# Firm Size and Compensation Dynamics with Insurance and Persistent Private Information*


Gerard Maideu Morera†


June 21, 2022


**Abstract**

I study a dynamic cash flow diversion model between a risk neutral lender and *risk averse* entrepreneur that has *persistent private information* about the firm's productivity. I show that firm size drifts downwards and the entrepreneur's compensation is smoothed but features immiseration. These results contrast equivalent models with risk neutrality, where firm size tends to increase over time and dividends are paid once the undistorted first best size is reached. Next, I use numerical simulations to study a third best implementation. With persistent shocks, the lender gives the entrepreneur a time-varying equity share, with i.i.d shocks a constant equity share suffices. Then, the shares are pledged as collateral to smooth consumption. The implementation suggests that the opposite firm size dynamics result from the equity share drifting upwards with risk-neutrality but downwards with risk aversion and persistence. Finally, I discuss the implications for the validity of the Modigliani-Miller theorem and the investment-cash flow sensitivity.



*I am grateful to Christian Hellwig and Nicolas Werquin for their advice and guidance. I have benefited very much from advice and discussions from Charles Brendon and Fabrice Collard. I would also like to thank Martin Beraja, Jonas Gathen, Eugenia Gonzalez-Aguado, David Martimort, Alessandro Pavan, Jean-Charles Rochet, Andreas Schaab, Stéphane Villeneuve and seminar participants at the TSE Macro Workshop.



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

# 1 Introduction

Financing constraints slow down firms' growth over their lifecycle. Theories of the origin of financing frictions are needed to understand whether financing constraints are efficient given some underlying agency frictions, or they could be corrected through policy. Dynamic contracting models have proved to be useful in answering these questions. The canonical setting in this literature is the cash flow diversion model: an entrepreneur needs funds from a lender to operate a project, but only the entrepreneur observes the project cash flows and can secretly divert them for consumption. A regular outcome of this class of models is that, in the optimal contract, the firm size drifts upwards, and the entrepreneur is compensated once the undistorted first best size is reached (Clementi and Hopenhayn (2006))[1].

The literature has typically assumed that the entrepreneur is risk neutral and that the shocks to the firm's cash flows are i.i.d (Clementi and Hopenhayn (2006), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007a), Biais *et al.* (2007)). However, by making these two assumptions, these models may abstract from first-order concerns for designing financial contracts. First, with risk neutrality, there is no need to smooth the entrepreneur's consumption, so the compensation can be backloaded at little cost. Second, the i.i.d assumption restricts the extent to which the entrepreneur may have more information about firm's future profitability.

In this paper, I study a dynamic contracting problem between a risk neutral lender and a *risk averse* entrepreneur that has *persistent private information* about the firm's productivity. I solve for the optimal contract and analyze the implied firm size and compensation dynamics. Together, risk aversion and persistence lead to remarkably different dynamics than models previously studied. I find that firm size (i.e. working capital invested) drifts downwards. Moreover, the entrepreneur's compensation is smoothed, but the variance of consumption is permanently increasing (i.e. it features immiseration Thomas and Worrall (1990), Atkeson and Lucas (1992)). These dynamics are shown theoretically but also illustrated with numerical simulations. Behind these different firm dynamics lies a fundamentally distinct role of the state variables of the problem with risk aversion and risk neutrality. As I show with the implementation, promised utility maps to the entrepreneur's private wealth with risk aversion, but it maps to his equity on the firm with risk neutrality. When productivity is persistent and the entrepreneur is risk averse, the equity share is instead related to another

[1]Empirically, it is not obvious that the firms' financing constraints are eventually relaxed as they age. For instance, in developing economies, where financing constraints are more stringent, we observe that old firms are relatively smaller than in developed economies (Hsieh and Klenow (2014)).

state variable, the promised insurance.

The firm's size depends only on the magnitude of the investment wedges. They capture the fact that higher capital increases information rents, which is more costly when the lender wants (or has promised) to provide more insurance to the entrepreneur. Therefore, higher wedges lower the implicit marginal product of capital and firm size. These wedges are tightly connected to labor wedges in dynamic Mirrlees models. Farhi and Werning (2013) and Makris and Pavan (2020) have shown that insurance (and so wedges) tends to increase over time. This is also the case in this model; with *both* persistence and risk aversion, investment wedges increase over time, so the firm's size will decrease. With i.i.d private information, the investment wedges are stationary, and so is firm size.

The entrepreneur's consumption process satisfies a Generalized Inverse Euler Equation (GIEE) similar to Hellwig (2021). As expected, with risk aversion, the entrepreneur's compensation is smoothed. In the GIIE, the martingale properties of the entrepreneur's marginal utility depend only on the sign of a savings wedge. This wedge captures how much savings affect information rents at periods $t$ and $t + 1$. Except (possibly) for very high persistence[2], the optimal contract features immiseration. That is, marginal utility drifts upwards and the variance of consumption increases over time without bound (Thomas and Worrall (1990), Atkeson and Lucas (1992)).

To further understand the compensation dynamics, I use numerical simulations and analyze (third best) implementations with simpler contracts. With i.i.d shocks, the following simple contract gets very close to the optimal allocation. The lender gives the entrepreneur a constant equity share on the firm's reported cash flows. Then the entrepreneur can pledge his shares as collateral and borrow to 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 the optimal contract[3].

With persistent private information, the principal's problem contains an extra state variable that captures the insurance promised to the agent. This state variable naturally maps to the equity share given to the entrepreneur. Thus, persistence can be accommodated by allowing

[2]With CARA utility and fixed capital, I show that there will be immiseration whenever there is some mean-reversion in the productivity process, consistent with the findings of Bloedel *et al.* (2018) and Bloedel *et al.* (2020). As I discuss, varying capital generates an extra force for immiseration.

[3]Pledging shares aligns the entrepreneur's consumption with the firm's value, but without having to sell shares and independently of dividend payout policies. In this model selling shares may not be optimal, as lowering the entrepreneur's stake on the firms increases his incentives to divert funds. This rationale is consistent with the primary motive for pledging shares estimated in Fabisik (2019): obtain liquidity while maintaining ownership.

for a time-varying equity share. Intuitively, because types $\theta'' > \theta'$ know they are expected to obtain higher cash flows at $t+1$ than $\theta'$, it is less attractive for them to give up equity. So, when the lender buys equity at $t+1$ to some type $\theta'$, it discourages the diversion of funds for types $\theta'' > \theta'$. That is, the lender optimal lowers the equity share at period $t+1$ (to an inefficient level once at $t+1$) because it helps screen types at $t$.

The implementation clarifies the discrepancy of the firm size dynamics with risk neutrality and risk aversion. With risk neutrality, it is optimal to reward the entrepreneur solely through a higher stake on the project to minimize diversion incentives. Therefore, the promised utility can be mapped to the entrepreneur's equity (Clementi and Hopenhayn (2006)). This is no longer the case with risk aversion. As I show, promised utility better maps to the entrepreneur's wealth, and promised marginal utility (or insurance) maps to the entrepreneur's equity share. Consequently, both models obtain a positive relation between equity and firm size[4]. However, with risk neutrality, the equity share drifts upwards, but with risk aversion and persistence, it drifts downwards. Therefore, breaking the tight relation between equity and firm size may be necessary to simultaneously be consistent with the firm size and compensation dynamics observed in the data. Otherwise, firm size converges to the first best level only if the entrepreneur becomes the firm's sole owner.

The distinction between promised utility and promised insurance in the implementation has broader implications. To illustrate this, I revisit the implications of the model for two classical questions in the corporate finance literature. The first one concerns the role of capital structure on the firm's value (the Modigliani-Miller theorem). With risk neutrality, the firm's value does depend on the promised utility given to the entrepreneur (Clementi and Hopenhayn (2006)). Instead, with risk aversion, numerical simulations show that firm value is approximately independent of promised utility, but it is decreasing on the amount of insurance promised. This observation corroborates the idea that, with risk aversion, promised utility maps to the entrepreneur's private wealth and is unrelated to the firm's capital structure. The second asks whether a firm's financing constraints can be inferred from the sensitivity of investment to cash flows (Fazzari *et al.* (1988), Kaplan and Zingales (1997)). Numerically, I find a slightly higher sensitivity for constrained firms, as found in the risk neutral model (Clementi and Hopenhayn (2006), DeMarzo and Fishman (2007a)), but only if I consider promised insurance as a measure of financing constraints.

I use two tools from the dynamic public finance literature to characterize the optimal contract

[4]When the entrepreneur has a high equity share, he has less incentives to divert funds, so the lender is willing to provide him with more capital.

while imposing minimal assumptions. The first is the first-order approach (FOA) as in Kapička (2013), Farhi and Werning (2013), Pavan, Segal and Toikka (2014) and Golosov *et al.* (2016a). It consists of solving a relaxed problem with the local IC constraints. The FOA is popular in dynamic public finance, but it is also used more broadly in dynamic mechanism design. The FOA allows solving the model with persistent private information. The second tool allows deriving analytical characterizations of the optimal allocation with risk aversion. This is the change of measure used in Hellwig (2021) for a Mirrlees taxation problem with non-separable preferences between consumption, leisure and type.

The challenge of introducing risk aversion in this model is that information rents depend on the entrepreneur's consumption. This is because marginal information rents in consumption units must be transformed into utils by multiplying the type's marginal utility. So if the principal increased consumption of some type $\theta$, his information rent would change. But then the information rents of all types $\theta' > \theta$ need to be adjusted nonlinearly to preserve incentive compatibility. A change of measure as in Hellwig (2021) reweights the density of types appropriately to account for these changes in information rents.

The resulting incentive-adjusted type distribution puts higher weight on lower types. The intuition is as follows. Lower types generate lower returns, so they obtain less information rents in consumption units and have higher marginal utility. This implies that perturbations in consumption change their information rents by more, which is costly for the principal. This effect reduces the lender's benefit of providing insurance to the entrepreneur, which in turn lowers the investment wedges.

Finally, the approach used in this paper can be useful more broadly in problems where there is no separation between insurance and information rents. Persistent private information can be challenging to handle in some settings. But the incentive-adjusted probability measures could allow using the FOA as in the dynamic public finance more often, where models with a large class of Markov processes can be studied. In Appendix E, I solve the sovereign debt model of Dovis (2019) with persistent private information by using the same type of change of measure.

**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). In particular, I contribute to the literature by studying a workhorse dynamic cash flow diversion model with risk aversion and persistent private information.

Models with risk aversion have been studied in He (2012) and Di Tella and Sannikov (2021). 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. 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. As I will show, some key effects of persistence on the optimal allocation, such as the downwards drifts in firm size, are only present with risk aversion. Fu and Krishna (2019) and Krasikov and Lamba (2021) show some interesting role of persistence on the dynamics of distortions. However, as in the i.i.d risk neutral models, they still find that distortions eventually disappear.

Throughout the paper, I use tools and insights from the dynamic Mirrlees literature[5]. I use the FOA and set up the principal's problem recursively as in Kapička (2013), Farhi and Werning (2013) or Golosov *et al.* (2016a). I also use incentive-adjusted probability measures as in Hellwig (2021)[6] to derive analytical characterizations of the optimal contract. The finding that firm size drifts 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 dynamic mechanism design with insurance (Makris and Pavan (2020)) and on insurance with persistent private information (Williams (2011), Bloedel *et al.* (2018) and Bloedel *et al.* (2020)).

**Outline** The rest of the paper is organized as follows. Section 2 describes the model, sets up the relaxed planning problem and presents the first best allocation. Section 3 discusses the incentive-adjusted measure and shows how it can be used to characterize the multipliers of the problem. Section 4 presents the main results on the optimal allocation and section 5 illustrates them with numerical simulations. Section 6 studies the third best implementation. Section 7 discusses the differences in models with risk neutrality and risk aversion and its implications. Finally, section 8 concludes.

[5] For a review of the literature see Stantcheva (2020). In some aspects, the model also resembles the setting of the dynamic taxation problems in Stantcheva (2017) and Brendon (2022).

[6] The incentive-adjusted measures have also been used in Hellwig and Werquin (2022).

# 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. At period $t$, the project generates a cash flow $f(k_t, \theta_t)$, where $\theta_t \in [\underline{\theta}, \overline{\theta}]$ is the entrepreneur's productivity. The agent's type history is denoted by $\theta^t = \{\theta_0, ..., \theta_t\}$ and is the agent's private information. $\theta_t$ follows a Markov process with conditional density $\varphi_t(\theta_t|\theta^{t-1})$.

The lender cannot observe the returns and instead relies on the entrepreneur's report. The entrepreneur can misreport and divert a fraction of the cash flow for his own consumption. There is a deadweight loss $(1-\phi) \in [0, 1)$ on diverted funds. 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 next period. The entrepreneur cannot save by himself[7], so the entrepreneur's period $t$ consumption if the true cash flow is $f(k_t, \theta_t)$ but he reports $f(k_t, \widetilde{\theta}_t)$ is

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

In particular, if the entrepreneur does not misreport returns he consumes $c_t = f(k_t, \theta_t) - b_t(\theta_t)$. The entrepreneur is risk averse, derives utility $u(c_t)$ from consumption, and discounts the future at rate $\beta$. Throughout the paper, I will use the following notation for the derivatives of the return 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 productivity process and the functions $f$ and $u$.

