Optimal Financial Contracting and the Effects of Firm's Size

Sandro Brusco, Giuseppe Lopomo, Eva Ropero, and Alessandro T. Villa

July 2020

WP 2024-18

https://doi.org/10.21033/wp-2024-18

FEDERAL RESERVE BANK of CHICAGO

Optimal Financial Contracting and the Effects of Firm's Size*

Sandro Brusco[§] Giuseppe Lopomo[¶] Eva Ropero[‖] Alessandro T. Villa[**]

Revised version – July 2020

Abstract

We consider the design of the optimal dynamic policy for a firm subject to moral hazard problems. With respect to the existing literature we enrich the model by introducing durable capital with partial irreversibility, which makes the size of the firm a state variable. This allows us to analyze the role of firm's size, separately from age and financial structure. We show that a higher level of capital decreases the probability of liquidation and increases the future size of the firm. Although analytical results are not available, we show through simulations that, conditional on size, the rate of growth of the firm, its variability and the variability of the probability of liquidation decline with age.

*This paper supersedes an earlier paper by Brusco and Ropero titled "Financing Constraints and Firm Dynamics with Durable Capital". We thank Marco Celentani, Gianluca Clementi, Vicente Cuñat, María Gutiérrez, Andrea Lanteri, Álvaro Lario, Eliseo Navarro, Josep Tribò and seminar participants at IE Business School Madrid, Universidad Carlos III de Madrid, the University of Miami, the University of Bath (UK), the Peking University HSBC Business School in Shenzen (China), the National Chi-Nan University (Taiwan) and the National Cheng-Kung University (Taiwan) for useful comments. A special thank goes to the two anonymous referees who helped immensely in improving both the content and the exposition. We remain responsible for errors and obscurities.

[§]Department of Economics and College of Business, Stony Brook University

[¶]Fuqua School of Business, Duke University

[‖]Departamento de Economía y Empresa, Universidad Europea de Madrid

[**]Department of Economics, Duke University

1 Introduction

In this paper we study the optimal dynamic policy of a firm subject to moral hazard problems. Our goal is to characterize theoretically the links between the size and age of the firm and variables such as the probability of survival, the rate of growth and its variability. The empirical literature has uncovered many regularities. Cooley and Quadrini [8] summarize early empirical results stating: "Conditional on age, the dynamics of firms (growth, volatility of growth, job creation, job destruction, and exit) are negatively related to the size of the firm." Similar conclusions are in Sutton [30], who cites studies by Evans [16], [17] and Dunne, Roberts and Samuelson [13], [14]. Sutton [30] observes: "For any given size of firm (or plant), the proportional rate of growth is smaller for older firms (or plants), but the probability of survival is greater."

More recent work has emphasized the role of age. Haltiwanger, Jarmin and Miranda [20] have shown[1] that "without age controls [there is] a strong inverse relationship between firm size and net employment growth". However "once we control for firm age, there is no systematic relationship between firm size and growth". They add that "young firms are more volatile and exhibit a higher rate of gross job creation and destruction that their older counterpart". Similarly, Decker et al. [9] observe that 'conditional on survival, younger firms have much higher rates of job growth'.

The theoretical literature has lagged behind, despite the recent rapid development of the dynamic contracting literature. Analyzing the impact of size requires introducing frictions which make the current level of capital relevant for the choice of optimal policies. We do this in a simple way, assuming partial irreversibility of investment: A firm can sell existing (used) capital only at a price lower than the price of new capital. The solution to the dynamic contracting problem yields policy functions that depend on the existing capital stock and on the value promised to the firm's owner (which, following the literature, we call equity).

In models in which equity is the only state variable and capital is chosen as a function of equity, as it is the case in Quadrini [25] and Clementi and Hopenhayn [6] (on which we build), there cannot possibly be a different behavior by *age* once we condition on *size*. It is only the presence of two state variables that allows firms with the same size to adopt different policies. As it will become clear, it is not age

[1] Haltiwanger, Jarmin and Miranda [20] use employment, rather than accumulated capital, as a measure of size. Their results can be interpreted through our framework if we assume that there is a fixed ratio of capital and labor.

per se which determines the different behavior of firms. Rather, firms of different ages have different distributions of equity *conditional on size*. While the optimal policy of the firm is determined only by the two state variables (value of equity and size), not by age, the probability distribution over the two state variables changes with age. This implies that when only one state variable, e.g. size, is observed, the conditional distribution will depend on age.

Our theoretical results can be summarized as follows. In general, optimal second-best policies prescribe inefficient liquidation and underinvestment with positive probability. However, these effects are smaller for larger firms. More precisely, other things equal, at higher levels of capital the optimal probability of liquidation is smaller, and future levels of capital are higher.

We do not have explicit theoretical results on the effect of age on the rate of growth and the variability of the rate of growth. To get some insights we have run simulations which show that, conditional on size, both the rate of growth and the variability of the rate of growth are negatively related to age. To see the intuition, consider firms with a low level of capital. Younger firms with low capital typically have lower levels of equity because they started with low equity and, even if they have been successful, they are still building up; such firms have more potential for growth. Old firms with a low level of capital typically had unsuccessful stories (this is the reason why their capital is low) and this reflects in a worsening distribution of equity. Thus, these firms tend to grow less.

The rest of the paper is structured as follows. In section 2 we discuss the literature. In section 3 we introduce the model and characterize the optimal policy absent agency problems. In section 4 we study the optimal contract with asymmetric information. Section 5 is dedicated to defining the empirical counterparts of our theoretical model. We next focus in section 6 on the sensitivity to size of the probability of survival, the investment policy and the rate of growth. Finally, section 7 explores the implication of the model for the effect of age. Section 8 contains concluding remarks. Appendix I collects the proofs and Appendix II describes the methodology used in the simulations.

