing cash. This portfolio is then constantly rebalanced as the firm value changes to maintain incentives. By contrast, here, because the agent's compensation comes solely from his ownership of the firm, letting him borrow against his stock is (approximately) sufficient to implement the required compensation. Moreover, the lender uses equity purchases to lower dynamic information rents, but in this cash flow diversion model, there is no need to rebalance the agent's portfolio every period to maintain incentives.

6 Comparison with risk neutral and equity dynamics

The quasi-implementation helps understand the different firm size dynamics with risk neutrality and risk aversion. In Table 3, I summarize the main features of the optimal contract and implementations with the different assumptions about the agent's utility and shock process. With risk neutrality, as long as the limited liability constraint ($c \geq 0$) is satisfied, increasing the agent's exposure to risk bears no cost. After high returns, it is optimal to compensate the entrepreneur with a higher stake in the project, i.e. by increasing his equity share. Therefore, with risk neutrality, the entrepreneur's promised utility maps to the value of equity, as shown in Clementi and Hopenhayn (2006).

If the entrepreneur is risk averse, increasing his exposure to risk through a higher equity share is costly. In the numerical simulations, we have seen that the entrepreneur's exposure to returns is independent of his promised utility. So with i.i.d types, a constant equity share and mapping the entrepreneur's promised utility to his private wealth gives a good approximation to the optimal allocation. With persistent types, the equity share should also be time-varying as in the risk neutral model, but the driving forces are different. With persistence, the lender has an incentive to lower equity below the efficient level at $t+1$ as it helps screen types at period t. Hence, when θ_0 is fixed and so the dynamic information rents decrease over time, the equity share of the entrepreneur also tends to decrease. Then, when the equity share is low, the entrepreneur has more incentives to divert funds, so the lender is less willing to provide capital.

In both models there is a positive relation between the entrepreneur's equity share and firm size. A lower equity share always increases the implicit lending costs because the incentives to divert funds are higher. However, without time-0 uncertainty, the equity share drifts in opposite directions: upwards with risk neutrality and downwards with risk aversion and persistence. With risk neutrality and i.i.d types, firm size converges to the first best level only because the entrepreneur's equity share goes to one (Clementi and Hopenhayn (2006)). That is, he becomes the sole owner of the firm, and the value of debt and outside equity go to zero. With persistent types and risk neutrality, the equity share does not necessarily have to converge to one for the firm's size to reach the first best (Fu and Krishna (2019)). However, the combination of equity and stock options also increases once the firm becomes unconstrained, and so it also tends to increase over time. These equity dynamics may be inconsistent with what is observed in the data. For example, in the venture capital industry, the founder's ownership is typically diluted over time as the firm's capital grows through multiple financing rounds (Sahlman (1990)).

By introducing uncertainty about the entrepreneur's initial type, we have seen that it is possible to have an upward drift in firm size even with risk aversion and persistent private information. However, this is achieved by making the information rents increase over time, which implies that the equity share should also increase in an implementation. Accordingly, to simultaneously explain firm size and equity dynamics, it may be necessary to break the tight link between equity and firm size that these models generate.

Table 3: Comparisons optimal contract and implementation across models

	Risk neutral & i.i.d (CH 2006)	Risk neutral & persistent (FK 2019)	Risk averse & i.i.d	Risk averse & persistent
Convergence to FB	Yes	Yes	No	No
Deferred compensation	Yes	Yes	No	No
Link firm size & compensation	Yes (strong)	Yes	Weak	No
Firm size drift w/out time-0 uncertainty	Increasing	Increasing	(Approximately) Constant	Decreasing
Implementation	Equity & cons. = div.	Equity + stock options & cons. = div. + option	Wealth + Equity (fixed) & cons. ≠ div.	Wealth + Equity (varying) & cons. ≠ div.

Notes: CH 2006 stands for Clementi and Hopenhayn (2006), FK 2019 for Fu and Krishna (2019), cons. for consumption and div. for dividends. Implementations of the optimal contract are generally not unique. So, in the table, I just describe the implementation in each of the corresponding papers.

7 Extensions

The problem studied throughout the paper is the simplest version of a cash flow diversion model with persistent private information and risk aversion. To focus on the role of persistence and risk aversion, I have imposed some assumptions and abstracted from other interesting margins. I explore three extensions in Appendix C.

Limited commitment. In Appendix C.1, I relax the assumption of full commitment of the entrepreneur. At every period, the entrepreneur can steal all the capital advanced by the lender and leave the contract. This friction also lowers firm size but adds an incentive to have promised utilities increase over time. So when the limited commitment is binding, it can generate dynamics where firm size increases over time, as in Albuquerque and Hopenhayn (2004), even with risk aversion and persistence.

Endogenous termination. I have assumed that the lender does not have the option to terminate the project. I discuss endogenous termination in Appendix C.2. Although termination may sometimes be optimal, I show that it does not affect any of the results presented. I also discuss what inefficiencies may cause termination with risk aversion and persistence: too low promised information rents or a combination of private information and limited commitment, similar to Dovis (2019).

Moreover, if at some point terminating with some probability $\alpha_t(\theta^t) \in (0,1)$ is optimal, $\alpha_t(\theta^t)$ increases with the persistence of the shocks. I show this in a simplified two-period and two-type version of the model. The intuition is similar to the equity purchases. Imagine that the principal increases the termination probability of type θ^t while increasing his payment after termination such that his ex-ante utility is held constant. With persistence, types (θ^{t-1}, θ') with $\theta' > \theta_t$ know they have higher expected returns at $t+1$ than θ_t, so a higher termination probability is relatively less attractive for them. Hence, a higher termination probability can discourage misreporting for types $\theta' > \theta_t$, and so it lowers the cost of screening types.

Divert funds before investing (screening). Finally, in Appendix C.3, I study a model where instead of diverting the returns, the entrepreneur can choose the fraction of available funds invested in the firm and divert the rest. Then, the lender can observe the project returns but not invested funds. The investment wedge can now be defined as the distortion to invested and diverted funds relative to the first best. Moreover, this model yields the same characterizations of the shadow cost of information rents, the GIEE, and the firm size dynamics.

8 Conclusion

In this paper, I revisited the firm size and compensation dynamics predicted by the optimal contracting solution of dynamic cash flow diversion models. I departed from the previous literature by allowing the entrepreneur to be risk averse and to have persistent private information about the firm's productivity.

Relaxing these assumptions leads to remarkably different dynamics than those of models with a risk neutral entrepreneur. First, the interaction between risk aversion and persistent private information decouples the dynamics of the firm's size and the entrepreneur's compensation. Second, the firm's size never converges to the first best, and its distortions inherit the autoregressive properties of the type process. Moreover, if there is no initial uncertainty about the entrepreneur's productivity –as assumed in the literature– the distortions tend to increase over time, so the firm's size tends to decrease. Third, the entrepreneur's compensation is smoothed intertemporally, but the variance of consumption increases over time. Finally, implementing the optimal contract requires separately keeping track of the entrepreneur's wealth and equity share in the firm.

I argue that canonical cash flow diversion models cannot simultaneously generate realistic firm size and equity share dynamics due to the embedded link between the two variables. Accordingly, an important avenue for future work is to study departures from this model that can break the tight between the firm's size and the entrepreneur's equity share. In particular, an empirical regularity that these models should be able to rationalize is the dilution of the entrepreneur's equity share as the firm grows.

References

Albuquerque, R. and Hopenhayn, H. A. (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies*, **71** (2), 285–315.

Atkeson, A. and Lucas, R. E. (1992). On efficient distribution with private information. *The Review of Economic Studies*, **59** (3), 427–453.

Biais, B., Mariotti, T., Plantin, G. and Rochet, J.-C. (2007). Dynamic security design: Convergence to continuous time and asset pricing implications. *The Review of Economic Studies*, **74** (2), 345–390.

—, —, Rochet, J.-C. and Villeneuve, S. (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica*, **78** (1), 73–118.

Bloedel, A., Krishna, R. and Strulovici, B. (2023a). *Persistent private information revisited*. Tech. rep., Technical report, Stanford University.

Bloedel, A. W., Krishna, R. V. and Leukhina, O. (2023b). Insurance and inequality with persistent private information. *Available at SSRN 3091457*.

Brendon, C. (2013). Efficiency, equity, and optimal income taxation.

Clementi, G. L., Cooley, T. F. and Di Giannatale, S. (2010). A theory of firm decline. *Review of Economic Dynamics*, **13** (4), 861–885.

— and Hopenhayn, H. A. (2006). A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics*, **121** (1), 229–265.

DeMarzo, P. M. and Fishman, M. J. (2007a). Agency and optimal investment dynamics. *The Review of Financial Studies*, **20** (1), 151–188.

— and — (2007b). Optimal long-term financial contracting. *The Review of Financial Studies*, **20** (6), 2079–2128.

—, —, He, Z. and Wang, N. (2012). Dynamic agency and the q theory of investment. *The Journal of Finance*, **67** (6), 2295–2340.

— and Sannikov, Y. (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *The Journal of Finance*, **61** (6), 2681–2724.

— and — (2016). Learning, termination, and payout policy in dynamic incentive contracts. *The Review of Economic Studies*, **84** (1), 182–236.

Di Tella, S. and Sannikov, Y. (2021). Optimal asset management contracts with hidden savings. *Econometrica*, **89** (3), 1099–1139.

Dovis, A. (2019). Efficient sovereign default. *The Review of Economic Studies*, **86** (1), 282–312.

Edmans, A., Gabaix, X., Sadzik, T. and Sannikov, Y. (2012). Dynamic ceo compensation. *The Journal of Finance*, **67** (5), 1603–1647.

Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The journal of industrial economics*, pp. 567–581.

Fabisik, K. (2019). Why do us ceos pledge their own company's stock? *Swiss Finance Institute Research Paper*, pp. 19–60.

Farhi, E. and Werning, I. (2013). Insurance and taxation over the life cycle. *Review of Economic Studies*, **80** (2), 596–635.

Fernandes, A. and Phelan, C. (2000). A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, **91** (2), 223–247.

Fu, S. and Krishna, R. V. (2019). Dynamic financial contracting with persistent private information. *The RAND Journal of Economics*, **50** (2), 418–452.

Golosov, M., Troshkin, M. and Tsyvinski, A. (2016a). Redistribution and social insurance. *American Economic Review*, **106** (2), 359–86.