**Assumptions**

**A1:** The conditional density $\varphi_t(\theta_t|\theta^{t-1})$ is differentiable with respect to the second argument and persiststent, i.e

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

[7]Note it straightforward to allow the agent to also save by himself. Let $d_t$ be divident payments, $w_t$ the agent's net worth and $B_t$ the funds advanced by the principal. Then we would have $c_t = d_t$, a LOM for the entrpreneur's net worth $w_{t+1} = f(k_t, \theta_t) - b_t - d_t + w_t$ and investment equal to $k_{t+1} = w_{t+1} + B_{t+1}$. If the agent net worth is observable, it is without loss to have the lender doing all the savings for the entrepreneur.

is non-decreasing in $\theta_t$

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

**A3:** The utility function satisfies $u'' < 0 < u'$, and the inada conditions $u'(0) = \infty$ and $\lim_{c\to\infty} u'(c) = 0$

The first assumption (A1) requires that 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:

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

This is the impulse response of $\theta^t$ to $\theta^{t-1}$ 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 autoregresive parameter $\rho$, 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 $\rho_t(\theta^t) = \rho$. 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.

### Discussion of the model

The model presented is the most simple version of a cash flow diversion model, but 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. First, I do not allow the principal the option to terminate the project. I analyze the model with endogenous termination in Appendix D.3. As is well known, if there are regions of the state space where the Pareto frontier is not concave, the principal may optimally randomize between shutting down the project and continuing. This may be the case in this model, but as I show, allowing for stochastic termination does not affect any of the following results.

Second, I have also assumed full commitment of the lender and entrepreneur. A model where the entrepreneur has limited commitment is studied in Appendix D.2. Althought there may be interesting interactions between the limited commitment and private information frictions (see Dovis (2019) in the context of sovereign debt), risk aversion and persistence do not alter in any significant way the effects of the limited commitment friction. So in the paper, I focus on the private information friction.

Finally, in Appendix D.1, I study a model where instead of diverting cash flows, 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 main results continue to hold in this setting. Moreover, the model is intuitive as the investment wedge that shows up in the firm size dynamics and the GIEE corresponds to the wedge between invested and diverted funds relative to the first best allocation.

## 2.1 Lender's problem

The lender is risk neutral and discounts the future at rate $q$. By the revelation principle, it is without loss to focus on direct mechanisms. At any history, the entrepreneur sends a report $r \in [\underline{\theta},\overline{\theta}]$ about $\theta_t$ to the lender. Define a reporting strategy by $\sigma = \{\sigma_t(\theta^t)\}$, it implies a history of reports $\sigma^t(\theta^t) = \{\sigma_1(\theta_0), ..., \sigma_t(\theta_t)\}$. The entrepeneur'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 $c(\theta^t) = f(k_t(\theta^{t-1}), \theta_t) - b_t(\theta^t)$. The continuation utility of type $\theta^t$ with reporting strategy $\sigma$ 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

$$c(\theta_t, \sigma^t(\theta^t)) = \phi\left(f(k_t(\sigma^{t-1}(\theta^{t-1})), \theta_t) + (1-\phi)f(k_t(\sigma^{t-1}(\theta^{t-1})), \sigma_t(\theta_t))\right) - b_t(\sigma^t(\theta^t))$$

Then, an allocation $\{k_{t+1}(\theta^t), b_t(\theta^t)\}$ is incentive compatible if for all $\theta^t$ and $\sigma$

$$w_t(\theta^t) \geq w_t^\sigma(\theta^t) \qquad (IC) \tag{5}$$

The lender problem consists of choosing an allocation $\{k_{t+1}(\theta^t),b_t(\theta^t)\}$ to minimize the cost of providing expected utility $v$ subject to the incentive compatibility constraints:

$$
\begin{aligned}
K(v) \quad = \quad & \min_{\{k_{t+1}(\theta^t), b_t(\theta^t)\}} \mathbb{E}_0 \left[ \sum_{t=1}^{\infty} q^t \left( k_{t+1}(\theta^t) - b_t(\theta^t) \right) \right] \\
& s.t \quad \mathbb{E}_0 \left[ w_1(\theta^1) \right] \geq v \\
& \{k_{t+1}(\theta^t), b_t(\theta^t)\} \in IC
\end{aligned} \tag{6}
$$

## 2.2 Relaxed problem

The problem is solved recursively, write entrepreneur's continuation utility under truth-telling as

$$w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \tag{7}$$

$$v_t(\theta^t) = \int w_{t+1}(\theta^{t+1}) \varphi_{t+1}(\theta_{t+1}|\theta^t) d\theta_{t+1} \tag{8}$$

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

$$\frac{\partial}{\partial \theta_t} w_t(\theta^t) = u'(c(\theta^t)) \phi f_\theta(k_t(\theta^{t-1}), \theta_t) + \beta \Delta_t(\theta^t) \tag{9}$$

$$\Delta_t(\theta^t) = \int w_{t+1}(\theta^{t+1}) \frac{\partial \varphi_{t+1}(\theta_{t+1}|\theta^t)}{\partial \theta^t} d\theta_{t+1} \tag{10}$$

The within period marginal information rent is $u'(c(\theta^t))\phi f_\theta(k_t(\theta^{t-1}), \theta_t)$, and it depends on 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 mimick the returns of the type right below him and divert the extra funds, he can then obtain $\phi f_\theta(k_t(\theta^{t-1}), \theta_t)d\theta_t$ extra consumption units. Because the entrepreneur is risk averse, 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 consumption of type $\theta_t$, his information rent changes. But then the information rents of all types $\theta' > \theta_t$ have to be adjusted in order to preserve incentive compatibility. In section 3, I will show in more detail how the incentive-adjusted probability measures developed in Hellwig (2021) can be used to characterize the solution to this problem.

[8] Global incentive compatibility can be verified numerically ex-post, as in Farhi and Werning (2013).

The variable $\Delta_t(\theta^t)$ captures the dynamic incentive commitments promised by the lender. Intuitively, it captures how much insurance the principal promises to provide in future periods. If types were independent over time, we would have $\Delta_t(\theta^t) = 0$. The state variables of the recursive problem are the promised utility, $v$, the dynamic incentive commitments (or promised insurance), $\Delta$, and the funds advanced at $t-1$, $k_t$. The principal solves a dynamic optimization problem where within every period, there is an optimal control problem. The relaxed problem is

$$K_t(v_{t-1}, \Delta_{t-1}, \theta^{t-1}, k_t) = \min \int \big(k_{t+1}(\theta^t) - b_t(\theta^t) + qK_{t+1}(v_t(\theta^t), \Delta_t(\theta^t), \theta^t, k_{t+1}(\theta^t))\big)\, \varphi_t(\theta_t|\theta^{t-1})d\theta_t$$

$$\begin{aligned} s.t \quad (PK) \quad & w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \qquad [\varphi_t(\theta_t|\theta^{t-1})\xi_t(\theta^t)] \\ & v_{t-1} = \int w_t(\theta^t)\varphi(\theta_t|\theta^{t-1})d\theta_t \qquad [\varphi_t(\theta_t|\theta^{t-1})\lambda_t] \\ (IC) \quad & \dot{w}_t(\theta) = u'(c(\theta^t))\phi f_\theta(k_t, \theta_t) + \beta\Delta_t(\theta^t) \qquad [\mu_t(\theta^t)] \\ & \Delta_{t-1} = \int w_t(\theta^t)\frac{\partial\varphi(\theta_t|\theta^{t-1})}{\partial\theta^{t-1}}d\theta_t \qquad [\varphi_t(\theta_t|\theta^{t-1})\gamma_t] \\ (Feasibility) \quad & c(\theta^t) = f(k_t, \theta_t) - b_t(\theta^t) \end{aligned} \tag{11}$$

where $\mu_t(\theta^t)$ is the co-state variable of the within period Hamiltonian. Note that I write inside square brackets the multipliers associated with each constraint. The Hamiltonian of this problem and the derivation of the optimality conditions can be found in Appendix B. To economize notation I will often write directly $u(\theta^t)$ and $f(\theta^t)$ instead of $u(c(\theta^t))$ and $f(k_t(\theta^{t-1}), \theta_t)$. The sequential problem (6) can be recovered by treating $\Delta_0$ and $k_1$ as free variables, $K(v_0) = \min_{\Delta_0, k_1} K(v_0, \Delta_0, \theta_0, k_1)$.

## 2.3 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 $\theta^t$, there is*

1. *No diversion of funds*

$$f(k_t, \widetilde{\theta}_t) = f(k_t, \theta_t) \tag{12}$$

*2. Full insurance and intertemporal consumption smoothing*

$$u'(c(\theta^t)) = \frac{\beta}{q} u'(c(\theta^{t+1})) \tag{13}$$

*3. No distortion of the project size*

$$\frac{1}{q} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})|\theta^t\right] \tag{14}$$

Because diverting funds is inefficient and by the revelation principle, in the second best there will also be no diversion of funds. However, as will be discussed in section 4, the points 2. and 3. of the proposition do not hold in the second best allocation. Therefore, in section 4 we will be interested on how consumption and firm size dynamics differ from the first best.

# 3 Incentive-adjusted probability measures

As discussed, the main challenge to characterize the optimal allocation in this problem is that the static marginal information rents, $u'(c(\theta^t))\phi f_\theta(k(\theta^t), \theta_t)$, depend on consumption. This is the same problem encountered in a Mirrlees taxation problem with general non-separable preferences. Hellwig (2021) solves this by applying a change of measure to the distribution of types. It consists of reweighting the density of types appropiately to preserve incentive compatibility. The following proposition shows how this change of measure can be used to characterize the shadow costs of insurance $\mu_t(\theta^t)$ and the multiplier on the promise-keeping constraint $\lambda_t$. This shadow cost captures the resource gain from redistributing consumption around $\theta^t$, while preserving incentive compatibility, promised expected utility ($v$) and prior incentive commitments ($\Delta$). The multiplier $\lambda_t$ will be used to derive the Generalized Inverse Euler Equation (GIEE) later.

**Proposition 2.** *(Hellwig (2021)) The shadow cost of insurance $\mu_t(\theta^t)$ and the multiplier $\lambda_t$ can be characterized as*

$$\frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} = MB(\theta^t) + \hat{\rho}(\theta^t)\frac{\beta}{q}\frac{\mu_{t-1}(\theta^{t-1})}{\varphi_{t-1}(\theta_{t-1}|\theta^{t-2})} \tag{15}$$

*With*

$$MB(\theta^t) = \frac{1-\hat{\Phi}_t(\theta_t|\theta^{t-1})}{\hat{\varphi}_t(\theta_t|\theta^{t-1})}\left\{\hat{\mathbb{E}}\left[\frac{1}{u'(\theta',\theta^{t-1})} \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^t)}|\theta^{t-1}\right]\right\} \tag{16}$$

$$\hat{\varphi}_t(\theta_t|\theta^{t-1}) \equiv \frac{\varphi_t(\theta_t|\theta^{t-1})m(\theta^t)}{\mathbb{E}[\varphi_t(\theta_t|\theta^{t-1})m(\theta^t)|\theta^{t-1}]} \quad where \quad \frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\phi f_\theta(\theta^t)}{u'(\theta^t)} < 0 \tag{17}$$

$$\hat{\rho}(\theta^t) \equiv \frac{1-\hat{\Phi}_t(\theta_t|\theta^{t-1})}{\hat{\varphi}_t(\theta_t|\theta^{t-1})}\left\{\hat{\mathbb{E}}\left[\mathcal{E}(\theta',\theta^{t-1}) \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \hat{\mathbb{E}}\left[\mathcal{E}(\theta_t,\theta^{t-1})|\theta^{t-1}\right]\right\} \tag{18}$$

*And*

$$\lambda_t = \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^t)} \mid \theta^{t-1}\right] - \gamma_t\hat{\mathbb{E}}\left[\mathcal{E}(\theta_t,\theta^{t-1}) \mid \theta^{t-1}\right] \tag{19}$$

Note the operator $\hat{\mathbb{E}}$ denotes expectations under the measure $\hat{\varphi}$. The proposition shows that it is enough to change the probability measure over types from $\varphi$ to the incentive adjusted measure $\hat{\varphi}$. This is the only adjustment needed to account for the fact that information rents depend on consumption. In the Mirrlees taxation problem with general non-separable preferences $U(\theta, c, y)$, where $y$ is the agent's income, the change of measure is with $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{U_{\theta C}(\theta^t)}{U_C(\theta^t)}$[9]. Except for the different change of measure, the characterization of $\mu_t(\theta^t)$ and $\lambda_t$ of proposition 2 are exactly the same. Because $\frac{u''(\theta^t)f_\theta(\theta^t)}{u'(\theta^t)} < 0$ this corresponds to the case where higher types have lower consumption needs, i.e $\frac{U_{\theta C}(\theta^t)}{U_C(\theta^t)} < 0$. So the effects of private information on the optimal allocation will compare to the case with $\frac{U_{\theta C}(\theta^t)}{U_C(\theta^t)} < 0$ in the Mirrlees taxation problem.

The literature on dynamic mechanism design with insurance often analyses models with separable preferences of the form

$$U(\theta, y, c) = u(c) - \psi(y, \theta) \tag{20}$$

where $y$ can represent the agent's income or effort. As discussed in Makris and Pavan (2020), this includes dynamic public finance models with separable preferences but also models of managerial compensation, among others. In these settings, the static marginal information rents are $\psi_\theta(y, \theta)$, which do not depend on consumption. Therefore, there is a complete separation between insurance (or redistribution) and information rents. This setting also

[9] The problem admits an alternative representation in terms of redistribution throught leisure with $\frac{\tilde{m}'(\theta^t)}{\tilde{m}(\theta^t)} = \frac{U_{\theta Y}(\theta^t)}{U_Y(\theta^t)}$.

admits the same characterization of $\mu(\theta^t)$ as proposition 2 but under the original measure $\varphi$ (Makris and Pavan (2020), Hellwig (2021), Brendon (2013))[10]. For $\lambda_t$, there is an extra term that emerges compared to the separable models that captures the interaction between the incentive-adjusted measure and the persistence of the process[11].