2 Discussion of the Literature

The impact of financial constraints on firms' investment policies has long been an important issue both in financial economics and industrial organization. One line of research has taken the financial constraints as given and analyzed implications for firms' dynamics. Cooley and Quadrini [8], for instance, show that exogenous borrowing constraints and persistent shocks can generate relationships between age and growth of the firm that are consistent with empirical evidence. Cao, Lorenzoni and Walentin [5] reach similar conclusions in a model with limited enforcement in which a continuum of firms are subject to aggregate economy-wide shocks and are exogenously restricted to issue state-dependent securities. More recently, Bolton, Chen and Wang [4] have analyzed the optimal policies with exogenous financial constraints in a continuous time model. In this line of research financial restrictions are exogenous, so the next question is where financial constraints come from. The research has therefore focused on the design of optimal dynamic financial contracts in the presence of asymmetric information or limited ability to enforce contracts.

Optimal Dynamic Contracts. With agency problems the first-best is not attainable. The main theme of this literature is that financing constraints are imposed to provide incentives for the borrower. These incentives typically take the form of threats to either transfer control (including liquidation) or to reduce future financing. In our model asymmetric information plays a central role.[2] We build on the work of Quadrini [25] and, more closely, Clementi and Hopenhayn [6]. In their models, the firm's revenue depends on capital investment and on random shocks[3]. They show that in an optimal contract both the amount of investment and the value of equity increase with high revenue shocks and decrease with low ones. The sensitivity of equity value to revenue shocks provides incentives to the entrepreneur to reveal the true value of the shock. Financing constraints tend to disappear when the value of equity becomes sufficiently large, which in turn happens as the firm approaches the optimal size. In Clementi and Hopenhayn [6] the capital invested in each period depreciates completely at the end of the period. Therefore size, defined as the amount of capital invested in the firm, has to be decided in every period and it is not a state variable. Clementi and Hopenhayn [6] use the value of equity as a proxy for size.

[2]See Albuquerque and Hopenhayn [2] for a model of optimal financial contracts and firm dynamics with limited enforcement. One important issue in these models is the punishment which can be inflicted on a defaulting entrepreneur. Cooley, Marimon and Quadrini [7] consider a model with limited enforcement and optimal financial contracts, assuming that a defaulting entrepreneur retains access to financial markets and can have a fresh start in the next period. Another important issue is that with limited enforcement durable capital may become more valuable since it is more easily sized in case of bankruptcy, see Rampini and Viswanathan [27].

[3]In Clementi and Hopenhayn [6] the sequence of productivity shocks is i.i.d.. Fu and Krishna [18] generalize the model allowing for Markovian shocks. DeMarzo et al. [12] also consider the case of persistent shocks, although their model is technically different.

This can be justified by the fact that in the optimal contract higher equity values are on average (not always) associated to higher investment. However, as previously pointed out, the model cannot make predictions on the effects of size which are distinct from predictions related to the financial structure of the firm.

Size as a State Variable. Size can have an impact on the optimal policy only with adjustment costs. If capital can be bought and sold at the same price and there are no other adjustment costs, as in Quadrini [25], then in each period the firm can choose the optimal amount of capital independently of the current level of capital.

We enrich the Clementi–Hopenhayn model by introducing durable capital, so that size becomes a relevant variable in the firm's decision problem. In our model capital depreciates at rate $d \in [0, 1]$ (the Clementi–Hopenhayn model corresponds to the special case $d = 1$). In each period the existing capital can be increased by new investment, or reduced by selling any part of it at a price that is lower than the investment's cost. Size is defined as the amount of capital existing at the beginning of the period, determined by past investment decisions and depreciation. The fact that used capital can be sold only at a discount is what makes size a state variable in our model, not just the fact that capital is durable. It is worth emphasizing that in Clementi-Hopenhayn the optimal policy would still depend on equity only if depreciation were partial *but* the price of used capital were the same as the price of new capital.

The assumption of decreasing returns to scale is also important in generating our main results. Recent work in dynamic financial contracting (De Marzo and Fishman [10], DeMarzo and Fishman [11] and DeMarzo et al. [12]) has analyzed the dynamic contracting problem under the assumption of constant returns to scale and convex adjustment costs[4]. These models generate interesting and important predictions on the optimal investment policy but are ill-suited to evaluate the effect of size, as the optimal policy depends on the ratio between firm's value and size (i.e. the value per unit of capital) rather than on the two variables separately. In the words of DeMarzo et al. [12]:

> "Because our model is based on a constant returns to scale investment technology ... it is not well suited to address questions relating firm size and growth. Indeed, if we control for past performance or financial slack, size does not matter in our framework."

[4] Ai, Kiku and Li [1] also have a constant returns model and durable capital but they assume limited liability for all parties and look at general equilibrium effects.

Our model has decreasing returns to scale and the optimal policy is based both on the value of the firm and its size – not on their ratio or any other single-valued function of the two. This is the minimal requisite for making predictions that identify the impact of size on a firm's decisions.

Evidence on Used Capital Discount. The assumption that used capital goods sell at a discount has ample empirical support. Ramey and Shapiro [26], in their study of the aerospace industry, state that "even after age-related depreciation is taken into account, capital sells for a substantial discount relative to replacement cost". The relevance of the market for used capital has been analyzed in many recent papers, see e.g. Eisfeldt and Rampini [15], Gavazza [19] and Lanteri [21]. One feature that is absent in our model is the dependence of the used capital price on the state of the economy. Our model could easily be generalized to make the price of used capital stochastic and dependent on the state of the economy. We could also allow the state of the firm to be correlated with the state of the economy. These complications however do not produce interesting insights about the optimal policy of the single firm. They should be taken into account when looking at the aggregate consequences of investment and divestment behavior, an issue we do not discuss.

3 The Model

Time is discrete and the horizon is infinite. At time 0 an entrepreneur has an idea for a project which requires funding to acquire an enabling asset (e.g. a license or land) at cost A and then capital at subsequent times. The entrepreneur has insufficient funds and thus needs financing from a lender. Both the lender and the entrepreneur are risk neutral and have a common discount factor $\delta \in (0, 1)$. To rule out trivial cases, A is lower than the value generated under the optimal financing contract that we characterize in Section 4.