—, Tsyvinski, A. and Werquin, N. (2016b). Recursive contracts and endogenously incomplete markets. In *Handbook of Macroeconomics*, vol. 2, Elsevier, pp. 725–841.

He, Z. (2012). Dynamic compensation contracts with private savings. *The Review of Financial Studies*, **25** (5), 1494–1549.

Hellwig, C. (2021). Static and dynamic mirrleesian taxation with non-separable preferences: A unified approach.

Kapička, M. (2013). Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *Review of Economic Studies*, **80** (3), 1027–1054.

Khan, A., Popov, L. and Ravikumar, B. (2020). Enduring relationships in an economy with capital and private information. *FRB St. Louis Working Paper*, (2020-34).

Krasikov, I. and Lamba, R. (2021). A theory of dynamic contracting with financial constraints. *Journal of Economic Theory*, **193**, 105196.

Makris, M. and Pavan, A. (2020). Wedge dynamics with evolving private information.

Ndiaye, A. (2020). *Flexible Retirement and Optimal Taxation*. Tech. rep.

Pavan, A., Segal, I. and Toikka, J. (2014). Dynamic mechanism design: A myersonian approach. *Econometrica*, **82** (2), 601–653.

Sahlman, W. A. (1990). The structure and governance of venture-capital organizations. *Journal of Financial Economics*, **27** (2), 473–521.

Stantcheva, S. (2017). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy*, **125** (6).

Strulovici, B. (2022). Renegotiation-proof contracts with persistent states.

Thomas, J. and Worrall, T. (1990). Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory*, **51** (2), 367–390.

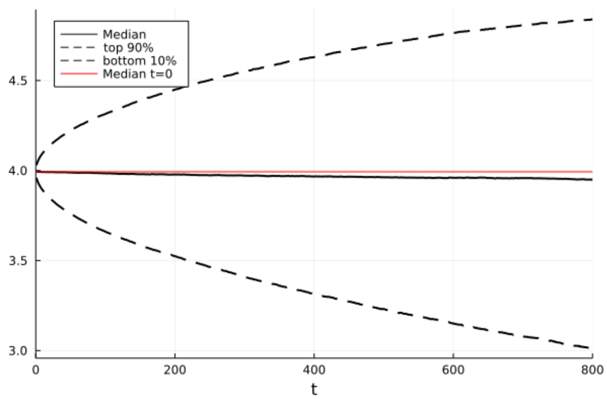
Williams, N. (2011). Persistent private information. *Econometrica*, **79** (4), 1233–1275.

Appendix

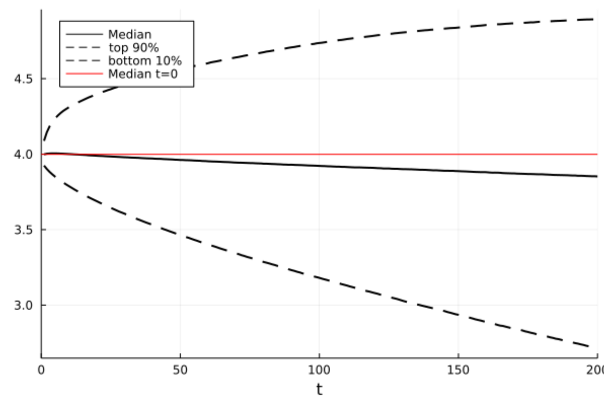
A Additional tables and figures

Figure 6: Immiseration in the long run

(a) Consumption i.i.d ($\rho = 0$)

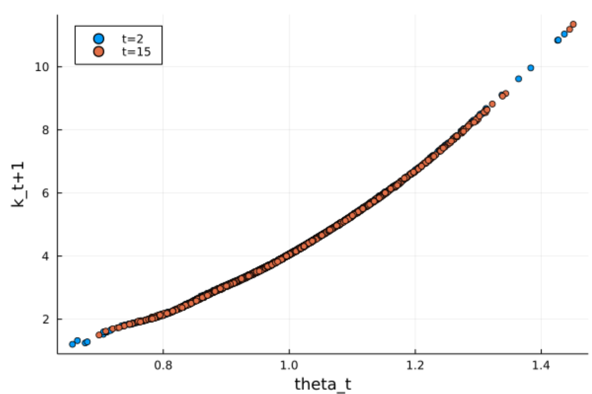


(b) Consumption with persistence ($\rho = 0.7$)



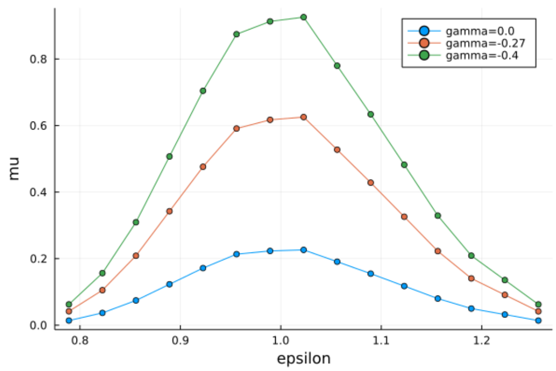
Note: The figures show the median, 10%, and 90% quantiles of the distribution of consumption at every period. For reference, the red line displays the median at period $t = 0$. The median monotonically decreases and the growth of the 90% quantile decreases over time. This implies that consumption will converge to its lower bound. However, we also observe that this convergence is very slow.

Figure 7: Relation $k_{t+1}(\theta_t)$ and θ_t



Note: For a random subsample of 1000 realizations, the plot shows the policy functions of k_{t+1} as a function of θ_t. The blue dots are policies at period $t = 2$, and the red dots at $t = 15$.

Figure 8: Shadow cost information rents μ at different γ



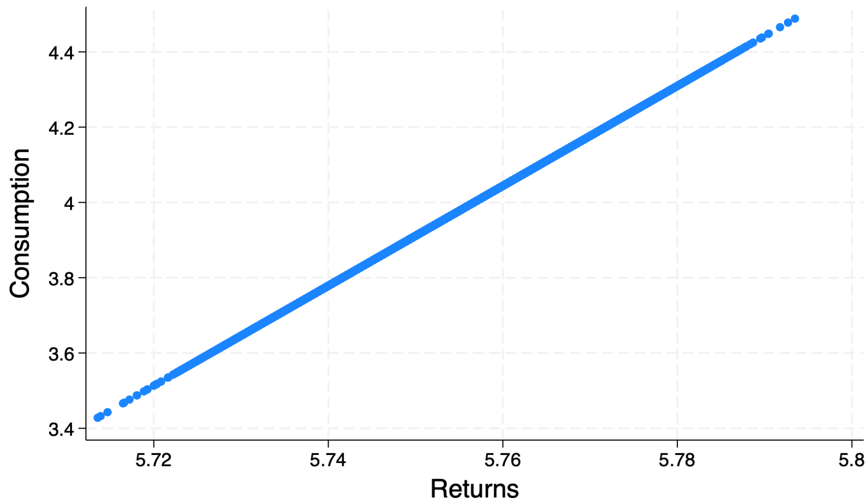
Note: For a fixed (λ_-, k, θ_-), the figure shows the shadow cost μ as a function of the shock ε for different γ_-. The dynamic information rent Δ_- is increasing in γ_-. So when the agent is promised lower information rents (low Δ_- and γ_-), the shadow costs μ are higher. The increase is more pronounced for the types in the middle.

Table 4: Regressions with i.i.d type process

	(1) c_t	(2) c_t	(3) c_t	(4) c_t	(5) c_t
$returns_t$	0.0489 (10558.17)	0.0511 (399.15)	0.0508 (108.58)	0.0704 (926.10)	
v_{t-1}	0.792 (14968.42)		0.790 (1459.73)	0.792 (15010.68)	
$returns_{t-5}$		0.0490 (382.92)			
$returns_t * v_{t-1}$			0.000386 (4.13)		
$returns_t^2$				-0.00185 (-283.17)	
c_{t-1}					0.998 (5300.13)
N	2400000	1900000	2400000	2400000	2300000
R^2	0.999	0.139	0.999	0.999	0.924

t statistics in parentheses

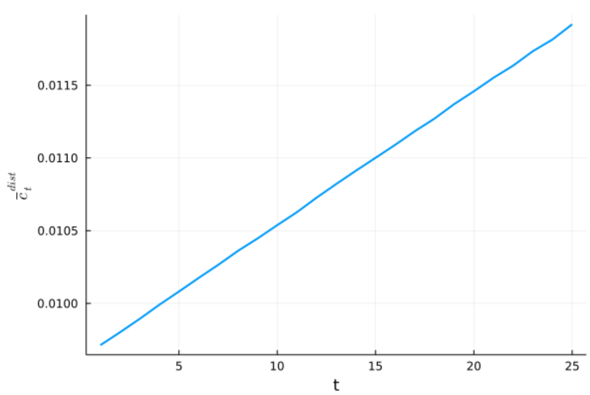
Figure 9: Consumption and returns residualized (i.i.d)



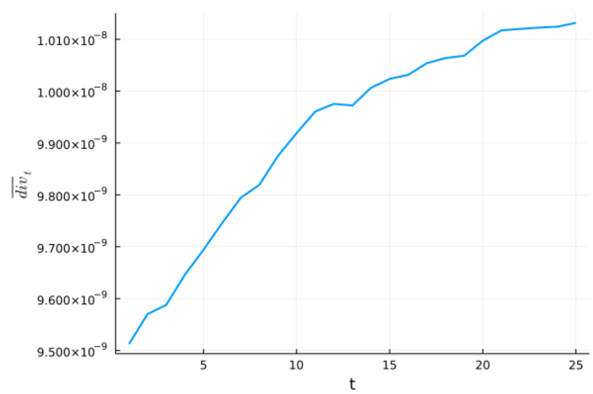
Note: Using simulated data, I residualize both consumption and returns on v_{t-1}.