Because $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\phi f_\theta(\theta^t)}{u'(\theta^t)} < 0$, the function $m(\theta^t) = e^{-\int_{\theta_t}^{\bar{\theta}} \frac{u''(\theta',\theta^{t-1})\phi f_\theta(\theta',\theta^{t-1})}{u'(\theta',\theta^{t-1})} d\theta'}$ is decreasing in $\theta_t$. So the principal puts more weight on lower types. That is, $\Phi_t(\cdot|\theta^{t-1})$ first-order stochastically dominates $\hat{\Phi}_t(\cdot|\theta^{t-1})$. Intuitively, lower types have lower returns, so they collect lower information rents in consumption units. Hence, their marginal utility is higher. When the principal redistributes consumption, information rents change more for types with high marginal utility. Therefore, the incentive-adjusted measure that guarantees incentive compatibility has to put more weight on lower types.

This higher sensitivity of the information rents implies that the cost of adjusting information rents to preserve incentive compatibility is higher for lower types. These higher costs can partly offset the direct resource gain of redistributing consumption to types with high marginal utility. Recall $MB(\theta^t)$ represents the resource gain of redistributing from types $\theta' > \theta_t$ to the types $\theta'' < \theta_t$, while preserving incentive compatibility and keeping expected utility constant. Imagine the principal redistributes $\delta u$ utils from some type $\theta'$ to type $\theta''$ with $\theta' > \theta''$. Because lower types have higher marginal utility, the principal obtains a direct resource gain proportional to $\Delta c(\theta')\delta u - \Delta c(\theta'')\delta u = \left(\frac{1}{u'(\theta')} - \frac{1}{u'(\theta'')}\right)\delta u > 0$.

In a separable environment with utility function as in (20), it would be enough to adjust $\delta u$ to keep expected utility constant. However, in this model, this redistribution changes information rents. The changes in consumption now need to be adjusted by the factors $m(\theta')$ and $m(\theta'')$ to preserve incentive compatibility. Because $m(\theta)$ is decreasing, the adjustment for type $\theta''$ is higher, i.e. $m(\theta')\delta u > m(\theta'')\delta u$. So after accounting for incentive compatibility, the resource gain is reduced

$$\left(\frac{1}{u'(\theta')}m(\theta') - \frac{1}{u'(\theta'')}m(\theta'')\right)\delta u < \left(\frac{1}{u'(\theta')} - \frac{1}{u'(\theta'')}\right)\delta u$$

Therefore, the shadow costs cost of insurance $\mu_t(\theta^t)$ will be small, especially if types are i.i.d as the second term in equation (15) zero. This will be verified in the numerical simulations in section (5). Small $\mu_t(\theta^t)$ implies that the entrepreneur is provided with little insurance

[10]Although these models also admit a representation of $\mu_t(\theta^t)$ and $\lambda_t$ with the incentive adjusted measure.
[11]Note that under the original type measure, we have $\mathbb{E}\left[\mathcal{E}(\theta_t, \theta^{t-1}) \mid \theta^{t-1}\right] = 0$.

against bad cash flow realizations but that the distortions to optimal firm size are also small (see section (4.1)).

The static version of this model is illustrative. In a dynamic model, the principal can provide information rents through higher current consumption or higher future promised utility. In a static model, the provision of information rents cannot be smoothed over time, so we must have $c'(\theta) = \phi f_\theta(\theta)$. This in turn implies that $m(\theta) = \frac{u'(\theta)}{u'(\bar{\theta})}$. Therefore, the higher cost of adjusting information rents fully offsets the direct resource gain of redistributing to types with high marginal utility. The only effects that survive this redistribution are from the changes in expected utility[12].

The same characterizations as in the static model can also be obtained by applying a new change of measure. This new measure accounts for the fact that the principal can provide part of the information rents by promising higher continuation utility. Rearranging the local IC we can obtain $\phi f_\theta(\theta^t) - c'(\theta^t) = \beta \frac{v'(\theta^t)}{u'(\theta^t)}$, where $v'(\theta^t) \equiv \mathbb{E}\left[\frac{\partial w(\theta^{t-1}, \theta_t, \theta_{t+1})}{\partial \theta_t} | \theta_t\right]$. This incentive-adjusted measure is defined as $\widetilde{\varphi}_t(\theta_t|\theta^{t-1}) = \frac{\varphi(\theta_t|\theta^{t-1})n(\theta^t)}{\mathbb{E}[n(\theta^t)|\theta^{t-1}]}$, where

$$\frac{n'(\theta^t)}{n(\theta^t)} = \frac{u''(\theta^t)}{u'(\theta^t)} \beta \frac{v'(\theta^t)}{u'(\theta^t)}$$

Incentive compatibility requires that, for any $\theta^t$, $v'(\theta^t) \geq 0$, so $\frac{n'(\theta^t)}{n(\theta^t)} = \frac{u''(\theta')}{u'(\theta')} \beta \frac{v'(\theta')}{u'(\theta')} < 0$. Therefore, we also have that $\Phi(\cdot|\theta^{t-1})$ first-order stochastically dominates $\widetilde{\Phi}(\cdot|\theta^{t-1})$. By the local IC, the incentive-adjustment terms $m(\theta^t)$ and $n(\theta^t)$ are related by $m(\theta^t) = \frac{u'(\theta^t)}{u'(\bar{\theta})} n(\theta^t)$. So the incentive-adjusted measures are related by

$$\hat{\varphi}(\theta_t|\theta^{t-1}) = \frac{\widetilde{\varphi}_t(\theta_t|\theta^{t-1}) u'(\theta^t)}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]}$$

The following proposition shows how under the measure $\widetilde{\varphi}$, one can obtain the same representation of $MB(\theta^t)$ as in a static model.

**Proposition 3.** *Under the incentive-adjusted measure $\widetilde{\varphi}$, the terms $MB(\theta^t)$ and $\lambda_t$ admit*

[12] There is a more straightforward reason why the effects cancel out. In a static model, it is not possible to provide insurance in an incentive compatible manner. However, this is also true in the screening model studied in section D.1. The reason is that the incentive constraints depend on the ordinal properties of the utility. So, if there is no participation constraint, any cardinalization with a particular utility function does not affect the optimal allocation (see Brendon (2013) for more discussion on this).

*the following representations*

$$MB(\theta^t) = \frac{1-\widetilde{\Phi}_t(\theta_t|\theta^{t-1})}{\widetilde{\varphi}_t(\theta_t|\theta^{t-1})}\frac{1}{u'(\theta^t)}\left[1-\frac{\widetilde{\mathbb{E}}\left[u'(\theta',\theta^{t-1})|\theta'>\theta_t,\theta^{t-1}\right]}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]}\right]$$

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

Therefore, by applying a change of measure that accounts for how the principal spreads continuation utilities. We can obtain a representation of the benefits of redistributing across types that only captures the effect on the changes in expected utility, as in the static model.

# 4 Optimal allocation

In this section, I present the two main results on the dynamics of the optimal allocation. The first subsection is on the firm size dynamics, the second on the dynamics of the entrepreneur's consumption (or compensation). Before presenting the results, it is useful to define the investment wedge, as it will show up in both the GIIE and the equation for the firm size size dynamics. Let

$$\tau^k(\theta^t) \equiv \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}u'(\theta^t)\phi\frac{f_{\theta k}(\theta^t)}{f_k(\theta^t)} \geq 0 \tag{21}$$

In a model where the entrepreneur can choose to divert funds before investing in the project, this wedge captures the distortion in invested and diverted funds relative to the first best. I discuss this model in more detail in appendix D.1.

## 4.1 Firm size dynamics

In this section, I look at the dynamics of the optimal size of the firm, i.e. $k_t$. 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, the firm size tends to increase and financing constraints are relaxed. The literature has tried to use dynamic contracting models to explain these patterns. Models of cash flow diversion with a risk-neutral agent and limited liability (Clementi and Hopenhayn (2006)) can qualitatively replicate the dynamics observed in data.

However, once risk aversion and persistent private information are introduced, this is no longer the case. The opposite dynamics emerge, the firm size tends to decrease over time, and the first best size is never reached. In the following proposition, I show the first-order condition for the optimal firm size.

**Proposition 4.** *At any history $\theta^t$, the optimal advancement of funds $k_{t+1}(\theta^t)$ satisfies*

$$\frac{1}{q} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1})(1 - \tau^k(\theta^{t+1}))|\theta^t\right] \tag{22}$$

The proposition shows that the FOC for $k_{t+1}(\theta^t)$ is the same as in the FB but with an extra investment wedge that lowers the implicit marginal product of capital. Because $\tau^k(\theta^{t+1}) \geq 0$, we have $k^{SB}_{t+1}(\theta^t) \leq k^{FB}_{t+1}(\theta^t)$. Besides the direct effect of the productivity process $\{\theta^t\}$, the dynamics of the firm's size $(k_{t+1}(\theta^t))$ depend only on the dynamics of the investment wedge $(\tau^k(\theta^{t+1}))$. Intuitively, because $f_{\theta k} > 0$, higher funds increases information rents. That is, it makes diversion of funds relatively more attractive for higher types. Increasing information rents is more costly when the expected shadow costs $\mu_t$ are high, i.e. when the planner has promised to provide more insurance.

It has been shown that with persistent private information, the shadow costs of insurance $\mu_t$ (and so wedges) tend to increase over time (Farhi and Werning (2013), Makris and Pavan (2020)[13]). Not surprisingly, this is also true in this model, which implies that if types are persistent firm size will tend to decrease over time. To see this, iterate backward on equation (15) to get

$$\frac{\tau^k(\theta^t)}{u'(\theta^t)} \frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)} = \sum_{\tau=0}^{t-1} \left(\frac{\beta}{q}\right)^\tau \prod_{s=0}^{\tau-1} \hat{\rho}_{t-s}(\theta^{t-s}) MB_{t-\tau}(\theta^{t-\tau}) \tag{23}$$

The right-hand side of this equation is the same as in Hellwig (2021), but with $\hat{\rho}$ computed under the incentive adjusted measure (17) instead of with $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{U_{\theta C}(\theta^t)}{U_C(\theta^t)}$. In a separable model with utility as in (20), one also obtains the same formula but with $\rho$, i.e. with impulse responses under the original type measure.

The formula shows that the incentive cost of increasing a type's consumption grows with the distance from the starting period. Intuitively, imagine the principal increases consumption

[13]The early papers attributed these wedge dynamics to the variance of the types increasing over time, which is the case if the type process follows a random walk. However, Makris and Pavan (2020) have clarified why this intuition is incomplete, as wedges can increase even the variance of the types decreases over time. As it will be shown in the numerical simulations, the wedges are initially increasing even with an AR(1) process.

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). However, as it will be shown in the numerical simulations, the wedges may converge to a stationary distribution. It is important to stress that, for every type $\theta_t$, 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 $\Delta_0$ freely. So $\Delta_0$ is set to not have any "extra" promised insurance. So for every $\theta_t$, the wedges will not be smaller than in the initial period.

The change of measure can amplify or dampen the persistence of the wedges. We have $\hat{\rho}_t(\theta^t) \gtreqless \rho_t(\theta^t)$ 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 $\theta$ (see proposition 3 in Hellwig (2021)). If we assume that the type process is AR(1) with autoregressive paramater $\rho$ (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 we have $\hat{\rho}_t(\theta^t) = \rho$ (resp. $\hat{\rho}_t(\theta^t) > \rho$ ) if the agent has CARA (resp. CRRA) utility.

It is also important to remark that both risk aversion and persistence are necessary to have investment wedges increasing over time. If the agent is risk-neutral we have $MB_{t-\tau}(\theta^{t-\tau}) = 0$. If the type process is not persistent we have $\hat{\rho}_t(\theta^t) = \rho_t(\theta^t) = 0$ and

$$\frac{\tau^k(\theta^t)}{u'(\theta^t)} \frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)} = MB_t(\theta^t) \tag{24}$$

so the wedges and firm size are stationary. As discussed in section 3, the higher incentive cost of redistributing to lower types makes $\mu_t$ smaller than in comparable models with separable preferences. Therefore, as will be verified in the numerical simulations, the wedges and distortions are small, especially when types are i.i.d.

The firm size dynamics generated by this model appear to be contradictory with what is regularly observed in the data. Firms usually start small and gradually grow over time. In Appendix D.2, I study a model where the entrepreneur has limited commitment. The firm dynamics induced by this type of model do not change in any meaningful way once risk aversion and persistent private information are introduced. So this type of friction can still generate dynamics where firm size increases over time (as found in Albuquerque and Hopenhayn (2004)). In section 7, I discuss in more detail why models with risk neutrality generate different firm dynamics and it's implications.

More generally, the firm lifecycle dynamics are driven by many different frictions. This model could generate more consistent firm dynamics in a straighfforward manner by allowing for a drift in the productivity process $\{\theta_t\}$. Then, this type of friction may act as a constraint on the size that firms can eventually reach rather than on the growth of young firms. This may then help explain other empirical facts. For instance, in developing economies, where financing frictions are more stringent, we observe fewer large firms (Hsieh and Klenow (2014)).

## 4.2 Compensation dynamics

The dynamics of the entrepreneur's marginal utility and consumption can be characterized by a Generalized Inverse Euler Equation (GIEE), as in Hellwig (2021). The only differences are the change of measure and the static wedges. As in the standard Inverse Euler Equation, the principal arbitrages between period $t$ inverse marginal utility and period $t+1$ discounted expected inverse marginal utility. However, expectations are taken with respect to the incentive-adjusted probability measure because consumption at $t+1$ has to be redistributed non-linearly 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 $t+1$ are passed as a cost at period $t$ at rate $\rho_{t+1}(\theta^{t+1})$. So the size and sign of the savings wedge depends on the persistence of the process. This wedge is then scaled by the cost of insurance provision at period $t$.