Once the enabling asset is acquired, the firm buys capital and uses it to generate cash flow at any time $t > 0$. At the beginning of the period the project can be continued or liquidated. If liquidated, the existing assets are sold at their 'scrap' value, to be specified shortly. If the project continues, the firm makes a capital adjustment decision, i.e. either sells part of its existing capital or buys new capital. The capital stock K_t then generates a cash flow of $\theta_t \cdot R(K_t)$, where θ_t is the realization of a random variable with support $\{0, 1\}$ and R is a production function. We make the following assumptions about the production process.

Assumption 1 *The random variables* $\{\widetilde{\theta}_t\}_{t=1}^{+\infty}$ *are i.i.d., with* $p \equiv \Pr(\widetilde{\theta}_t = 1) \in (0,1)$. *The function* $R(\cdot)$ *is defined on* $[0, +\infty)$, *bounded, continuously differentiable, strictly increasing, strictly concave and such that* $R(0) = 0$ *and* $p \cdot R'(0) > 1$.

Capital depreciates at rate $d \in (0,1)$. Denoting I_t the change in capital through investment or divestment (sale), at time t, the law of motion is

$$K_0 = 0 \quad \text{and} \quad K_t = (1-d)K_{t-1} + I_t, \qquad t = 1, 2...$$

The unit price of new capital is 1, while the selling price of 'used' capital is $q \in (0,1)$. Thus the cost of adjusting capital is given by the function

$$I(K_t, K_{t-1}) = \begin{cases} K_t - (1-d)K_{t-1}, & \text{if} \quad K_t \geq (1-d)\, K_{t-1} \\ \\ q(K_t - (1-d)K_{t-1}), & \text{if} \quad K_t < (1-d)K_{t-1}. \end{cases}$$

We make the following assumption on the liquidation ('scrap') value $S(K_{t-1})$.

Assumption 2 *The liquidation value at time* t *is* $S(K_{t-1}) = S + q(1-d)K_{t-1}$, *with* $0 < S < A$.

The variable S denotes the liquidation value of the enabling asset. If K_{t-1} is the amount of capital available at the beginning of period $t-1$, then the amount of capital left at the end of period $t-1$ is $(1-d)K_{t-1}$. Selling that amount of capital at the beginning of period t at a unit price q generates revenue $q(1-d)K_{t-1}$. The inequality $S < A$ implies that it cannot be optimal to buy the enabling asset only to liquidate the firm.

As in Clementi and Hopenhayn [6], we assume that the entrepreneur is protected by limited liability, which implies that the monetary transfer from the firm to the lender in period t cannot exceed the cash flow $\theta_t R(K_t)$. The cash flow is privately observed by the entrepreneur and not verifiable, so the contract has to provide incentives to report correctly the realization of $\widetilde{\theta}_t$.

As a benchmark, consider the case of verifiable cash flow (no agency problem). Let $W^{sym}(K)$ denote the present value of the project under symmetric information, when the capital stock at the beginning of the previous period was K. Since we assume that A is lower than the value generated under the second best policy, we have $A < W^{sym}(0)$. By Assumption 2 we have $S < A$, so $S < W^{sym}(0)$. The inequality in turn implies

$$S + q(1-d)K < W^{sym}(0) + q(1-d)K. \quad (1)$$

Since the firm can always sell all the existing capital $(1-d)K$ at unit price q and then follow the optimal investment policy starting at $K = 0$, we also have

$$W^{sym}(0) + q(1-d)K \leq W^{sym}(K) \quad (2)$$

Combining 1 and 2 yields

$$S + q(1-d)K < W^{sym}(K) \quad (3)$$

so *continuation is always optimal*. Therefore the first-best problem can be stated as

$$\max_{\{K_t\}_{t=1}^{\infty}} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^t \left(\widetilde{\theta}_t R(K_t) - I(K_t, K_{t-1})\right)\right] \quad (4)$$

with initial condition $K_0 = 0$. Proposition 1 establishes that the optimal policy is characterized by the amount of capital K^* determined by the equation

$$p\, R'\left(K^*\right) = 1 - \delta\left(1-d\right). \quad (5)$$

Proposition 1 *With no asymmetric information, it is optimal to buy K^* in period 1 and dK^* in every subsequent period.*

We omit the proof, which is standard. The left-hand side of equation (5) is the marginal benefit of increasing capital, i.e. the expected marginal product of capital. The right-hand side is the net marginal cost of new capital – any unit bought in period t costs one, but reduces by $(1-d)$ the amount of capital that will be purchased in period $t+1$. Along the optimal path, firm's size remains constant at K^*. Without asymmetric information *higher current size does not imply higher future size.* If the firm reaches any level $\hat{K} > K^*$, something that can only happen when it deviates momentarily from the optimal policy, capital will "revert to the mean" K^*, at a speed that depends on the resale price q.

4 Optimal Contracts under Asymmetric Information

In this section we study the optimal contract under asymmetric information. We first describe the set of feasible contracts and then discuss optimality.