Figure 10: Simulations implementation i.i.d

(a) Average distance consumption SB and implementation



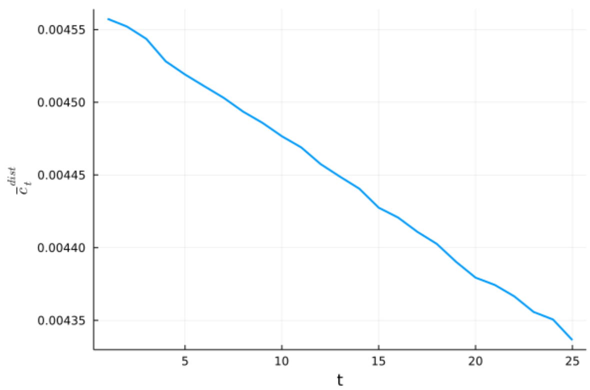
(b) Average fraction of diverted funds



Note: The left figure shows, for every period, the average distance between consumption in the optimal contract (c^{SB}) and the implementation (c^I), i.e. $\bar{c}_t^{dist} = \frac{1}{N}\sum_i \sqrt{(c_t^{SB}(\{\varepsilon_{i,\tau}\}_{\tau=1}^t) - c_t^I(\{\varepsilon_{i,\tau}\}_{\tau=1}^t))^2}$. The right figure shows the average of the diverted funds as a fraction of total returns, i.e. $\overline{div}_t = \frac{1}{N}\sum_i \frac{f(k_{SB},\theta_i^t) - f(k_{SB},\tilde{\theta}_t(\theta_i^t))}{f(k_{SB},\theta_i^t)}$

Figure 11: Simulations implementation i.i.d with log utility ($\sigma = 1$)

(a) Average distance consumption SB and implementation



(b) Average fraction of diverted funds

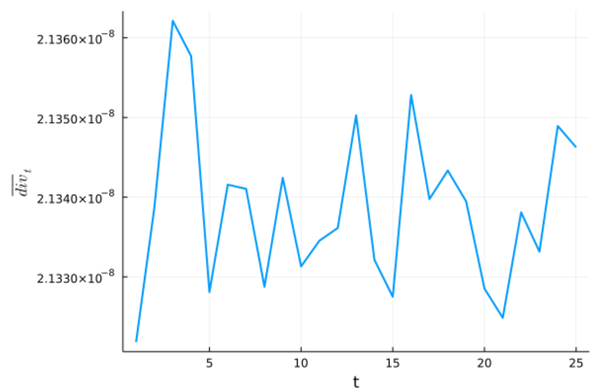
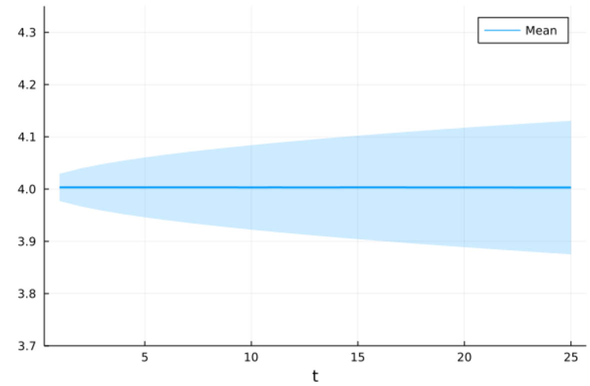
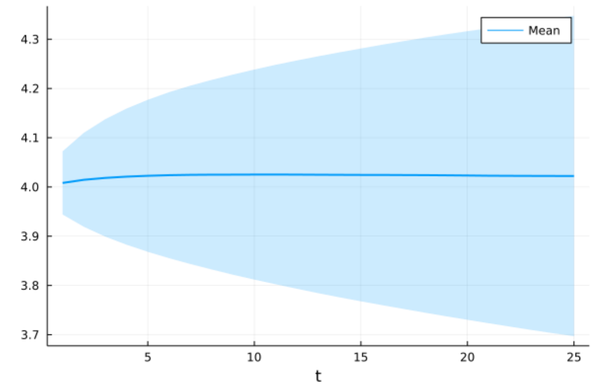


Figure 12: Log utility ($\sigma = 1$)

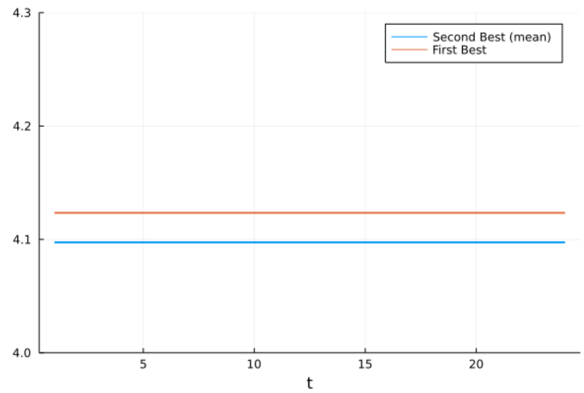
(a) Consumption i.i.d ($\rho = 0$)



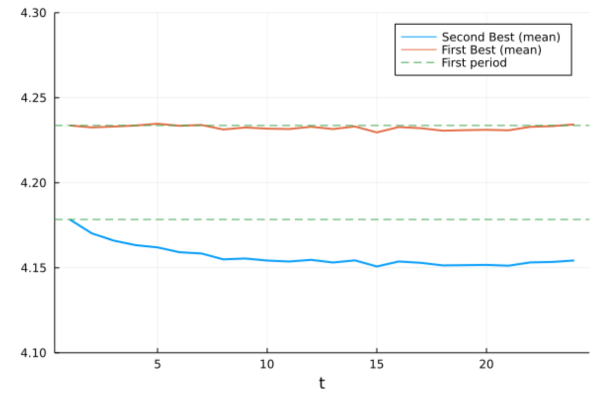
(b) Consumption with persistence ($\rho = 0.7$)



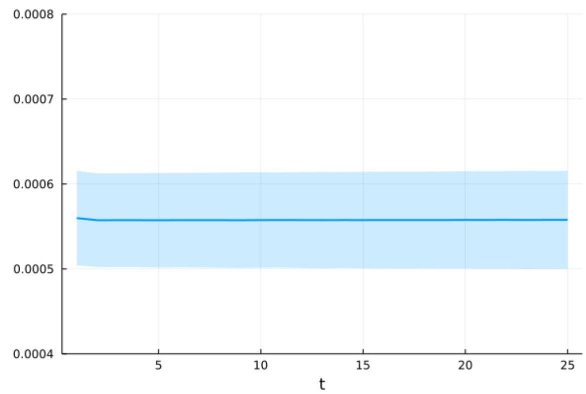
(c) Firm size i.i.d ($\rho = 0$)



(d) Firm size with persistence ($\rho = 0.7$)



(e) Investment wedge i.i.d ($\rho = 0$)



(f) Investment wedge with persistence ($\rho = 0.7$)

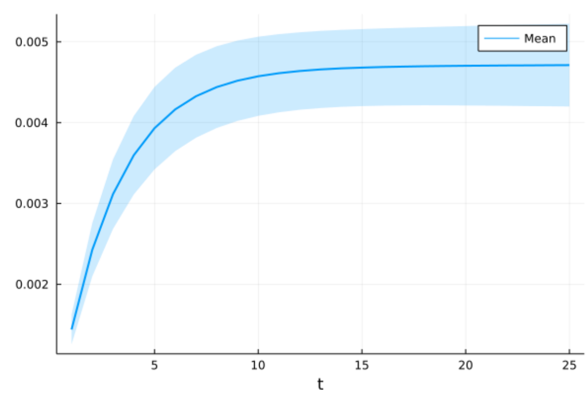
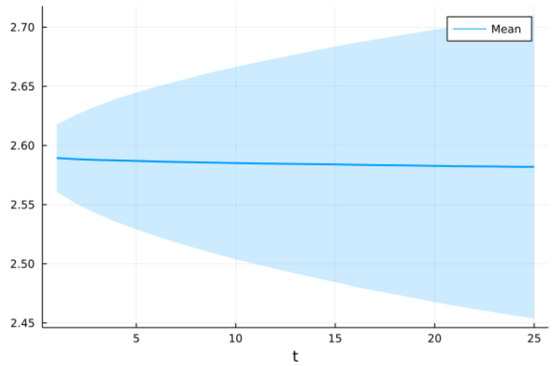
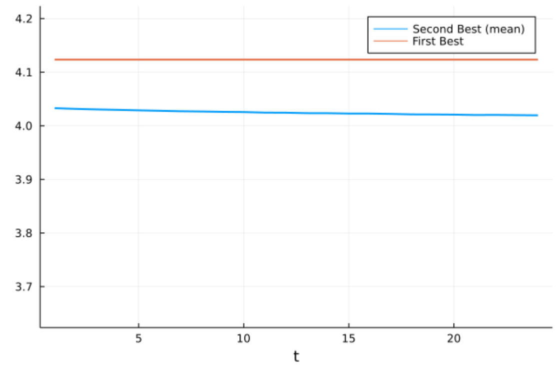


Figure 13: CARA utility

(a) Consumption i.i.d ($\rho = 0$)



(b) Firm size i.i.d ($\rho = 0$)



B Derivations and proofs

The Hamiltonian of the recursive principal's problem is

$$\begin{aligned}\mathcal{H} = \left[k_{t+1}(\theta^t) - b_t(\theta^t) + qK_{t+1}(v_t(\theta^t), \Delta_t(\theta^t), \theta_t, k_{t+1}(\theta^t))\right] \varphi_t(\theta_t|\theta_{t-1}) \\ -\lambda_t \varphi_t(\theta_t|\theta_{t-1}) \left[w_t(\theta^t) - v_{t-1}\right] - \gamma_t \varphi_t(\theta_t|\theta_{t-1}) \left[w_t(\theta^t)\mathcal{E}(\theta_t, \theta_{t-1}) - \Delta_{t-1}\right] \\ +\mu_t(\theta^t) \left[u'(f(k_t, \theta_t) - b_t(\theta^t))\iota f_\theta(k_t, \theta_t) + \beta \Delta_t(\theta^t)\right] \\ +\xi_t(\theta^t)\varphi_t(\theta_t|\theta_{t-1}) \left[w_t(\theta^t) - u(f(k_t, \theta_t) - b_t(\theta^t)) - \beta v_t(\theta^t)\right]\end{aligned}$$

The optimality conditions are

$b_t(\theta^t)$:

$$\xi_t(\theta^t) = \frac{1}{u'(\theta^t)} \left[1 + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} \iota f_\theta(\theta^t) u''(\theta^t)\right] \tag{38}$$

The envelope conditions are

$$\frac{\partial K_{t+1}}{\partial v_t(\theta^t)} = \lambda_{t+1}(\theta^t) \tag{39}$$

$$\frac{\partial K_{t+1}}{\partial \Delta_t(\theta^t)} = \gamma_t(\theta^t) \tag{40}$$

$$\frac{\partial K_{t+1}}{\partial k_{t+1}(\theta^t)} = \mathbb{E}\left[-\xi_{t+1}(\theta^{t+1})u'(\theta^{t+1})f_k(\theta^{t+1})|\theta_t\right] + \tag{41}$$
$$\mathbb{E}\left[\frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)}\left(u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})f_k(\theta^{t+1}) + u'(\theta^{t+1})\iota f_{\theta k}(\theta^{t+1})\right)|\theta_t\right].$$