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

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{1}{u'(\theta^t)}(1+s(\theta^t)) \tag{25}$$

*where*

$$s(\theta^t) = \left(\frac{\phi f_\theta(\theta^t)u''(\theta^t)}{u'(\theta^t)} - \hat{\mathbb{E}}\left[\rho_{t+1}(\theta^{t+1})\frac{\phi f_\theta(\theta^{t+1})u''(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right]\right)\frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\tau^k(\theta^t) \tag{26}$$

The entrepreneur's marginal utility process can follow a sub- or super- martingale. But as I now show, this only depens on the sign of the savings wedge $s(\theta^t)$. For exposition, set $\frac{q}{\beta}=1$, then if peristence (i.e. $\rho_{t+1}(\theta^{t+1})$) is not too high, we have $s(\theta^t)<0$. So

$$\frac{1}{u'(\theta^t)} > \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right]$$

this then implies that marginal utility follows a sub-martingale[14]

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

Thus, we have the well-known immiseration dynamics recurrent in private information models with risk aversion. If persistence is high enough, we may have $s(\theta^t) \geq 0$ for some types. So inverse marginal utility follows a sub-martingale under the incentive-adjusted measure, i.e.

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

The implications for the dynamics under the original type measure are now less direct. Multiplying $m(\theta^{t+1})$ by $\frac{u'(\theta^{t+1})}{u'(\theta^{t+1})}$ this inequality can be rewritten as

$$u'(\theta^t) \geq \mathbb{E}\left[u'(\theta^{t+1})|\theta^t\right] + \frac{1}{\mathbb{E}\left[\frac{m(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right]} cov\left(u'(\theta^{t+1}), \frac{m(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right)$$

Therefore, the dynamics are preserved under the original type measure if the covariance term is non-negative. Because $u'(\theta^{t+1})$ is decreasing, the covariance is non-negative if $\frac{m(\theta^{t+1})}{u'(\theta^{t+1})}$ is weakly decreasing $\theta_{t+1}$. Differentiating it is easy to see that this is the case if $f_\theta(\theta^{t+1}) - c'(\theta^{t+1}) \geq 0$, or equivalently if $v'(\theta^{t+1}) \geq 0$, which is the case by the IC constraint.

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 $f(k, \theta) = \theta k^\alpha$, then

$$s(\theta^t) = -\sigma\phi\theta_t \times \tau^k(\theta^t) \times (k_t^\alpha - \rho k_{t+1}(\theta^t)^\alpha)$$

hence $s(\theta^t) < 0$ if $\rho < \left(\frac{k_t}{k_{t+1}(\theta^t)}\right)^\alpha$. With fixed capital ($k_t = k$) and $\phi = 1$ this models nests a hidden endowment model. Moreover, with CARA utility, it is also equivalent to a taste shocks model as in Atkeson and Lucas (1992). In this case, $s(\theta^t) = 0$ and marginal utility follows a supermartingale if and only if the type process has a unit root ($\rho = 1$). This result has been shown for more general utility functions in Bloedel *et al.* (2018) and Bloedel *et al.* (2020), which have corrected the findings in Williams (2011) and shown there is immiseration whenever there is some mean-reversion in the type process. Thus, for CARA utility, the GIEE provides a very direct characterization of the effect of persistence on consumption dynamics.

[14]To see this, first $\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] \geq \mathbb{E}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right]$ because $\Phi$ first-order stochastically dominates $\hat{\Phi}$ and $\frac{1}{u'(\theta^t)}$ is increasing in $\theta_t$. And then $\mathbb{E}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] \geq \frac{1}{\mathbb{E}[u'(\theta^{t+1})|\theta^t]}$ by Jensen's inequality.

Note also that time-varying capital should generate an extra force towards immiseration. As we should have $\left(\frac{k_t}{k_{t+1}(\theta^t)}\right)^\alpha$ decreasing in $\theta_t$[15], given some high enough $\rho$ 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. 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, unless we consider a unit root process and assume that capital is fixed, we should expect the marginal utility process to follow a sub-martingale. As I will next show, this is the case in all the numerical simulations performed. In practice, the sub-martingale process implies that, on average, the marginal utility will tend to increase over time and that the variance of consumption and marginal utility will increase over time without bound.

# 5 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 the next section. I assume the agent has log-utility

$$u(c) = \log(c)$$

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 $\alpha = 3/4$ 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

[15] This would not be the case, if for some types $\theta_t' > \theta_t''$, the effect of higher wedges at $t+1$ for type $\theta_t'$ is stronger than from the higher expected productivity. The numerical simulations verify that $k_{t+1}(\theta^t)$ is indeed increasing in $\theta_t$, see figure 6 in Appendix A.

$\sigma_\varepsilon^2 = 0.01$. The comparative statics of this section focus on the effect of the persistence $\rho$, the model is solved with $\rho = 0$ (i.i.d types) and $\rho = 0.7$. For the i.id case, I also solve the model with different parametrizations of the utility function (CRRA with higher risk aversion and CARA), the results can be found in Appendix A. Details on the solution method and algortihm can be found in Appendix C. After solving for the value functions ($K$, $v$ and $\Delta$), the policy functions ($c_t$, $\lambda_{t+1}$, $\gamma_{t+1}$ and $k_{t+1}$) and the costate ($\mu_t$), I run a montecarlo simulation with $10^6$ draws over 25 periods each.

Figure 1 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 log utility and i.i.d types average consumption is extacly constant. If the relative risk aversion is higher, average consumption slowly decreases over time (see figure 10 in Appendix A). We can observe that with persitence, there is also a small 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 should be driven by the initial increase in the investment wedge (see figure 2).

To visualize the immiseration dynamics, in figure 5 in Appendix A I plot the mean and variance of the marginal utility of consumption over a long time horizon. Even if average consumption is constant, average marginal utility increases over time because the agent is risk averse. Moreover, average marginal utility increases very slowly, such that it may be irrelevant for the usual lifespan of a firm.

Figure 1: Consumption dynamics

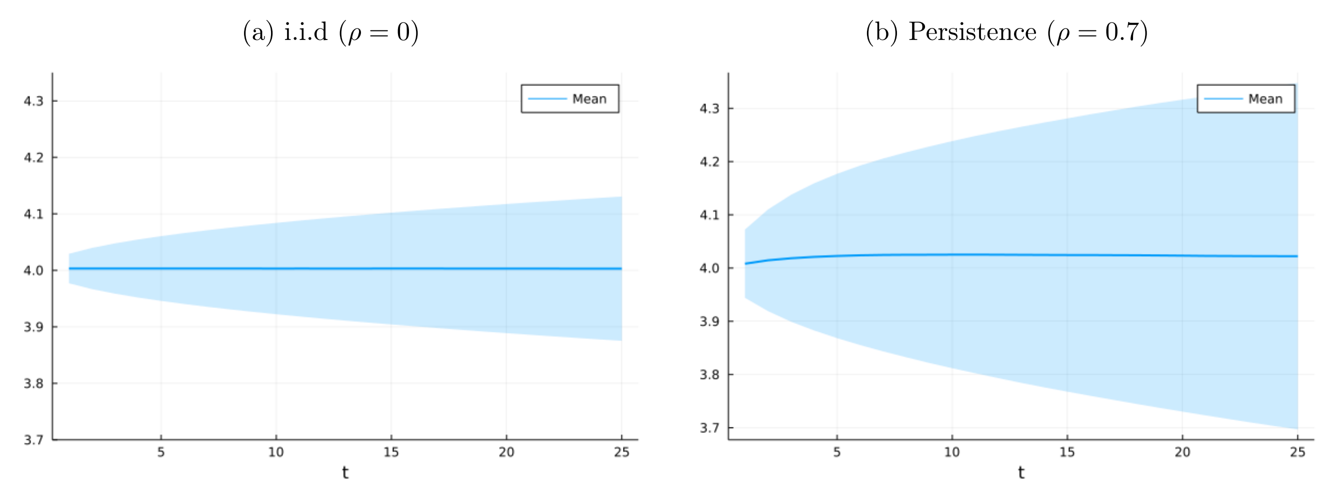
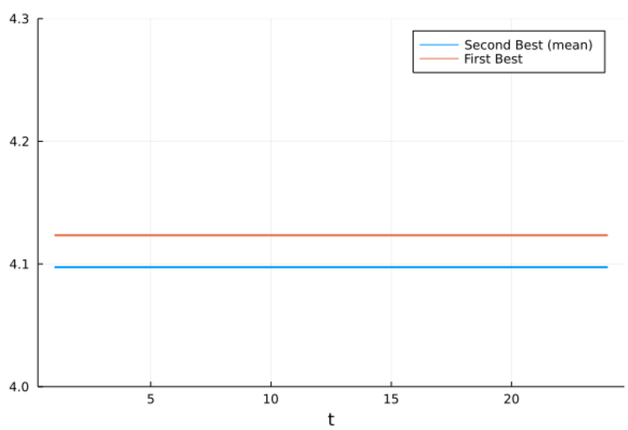


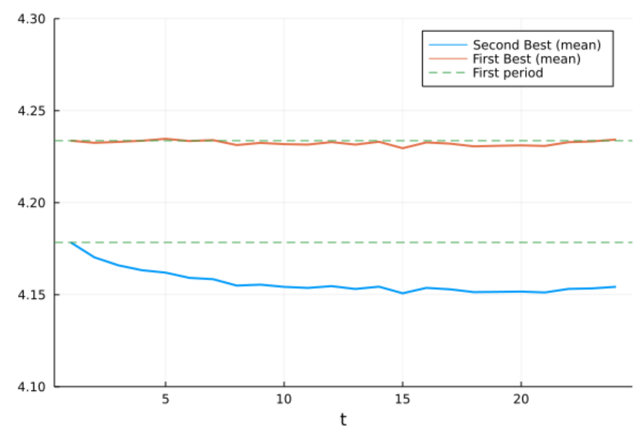
Figure 2 shows the firm size and investment wedge dynamics. In both cases, the firm size closely follows the dynamics of the investment wedge. With i.i.d shocks, the wedges are stationary and so firm size is constant. Moreover, as discussed, the wedges are small, so firm size is also very close to the first best level. If the process is persistent, at the first best, firm size is stationary and all variation is driven only by differences in expected returns. At the second best, we can observe that on average, the wedges tend to increase over time and firm size tends to decrease. However, the wedges do not increase indefinitely, over time they also converge to a stationary distribution. Overall the decrease in firm size does not appear to be large, as the wedges remain small. The decrease would be larger with higher risk aversion or persistence.

Figure 2: Firm size and investment wedge dynamics

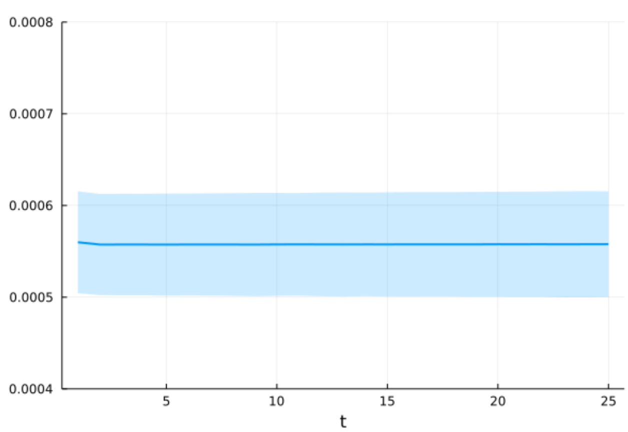
(a) Size (k) i.i.d ($\rho = 0$)



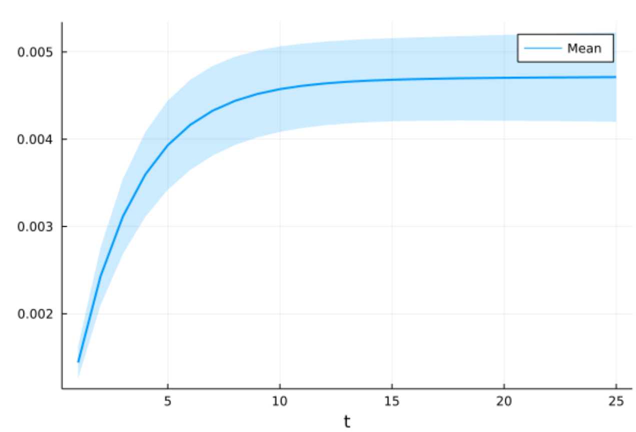

(b) Size (k) with persistence ($\rho = 0.7$)




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




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



# 6 Third best implementation

This section aims to look for simpler (third best) contracts that can get very close to the second best optimal contract studied. First, I use regressions with the model simulated data to better understand the compensation dynamics. Then, I propose a contract and use the simulated data and regression estimates to calibrate key parameters of the contract. Finally, I solve the entrepreneur's problem under the third best contract and compare consumption dynamics with the second best.

I start by studying the model with an i.i.d type process. I show how giving a constant equity share and allowing him to pledge his shares as collateral and borrow to smooth consumption gives a very close approximation to the optimal contract. Then I discuss how, with persistent types, the contract has to allow for a time-varying equity share.