4.1 Feasible Contracts

A financing contract specifies a monetary transfer from the lender to the borrower at $t = 0$ and, for each subsequent period, a probability of liquidation, an investment level, and payments between the two parties as functions of history. At $t = 0$ the only possible activity is the acquisition of the enabling asset at price A. At $t \geq 1$, if the firm is still active, the sequence of events is as follows. First, the firm is liquidated with probability $\lambda_t \in [0, 1]$. If the firm is liquidated, the borrower is paid Q_t, the lender keeps the remaining part of the scrap value $S(K_{t-1}) - Q_t$ and the relationship ends. If the firm remains active, an investment decision is made. Let K_t denote the amount of capital available at period t. The borrower privately observes the outcome $\theta_t\, R(K_t)$, reports[5] $\widehat{\theta}_t \in \{0, 1\}$ to the lender and pays an amount τ_t which may depend on the current message, the level of capital K_t chosen and the history. At time t the outcome is $a_t = \left\{K_t, \widehat{\theta}_t\right\}$ where $\widehat{\theta}_t$ is the message issued by the borrower. A *history up to* t is a collection $h_t = \{a_s\}_{s=0}^{t}$ and H_t is the set of all possible h_t. A *financing scheme* specifies a vector $(\lambda_t(h_{t-1}),\, Q_t(h_{t-1}),\, \kappa_t(h_{t-1}),\, \tau_t(h_t))$ for each possible history, where κ_t denotes a probability distribution[6] over $\mathbb{R}_+$. The scheme is *feasible* if for each h_t we have $\lambda_{t+1}(h_t) \in [0, 1]$, $\operatorname{supp} \kappa_{t+1}(h_t) \subset [0, +\infty)$, $Q_{t+1}(h_t) \geq 0$ and $\tau_t(h_t) \leq \widehat{\theta}_t R(K_t)$, the last two inequalities being the consequence of borrower's limited liability. The contract is individually rational if the present expected value of payments is larger than the amount invested at time 0 by both parties, where the expected value is computed under the assumption that the borrower reports the true value θ_t at each period. Let $\widehat{r}_t : H_t \to \Theta$ be a reporting strategy at time t for the borrower and $\widehat{\mathbf{r}} = \{\widehat{r}_t\}_{t=0}^{+\infty}$ be a reporting strategy for all periods. Denote by $\mathbf{r}$ the truth-telling strategy, that is $r_t\,(h_{t-1}, (K_t, \theta_t)) = \theta_t$ for each $(h_{t-1}, (K_t, \theta_t))$. Let $V_t^{\widehat{\mathbf{r}}}(h_{t-1})$ be the present value of the expected payment to the entrepreneur from time t on, given history h_t and reporting strategy $\widehat{\mathbf{r}}$. A contract is *incentive compatible* if $V_t^{\mathbf{r}}(h_t) \geq V_t^{\widehat{\mathbf{r}}}(h_t)$ for each history h_t and reporting strategy $\widehat{\mathbf{r}}$. Since we look at contracts inducing truth-telling, it will be convenient to simplify notation by setting $V_t(h_t) \equiv V_t^{\mathbf{r}}(h_t)$.

[5]By the revelation principle, allowing for other sets of feasible messages is inconsequential.

[6]Using a probability distribution over K_t is useful to establish some concavity properties of the value function. The matter is discussed in the proof of Proposition 2.

4.2 Efficient Contracts

To analyze the incentive–efficient contract, we adopt a recursive formulation[7] similar to Clementi and Hopenhayn [6]. In our model, besides the entrepreneur's promised utility, the current level of capital must also be treated as a state variable.

We introduce the following functions. At the beginning of period t, $V(h_{t-1})$ denotes the expected continuation payoff for the entrepreneur. If the firm is not liquidated, then a level of capital K_t is decided as realization of the random variable $\kappa_t(h_{t-1})$. Then the entrepreneur observes $\theta_t R(K_t)$ and reports $\widehat{\theta}_t$. The expected continuation payoff for the entrepreneur after history h_{t-1}, capital choice K_t, shock realization θ_t and report $\widehat{\theta}_t$ is given by

$$\widehat{V}(\theta_t, \widehat{\theta}_t, K_t, h_{t-1}) = \theta_t\, R(K_t) - \tau(\widehat{h}_t) + \delta\, V(\widehat{h}_t), \tag{6}$$

where $\widehat{h}_t = (h_{t-1}, (K_t, \widehat{\theta}_t))$. The functions V and $\widehat{V}$ are linked by the 'promise keeping' constraints

$$V(h_{t-1}) = \lambda_t\, Q_t + (1-\lambda_t)\, \mathbb{E}_{\kappa_t} \left[p\, \widehat{V}(1,1,K_t,h_{t-1}) + (1-p)\, \widehat{V}(0,0,K_t,h_{t-1}) \right] \quad \forall h_{t-1} \in H_{t-1} \tag{7}$$

where λ_t, Q_t and κ_t are functions of h_{t-1}. Incentive compatibility requires

$$\widehat{V}(\theta_t, \theta_t, K_t, h_{t-1}) \geq \widehat{V}(\theta_t, \widehat{\theta}_t, K_t, h_{t-1}) \quad \forall K_t \in \operatorname{supp} \kappa, \quad \forall \theta_t, \widehat{\theta}_t \tag{8}$$

and limited liability implies

$$\tau(h_{t-1}, (K_t, 0)) \leq 0 \tag{9}$$

To satisfy the participation constraint of the lender there must be histories such that the entrepreneur pays a strictly positive amount when $\theta_t = 1$. In order to convince the entrepreneur to truthfully report $\theta_t = 1$ the incentive compatibility constraint

$$\widehat{V}\,(1,1,K_t,h_{t-1}) \geq R(K_t) + \delta V\,(h_{t-1}, (K_t, 0)) \tag{10}$$

has to be satisfied[8]. Since $\widehat{V}\,(1,1,K_t,h_{t-1}) = R(K_t) - \tau\,(h_{t-1},(K_t,1)) + \delta V\,(h_{t-1},(K_t,1))$, the inequality in (10) is equivalent to

$$\tau\,(h_{t-1},(K_t,1)) \leq \delta \left[V\,(h_{t-1},(K_t,1)) - V\,(h_{t-1},(K_t,0)) \right]. \tag{11}$$

We will use this form of the incentive compatibility constraint in our analysis of the efficient contract.

[7]The idea of using an agent's continuation utility was first introduced by Spear and Srivastava [28]. Atkeson and Lucas [3] provide a justification of the recursive approach in dynamic moral hazard problems.

[8]Notice that we have assumed $\tau\,(h_{t-1},(K_t,0)) = 0$, since negative values (equivalent to giving extra cash to the entrepreneur when $\theta = 0$ is announced) can never be optimal as they only worsen the incentive problem.