Using the envelope conditions (39) and (40) we get

$v_t(\theta^t)$:

$$\lambda_{t+1}(\theta^t) = \frac{\beta}{q}\xi_t(\theta^t) \tag{42}$$

$\Delta_t(\theta^t)$:

$$\gamma_{t+1}(\theta^t) = -\frac{\beta}{q}\frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta_{t-1})}. \tag{43}$$

Substituting (38) and (41) into the FOC for $k_{t+1}(\theta^t)$ we get

$$\frac{1}{q} = \mathbb{E}\left[f_k(\theta^{t+1}) - \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)}u'(\theta^{t+1})\iota f_{\theta k}(\theta^{t+1})|\theta_t\right]. \tag{44}$$

Finally, the law of motion for the co-state is

$$\dot{\mu}_t(\theta^t) = -\left[\xi_t(\theta^t) - \lambda_t - \gamma_t\mathcal{E}(\theta_t, \theta_{t-1})\right]\varphi_t(\theta_t|\theta_{t-1}) \tag{45}$$

Proof of Proposition 1. Set $\mu_t(\theta^t) = 0$ for all θ^t, then from equation (44) we obtain point 3. For point 2, note that with $\mu_t(\theta^t) = 0$, equation (45) becomes

$$\xi_t(\theta^t) = \lambda_t.$$

From equation (38),

$$\frac{1}{u'(\theta^t)} = \xi_t(\theta^t)$$

and using (42) gives point 2. Point 1 holds in the first best and second best allocations.

Proof of Proposition 2. From the FOC for $k_{t+1}(\theta^t)$ (equation (44)), multiplying the second term inside the expectation by $\frac{f_k(\theta^{t+1})}{f_k(\theta^{t+1})}$ and letting

$$\tau^k(\theta^{t+1}) = \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)} u'(\theta^{t+1}) \iota \frac{f_{\theta k}(\theta^{t+1})}{f_k(\theta^{t+1})}, \tag{46}$$

we have $\frac{1}{q} = \mathbb{E}[f_k(\theta^{t+1})(1 - \tau^k(\theta^{t+1}))|\theta_1]$. Combining with the definition of $\tilde{\tau}^k(\theta^t)$,

$$(1 - \tilde{\tau}^k(\theta^t))\mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})|\theta^t\right] = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})\left(1 - \tau^k(\theta^{t+1})\right)|\theta^t\right],$$

or

$$\tilde{\tau}^k(\theta^t) = \frac{\mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})\tau^k(\theta^{t+1})|\theta^t\right]}{\mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})|\theta^t\right]}$$

Finally, to write the return-dependent wedge as in the proposition $\tau^k(\theta^{t+1})$ multiplying by $\frac{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}$ and rearrange terms to get

$$\tau^k(\theta^{t+1}) = \iota \frac{f_{\theta k}(\theta^{t+1})}{f_k(\theta^{t+1})} \frac{1 - \Phi_{t+1}(\theta_{t+1}|\theta_t)}{\varphi_{t+1}(\theta_{t+1}|\theta_t)} \times \frac{\mu_{t+1}(\theta^{t+1})}{1 - \Phi_{t+1}(\theta_{t+1}|\theta_t)} u'(\theta^{t+1}).$$

Proof of Proposition 3. The proof follows similar steps as Proposition 1 in Hellwig (2021). Substitute $\xi_t(\theta^t)$ in the LOM of the co-state (45)

$$\dot{\mu}_t(\theta^t) + \mu_t(\theta^t)\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} = -\left[\frac{1}{u'(\theta^t)} - \lambda_t - \gamma_t \mathcal{E}(\theta_t, \theta_{t-1})\right]\varphi_t(\theta_t|\theta_{t-1}),$$

substitute $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}$, using the boundary conditions $\mu_t(\underline{\theta}) = 0$ and $\mu_t(\overline{\theta}) = 0$ and integrating upwards

$$-\mu_t(\theta^t)m(\theta^t) = \int_{\theta_t}^{\overline{\theta}} \left[\lambda_t + \gamma_t \mathcal{E}(\theta', \theta_{t-1}) - \frac{1}{u'(\theta', \theta^{t-1})}\right]\varphi_t(\theta'|\theta_{t-1})m(\theta^t)d\theta'.$$

Using the definition of the incentive-adjusted measure

$$\begin{aligned} \mu_t(\theta^t)m(\theta^t) =& (1 - \Phi(\theta_t|\theta_{t-1}))\mathbb{E}\left[m(\theta^{t-1}, \theta')|\theta' > \theta_{t+1}, \theta_t\right] \\ &\times \left\{\hat{\mathbb{E}}\left[\frac{1}{u'(\theta', \theta^{t-1})} \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \gamma_t \hat{\mathbb{E}}\left[\mathcal{E}(\theta', \theta_{t-1}) \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \lambda_t\right\}. \end{aligned} \tag{47}$$

To get λ_t, note that using the boundary conditions we have

$$0 = \int_{\underline{\theta}}^{\overline{\theta}} \left[\lambda_t + \gamma_t \mathcal{E}(\theta', \theta_{t-1}) - \frac{1}{u'(\theta', \theta_{t-1})} \right] \varphi_t(\theta_t|\theta^{t-1}) m(\theta^t) d\theta'$$

or

$$\lambda_t = \hat{\mathbb{E}} \left[\frac{1}{u'(\theta^t)} \mid \theta^{t-1} \right] - \gamma_t \hat{\mathbb{E}} \left[\mathcal{E}(\theta_t, \theta_{t-1}) \mid \theta^{t-1} \right].$$

Substituting back λ_t into Equation (47), using the definition of $\hat{\rho}(\theta^t)$ (Equation (24)), multiply both sides by $u'(\theta^t)$, collecting terms and using $\widetilde{\mu}_t(\theta^t) = \frac{\mu_t(\theta^t)}{1-\Phi_t(\theta_t|\theta_{t-1})} u'(\theta^t)$ we get the equation in the propositon. Finally, the inequalities $\hat{MB}(\theta^t) \geq 0$ and $\hat{\rho}_t(\theta^t) \geq 0$ follow by using that $\frac{1}{u'(\theta')}$ and $\mathcal{E}(\theta', \theta_{t-1})$ are non-decreasing due to incentive compatibility and assumption A.1, respectively.

Proof of Proposition 4. Equation (47) also holds for period $t = 1$, hence we only need to characterize $\gamma_1(\theta_0)(\equiv -MB_0(\theta_0))$ from the time 0 problem. I start with the case $\kappa = 1$. Using $\dot{w}_0(\theta_0) = \beta \dot{v}_0(\theta_0)$, the Hamiltonian writes

$$\begin{aligned} \mathcal{H} &= [k_1(\theta_0) + qK_1(v_0(\theta_0), \Delta_0(\theta_0), \theta_0, k_1(\theta_0))]\, h(\theta_0) \\ &\quad - \lambda_0 h(\theta_0) [\beta v_0(\theta_0) - v_-] + \mu(\theta_0) \Delta_1(\theta) \end{aligned}$$

Combining the optimality condition for $\Delta_0(\theta_0)$ and the Envelope condition (40) gives:

$$\gamma_1(\theta_0) = -\frac{1}{q} \frac{\mu(\theta_0)}{h(\theta_0)} \tag{48}$$

Using the Envelope condition (39), the LOM of the co-state satisfies

$$\dot{\mu}(\theta_0) = -[q\lambda_1(\theta_0) - \beta\lambda_0]\, h(\theta_0).$$

Integrating, we get $q\mathbb{E}_h[\lambda_1(\theta_0)] = \beta\lambda_0$, and

$$\frac{1}{q} \frac{\mu(\theta_0)}{h(\theta_0)} = \frac{1 - H(\theta_0)}{h(\theta_0)} \{\mathbb{E}_h[\lambda_1(\theta_0')|\theta_0' > \theta_0] - \mathbb{E}_h[\lambda_1(\theta_0)]\} = MB_0(\theta_0)$$

where the last equality uses (48). The inequality $MB_0(\theta_0) \geq 0$ holds because $v_0(\theta_0)$ is increasing in θ_0 (by the IC constraint) and the multiplier $\lambda_1(\theta_0)$ is increasing in $v_0(\theta_0)$.

For the case $\kappa = 0$, notice we can set $\lambda_0 = 0$ and because $w_0(\underline{\theta}_0)$ is not a free variable we only have the boundary condition $\mu(\overline{\theta}_0) = 0$. Integrating the LOM:

$$\mu(\theta_0) = \mu(\underline{\theta}_0) - qH(\theta_0)\mathbb{E}\left[\lambda_1(\theta_0')|\theta_0' < \theta_0\right].$$

So we have

$$\mu(\underline{\theta}_0) = \mathbb{E}\left[\lambda_1(\theta_0)\right]$$

and

$$\frac{1}{q}\frac{\mu(\theta_0)}{h(\theta_0)} = \frac{1-H(\theta_0)}{h(\theta_0)}\mathbb{E}\left[\lambda_1(\theta_0')|\theta_0' > \theta_0\right].$$

Finally, if θ_0 is fixed we can drop the incentive constraint so $\mu(\theta_0) = 0$ and $MB_0(\theta_0) = 0$.