## 6.1 i.i.d types

Because lending is approximately constant with an i.i.d type process, I fix capital to the second best level $k_{SB}$ from figure 2 and focus on implementing the compensation dynamics. I use the simulated data from section 5 to run regressions of consumption on returns and promised utility. The results can be found in table 1. The regressions give three key observations :

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

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

3. The effect of returns on compensation is close to linear. Note in column $returns_t^2$ is close to 0 in column 4.

Points 2. and 3. suggest that a constant equity share can be a good approximation. 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), Brendon (2022)). Let $W_t$ denote the agent's wealth and $\chi$ the equity share, i.e. the portion of cash flows accruing

Table 1: Regressions with i.i.d type process

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | $c_t$ | $c_t$ | $c_t$ | $c_t$ | $c_t$ |
| $returns_t$ | 0.0481*** | 0.0484*** | 0.0482*** | 0.0553*** | |
| | (15830.95) | (415.74) | (139.87) | (1050.38) | |
| $v_{t-1}$ | 0.199*** | | 0.199*** | 0.199*** | |
| | (29257.84) | | (2770.02) | (29265.77) | |
| $returns_t$ | | 0.0475*** | | | |
| | | (408.20) | | | |
| $returns_t * v_{t-1}$ | | | -0.00000361 | | |
| | | | (-0.29) | | |
| $returns_t^2$ | | | | -0.000616*** | |
| | | | | (-136.28) | |
| $c_{t-1}$ | | | | | 1.000*** |
| | | | | | (10906.36) |
| $N$ | 4900000 | 4400000 | 4900000 | 4900000 | 4800000 |
| $R^2$ | 0.999 | 0.072 | 0.999 | 0.999 | 0.961 |

$t$ statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

to the entrepreneur . Let $\overline{f}(k_{SB}) = \mathbb{E}[f(k_{SB}, \theta_t)]$ denote the expected returns if capital is $k_t$. The entrepreneur also receives initial cash $W_0$[16]. 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 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 LOM of the entrepreneur's wealth follows

$$c_t + W_{t+1} = \frac{1}{q} W_t + \chi\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right) \equiv C(W_t, \theta_t) \tag{27}$$

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

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

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 $q$.

The next step for the numerical implementation is to obtain a value for $\chi$. 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 $\chi$ can be identified from the regressions as $\hat{\chi} = \frac{\beta_{returns}}{(1-q)} = \frac{0.0481}{0.05} = 0.962 \approx \phi$, where $\beta_{returns}$ is the regression coefficient on returns in column (1) of table 1. So I set directly $\hat{\chi}=\phi$. Then, given $\hat{\chi}$, the entrepreneur recursive problem with wealth $W_t$ and productivity $\theta_t$ is

$$\begin{aligned}
\mathcal{W}(W_t,\theta_t) &= \max_{\widetilde{\theta}} u(\widetilde{c}_t) + \beta\mathcal{V}(W_{t+1}) \\
s.t \qquad W_{t+1} &= qC(W_t,\widetilde{\theta}_t) \\
c_t &= (1-q)C(W_t,\widetilde{\theta}_t) \\
\widetilde{c}_t &= c_t + \phi(f(k_{SB},\theta) - f(k_{SB},\widetilde{\theta}))
\end{aligned} \tag{28}$$

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 $\mathcal{V}(W_0) = v_1$, i.e. the promised utility under the direct mechanism. Notice that, 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 $(W_t,\widetilde{\theta}_t)$. 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 (28). Then, I run the same montecarlo simulation as for the optimal allocation and compare the results[17]. Table (8) in Appendix A shows that the consumption and repayment are very close to that of the optimal allocation and that this contract induces very little diversion of funds. The assumption of log utility also simplifies the implementation as in the optimal allocation average consumption is constant. With CRRA utility and higher risk aversion average consumption slowly decreases over time. So in this case, for a more accurate implementation would require an extra wedge between the market interest rate $q$ and the rate given to the entrepreneur. However, the same contract still delivers a good approximation, see figure 9 in Appendix A.

## 6.2 Persistent types

With persistent private information, there is an extra state variable in the recursive planning problem (11), $\Delta_{t-1}$. This state variable captures how much insurance is provided to the agent, as equation (10) can be written as

$$\Delta_{t-1} = \mathbb{E}\left[\rho(\theta^t)\dot{w}(\theta^t)|\theta^{t-1}\right] \tag{29}$$

Therefore, given a level of persistence, a lower $\Delta_{t-1}$ implies more insurance is provided to the agent. In this implementation, the level of insurance provided to the entrepreneur naturally maps to the equity share. Thus, the implementation with i.i.d types of section (6.1) may also be a good approximation of the optimal contract with persistence if augmented with a time-varying equity share. Intuitively, lowering the entrepreneur's equity is beneficial as it increases insurance, but it also comes at the cost of increasing the entrepreneur's 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.

This can also be verified in the regressions with model simulated data. In the regression table 2 in column (1), we can observe that the coefficient on interaction term $\Delta_{t-1} \times \theta_t$ is positive. So when the lender has promised high insurance (i.e low $\Delta_{t-1}$), the entrepreneur's compensation is less sensitive to the type realization. However, as discussed below, with persistence, it is less straightforward to infer the equity share from these regressions.

[17] To have accurate comparisons, in the montecarlo simulations, for each realization of the shock process $\{\varepsilon_t\}_{t=1}^{25}$ I compute consumption and repayment 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 $\overline{c}_t^{dist} = \sum_i \sqrt{\left(c_t^{SB}(\{\varepsilon_{i.\tau}\}_{\tau=1}^t) - c_t^{TB}(\{\varepsilon_{i.\tau}\}_{\tau=1}^t)\right)^2}$, where $c^{SB}$ is the consumption under the optimal allocation and $c^{TB}$ under the implementation, and similarly for repayment $b$.

Table 2: Regressions with persistent type process

| | (1) $c_t$ | | (2) $\Delta_t$ | | (3) $\gamma_t$ | |
|---|---|---|---|---|---|---|
| $\theta_t$ | -0.179*** | (-115.01) | | | | |
| $\Delta_t$ | -1.139*** | (-262.03) | | | | |
| $\theta_t \times \Delta_t$ | 0.437*** | (571.15) | | | | |
| $\Delta_{t-1}$ | | | 0.470*** | (856.18) | | |
| $\theta_{t-1}$ | 0.856*** | (125.27) | 1.736*** | (3727.13) | 0.0112*** | (244.54) |
| $v_t$ | 0.141*** | (384.54) | | | | |
| $v_{t-1}$ | | | -0.0504*** | (-943.57) | | |
| $\theta^2_{t-1}$ | | | -0.113*** | (-497.06) | -0.0178*** | (-801.45) |
| $\theta_{t-2}$ | | | -0.652*** | (-653.63) | 0.00517*** | (666.30) |
| $\gamma_{t-1}$ | | | | | 0.689*** | (5130.61) |
| $N$ | 2300000 | | 2200000 | | 2200000 | |
| $R^2$ | 0.997 | | 0.998 | | 0.985 | |

$t$ statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To gain intuition, imagine that, at period $t$, the lender offers to buy some equity to type $(\theta^{t-1}, \theta')$. Assume also that the lender offers to pay him the certainty equivalent 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 $(\theta^{t-1}, \theta'')$ with $\theta'' > \theta'$ have higher expected returns at period $t+1$. So it is not attractive for them to sell equity at price $P_{\Delta\chi}((\theta^{t-1}, \theta'))$. Therefore, the lender can use equity purchases, which inefficiently lower the equity share, to better screen types.

More formally, this intuition is related to the Atkinson and Stiglitz (1976) result for commodity taxation. With i.i.d shocks, less productive entrepreneurs are also more willing to sell equity as they have higher marginal utility. But in this case, the willingness to sell equity does not reveal any information to the lender that is not already contained in reported returns. With persistence, lower types would be more willing to sell equity even if they had the same marginal utility as higher types. So the lender optimally distorts the equity share as it directly reveals information about the entrepreneur's productivity.

(*Note: The implementation with persistence is still work in progress, but I write some initial steps here*)

The first step toward the implementation is understanding the stochastic process of the equity share. For this, it is easier to focus on the multiplier of the constraint (29), denoted by $\gamma_t$, instead of $\Delta_{t-1}$ directly. Combining the FOC for $\Delta_t(\theta^t)$ and equation (15), we obtain

$$\gamma_{t+1}(\theta^t) = \frac{q}{\beta}\hat{\rho}(\theta^t)\gamma_t(\theta^{t-1}) - \frac{q}{\beta}MB(\theta^t)$$

which resembles an autoregressive process with innovations given by $MB(\theta^t)$, see figure 2. The implementation is also complicated by the inverse u-shape of $\gamma_{t+1}(\theta^t)$ over $\theta_t$. To see this, note the first order condition for $\Delta_t(\theta^t)$ is

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

and the co-state $\mu_t$ goes to zero as $\theta_t \to \overline{\theta}, \underline{\theta}$, see figure 7 in Appendix A. So, unless the shocks are Pareto distributed, there should be no distortion at the top and bottom in the equity share at $t+1$. The second step is understanding the LOM of the entrepreneur's wealth with persistence and share purchases and sales. The issue is that the equity share cannot be identified from the regressions as for the i.i.d case, as the equity purchases and sales also affect the entrepreneur's wealth and consumption. To start, consider first the case with persistence but fixed equity share. Denote

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

then the entrepreneurs cash on hand at period $t$ if he reports type $\widetilde{\theta}_t$ and his past type report was $\widetilde{\theta}_{t-1}$ is

$$C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) = \frac{1}{q}W_t + \chi\left(f_t(k_t(\widetilde{\theta}^{t-1}), \widetilde{\theta}_t) + q\overline{f}_{t+1}(\widetilde{\theta}_t) - \overline{f}_t(\widetilde{\theta}_{t-1})\right)$$

Note that with persistence, after a high report, the entrepreneur also obtains a capital gain because the net present value of the firm's cash flows increases. Now consider that the lender can buy (sell) equity $\Delta\chi_{t+1} = \chi_{t+1} - \chi_t < 0$ ($> 0$) at per-unit price $P_{\Delta\chi_{t+1}}(\widetilde{\theta}_t) > 0$. Then we have

$$\begin{aligned} C(W_t, \chi_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) = &\frac{1}{q}W_t + \chi_t\left(f_t(k_t(\widetilde{\theta}^{t-1}), \widetilde{\theta}_t) + q\overline{f}_{t+1}(\widetilde{\theta}_t) - \overline{f}_t(\widetilde{\theta}_{t-1})\right) \\ &- \Delta\chi_{t+1}(\widetilde{\theta}_t)\left(P_{\Delta\chi_{t+1}}(\widetilde{\theta}_t) - q\overline{f}_{t+1}(\widetilde{\theta}_t)\right) \end{aligned}$$

There are now four terms that depend on the current period type, compared to only one in the i.i.d case. $\Delta\chi_{t+1}(\widetilde{\theta}_t)$ could potentially be backed out from the the difference $\gamma_{t+1}(\theta^t)-\gamma_t$. $\overline{f}_{t+1}(\widetilde{\theta}_t)$ can be computed numerically. The equity prices $P_{\Delta\chi_{t+1}}(\widetilde{\theta}_t)$ may be more challenging, one approach could be to restrict it to be the price such that the entrepreneur is indifferent, and then try to approximate the risk premium.

# 7 Comparison with risk-neutral and equity dynamics

The implementation helps understand the different firm size dynamics with risk neutrality and risk aversion. With risk neutrality, as long as the limited liability constraint 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 on 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 with CRRA utility, I find 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 optimall allocation. With persistent types, the equity share is also time-varying as for 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, over time, the equity share of the entrepreneur tends to decrease. When the equity share is low, the entrepreneur has more incentives to divert funds, so the lender is less willing to lend high capital.

Consequently, both models obtain a positive relation between equity and firm size. However, equity drifts in opposite directions. With risk neutrality, equity drifts upwards, but with risk aversion and persistence, equity drifts downwards. The model with risk neutrality obtains that firm size converges to the first best level only because the entrepreneur's equity share goes to one. That is, he becomes the sole owner of the firm and the value of debt and outside equity go to zero. These equity dynamics are inconsistent with what is observed in the data. 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.

In the numerical simulations, it turned out that a linear compensation with a constant equity share gave a very close approximation. More generally, this may not always be the case, and if compensation is more nonlinear extra instruments such as stock options would be needed (see Fu and Krishna (2019)). However, the main point should still stand. The relevant state variable for an implementation with these instruments would be promised insurance ($\Delta_{t-1}$), not promised utility ($v_{t-1}$).

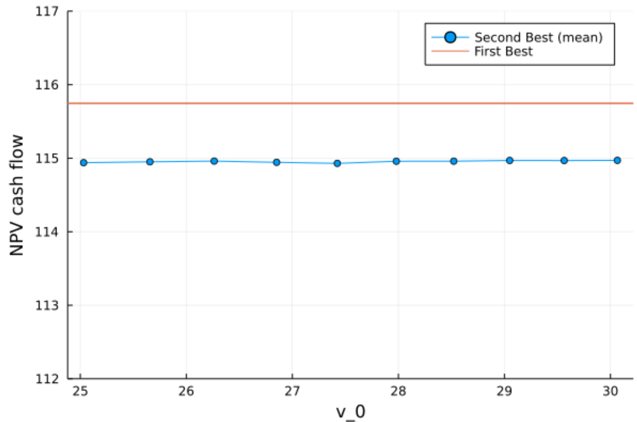
In what follows, I analyze the implications of the model for the role of the capital structure on the firm's value (the Modigliani-Miller theorem) and the link between the financing constraints and the investment-cash flow sensitivity. In both cases, I find similar results to the risk neutral model only if promised insurance is considered to be the relevant variable for the firm's capital structure and financing constraints. This would not be the case if we took promised utility as the relevant state variable.

**Modigliani-Miller and promised utility** With risk-neutrality, the firm's value depends on the value of equity (or promised utility), so the Modigliani-Miller theorem does not hold (Clementi and Hopenhayn (2006)). Interestingly, I find in the numerical simulations that, with risk aversion, the value of the firm does not vary with the initial promised utility given to the lender. So in this sense, Modigliani-Miller does hold "over promised" utility. This observation corroborates the idea of the implementation with risk aversion presented in the previous section. Promised utility does not map properly to the entrepreneur's equity, instead, it maps to the entrepreneur's private wealth.