4.3 The Value of Equity and the Value of the Firm

Any non–negative value V can be attained liquidating the firm and giving $Q = V$ to the borrower. Negative values cannot be implemented due to the limited liability of the borrower. Thus, the set of possible values for V is $[0, +\infty)$. Without loss of generality we assume that policy depends on history h_{t-1} only via the promised equity V and capital K. Introducing additional variations cannot increase the value of the firm.

At each state (V, K), let $W(V, K)$ be the value of the firm. The problem of finding the value function $W(V, K)$ can be decomposed in two parts. First, we can compute the value of the firm when continuation is imposed. Second, once that value has been obtained, we can use the continuation value and the liquidation value to compute the optimal liquidation policy.

Let $W_c(V_c, K)$ be the highest value of the firm that can be achieved when continuation is imposed. To simplify notation, let $V^H(K')$ ($V^L(K')$) be the value of promised equity when capital K' is chosen and the report is $\widehat{\theta}_t = 1$ ($\widehat{\theta}_t = 0$). Therefore $W_c(V_c, K)$ is obtained solving the problem

$$W_c(V_c, K) = \max_{\kappa, \tau, V^H, V^L} \mathbb{E}_\kappa \left[pR(K') - I(K', K) + \delta p W\left(V^H(K'), K'\right) + (1-p) W\left(V^L(K'), K'\right) \right] \quad (12)$$

subject to

$$V_c = E_\kappa \left[p\left(R(K') - \tau(K')\right) + \delta\left(pV^H(K') + (1-p)V^L(K')\right) \right] \quad (13)$$

$$\tau(K') \leq \delta\left(V^H(K') - V^L(K')\right), \qquad \tau(K') \leq R(K') \quad (14)$$

$$V^H(K') \geq 0, \quad V^L(K') \geq 0 \text{ each } K' \in \text{supp}\kappa \quad (15)$$

Notice that $W_c(V_c, K)$ is computed taking as given the function $W(V, K)$. Once we have the function W_c, we can rewrite the maximization problem as:

$$W(V, K) = \max_{\lambda, Q, V_c} \quad \lambda S(K) + (1-\lambda) W_c(V_c, K) \quad (16)$$

subject to $V = \lambda Q + (1-\lambda) V_c$, $Q \geq 0$, $1 \geq \lambda \geq 0$.

Standard results in dynamic programming imply that the solution to the functional equation (16) is unique (see Quadrini [25] for details). Inspecting problem (16) we can make a few simple observations. First, when the optimal policy requires $\lambda(V, K) = 0$ then $W(V, K) = W_c(V, K)$. Second, the only efficient way to give $V = 0$ is to liquidate immediately. The next proposition provides further properties.

Proposition 2 *$W(V,K)$ and $W_c(V,K)$ are non-decreasing in both arguments. For each K, $W(\cdot,K)$ and $W_c(\cdot,K)$ are concave and the partial derivatives with respect to V are defined almost everywhere. For each K there is a value $V_{(K)}$ such that the W is linear in V on $[0, V_{(K)}]$. The first best policy is achievable if and only if $V \geq \frac{pR(K^*)}{1-\delta}$, where K^* is the first-best level of capital defined in (5).*

The proposition shows that the results of Clementi and Hopenhayn [6] and Quadrini [25] continue to hold, with obvious modifications. For a given K the value of the firm is non-decreasing and concave in the value of equity. The linear part of the value function corresponds to the case in which liquidation occurs with positive probability: When $V < V_{(K)}$ liquidation probability is $\lambda = 1 - \frac{V}{V_{(K)}}$. When $V \geq V_{(K)}$ the firm continues with probability 1. The important difference is that the threshold values for which liquidation occurs now depend on K.

5 The Empirical Distribution of Size

In the rest of the paper we will analyze the predictions of our model on the effects of size. We will pay particular attention to comparing the predictions of the model with (a version of) Clementi-Hopenhayn, henceforth CH. Specifically, since the assumptions that make size a state variable in our model is that $q < 1$, we will compare our model to a version of CH in which capital is durable *but* $q = 1$, so that capital is not a state variable[9]. CH assumes $d = 1$, but their model would remain essentially the same with $d < 1$.

Theoretical results will be used whenever possible and simulations will be used when we cannot provide analytical results. The first task is to establish the theoretical implications for the type of data that we actually observe.

5.1 The Joint Distribution of V and K

When a firm is created, the state is $(V_0, 0)$, where V_0 is the value assigned to the entrepreneur when the process is started. The value depends on the amount of self-financing that the entrepreneur can bring and on the bargaining power of the two parties. We can see V_0 as the realization of a random variable $\widetilde{V_0}$ with distribution $f_0(V)$. Once V_0 is realized, the optimal policy induces a stochastic process

[9] These are actually the assumptions used in Quadrini [25]. He uses different assumptions on the distribution of $\tilde{\theta}$.

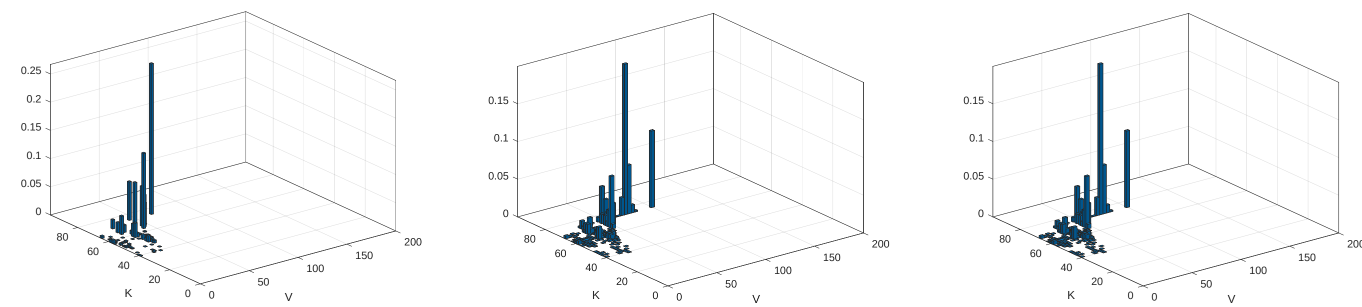
$\left\{\left(\widetilde{V}_t, \widetilde{K}_t\right)\right\}_{t=1}^{\infty}$ that has two absorbing states: $(0,0)$, reached when liquidation occurs, and (V^*, K^*), where $V^* \equiv \frac{pR(K^*)}{1-\delta}$ is the equity value at which efficient investment is made in each period and the agent never pays.[10] From the stochastic process one can compute the distribution at time t:

$$f_t(V,K) \equiv \Pr(\widetilde{V}_t = V, \widetilde{K}_t = K).$$

If many firms are born at time 0 then the distribution of their values after t periods should be close to the theoretical distribution $f_t(V,K)$. However, what we typically have in the data is the distribution of firms of age t *conditional on survival*, given by

$$h_t(V,K) = \frac{f_t(V,K)}{1 - f_t(0,0)} \quad \text{for each pair } (V,K) \neq (0,0).$$

We have solved the dynamic programming problem numerically, assuming $R(K) = \frac{1}{\alpha}K^{\alpha}$, $S(K) = a + bK$. Appendix II provides a detailed explanation of our methodology as well as all parameters' values. With our parametrization, we have $K^* = 81.5$ and $V^* = 722.3$. The values of V_0 were drawn from a uniform distribution on the interval $\left[\underline{V}, \overline{V}\right]$, where $\underline{V} = V^*/50 \simeq 14.5$ and $\overline{V} = V^*/40 \simeq 18$.



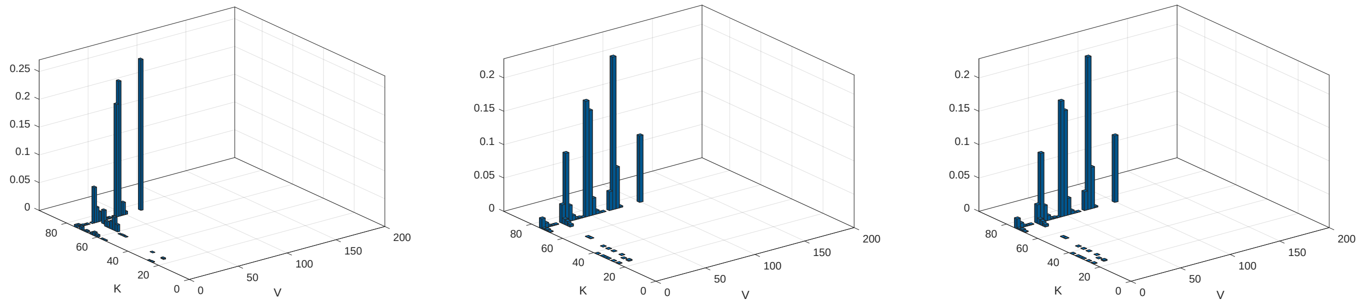
(a) $h_{10}(V,K)$ (b) $h_{15}(V,K)$ (c) $h_{25}(V,K)$

Figure 1: Plots of $h_t(V,K)$ at different ages when $q = 0.8$.

Figure 1 shows the joint distribution of V and K when $q = 0.8$, so that K is a state variable. The distributions show quite a lot of variability in both dimensions, although it is clear that over time the surviving firms tend to move towards bigger sizes.

In the version of CH with $q = 1$ the optimal choice of size only depends on V and K is not a state variable. When we run the same simulations with $q = 1$ the resulting joint distribution of V and K tend to show much less variability with respect to age. It is far more frequent and faster to achieve the optimal size and in general there is much less variation in K.

[10]Notice that points of the form $(0,K)$ with $K > 0$ have zero probability, since it is impossible to give $V = 0$ to the agent when K is strictly positive.



(a) $h_{10}(V, K)$ (b) $h_{15}(V, K)$ (c) $h_{25}(V, K)$

Figure 2: Plots of $h_t\,(V, K)$ at different ages when $q = 1$.

The results of the simulation are shown in Figure 2. Surviving firms reach the optimal capital size K^* for relatively low values of V, well below V^*. This implies that at K^* there is a long queue of V, since the optimal policy prescribes that V must increase for every positive realization of θ until V^* is reached.

Comparing Figure 1 and Figure 2 we can see how the presence of a friction, in the form of partial irreversibility, generates more variability for any given age. Although eventually the distribution, for any value of q, must converge to the absorbing points, the convergence is much slower with partial irreversibility. In particular this means that, conditional on size, we expect more variability in the policy function. The reason is that conditional on K there is more variability in V under partial irreversibility than under the case $q = 1\cdot$

Remark. One unrealistic feature of the model is that the limiting distribution of size for surviving firms is concentrated on K^*. Thus, all heterogeneity must come from younger entering firms. Li [22] discusses some modifications that would avoid this undesirable feature. If the borrower is less patient than the lender then payments to the lender would be limited and the growth of V would be slower; the firm would never reach the first best level and every firm would have a positive probability of (eventual) death. A similar effect applies if the lender has limited commitment. Such features could easily be introduced in our model. Another alternative is to introduce an exogenous death rate due to circumstances outside the model, as in Pugsley, Sedláček and Stern et al. [24].