Proof of Proposition 5. This proof also follows similar steps to Theorem 1 in Hellwig (2021). Using the characterization of λ_t in Proposition 3 and substitute the multipliers $\lambda_{t+1}(\theta^t)$ and $\gamma_{t+1}(\theta^t)$ from the optimality conditions (42) and (43), and using equation (38) to substitute for ξ_t:

$$\frac{1}{u'(\theta^t)} + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta_{t-1})}\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} = \frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] + \frac{\mu(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\hat{\mathbb{E}}\left(\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right), \quad (49)$$

where we can rewrite

$$\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] = \hat{\mathbb{E}}\left[\rho(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right].$$

To show this, note that

$$\begin{aligned}\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] &= \int_{\underline{\theta}}^{\overline{\theta}} \mathcal{E}(\theta_{t+1},\theta_t)\frac{\varphi(\theta_{t+1}|\theta_t)m(\theta^{t+1})}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}d\theta_{t+1} \\ &= \frac{1}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}\int_{\underline{\theta}}^{\overline{\theta}}\left(-\int_{\theta_{t+1}}^{\overline{\theta}}\mathcal{E}(\theta',\theta_t)\varphi(\theta'|\theta_t)d\theta'\right)' m(\theta^{t+1})d\theta_{t+1}.\end{aligned}$$

Integrate by parts and use $\mathbb{E}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta_t\right] = \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial\varphi(\theta_{t+1}|\theta_t)}{\partial\theta_t}d\theta_{t+1} = 0$. Then using the definition of $\rho(\theta^{t+1})$ and $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}$,

$$\begin{aligned}
\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] &= \int_{\underline{\theta}}^{\overline{\theta}}\int_{\theta_{t+1}}^{\overline{\theta}} \mathcal{E}(\theta_{t+1},\theta_t)\varphi_{t+1}(\theta'|\theta^t)d\theta' \frac{m'(\theta^{t+1})}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}d\theta_{t+1} \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\varphi_{t+1}(\theta_{t+1}|\theta^t)}\int_{\theta_{t+1}}^{\overline{\theta}} \mathcal{E}(\theta',\theta^t)\varphi_{t+1}(\theta'|\theta^t)d\theta' \frac{m'(\theta^{t+1})}{m(\theta^{t+1})}\frac{m(\theta^{t+1})}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1} \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \rho(\theta^{t+1})\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}\hat{\varphi}_{t+1}(\theta_{t+1}|\theta^t)d\theta_{t+1}.
\end{aligned}$$

Substitute back and use Equation (46) to substitute $\frac{\mu_t(\theta^t)}{\varphi(\theta_t|\theta^{t-1})}$:

$$\frac{1}{u'(\theta^t)}+\frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\frac{\tau^k(\theta^t)}{u'(\theta^t)}\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} = \frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right]+\frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\frac{\tau^k(\theta^t)}{u'(\theta^t)}\hat{\mathbb{E}}\left[\rho(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right].$$

Finally, rearranging we get

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \left\{1+\underbrace{\left[\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} - \hat{\mathbb{E}}\left[\rho(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right]\right]\frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\tau^k(\theta^t)}_{\equiv s(\theta^t)}\right\}\frac{1}{u'(\theta^t)}.$$

Proof of Proposition 6. If $s(\theta^t) \geq 0$,

$$\frac{1}{u'(\theta^t)} \leq \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{\mathbb{E}\left[M(\theta^{t+1})|\theta_t\right]}{\mathbb{E}\left[u'(\theta^{t+1})M(\theta^{t+1})|\theta_t\right]},$$

where $M(\theta^{t+1}) = \frac{m(\theta^{t+1})}{u'(\theta^{t+1})}$, rearranging

$$\frac{\mathbb{E}\left[u'(\theta^{t+1})M(\theta^{t+1})|\theta_t\right]}{\mathbb{E}\left[M(\theta^{t+1})|\theta_t\right]} \leq u'(\theta^t).$$

Because $u'(\theta^{t+1})$ is decreasing, to show $\mathbb{E}\left[u'(\theta^{t+1})|\theta^t\right] \leq u'(\theta^t)$ we only need to show that $M(\theta^{t+1})$ is weakly decreasing. Differentiating

$$\frac{d}{d\theta_{t+1}}\left(M(\theta^{t+1})\right) = \underbrace{M(\theta^{t+1})\frac{u''(\theta^{t+1})}{u'(\theta^{t+1})}}_{\leq 0}\left(\iota f_\theta(\theta^{t+1}) - c'(\theta^{t+1})\right),$$

from the the local IC constraint $\frac{dw(\theta^{t+1})}{d\theta_{t+1}} = \frac{\partial w(\theta^{t+1})}{\partial \theta_{t+1}}$ we have $c'(\theta^{t+1}) + \beta \frac{\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right)}{u'(\theta^{t+1})} = \iota f_\theta(\theta^{t+1})$. Then, the assumption $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$ implies $\iota f_\theta(\theta^{t+1}) - c'(\theta^{t+1}) \geq 0$, and so $M(\theta^{t+1})$ is weakly decreasing. For the second part, assume $s(\theta^t) \geq 0$ for all θ^t, then u' follows a non-negative super-martingale. By Doob's super-martingale convergence theorem, u' converges almost surely to a finite limit. By contradiction, assume u' converges to a positive limit $\overline{u'} > 0$. Then, almost sure convergence implies that for some τ we have $u'(\theta^\tau) = u'(\theta^\tau, \theta_{\tau+1}) = ... = \overline{u'}$, which would violate incentive compatibility.[31] Hence, we must have $u' \to 0$ almost surely.

C Extensions

C.1 Limited commitment

In this section, I relax the assumption of full commitment of the entrepreneur. Limited commitment leads to firm size and compensation dynamics that are very different from those with the private information friction. The limited commitment works as follows. At every period, before knowing the realization of his productivity, the entrepreneur can divert and consume all the funds advanced by the lender and terminate the project. In this case, I assume the entrepreneur would obtain utility $h(k_{t+1}(\theta^t))$, where h is increasing and concave. Therefore, the agent will not terminate the project at period $t+1$ if $h(k_{t+1}(\theta^t)) \leq v_t(\theta^t)$.[32] This limited commitment constraint can be added directly to the planning problem (15). Because the limited commitment constraint does not affect the within-period insurance and incentives trade-off, the characterization of the shadow cost of information rents (Proposition 3) is not affected by the limited commitment assumption.

However, the limited commitment constraint does modify the consumption and firm size dynamics. Let $\eta_t(\theta^t)$ be the multiplier on the limited commitment constraint. Then, the GIEE is given by

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{1}{u'(\theta^t)}(1 + s(\theta^t)) + \frac{\eta_t(\theta^t)}{\beta}.$$

[31] Formally, this is statement is true because the proposition assumes fixed capital. Otherwise, this would be incentive compatible by setting $k_{t+1}(\theta^t) = 0$, which is generally not optimal because of the Inada condition $\lim_{k\to 0} f_k(k,\theta) = \infty$.

[32] A natural specification of the function h is $\frac{u((1-\iota)(1-q)k_{t+1})}{1-\beta}$, this is the value that the agent would obtain if he could keep a fraction $(1-\iota)$ of the capital and then save outside the contract at rate $\frac{1}{q}$.

Because $\eta_t(\theta^t) \geq 0$, the limited commitment gives a force to have a downward drift in marginal utilities. As is well known, in models with only limited commitment, the agent's consumption is backloaded, and consumption follows a sub-martingale. Therefore, the private information and limited commitment frictions will generally have opposite effects on consumption dynamics.

The investment wedge is now given by

$$\tau^{k,LC}(\theta^{t+1}) = \tau^k(\theta^t) + \eta_t(\theta^t)\frac{h'(k_{t+1}(\theta^t))}{f_k(k_{t+1}(\theta^t), \theta_t)} \geq 0,$$

where $\tau^k(\theta^t)$ is the wedge from the private information friction in Proposition 2. Because $\eta_t(\theta^t)h'(k_{t+1}(\theta^t)) \geq 0$, the limited commitment friction also lowers firm size relative to the first. However, if the promised utility increases over time, the limited commitment constraint will eventually not bind ($\eta_t(\theta^t) = 0$). Therefore, this friction still gives an incentive to have firm size increasing over time.

C.2 Endogenous termination

In this section, I show how the model can be extended to allow for endogenous termination of the contract. As is well known, in regions of the state space where the contract becomes very inefficient, the principal may be better off terminating the project or randomizing between terminating and continuing the contract at an efficient point. I assume that after termination, the lender receives a scrap value S. At period t, based on θ^t, the lender can choose a probability $\alpha_{t+1}(\theta^t)$ of termination at $t+1$. In that event, the principal can also give the entrepreneur a compensation of $Q_{t+1}(\theta^t)$. In case of no termination at period t the objective of the principal is

$$\int \left[-b(\theta^t) + \alpha_{t+1}(\theta^t)q\left(S - Q_{t+1}(\theta^t)\right) + \left(1-\alpha_{t+1}(\theta^t)\right)\left(k_{t+1}(\theta^t) + qK_{t+1}(v_t(\theta^t), \Delta_t(\theta^t), \theta^t, k_{t+1}(\theta^t))\right)\right]\varphi_t(\theta_t|\theta_{t-1})d\theta_t.$$

I assume that after terminating the contract, the entrepreneur can freely save $Q_{t+1}(\theta^t)$ and obtains a per period gross return $\frac{1}{q}$. Then, his value after terminating the contract is $\frac{u\left((1-q)Q_{t+1}(\theta^t)\right)}{(1-q)}$. The continuation utility now becomes

$$w_t(\theta^t) = u(c(\theta^t)) + \beta\left[\alpha_{t+1}(\theta^t)\frac{u\left((1-q)Q_{t+1}(\theta^t)\right)}{(1-q)} + (1-\alpha_{t+1}(\theta^t))v_t(\theta^t)\right],$$

and the local IC

$$\dot{w}_t(\theta^t) = u'(c(\theta^t))\iota f_\theta(k_t, \theta_t) + \beta(1 - \alpha_{t+1}(\theta^t))\Delta_t(\theta^t).$$

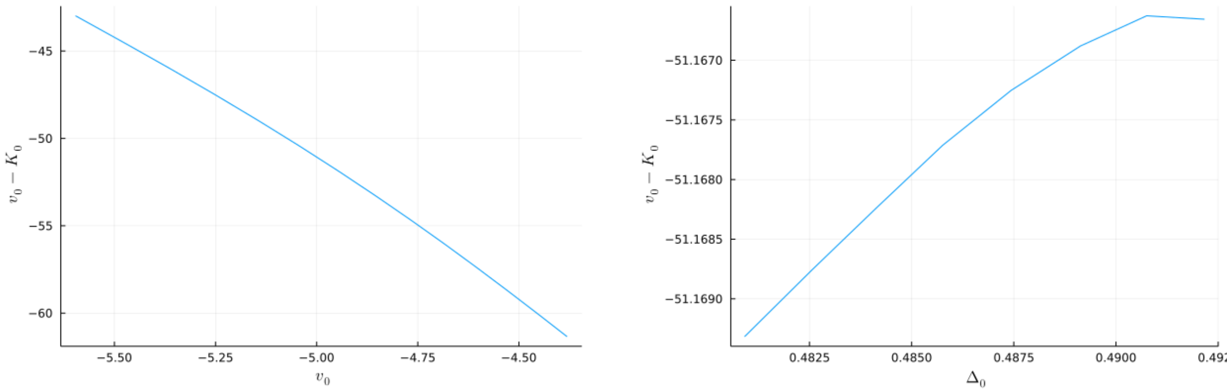
It is then easy to see that the optimality conditions for $b(\theta^t)$, $k_{t+1}(\theta^t)$,$v_t(\theta^t)$, $\Delta_t(\theta^t)$ and $w_t(\theta^t)$ are the same as in the main model. Therefore, although it may be optimal to terminate the contract, the characterizations of the optimal allocation presented in the paper do not rely on the assumption of no termination.