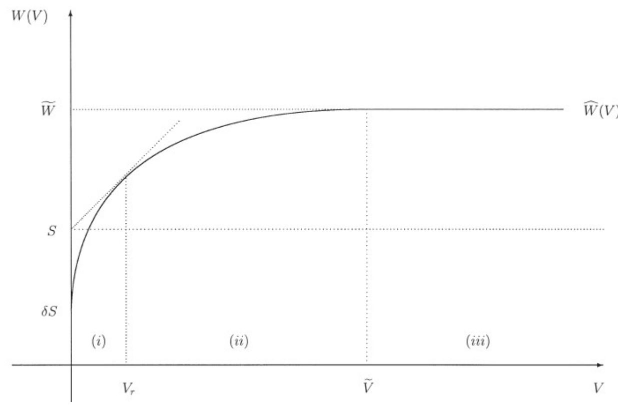
I illustrate these differences in panels (a) and (b) of figure 3. Panel (a) shows the net present value of the firm cash flows for different initial levels of promised utility. As we can observe, the line is approximately flat. Panel (b) shows a plot of the firm's value also as a function of promised utility with risk neutrality from Clementi and Hopenhayn (2006). By contrast, the firm's value is now increasing in promised utility until the region where the firm reaches it's first best value . Panel (c) and (d) also show the net present value of the firm but with persistent types. In (c), I fix the promised insurance $\Delta_{t-1}$ and vary promised utility $v_{t-1}$. Similar to the i.i.d case, the firm's value does not vary much with promised utility. In (d), I instead fix $v_{t-1}$ but vary $\Delta_{t-1}$, now similar to (b) the firm's value decreases more as $\Delta_{t-1}$ decreases. This supports the idea that with risk aversion, it is promised insurance what maps to the entrepreneur's equity share and so what affects the firm's capital structure.

Figure 3: Value of the firm over initial promised utility or insurance: Risk averse vs Risk neutral

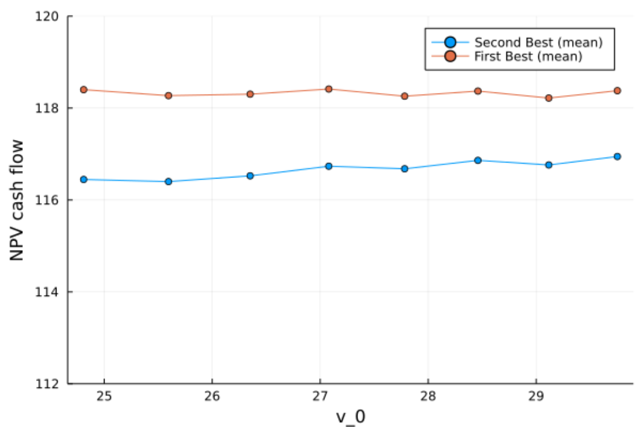
(a) NPV cash flow with risk aversion and i.i.d



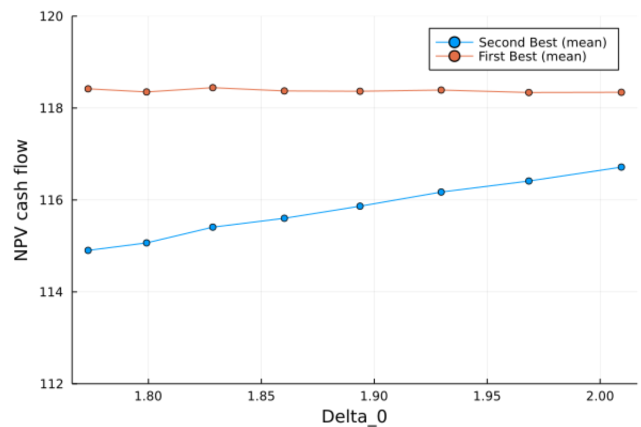

(b) Value of the firm with risk neutrality (from Clementi and Hopenhayn (2006))




(c) NPV cash flow with risk aversion and persistence (over $v_0$)




(d) NPV cash flow with risk aversion and persistence (over $\Delta_0$)



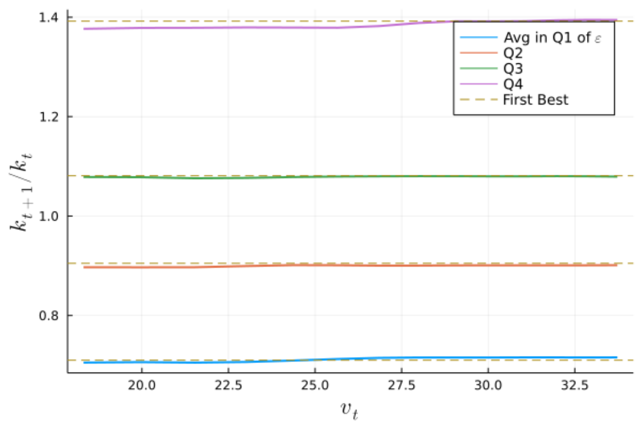

**Investment-cash flow sensititivty** Starting with Fazzari *et al.* (1988), an extensive empirical literature used the sensitivity of the firm's investment to its cash flow as a measure of financing constraints. Later on, Kaplan and Zingales (1997) provided convincing evidence that there is no relation between the investment-cash flow sensitivity and financing constraints. Setting the empirical debate aside, we may ask whether, in an environment where financing constraints are an endogenous outcome of the optimal contract, do we observe higher investment-cash flow sensitivity for financially constrained firms? The answer is positive with risk neutrality (Clementi and Hopenhayn (2006), DeMarzo and Fishman (2007a)).

Panel (b) of figure 4 (from Clementi and Hopenhayn (2006)) shows that investment responds more to the cash flow realization when the value or equity (or promised utility) is low.

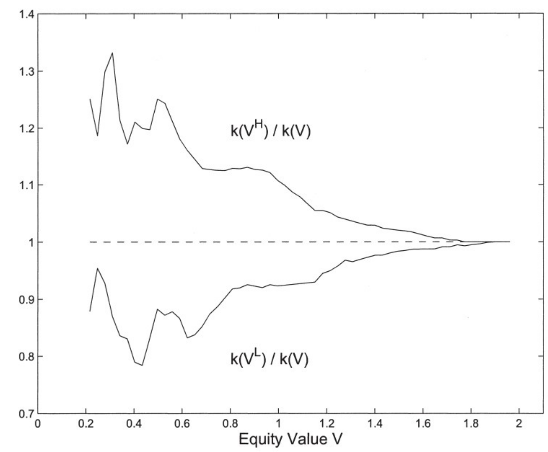
The answer to this question with risk aversion and persistence will depend again on what state variable of the optimal contract best relates to financing constraints. In panel (a), I plot the average growth rate of capital $\frac{k_{t+1}(\theta_t)}{k_t}$ at the different quartiles of the distribution of the shock $\varepsilon_t$ over different values of promised utility. Mechanically, we will observe some investment-cash flow sensitivity when shocks are persistent even without financing constraints. For this reason, the figure also contains the growth rates in the first best (dashed lines). As we can observe, the sensitivity of investment now does not depend on the promised utility. This observation is again consistent with the idea that with risk aversion, promised utility is related to the entrepreneur's private wealth but not to the financing constraints that the firm faces. Panel (c) contains the same type of plot but now varies the promised insurance ($\Delta_t$). Although the effects are minimal, as $\Delta_t$ decreases, the growth rate $\frac{k_{t+1}(\theta_t)}{k_t}$ is relatively higher for the high cash flow realization. So if $\Delta_t$ is the relevant measure of financing constraints, then there is some positive relation between investment-cash flow sensitivity and financing constraints, but the effect appears to be very small.

Figure 4: Growth rate capital (firm size) by type realization

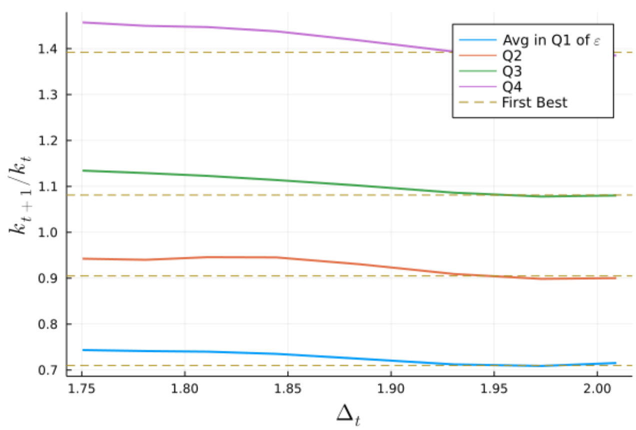
(a) With risk aversion and persistent types (over promised utility $\lambda_t$)




(b) With risk neutrality and i.i.d types (from Clementi and Hopenhayn (2006))




(c) With risk aversion and persistent types (over promised insurance $\gamma_t$)




# 8 Conclusion

In this paper, I have studied a dynamic cash flow diversion model with a risk averse agent that has persistent private information about the firm's productivity. I have used the first order approach and a change of measure to solve and derive analytical characterizations of the optimal contract. The firm size and compensation dynamics differ significantly from models with risk neutrality. Most notably, firm size tends to decrease over time, as opposed to models with risk neutrality. The implementation helps understand the opposite size dynamics. Equity drifts upwards with risk aversion and (in the implementation) downwards with risk aversion and persistence. These findings suggest that it may be challenging for this type of models to generate realistic firm size and equity dynamics. As in the risk neutral case, firm size converges to the first best only because the entrepreneur's equity share goes to one.

The implementation section is still work in progress. The next steps include solving the third best implementation with persistence numerically, and trying to derive a full second best implementation with a specific parametrization such as CARA utility and a unit root process.

## 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.

Atkinson, A. B. and Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of public Economics*, **6** (1-2), 55–75.

Battaglini, M. and Lamba, R. (2019). Optimal dynamic contracting: The first-order approach and beyond. *Theoretical Economics*, **14** (4), 1435–1482.

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. (2020). *Persistent private information revisited*. Tech. rep., Technical report, Stanford University.

Bloedel, A. W., Krishna, R. V. and Leukhina, O. (2018). Insurance and inequality with persistent private information. *Available at SSRN 3091457*.

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

— (2022). Optimal nonlinear savings taxation.

Chi, C.-K. and Choi, K. J. (2017). The impact of firm size on dynamic incentives and investment. *The RAND Journal of Economics*, **48** (1), 147–177.

Clementi, G. L. 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.

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.

Fazzari, S. M., Hubbard, R. G., Petersen, B. C., Blinder, A. S. and Poterba, J. M. (1988). Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, **1988** (1), 141–206.

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., Kocherlakota, N. and Tsyvinski, A. (2003). Optimal indirect and capital taxation. *The Review of Economic Studies*, **70** (3), 569–587.

—, 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.

— and Werquin, N. (2022). A fair day's pay for a fair day's work: Optimal tax design as redistributional arbitrage.

Hsieh, C.-T. and Klenow, P. J. (2014). The life cycle of plants in india and mexico. *The Quarterly Journal of Economics*, **129** (3), 1035–1084.

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.

Kaplan, S. N. and Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *The quarterly journal of economics*, **112** (1), 169–215.

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

Leary, M. T. and Michaely, R. (2011). Determinants of dividend smoothing: Empirical evidence. *The Review of Financial Studies*, **24** (10), 3197–3249.

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

— and — (2021). Taxation under learning by doing. *Journal of Political Economy*, **129** (6), 1878–1944.

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.

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

— (2020). Dynamic taxation. *Annual Review of Economics*, **12**, 801–831.

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.

# A Additional tables and figures

Figure 5: Immiseration in the long run

(a) Marginal utility i.i.d ($\rho = 0$)

(b) Marginal utility with persistence ($\rho = 0.7$)