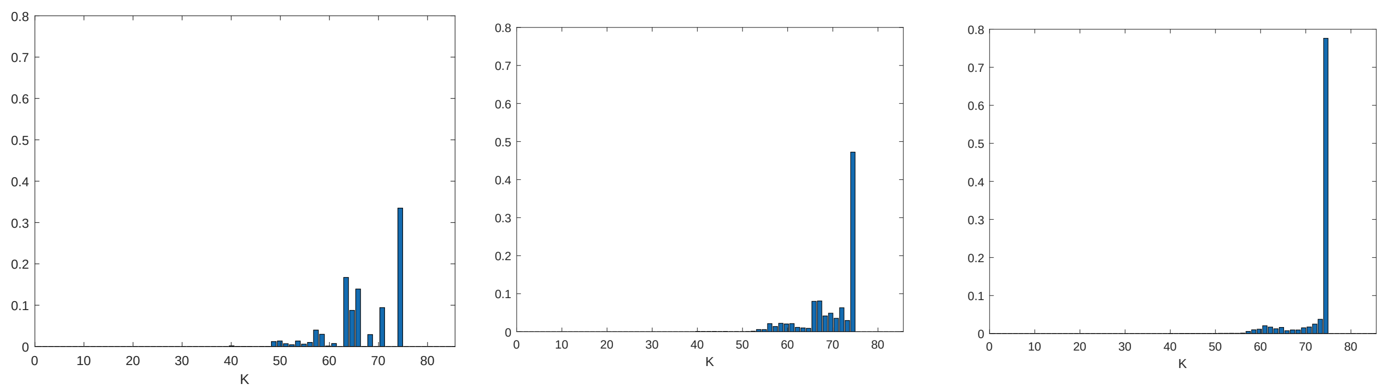
5.2 The Marginal Distribution of K

While empirical counterparts for K are relatively easy to identify, this is not so for V. In general V is the value given to the original entrepreneur and founder of the firm. We will later discuss a possible interpretation that follows Clementi and Hopenhayn [6], V as firm's equity and $W(V,K)-V$ as firm's debt. However, even adopting this interpretation, the market value of equity V may be difficult to observe, especially for small and medium-sized firms.

For this reason we will also look at distributions over K, obtained as the marginal of the corresponding distributions over (V,K). Let $\mathcal{V}$ be the set of all possible values of V that can be generated by the stochastic process $\left\{\widetilde{V}_t, \widetilde{K}_t\right\}_{t=0}^{+\infty}$. Then we can define

$$\gamma_t\left(K\right)=\sum_{V\in\mathcal{V}} h_t\left(V,K\right)$$

as the probability distributions over size of firms surviving after t periods. The marginal distributions obtained from Figure 1 are shown in Figure 3.

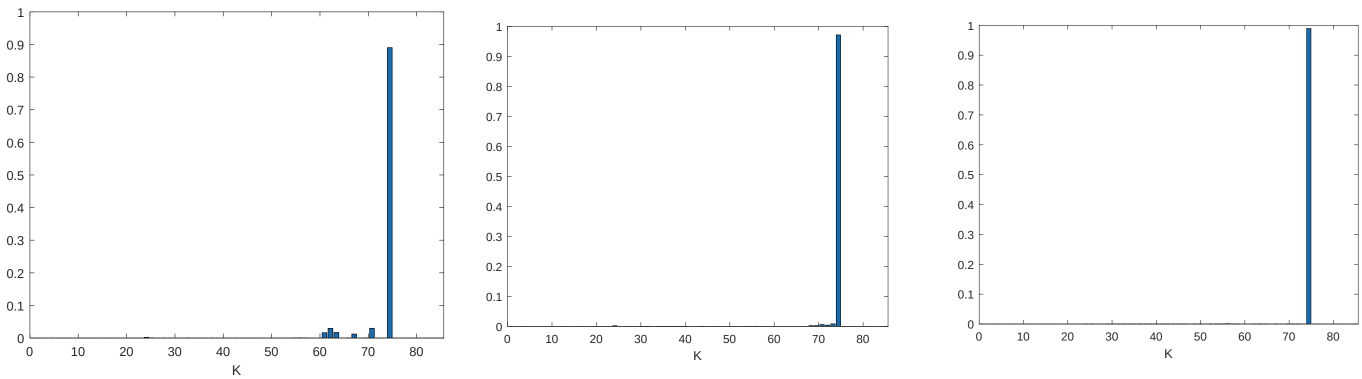


(a) $\gamma_{10}\left(K\right)$ (b) $\gamma_{15}\left(K\right)$ (c) $\gamma_{25}\left(K\right)$

Figure 3: Plots of $\gamma_t\left(K\right)$ at different ages when $q=0.8$.

As previously pointed out, partial irreversibility generates much more dispersion in size compared to the case $q=1$. This can be confirmed looking at Figure 3 and comparing it to Figure 4, obtained as the marginal over K of Figure 2.

Conditional on survival the distribution of size tends to concentrate more on K^*, the optimal size, as age increases. However the convergence is much slower with partial irreversibility. The message is that partial irreversibility plays an important role in making age an important variable when analyzing the evolution of firms. On one hand the slower convergence generates more dispersion in size for younger firms. On the other hand, under partial irreversibility firms tend to be more heterogeneous *even conditional on* K. We can therefore expect greater volatility in policy for younger firms than for older firms when partial irreversibility is introduced.



(a) $\gamma_{10}(K)$ (b) $\gamma_{15}(K)$ (c) $\gamma_{25}(K)$

Figure 4: Plots of $\gamma_t(K)$ at different ages when $q = 1$.

6 The Effect of Size on Survival and Growth

In this section we focus on the effect of size on liquidation and investment.