It is interesting to understand when and where (in the state space) termination may occur in this model and how it compares with the risk neutral case. With risk neutrality (Clementi *et al.* (2010)) the Pareto frontier is increasing in the promised utility. Consequently, termination occurs when the promised utility is low. Moreover, because utilities drift upward, the termination probability tends to decrease over time. With risk aversion, the Pareto frontier is decreasing in promised utility (see Figure 14a). So, the motives for termination differ from the risk neutral case.

Two other sources of inefficiencies could motivate the termination of the firm. First, low promised information rents (i.e. low Δ_0) is inefficient. Panel 14b shows that the Pareto frontier is increasing in Δ_0. Over time, if the lender has promised too low information rents to the entrepreneur, termination could potentially become optimal. Second, introducing a limited commitment constraint as in Section C.1 could also generate endogenous termination. If $v_t(\theta^t)$ decreases sufficiently, the limited commitment constraint may require that $k_{t+1}(\theta^t) \to 0$. Then, as shown in Dovis (2019), the Inada condition $\lim_{k\to 0} f_k(k, \theta) = \infty$ implies that in this region, the Pareto frontier is increasing in v. When the frontier is increasing, there may be a range of scrap values S where it is optimal for the lender to randomize between termination and continuing at a higher v. Because the shadow costs of information rents and the variance of promised utility tend to increase, both inefficiencies should imply that the termination probabilities tend to increase over time. Again, these are the opposite dynamics of what is found with risk neutrality.

Figure 14: Pareto frontier

(a) As a function of promised utility v_0 (b) As a function of promised information rent Δ_0



Note: For Panel (a), the promised information rent is set at the optimal level at $t = 0$, i.e. I set $\gamma_0 = 0$. For Panel (b), λ_0 is adjusted for every value of Δ_0 so that v_0 is kept fixed.

Persistent shocks and optimal termination probabilities. Another interesting observation is that whenever termination is optimal, the lender may have more incentives to increase the termination probabilities when the persistence of the shocks is higher. The reason is that a higher termination probability decreases the dynamic information rents. The intuition is similar to the equity purchases and the distortions in firm size. Imagine that, at history θ^{t-1}, the lender increases the termination probability of type θ_t and compensates him by increasing $Q(\theta^t)$ such that his ex-ante continuation utility is kept constant. Types $\theta' > \theta_t$ know they are expected to obtain higher returns at $t+1$, so they have a relatively higher preference for continuing to operate the firm. Therefore, the increase in $\alpha(\theta^t)$ makes deviations less attractive for $\theta' > \theta_t$, and so it lowers the cost of screening types.

I show this intuition more formally in a simplified two-period and two-type version of the model. Assume the entrepreneur's productivity can take values $\{\theta^H, \theta^L\}$ with $\theta^H > \theta^L$. In the first period, $P(\theta_1 = \theta^H) = p^1$, and in the second one, $P(\theta_2 = \theta^H|\theta_1 = \theta^H) = p^H$ and $P(\theta_2 = \theta^H|\theta_1 = \theta^L) = p^L$. Let $\rho = p^H - p^L \geq 0$, if $\rho > 0$ we say types are persistent. We assume the production function is of the form $f(\theta, k) = \theta$, so we can abstract away from the choice of firm size. In the second (and last) period, we assume that the entrepreneur consumes all its endowment, so there is no repayment. The principal's objective is

$$K(v) = p^1 \left[-b^H + q\alpha^H(S - Q^H)\right] + (1 - p^1)\left[-b^L + q\alpha^L(S - Q^L)\right].$$

The values of the high and low types are

$$w^H = u(\theta^H - b^H) + \beta\left[\alpha^H u(Q^H) + (1-\alpha^H)\mathbb{E}^H\left(u(\theta_2)\right)\right]$$
$$w^L = u(\theta^L - b^L) + \beta\left[\alpha^L u(Q^L) + (1-\alpha^L)\mathbb{E}^L\left(u(\theta_2)\right)\right],$$

where for $j \in \{H, L\}$, $\mathbb{E}^j\left(u(\theta_2)\right) = p^j u(\theta^H) + (1-p^j)u(\theta^L)$. The participation constraint is

$$p^1 w^H + (1-p^1)w^L = v$$

and the IC constraint can be written as

$$w^H = w^L + \underbrace{u(\theta^H - b^L) - u(\theta^L - b^L)}_{\text{static info rent}} + \underbrace{(1-\alpha^L)\beta\rho(u(\theta^H) - u(\theta^L))}_{\text{Dynamic info rent}}.$$

Notice that the dynamic information rent is increasing in ρ and decreasing in α^L. I directly assume that the parameters are such that $\alpha^L \in (0,1)$ is optimal and show that the principal increases the termination probability when the persistence increases.

Proposition 7. *If $\alpha^L \in (0,1)$ is optimal, the optimal contract is such that $\frac{\partial \alpha^L}{\partial \rho} > 0$.*

Proof. The proof is as follows. Starting from the optimal contract, we consider a perturbation where we increase α^L while preserving the IC and PK constraints, and show that the resource gains are increasing in ρ . To this end, let $\Delta\alpha^L = \varepsilon > 0$, for ε small. We perturb the allocation along the low type's indifference curve, so to keep w^L constant, we increase Q^L by

$$Q^L = \frac{\left[u(Q^L) - \mathbb{E}^L\left(u(\theta_2)\right)\right]}{\alpha^L u'(Q^L)}\varepsilon^L.$$

The perturbation lowers the dynamic information rents, and so it relaxes the IC constraint. This allows us to lower the high type's period one utility by

$$\Delta u^H = -\beta\rho(u(\theta^H) - u(\theta^L))\varepsilon.$$

Because w^L is kept fixed, this changes the ex-ante utility by $p^1\Delta u^H$. Then, to satisfy the PK constraint, we increase the period one utility of both types in an incentive-compatible manner. Because information rents depend on consumption, increasing utilities uniformly would not be incentive compatible. If we increase the low type's utility by Δu^L, the IC constraint requires increasing the utility of the high type by

$$\Delta u^{H,IC} = \frac{u'(\theta^H - b^L)}{u'(\theta^L - b^L)} \Delta u^L.$$

The ex-ante utility is kept fixed if

$$p^1 \frac{u'(\theta^H - b^L)}{u'(\theta^L - b^L)} \Delta u^L + (1 - p^1)\Delta u^L = -p^1 \Delta u^H,$$

which implies

$$\Delta u^L = -\frac{p^1 u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1 - p^1)u'(\theta^L - b^L)} \Delta u^H.$$

Therefore, the total change in the high type utility is

$$\begin{aligned} \Delta u^{H,TOT} &= \Delta u^H + \Delta u^{H,IC} \\ &= (1 - p^1) \frac{u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1 - p^1)u'(\theta^L - b^L)} \Delta u^H. \end{aligned}$$

Finally, the resource gain from this perturbation is

$$\begin{aligned} \frac{\Delta K}{\varepsilon} &\approx p^1 \left[\frac{1}{u'(\theta^H - b^H)} \Delta u^{H,TOT} \right] + (1 - p^1) \left[\frac{1}{u'(\theta^L - b^L)} \Delta u^L \right] + \Omega \\ &\approx \left[\frac{1}{u'(\theta^L - b^L)} - \frac{1}{u'(\theta^H - b^H)} \right] \frac{u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1 - p^1)u'(\theta^L - b^L)} (1 - p^1)p^1 \beta\rho(u(\theta^H) - u(\theta^L)) + \Omega \end{aligned}$$

where Ω collects all the terms that do not depend on ρ. Because the initial allocation is optimal, $\left[\frac{1}{u'(\theta^L - b^L)} - \frac{1}{u'(\theta^H - b^H)} \right] < 0$ and $u(\theta^H) - u(\theta^L) > 0$. Therefore, the principal's resource gain from this perturbation is increasing in ρ, i.e. $\frac{\partial \frac{\Delta K}{\varepsilon}}{\partial \rho} < 0$, which implies that $\frac{\partial \alpha^L}{\partial \rho} > 0$ is optimal. □

C.3 Screening model: divert funds before investing

In this section, I study a screening version of the model where the entrepreneur can choose what fraction of the funds available he invests in the project. The remaining funds are secretly diverted for consumption. Now, the lender can observe the entrepreneur's returns but not the entrepreneur's productivity nor invested and diverted funds. In this sense, the investment decision is similar to the labor/leisure choice in the Mirrlees taxation problem.

This model yields the same characterization of the shadow costs μ_t, the GIEE, and the firm size dynamics. Moreover, we can directly define the investment wedge $\tau^k(\theta^t)$ as the wedge between invested and diverted funds relative to the first best.