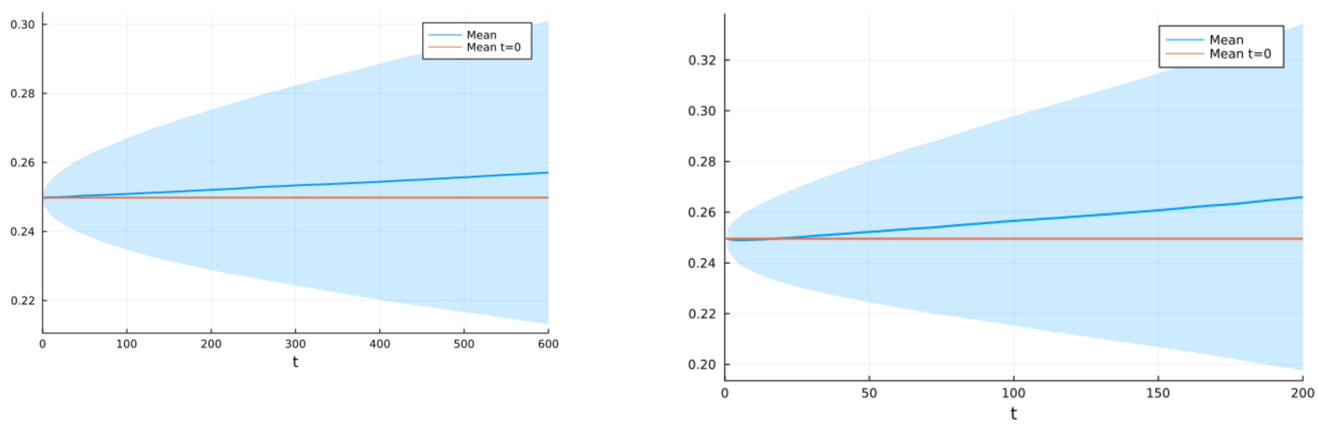


Figure 6: Relation $k_{t+1}(\theta_t)$ and $\theta_t$

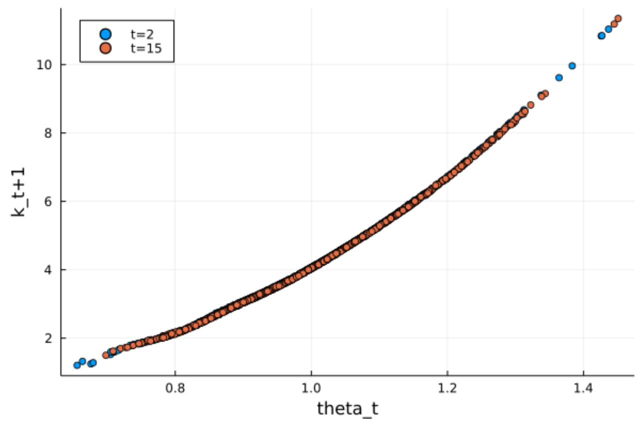



Figure 7: Shadow cost insurance $\mu$ at different $\gamma$

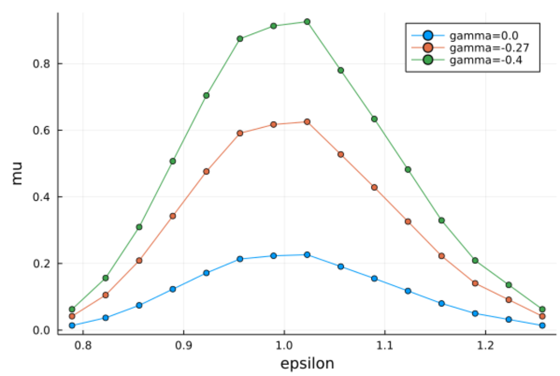



Figure 8: Simulations implementation i.i.d

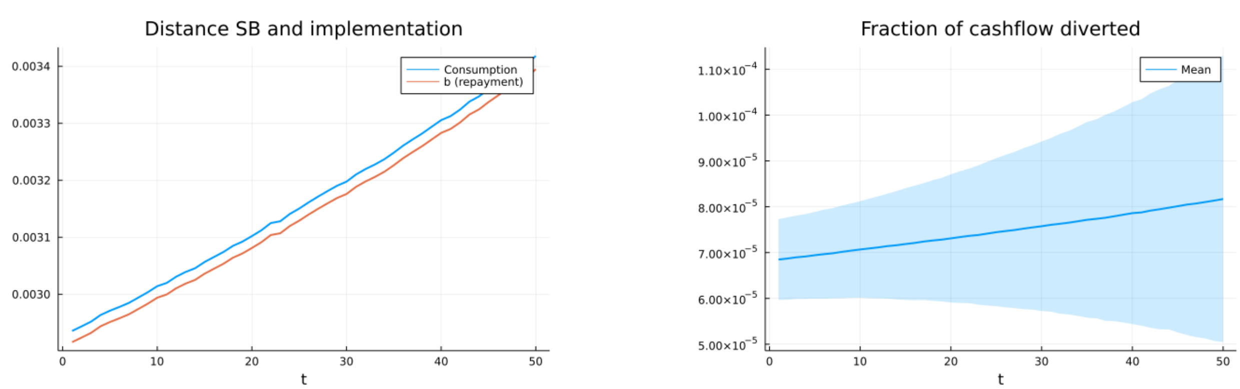


Figure 9: Simulations implementation i.i.d and CRRA with $\sigma = 2$

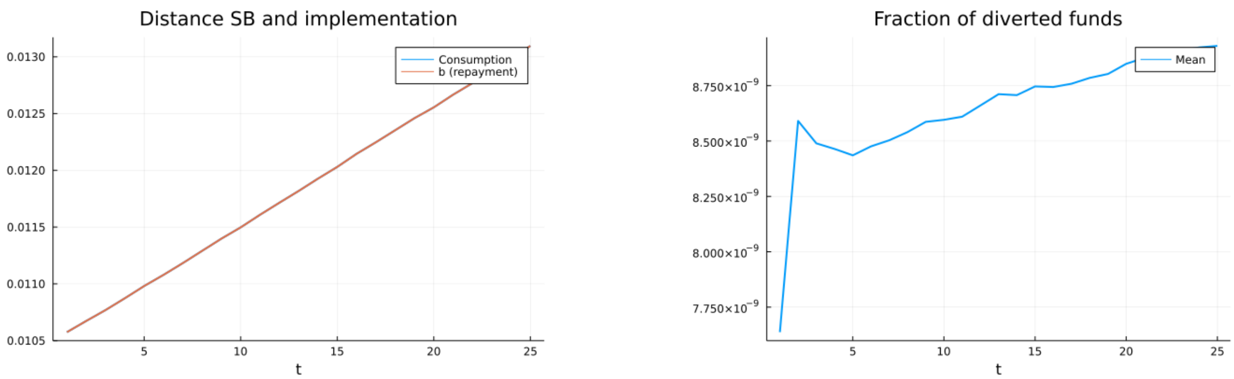
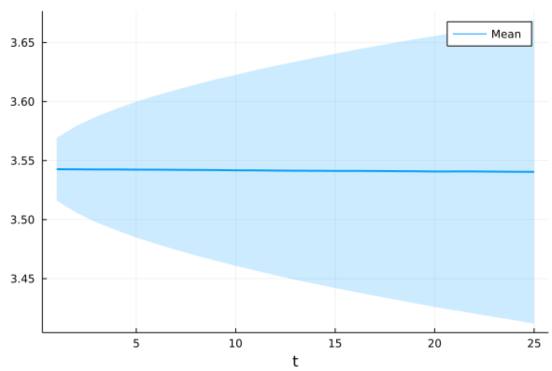



Figure 10: CRRA utility with $\sigma = 2$

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




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

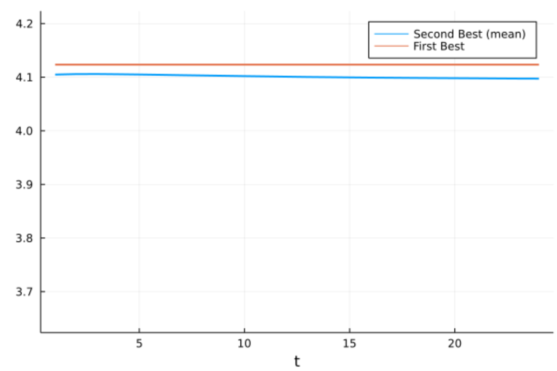
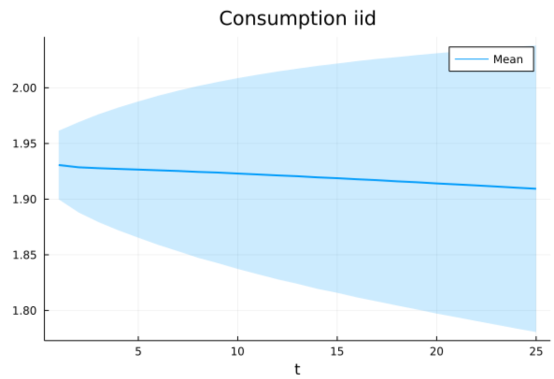
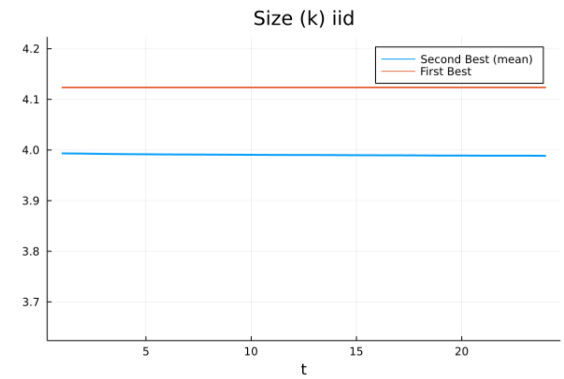


Figure 11: CARA utility

(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

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

The first order conditions are

$b_t(\theta^t)$ :

$$\xi_t(\theta^t) = \frac{1}{u'(\theta^t)} \left[1 + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\phi f_\theta(\theta^t)u''(\theta^t)\right] \tag{30}$$

The envelope conditions are

$$\frac{\partial K_{t+1}}{\partial v_t(\theta^t)} = \lambda_{t+1}(\theta^t) \tag{31}$$

$$\frac{\partial K_{t+1}}{\partial \Delta_t(\theta^t)} = \gamma_t(\theta^t) \tag{32}$$

$$\frac{\partial K_{t+1}}{\partial k_{t+1}(\theta^t)} = \tag{33}$$

$$\mathbb{E}\left[-\xi_{t+1}(\theta^{t+1})u'(\theta^{t+1})f_k(\theta^{t+1}) + \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta^t)}\left(u''(\theta^{t+1})\phi f_\theta(\theta^{t+1})f_k(\theta^{t+1}) + u'(\theta^{t+1})\phi f_{\theta k}(\theta^{t+1})\right)|\theta^t\right.$$

Using the envelope conditions (31) and (32) we get

$v_t(\theta^t)$ :

$$\lambda_{t+1}(\theta^t) = \frac{\beta}{q}\xi_t(\theta^t) \tag{34}$$

$\Delta_t(\theta^t)$ :

$$\gamma_{t+1}(\theta^t) = -\frac{\beta}{q}\frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} \tag{35}$$

Substituting (30) and (33) into the FOC for $k_{t+1}(\theta^t)$ we get

$$\frac{1}{q} = \mathbb{E}\left[f_k(\theta^{t+1}) - \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta^t)}u'(\theta^{t+1})\phi f_{\theta k}(\theta^{t+1})|\theta^t\right] \tag{36}$$

using $\tau^k(\theta^t) = \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}u'(c(\theta^t))\phi\frac{f_{\theta k}(k(\theta^t),\theta_t)}{f_k(k(\theta^t),\theta_t)}$ we have

$$\frac{1}{q} = \mathbb{E}\left[f_k(\theta^{t+1})(1-\tau^k(\theta^{t+1}))|\theta^t\right]$$

which proves **Proposition 5**. Finally the LOM for the co-state is

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

**Proof Proposition 1** Set $\mu_t(\theta^t) = 0$, then from equation (36) we obtain point 3. For point 2, note that with $\mu_t(\theta^t) = 0$ the LOM of the co-state becomes

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

From equation (30),

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

and using (34) gives point 2. Point 1 holds in the first best and second best allocations.

**Proof Proposition 2** These are the same steps as proposition 1 in Hellwig (2021). Substitue $\xi^J_t(\theta^t)$ in the LOM of the co-state:

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

substitute $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\phi f_\theta(\theta^t)}{u'(\theta^t)}$, 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

$$\frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} = \frac{1-\hat{\Phi}_t(\theta_t|\theta^{t-1})}{\hat{\varphi}_t(\theta_t|\theta^{t-1})}\left\{\hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta', \theta^{t-1}\right)} \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \gamma_t\hat{\mathbb{E}}\left[\mathcal{E}(\theta', \theta^{t-1}) \mid \theta' \geq \theta_t, \theta^{t-1}\right] - \lambda_t\right\} \tag{37}$$

To get $\lambda_t$, 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

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

Substiuting back $\lambda_t$ into equation (37) and using the definition of $\hat{\rho}(\theta^t)$ (equation (18)) we get the solution.

**Proof Proposition 3** Using $\hat{\varphi}(\theta_t|\theta^{t-1}) = \frac{\widetilde{\varphi}_t(\theta_t|\theta^{t-1})u'(\theta^t)}{\widetilde{\mathbb{E}}[u'(\theta^t)|\theta^{t-1}]}$, we have the following equivalences

$$\hat{\Phi}_t(\theta_t|\theta^{t-1}) = \widetilde{\Phi}(\theta_t|\theta^{t-1})\frac{\widetilde{\mathbb{E}}\left[u'(\theta')|\theta' \leq \theta_t, \theta^{t-1}\right]}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]} \tag{38}$$

$$\hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta^t\right)} \mid \theta^{t-1}\right] = \frac{1}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]} \tag{39}$$

$$\hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta'\right)} \mid \theta' \geq \theta^t, \theta^{t-1}\right] = \frac{1-\widetilde{\Phi}(\theta_t|\theta^{t-1})}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right] - \widetilde{\Phi}(\theta_t|\theta^{t-1})\widetilde{\mathbb{E}}\left[u'(\theta')|\theta' \leq \theta_t, \theta^{t-1}\right]} \tag{40}$$

Subsituting these into equation (16) and rearranging we get

$$MB(\theta^t) = \frac{1}{\widetilde{\varphi}_t(\theta_t|\theta^{t-1})}\frac{1}{u'(\theta^t)}\left\{1 - \widetilde{\Phi}(\theta_t|\theta^{t-1}) - 1 - \widetilde{\Phi}(\theta_t|\theta^{t-1})\frac{\widetilde{\mathbb{E}}\left[u'(\theta')|\theta' \leq \theta_t, \theta^{t-1}\right]}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]}\right\}$$
$$= \frac{1 - \widetilde{\Phi}_t(\theta_t|\theta^{t-1})}{\widetilde{\varphi}_t(\theta_t|\theta^{t-1})}\frac{1}{u'(\theta^t)}\left[1 - \frac{\widetilde{\mathbb{E}}\left[u'(\theta', \theta^{t-1})|\theta' > \theta_t, \theta^{t-1}\right]}{\widetilde{\mathbb{E}}\left[u'(\theta^t)|\theta^{t-1}\right]}\right]$$

Finally, we also have

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

subistituting this and equation (39) into (19) shows the second equation of the proposition.