Denote by B_t the funds advanced by the lender. The entrepreneur can use these funds to invest in the project k_t, but he can also divert a portion a_t of the funds for his consumption. Therefore, invested and diverted funds are subject to the constraint

$$k_t + a_t \leq B_t. \tag{50}$$

The lender now observes returns $f(k_t, \theta_t)$ but not productivity θ_t and how funds are used, i.e. k_t and a_t. Diverted funds are converted into consumption units according to the function $g(a_t)$, with $g'' \leq 0 < g'$, so the entrepreneur's consumption is

$$c_t = f(k_t, \theta_t) - b_t + g(a_t). \tag{51}$$

The principal's within period objective now is $B_t - b_t$. The envelope condition of the agent's problem is

$$\frac{\partial}{\partial \theta_t} w_t(\theta^t) = u'(c_t(\theta^t)) f_\theta(k_t(\theta^t), \theta_t) + \beta \Delta_t(\theta^t).$$

Now the investment wedge can be defined explicitly as the distortion in invested and diverted funds relative to the first best (where we would have $f_k(k_t(\theta^t), \theta_t) = g'(a_t(\theta^t))$. Define

$$\tau^k(\theta^t) \equiv 1 - \frac{g'(a(\theta^t))}{f_k(k(\theta^t), \theta_t)}.$$

The rest of the planning problem is the same but with the extra flow of funds constraint (50). The optimality condition for diverted funds is

$$\zeta_t(\theta^t) = g'(a_t(\theta^t)),$$

where $\zeta_t(\theta^t)$ is the multiplier on constraint (50). The FOC for investment is

$$\zeta_t(\theta^t) = f_k(k_t(\theta^t), \theta_t) - \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} u'(\theta^t) f_{\theta k}(k_t(\theta^t), \theta_t)$$

Then, combining the two optimality conditions, we get

$$\tau^k(\theta^t) = \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\frac{f_{\theta k}(\theta^t)}{f_k(\theta^t)}u'(\theta^t) > 0,$$

which is the same as in Proposition 2. Because $\tau^k(\theta^t) > 0$, there is more cash diversion than in the first best. This is the standard screening result; the principal distorts effort (here investment k_t) downwards to screen types at a lower cost. When shadow costs ($\mu_t(\theta^t)$) are high, the principal increases distortions to reduce the costs of screening types. Moreover, this wedge also captures the distortions to firm size as in the cash flow diversion model. Combining the FOC for $B_{t+1}(\theta^t)$ and the envelope condition, we get

$$\begin{aligned}\frac{1}{q} &= \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1}) - \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_{t+1})}u'(\theta^{t+1})f_{\theta k}(k_{t+1}(\theta^t), \theta_{t+1})|\theta_t\right] \\ &= \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})\left(1 - \tau^k(\theta^{t+1})\right)|\theta_t\right],\end{aligned}$$

which is exactly how the wedges were defined for the cash flow diversion model in equation (19). Then, it is also easy to verify that this model yields the same characterization for the shadow costs $\mu_t(\theta^t)$ and the GIEE.

D Full implementation in two-period model

In this section, I study a full implementation in a simplified two-period version of the model. Now, the entrepreneur can freely choose between paying himself dividends or reinvesting in the firm. This implementation focuses on another way the principal can lower dynamic information rents. Instead of buying equity, the principal discourages reinvesting profits to distort firm size. Because lower future capital is less attractive for higher types if shocks are persistent, this margin can be used to screen types. This implementation of the optimal allocation involves either a (nonlinear) subsidy on dividend payouts or a (nonlinear) tax on reinvested funds so that capital is lowered in period 2 relative to the first best. The magnitudes of the marginal subsidy (or tax) are increasing in the persistence of productivity.

There are two periods $t = 1, 2$. I assume the production function is of the form $f(k, \theta) = \theta f(k)$; for the first period, I normalize $f(k_1) = 1$ and assume there is no cost of diverting funds $\iota = 1$. The entrepreneur is risk neutral in the second period and there is no repayment.

So the entrepreneur's utility if he is type θ_1 and reports $\widetilde{\theta}_1$ is

$$w(\theta_1, \widetilde{\theta}_1) = u(\theta_1 - b(\widetilde{\theta}_1)) + \beta \mathbb{E}\left[\theta_2|\theta_1\right] f(k_2(\widetilde{\theta}_1)).$$

Solving the lender's problem, we obtain

$$u'(\theta_1 - b(\theta_1)) = \beta \mathbb{E}\left[\theta_2|\theta_1\right] f'(k(\theta_1)) \left(1 - \tau^k(\theta_1)\right),$$

where

$$\tau^k(\theta_1) = \widetilde{\mu}(\theta_1) \frac{1 - \Phi(\theta_1)}{\varphi(\theta_1)} \left(\frac{\frac{\partial \mathbb{E}[\theta_2|\theta_1]}{\partial \theta_1}}{\mathbb{E}\left[\theta_2|\theta_1\right]} - \frac{u''(\theta_1 - b(\theta_1))}{u'(\theta_1 - b(\theta_1))} \right) > 0.$$

The wedge to investment is increasing in the persistence of the process, as with higher persistence, higher types have an even higher preference for future capital. But now, it also depends on the absolute risk aversion.[33] We now turn to the implementation. The entrepreneur can freely use his returns to pay dividends d or reinvest in the firm I. The optimal can be implemented with either a tax T on investment such that $k_2 = T(I)$ or a subsidy on dividend payments $c = S(d)$. Here, I consider only the subsidy on dividends, so $I = k_2$. Then the entrepreneur's problem is

$$\begin{aligned} w(\theta_1) &= \max_{d, k_2} u(S(d)) + \beta \mathbb{E}\left[\theta_2|\theta_1\right] f(k_2) \\ s.t \quad & d + k_2 = \theta_1. \end{aligned}$$

The marginal subsidy on dividend payments that implements the optimum is

$$S'(d(\theta_1)) = \frac{1}{1 - \tau^k(\theta_1)}.$$

Moreover, we have $S'(d(\overline{\theta})) = S'(d(\underline{\theta})) = 1$ and $S'(d(\theta)) > 1$ for $\theta \in (\underline{\theta}, \overline{\theta})$, so the marginal subsidy is inverse U-shaped. Because $\tau^k(\theta_1)$ is increasing in the persistence of productivity, the marginal subsidy will also be increasing in the persistence.

[33] Because there is no repayment b_2 in the second period, capital also plays a similar role as savings for the entrepreneur.

E Details numerical simulations

I follow a similar procedure as Farhi and Werning (2013), Stantcheva (2017) and Ndiaye (2020). In these papers (and in Kapička (2013) and Golosov *et al.* (2016a)), the model is solved with a geometric random walk process. This allows to normalize the principal's optimization problem and drop θ_{t-1} as a state variable. Here, the problem can also be normalized if the production function is assumed to be of the form $f(k,\theta) = z\theta^{1-\alpha}k^{\alpha}$. However, I am interested in performing comparative statics with respect to the persistence of the process (ρ). Therefore, I solve the full problem without renormalizing.

It is convenient to transform the problem to write the Hamiltonian as a function of the current shock ε_t instead of the current productivity θ_t. Denote the density function of the shock by $g_\varepsilon(\varepsilon_t)$, then it follows that

$$\varphi\left(\theta_t \mid \theta_{t-1}\right) = \frac{g_\varepsilon(\varepsilon_t)}{\theta_{t-1}^{\rho}}.$$

Moreover, we also have that

$$\frac{\partial \varphi\left(\theta_t \mid \theta_{t-1}\right)}{\partial \theta_{t-1}} = -\frac{\rho}{\varepsilon_t \theta_{t-1}^{1+\rho}} \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \frac{\left(\log \theta_t - \rho \log \theta_{t-1} - \mu\right)}{\sigma_\varepsilon^2} \exp\left\{ -\frac{\left(\log \theta_t - \rho \log \theta_{t-1} - \mu\right)^2}{2\sigma_\varepsilon^2} \right\}$$

and

$$\frac{\partial g_\varepsilon(\varepsilon_t)}{\partial \varepsilon_t} = -\frac{1}{\varepsilon_t^2 \sigma_\varepsilon \sqrt{2\pi}} \frac{\left(\log \varepsilon_t - \mu\right)}{\sigma_\varepsilon^2} \exp\left\{ -\frac{\left(\log \varepsilon_t - \mu\right)^2}{2\sigma_\varepsilon^2} \right\}.$$

Therefore,

$$\tilde{g}_\varepsilon(\varepsilon_t) \equiv g_\varepsilon(\varepsilon_t) + \varepsilon \frac{\partial g_\varepsilon(\varepsilon_t)}{\partial \varepsilon_t} = \frac{\theta_{t-1}^{1+\rho}}{\rho} \frac{\partial \varphi\left(\theta_t \mid \theta_{t-1}\right)}{\partial \theta_{t-1}}.$$

Then note that $d\theta_t = \theta_{t-1}^{\rho} d\varepsilon_t$ implies

$$\varphi\left(\theta_t \mid \theta_{t-1}\right) d\theta_t = g_\varepsilon(\varepsilon_t) d\varepsilon_t,$$

and

$$\frac{\partial \varphi\left(\theta_t \mid \theta_{t-1}\right)}{\partial \theta_{t-1}} d\theta_t = \rho \frac{\tilde{g}_\varepsilon(\varepsilon_t)}{\theta_{t-1}} d\varepsilon_t.$$

The planning problem over the shock ε_t is

$$
\begin{aligned}
K(v_{t-1},\Delta_{t-1},k_t,\theta_{t-1}) = \min \int & \left(k_{t+1}(\varepsilon_t) - b_t(\varepsilon_t) + qK(v_t(\varepsilon_t),\Delta_t(\varepsilon_t),k_{t+1}(\varepsilon_t),\theta^{\rho}_{t-1}\varepsilon_t)\right) g_\varepsilon(\varepsilon_t)d\varepsilon_t \\
s.t \quad (PK) \quad & w_t(\varepsilon_t) = u(c_t(\varepsilon_t)) + \beta v_t(\varepsilon_t) \qquad [g_\varepsilon(\varepsilon_t)\xi_t(\varepsilon_t)] \\
& v_{t-1} = \int w_t(\varepsilon_t) g_\varepsilon(\varepsilon_t) d\varepsilon_t \qquad [g_\varepsilon(\varepsilon_t)\lambda_{t-1}] \\
(IC) \quad & \dot{w}_t(\varepsilon_t) = \theta^{\rho}_{t-1}\left(u'(c(\varepsilon_t))\iota f_\theta(k_t,\theta^{\rho}_{t-1}\varepsilon_t) + \beta\Delta_t(\varepsilon_t)\right) \qquad [\mu_t(\varepsilon_t)] \\
& \Delta_{t-1} = \int w_t(\varepsilon_t)\frac{\rho}{\theta_{t-1}}\tilde{g}_\varepsilon(\varepsilon_t)d\varepsilon_t \qquad [g_\varepsilon(\varepsilon_t)\gamma_{t-1}] \\
& (Feasibility) \quad c_t(\varepsilon_t) = f(k_t,\theta^{\rho}_{t-1}\varepsilon_t) - b_t(\varepsilon_t)
\end{aligned}
$$

The optimality conditions are

$$
\frac{q}{\beta}\lambda_t(\varepsilon_t) = \frac{1}{u'(c_t(\varepsilon_t))}\left[1 + \frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)}\theta^{\rho}_{t-1}\iota f_\theta(k_t,\theta^{\rho}_{t-1}\varepsilon_t)u''(c(\varepsilon_t))\right] \tag{52}
$$

$$
\gamma_t(\varepsilon_t) = -\frac{\beta}{q}\theta^{\rho}_{t-1}\frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)} \tag{53}
$$

and the two LOM

$$
\dot{\mu}(\varepsilon_t) = -\left[\frac{q}{\beta}\lambda_t(\varepsilon_t) - \lambda_{t-1} + \gamma_{t-1}\frac{\rho}{\theta_{t-1}}\frac{\tilde{g}_\varepsilon(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)}\right] g_\varepsilon(\varepsilon_t) \tag{54}
$$

$$
\dot{w}_t(\varepsilon_t) = \theta^{\rho}_{t-1}\left(u'(c(\varepsilon_t))\iota f_\theta(k_t,\theta^{\rho}_{t-1}\varepsilon_t) + \beta\Delta_t(\varepsilon_t)\right). \tag{55}
$$

I truncate the distribution of ε_t at the 0.01 and 0.99 percentiles, the boundary conditions then need to be adjusted to $\mu(\overline{\varepsilon}) = -\gamma_{t-1}\frac{\rho}{\theta_{t-1}}\overline{\varepsilon}g_\varepsilon(\overline{\varepsilon})$ and $\mu(\underline{\varepsilon}) = -\gamma_{t-1}\frac{\rho}{\theta_{t-1}}\underline{\varepsilon}g_\varepsilon(\underline{\varepsilon})$.

To solve the model, the state space is modified to $(\lambda_-,\gamma_-,k,\theta_-)$, so the multipliers λ_- and γ_- are used instead of v_- and Δ_-, respectively. I use 20 grid points for λ_-, 14 for γ_-, 25 for k and 15 for θ_-. I interpolate on K, v and Δ with cubic splines and allow to extrapolate. To solve the model with an i.i.d type process, the algorithm is the same but with $\Delta = 0$ and without the state variables γ_- and θ_-.

Algorithm

Step 0: Guess the value function K', promised utility v' and promised marginal utility Δ' on the grid $(\lambda_-, \gamma_-, k, \theta_-)$

Step 1: Compute the policy functions for k_+ on a grid $(\lambda_{pol}, \gamma_{pol}, \theta)$ by minimizing

$$k_+ + qK'(\lambda_{pol}(i), \gamma_{pol}(i), k_+, \theta(i))$$

(Note: k_+ needs to be computed multiple times at every step while solving the ODE. But to improve speed, we can solve before the policies on a dense grid and then interpolate when solving the ode).

Step 2: For each point in $(\lambda_-, \gamma_-, k, \theta_-)$, solve the optimal control problem with a shooting method.

- a) Guess continuation utility of lowest type $w(\underline{\varepsilon}) = \underline{w}$.
- b) For each ε, solve $\lambda(\varepsilon)$ in equation (52) and $\gamma(\varepsilon)$ in equation (53). To compute $c(\varepsilon)$, first compute $k_+(\varepsilon)$ by interpolationg the array of policies on $(\lambda(\varepsilon), \gamma(\varepsilon), \theta_-^\rho \varepsilon)$. Then obtain $v(\varepsilon)$ by interpolation of v' on $(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^\rho \varepsilon)$ and solve

  $$c(\varepsilon) = u^{-1}\left(w(\varepsilon) - \beta v(\varepsilon)\right).$$

  With these solutions solve the differential equations (54) and (55). Note when solving (54) also need to interpolate Δ' on $(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^\rho \varepsilon)$.
- c) Check the boundary condition $\mu(\bar{\varepsilon}) = -\gamma_- \frac{\rho}{\theta_-} \bar{\varepsilon} g_\varepsilon(\bar{\varepsilon})$. If it does not satisfy the tolerance, go back to step a).

Step 3: Given the solution $(\mu(\varepsilon), w(\varepsilon))$, repeat step b) to obtain all policy functions on a grid $(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, also compute $b(\varepsilon) = f(k, \theta_-^\rho \varepsilon) - c(\varepsilon)$.

Step 4: Compute the lender's value function, promised utility and expected marginal utility at every grid point

$$v(\lambda_-, \gamma_-, k, \theta_-) = \int w(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) g_\varepsilon(\varepsilon_t) d\varepsilon_t$$

$$\Delta(\lambda_-, \gamma_-, k, \theta_-) = \int w(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) \frac{\rho}{\theta_-} \tilde{g}_\varepsilon(\varepsilon_t) d\varepsilon_t$$

$$K(\lambda_-, \gamma_-, k, \theta_-) = \int \left(k_+(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) - b(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) + qK'(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^\rho \varepsilon)\right) g_\varepsilon(\varepsilon_t) d\varepsilon_t.$$

Calculate the distance with previous guess of K', v' and Δ' , and repeat from **Step 1** until the convergence criteria is satisfied.

When solving the model without time-0 screening, for the Montecarlo simulation, the initial λ_0 and θ_0 can be fixed at arbitrary values. Because Δ_0 is a free variable, we must set $\gamma_0 = 0$. Then k_1 is chosen optimally given (λ_0, θ) and $\gamma_0 = 0$.

E.1 Time-0 problem

I also solve the time-0 problem with a shooting method by iterating on $v_0(\underline{\theta})$. To solve the problem, we use the value functions $v(\lambda, \gamma, \theta, k)$ and $\Delta(\lambda, \gamma, \theta, k)$ and the policy function $k(\lambda, \gamma, \theta)$ from the solution of the recursive problem.

Algorithm.

1. Guess the lowest type continuation utility $v_0(\underline{\theta})$
2. Solve the ODEs $\dot{v}(\theta_0) = \Delta_0(\theta_0)$ and $\dot{\mu}(\theta_0) = -\left[\lambda_1(\theta_0) - \lambda_0\right] g(\theta_0)$
   (a) At every step for every $(v(\theta_0), \mu(\theta_0))$
      i. Get $\gamma_1(\theta_0)$ from: $-\gamma_1(\theta_0) = \frac{\mu(\theta_0)}{g(\theta_0)}$
      ii. With $\gamma_1(\theta_0)$, θ_0 and $v_0(\theta_0)$ can back out the corresponding $\lambda_1(\theta_0)$. Note, v will be computed on a grid $(\lambda_1(\theta_0), \gamma_1(\theta_0), \theta_0, k_1(\theta_0))$, so for every $\lambda_1(\theta_0)$ need to get the optimal $k_1(\theta_0)$ (from the policy functions in the recursive problem) and then use this to get v until solve the root.
      iii. With $(\lambda_1(\theta_0), \gamma_1(\theta_0), \theta_0)$ can in turn also back out $\Delta_0(\theta_0)$
      iv. Update the LOM
3. Check $\mu(\overline{\theta}) = 0$, if it doesn't satisfy tolerance go back to 1. with a new guess
4. Repeat (a) to recover all the policy functions on a grid for θ_0

E.2 Check global IC constraints

The first-order approach consists of solving a relaxed problem where only the local incentive constraints are considered. A priori global incentive constraints may bind, in which case the solutions of the relaxed program (15) and the full program (9) would not coincide. I follow the approach outlined in Kapička (2013) and Farhi and Werning (2013) to verify ex-post that only the local incentive constraints bind.

The procedure is the following. First, after solving numerically the relaxed problem, we have obtained the policy functions $\lambda(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, $\gamma(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, $k_+(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$ and $b(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$ and the value function $v(\lambda_-, \gamma_-, k, \theta_-)$. Let $\widetilde{\varepsilon}$ denote the agent's report about the innovation to the productivity. Then we consider a problem where the entrepreneur takes as given the policy functions and can report any $\widetilde{\varepsilon} \in [\underline{\varepsilon}, \varepsilon]$ and verify that for every $(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$

$$
\begin{aligned}
\varepsilon = \arg\max_{\widetilde{\varepsilon} \in [\underline{\varepsilon}, \varepsilon]} & u(\widetilde{c}(\lambda_-, \gamma_-, k, \theta_-, \varepsilon, \widetilde{\varepsilon})) \\
& + \beta v(\lambda(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), \gamma(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), k_+(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), \theta_-^\rho \widetilde{\varepsilon}) \\
s.t \quad & \widetilde{c}(\lambda_-, \gamma_-, k, \theta_-, \varepsilon, \widetilde{\varepsilon}) = \iota f(k, \varepsilon) + (1-\iota) f(k, \widetilde{\varepsilon}) - b(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}).
\end{aligned}
$$

E.3 Solution implementation

With persistent shocks and constant equity, the entrepreneur's problem in the quasi-implementation is

$$
\begin{aligned}
\mathcal{W}(W_t, \theta_{t-1}, \widetilde{\theta}_{t-1}, \varepsilon_t) = & \max_{\widetilde{\theta}_t \leq \theta_{t-1}^\rho \varepsilon_t} u(\widetilde{c}_t) + \beta \mathcal{V}(W_{t+1}, \theta_{t-1}^\rho \varepsilon_t, \widetilde{\theta}_t) \\
s.t \qquad & W_{t+1} = qC(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) \\
& c_t = (1-q) C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) \\
& \widetilde{c}_t = c_t + \iota(f(k_{SB}, \theta_{t-1}^\rho \varepsilon_t) - f(k_{SB}, \widetilde{\theta}_t))
\end{aligned}
$$

where

$$
C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) = \frac{1}{q} W_t + \chi(f(k_{SB}, \widetilde{\theta}_t) + q\overline{f}(k_{SB}, \widetilde{\theta}_t) - \overline{f}(k_{SB}, \widetilde{\theta}_{t-1}))
$$

$$\overline{f}(k_{SB}, \theta_t) = \mathbb{E}\left[\sum_{\tau=1}^{\infty} q^{\tau-1} f(k_{SB}, \theta_{t+\tau}) | \theta_t\right]$$

and

$$\mathcal{V}(W_{t+1}, \theta_t, \widetilde{\theta}_t) = \int \mathcal{W}(W_{t+1}, \theta_t, \widetilde{\theta}_t, \varepsilon_{t+1}) g_{\varepsilon}(\varepsilon_{t+1}) d\varepsilon_{t+1}.$$

With i.i.d shocks, the problem is the same but without θ_{t-1} and $\widetilde{\theta}_{t-1}$ as state variables and with $\overline{f}(k_{SB})$ independent of θ_t. The problem is solved with standard value function iteration, and to have the closest comparison with the solutions of the optimal allocation, $\mathcal{V}(W_{t+1}, \theta_t, \widetilde{\theta}_t)$ is computed with numerical integration.