**Proof Proposition 4** This proof also follows similar steps to Theorem 1 in Hellwig (2021). Using the characterization of $\lambda_t$ in Proposition 2 and substitute the multipliers $\lambda_{t+1}(\theta^t)$ and $\gamma_{t+1}(\theta^t)$ from the optimality conditions (34) and (35), and using equation (30) to substitute for $\xi_t$:

$$\frac{1}{u'(\theta^t)} + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\frac{\phi f_\theta(\theta^t)u''(\theta^t)}{u'(\theta^t)} = \frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] + \frac{\mu(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\hat{\mathbb{E}}\left(\mathcal{E}(\theta_{t+1}, \theta^t)|\theta^t\right) \quad (41)$$

where we can rewrite

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

To show this, note we can write

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

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 $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\phi f_\theta(\theta^t)}{u'(\theta^t)}$

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

Substitute back and use the definition of the investment wedge to substitute $\frac{\mu_t(\theta^t)}{\varphi(\theta_t|\theta^{t-1})}$

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

Rearranging we get

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

# C 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 $\theta_{t-1}$ 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 ($\rho$). Therefore, I solve the full problem without renormalizing.

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

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

and

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

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 $\varepsilon_t$ is

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

The optimality conditions are

$$\frac{q}{\beta}\lambda_t(\varepsilon_t) = \frac{1}{u'(c_t(\varepsilon_t))}\left[1 + \frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)}\theta_{t-1}^{\rho}\phi f_\theta(k_t, \theta_{t-1}^{\rho}\varepsilon_t)u''(c(\theta))\right] \tag{42}$$

$$\gamma_t(\varepsilon_t) = -\frac{\beta}{q}\theta_{t-1}^{\rho}\frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)} \tag{43}$$

And the two LOM

$$\dot{\mu}(\varepsilon_t) = -\left[\frac{q}{\beta}\lambda_t(\varepsilon_t) - \lambda_{t-1} + \gamma_{t-1}\frac{\rho}{\theta_{t-1}}\frac{\tilde{g}_\varepsilon(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)}\right] g_\varepsilon(\varepsilon_t) \tag{44}$$

$$\dot{w}_t(\varepsilon_t) = \theta_{t-1}^{\rho}\left(u'(c(\varepsilon_t))\phi f_\theta(k_t, \theta_{t-1}^{\rho}\varepsilon_t) + \beta\Delta_t(\varepsilon_t)\right) \tag{45}$$

I truncate the distribution of $\varepsilon_t$ 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 modificed to $(\lambda_-, \gamma_-, k, \theta_-)$ , so the multipliers $\lambda_-$ and $\gamma_-$ are used instead of $v_-$ and $\Delta_-$, respectively. I use 14 grid points for $\lambda_-$, 8 for $\gamma_-$, 20 for $k$ and 10 for $\theta_-$. I interpolate on $K$, $v$ and $\Delta$ 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 $\gamma_-$ and $\theta_-$.

## Algortihm

**Step 0:** Guess the value function $K'$, promised utility $v'$ and promised marginal utility $\Delta'$ 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

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

(Note: $k_+$ needs to be computed multiple times at every step while solving the ODE. But to improve speed, can solve before the policies on a dense grid and then interpolate when solve 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 $\varepsilon$, solve $\lambda(\varepsilon)$ in equation (42) and $\gamma(\varepsilon)$ in equation (43). 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 (44) and (45). Note when solving (44) also need to interpolate $\Delta'$ on $(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^\rho \varepsilon)$.

- c) Check the boundary condition $\mu(\overline{\varepsilon}) = -\gamma_- \frac{\rho}{\theta_-} \overline{\varepsilon} g_\varepsilon(\overline{\varepsilon})$. 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 previouss guess of $K'$, $v'$ and $\Delta'$ , and repeat from **Step 1** until the convergence criteria is satisfied.

## C.1 Solution implementation

# D Extensions

## D.1 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 $\mu_t$, the GIIE and the firm size dynamics. Moreover, the investment wedge $\tau^k(\theta^t)$ is also 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 part $a_t$ of the funds for his consumption. Therefore, invested and diverted funds are subject to the flow of funds constraint

$$k_t + a_t \leq B_t \tag{46}$$

The lender now observes returns $f(k_t, \theta_t)$ but not productivity $\theta_t$ 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 $g'' < 0 < g'$, so the entrepreneur's consumption is

$$c_t = f(k_t, \theta_t) - b_t + g(a_t) \tag{47}$$

The principal within period objective now is $B_t - b_t$. The envelope condition is

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

The rest of the planning problem is the same but with the extra flow of funds constraint (46). The optimality condition for diverted funds is

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

where $\lambda_t(\theta^t)$ is the multiplier on the flow of funds constraint. The FOC for investment is

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

Now the investment wedge can be defined explicitely 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

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

Then combining the two optimality conditions we get

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

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. Then, it is easy to verify that this model yields the same characterization for the shadow costs $\mu_t(\theta^t)$, the GIIE and the project size dynamics as the cash flow diversion model studied in the main text.

## D.2 Limited commitment

In this section, I relax the assumption of full commitment of the entrepreneur. Limited commitment leads to very different firm size and compensation dynamics than 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 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 $h(k_{t+1}(\theta^t)) \leq v(\theta^t)$. This limited commitment constraint can be added directly to the planning problem (11). Because the limited commitment constraint does not affect the within period insurance and incentives trade-off, the characterization of the shadow cost of insurance (Proposition 2) is not affected by the limited commitment assumption.

However, the limited commitment constraint does modify the consumption dynamics (Proposition 5) and the project size dynamics (Proposition 4). Let $\eta_t(\theta^t)$ be the multiplier on the limited commitment constraint, then the GIEE is given by

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

Because $\eta_t(\theta^t) \geq 0$, the limited commitment gives a force to have a downwards drift in marginal utilities. As is well know, 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 in general, have opposite effects on consumption dynamics.

The project size dynamics are now given by

$$\frac{1+\eta_t(\theta^t)h'(k_{t+1}(\theta^t))}{q} = \mathbb{E}\left[f_k(k_t(\theta^t),\theta_t)(1-\tau^k(\theta^t))|\theta^t\right]$$

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 promised utility increases over time, the limited commitment constraint will eventually not bind ($\eta_t(\theta^t) = 0$). Therefore, this friction still gives a force towards having firm size increasing over time.

## D.3 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 Pareto frontier is not concave, the principal may optimally randomize between terminating the project or continuing. I assume that after termination, the lender receives a scrap value $S$. At period $t$, based on $\theta^t$, 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 $Q_{t+1}(\theta^t)$. In case of no termination at period $t$ the objective of the principal is

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

I assume that after terminating the contract, the entrepreneur can freely save $Q_{t+1}(\theta^t)$ and obtains a per period gross return $2-q$. So in this scenario, his continuation utility 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

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

It is then easy to see that the optimality condtions 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, althought it may be optimal to exit the project, the characterizations of the optimal contract presented in the paper do not rely on the assumption of no termination.

### D.4 Costly effort

# E Application Sovereign Debt: Dovis (2019)

In this section, I will show how these techniques can be used to solve the sovereign debt model in Dovis (2019). To use the FOA I assume there is a continuum of types, instead of two as in the paper. I also allow for persistent private information. The rest of the model is the same as in the paper.

The foreign lenders (the principal) lend $m$ units of the intermediate good to the domestic government (the agent) and recieve $x$ exports in return. A benevolent domestic government can use the intermediates to produce domestic good $c$ or exports $x$. The agent's type $\theta^t$ is now the relative productivity of the domestic good. With constant returns to scale, we can write the country's aggregate resource constraint as

$$\frac{c_t}{\theta_t} + x_t \leq f(m_t) \qquad (RC) \tag{48}$$

The principal can observe exports $x$ and how inputs are used but not $\frac{c_t}{\theta_t}$. The domestic government maximizes

$$w(\theta_0) = \mathbb{E}\big[\sum_{t=0}^{\infty} \beta^t u(c_t)|\theta_0\big]$$
$$s.t \quad RC$$

There are also two limited commitment constraints

$$u(c(\theta^t)) + \beta v_t(\theta^t) \geq u(\theta_t f(m_{t-1})) + \beta v_a$$
$$v_t(\theta^t) \geq v_a$$

where $v_a$ is the value of autarky. The local IC constraint is

$$\frac{\partial}{\partial\theta_t}w(\theta^t) = u'(c(\theta^t))\left[f(m(\theta^{t-1}) - x(\theta^t)\right] + \beta\int w(\theta^t)\frac{\partial\varphi_{t+1}(\theta_{t+1}|\theta^t)}{\partial\theta^t}d\theta_{t+1}$$
$$= u'(c(\theta^t))\frac{c(\theta^t)}{\theta^t} + \beta\int w(\theta^t)\frac{\partial\varphi_{t+1}(\theta_{t+1}|\theta^t)}{\partial\theta^t}d\theta_{t+1}$$

As in the corporate finance model, static marginal information rents, here $u'(c(\theta^t))\frac{c(\theta^t)}{\theta^t}$, depend on consumption. So the incentive-adjusted probability measure can be used to characterize the optimal allocation. The relaxed planning problem then is

$$K_t(v,\Delta,\theta^{t-1},m_{t-1}) = \min\int\left[m_t(\theta^t) - x_t(\theta^t) + qK_{t+1}(v_t(\theta^t),\Delta_t(\theta^t),\theta^t,m_t(\theta^t))\right]\varphi_t(\theta_t|\theta^{t-1})d\theta_t$$

$$\begin{aligned}
s.t\quad (PK)\quad & w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \qquad [\xi_t^J(\theta^t)] \\
& v = \int w_t(\theta^t)\varphi_t(\theta_t|\theta^{t-1})d\theta_t \qquad [\xi_t^{PK}] \\
(IC)\quad & \dot{w}(\theta) = u'(c(\theta^t))\left[f(m(\theta^{t-1}) - x(\theta^t)\right] + \beta\Delta_t(\theta^t) \qquad [\mu_t(\theta^t)] \\
& \Delta = \int w_t(\theta^t)\frac{\partial\varphi_t(\theta_t|\theta^{t-1})}{\partial\theta^{t-1}}d\theta_t \qquad [\xi_t^{IC}] \\
(Feasibility)\quad & x(\theta^t) = f(m_{t-1}) - \frac{c(\theta^t)}{\theta^t} \qquad [\lambda_t(\theta^t)\varphi_t(\theta_t|\theta^{t-1})] \\
(LC)\quad & u(c(\theta^t)) + \beta v_t(\theta^t) \geq u(\theta_t f(m_{t-1})) + \beta v_a \qquad [\eta_t(\theta^t)\varphi_t(\theta_t|\theta^{t-1})] \\
& v_t(\theta^t) \geq v_a \qquad [\eta_t^{PK}(\theta^t)\varphi_t(\theta_t|\theta^{t-1})]
\end{aligned}$$

The optimality conditions are:

Combining the FOCs of $x_t(\theta^t)$ and $c_t(\theta^t)$:

$$\xi^J(\theta^t) = \frac{1}{u'(c(\theta^t))}\left[\frac{\lambda_t(\theta^t)}{\theta^t} + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}u''(c(\theta^t))\left[f(m(\theta^{t-1}) - x(\theta^t)\right]\right] - \eta_t(\theta^t)$$

$v_t(\theta^t)$:

$$\frac{q}{\beta}\xi_{t+1}^{PK}(\theta^t) = \xi_t^J(\theta^t) + \eta_t(\theta^t) + \frac{\eta_t^{PK}(\theta^t)}{\beta}$$

$\Delta_t(\theta^t)$:

$$\frac{q}{\beta}\xi_{t+1}^{IC}(\theta^t) = \frac{\beta}{q}\frac{\mu(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}$$

Using the FOC for $m_t(\theta^t)$ and subsituting $\lambda_t(\theta^t)$ from the FOC for $x_t(\theta^t)$:

$$\frac{1}{q} = f'(m_t(\theta^t)) \left[1 - \mathbb{E}\left[\eta_{t+1}(\theta^{t+1}) u'(\theta_{t+1} f(m_t(\theta^t)) | \theta^t\right]\right]$$

Substituting $\xi_t^J(\theta^t)$ in the LOM of the co-state

$$\dot{\mu}_t(\theta^t) + \mu_t(\theta^t) \frac{1}{\theta_t} \left(1 + \frac{u''(\theta^t)}{u'(\theta^t)} c(\theta^t)\right) = \left[\xi^{PK} + \xi^{IC} \mathcal{E}(\theta_t, \theta^{t-1}) - \frac{1}{u'(\theta^t)\theta_t}\right] \varphi_t(\theta_t | \theta^{t-1})$$

Now the change of measure is with $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{1}{\theta_t}\left(1 + \frac{c(\theta^t)u''(\theta^t)}{u'(\theta^t)}\right)$, so the sign depends on the relative risk aversion (RRA). With log utility, marginal information rents do not depend on consumption $u'(c(\theta^t))\frac{c(\theta^t)}{\theta^t} = \frac{1}{c(\theta^t)}\frac{c(\theta^t)}{\theta^t} = \frac{1}{\theta^t}$. Therefore, $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{1}{\theta_t}(1-1) = 0$ and the solution can be characterized under the original type measure as in a Mirrlees model with separable preferences. If the RRA is bigger than one, $\frac{m'(\theta^t)}{m(\theta^t)} < 0$ the incentive-adjusted measure puts higher weight on lower types. And conversely if the RRA is smaller than one. Dovis (2019) discusses the role of the RRA in the model. The intuition is that the strength of the income and substitution effects determine weather the high or low types want to export more.

Using $\frac{\mu_t(\theta^t)}{\varphi(\theta_t|\theta^{t-1})} = \frac{\lambda(\theta^t)-1}{u'(\theta^t)}$ from the FOC of $x_t(\theta^t)$ we get the following GIEE

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

$$s(\theta^t) = \left(\lambda(\theta^t) - 1\right)\left[\frac{m'(\theta^t)}{m(\theta^t)} - \hat{\mathbb{E}}\left[\rho_{t+1}(\theta^{t+1})\frac{m'(\theta^{t+1})}{m(\theta^{t+1})} | \theta^t\right]\right]$$