6.1 Size and Survival

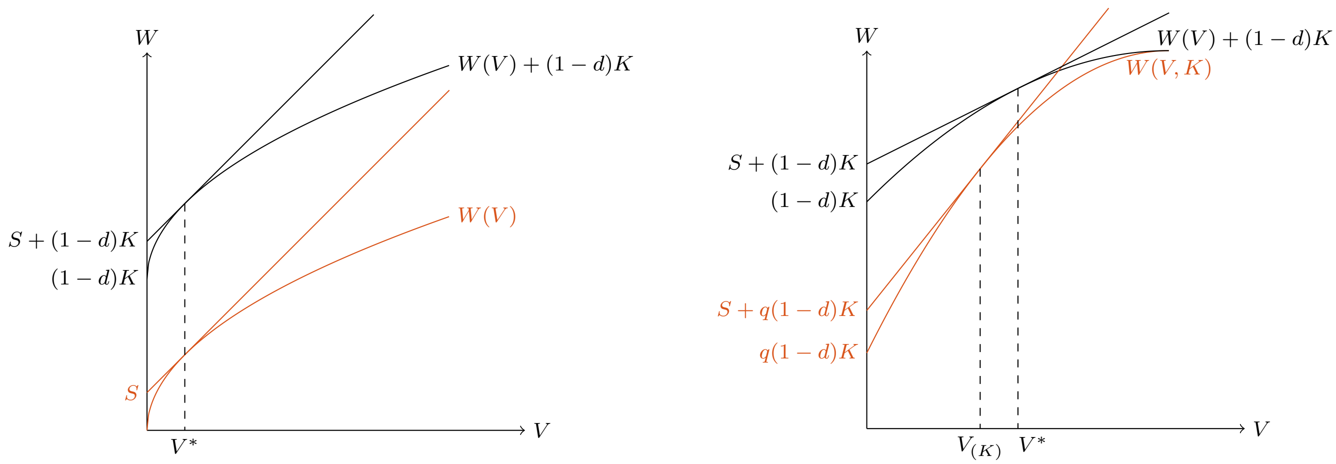
In our model the probability of liquidation is $\lambda(V, K) \equiv \max\left\{1 - \frac{V}{V_{(K)}}, 0\right\}$, where $V_{(K)}$ is the threshold value below which liquidation occurs with positive probability. The probability of survival $1 - \lambda(V, K)$ is therefore

$$\beta(V, K) \equiv \min\left\{\frac{V}{V_{(K)}}, 1\right\}. \tag{17}$$

From (17) we can immediately conclude that, for each given K, the probability of survival increases in V. The next result establishes that, conditional on V, the probability of survival is increasing in K.

Proposition 3 *The probability of survival $\beta(V, K)$ is non-decreasing in K for each V and it is strictly increasing whenever $V < V_{(K)}$.*

The basic intuition for the result in Proposition 3 is that an increase of capital increases the value of continuation more than the value of liquidation. This is easier to see when the optimal policy prescribes strictly positive investment with probability 1 under continuation. In that case a small additional amount of capital I increases the value of continuation by $(1 - d)I$ (since it reduces the necessary investment by that amount), while the value of liquidation increases only by $q(1 - d)I$ (the value at which capital can be sold). This happens only because of partial irreversibility.



(a) Value function when $q = 1$

(b) Value function when $q = 0.8$

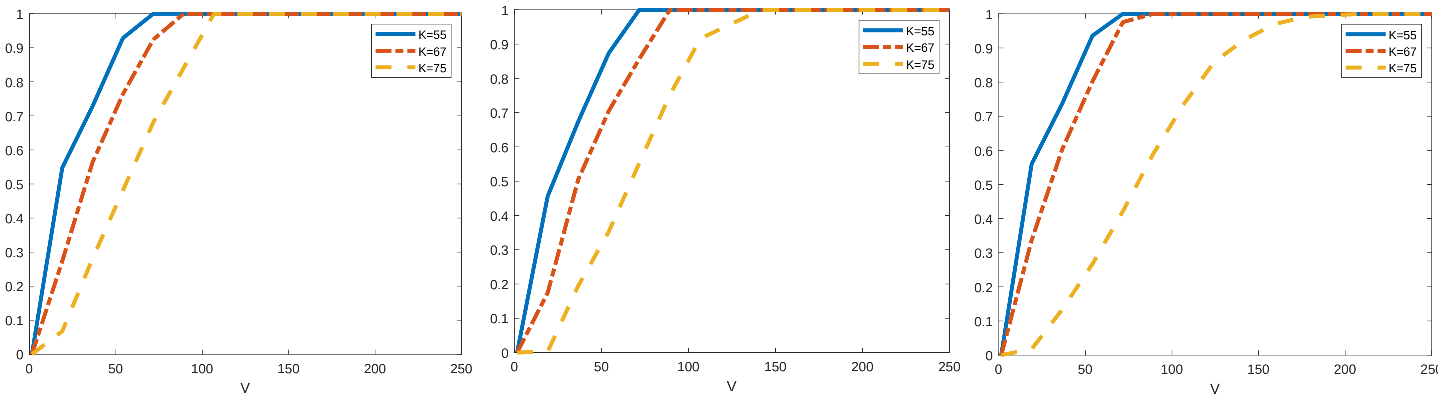
Figure 5: Effect of a change in q on the optimal liquidation policy.

When $q = 1$ the liquidation value can be written as $S(K) = S + (1-d)K$. Since both the continuation value and the liquidation value are translated upward by $(1-d)K$, the value V^* below which liquidation occurs with strictly positive probability is the same for each level K. Thus, size does not change the probability of liquidation, a crucial difference with the partial irreversibility case.

Proposition 3 provides a positive relation between size and probability of survival *conditional on* V. In empirical work it is difficult to control for V and the main existing results look at the unconditional relation. In other words, what we actually observe in empirical work is not the survival function $\beta(V,K)$ but the function

$$\beta^*(K) = E_V\left[\beta(V,K)|\, K\right] = \sum_{V\in\mathcal{V}} h\left(V|\, K\right)\beta(V,K),$$

where $h\left(V|\, K\right) = h\left(V,K\right)/\gamma\left(K\right)$. Even if $\beta(V,K)$ is non-decreasing in K for each V, it may still be the case that $\beta^*(K)$ is decreasing. However, since $\beta(V,K)$ is increasing both in V and K, the function $\beta^*(K)$ is increasing in K if the probability distribution $h\left(V|\, K\right)$ is increasing in K in the sense of first-order stochastic dominance. While we cannot provide a full theoretical characterization of the joint probability distribution, our simulations support the idea that $h\left(V|\, K\right)$ is FOSD-increasing in K. Let $H_t\left(V|\, K\right)$ be the cumulative distribution function of $h_t\left(V|\, K\right)$. Figure 6 shows how H_t changes when K increases, for $t = 10$, $t = 15$ and $t = 25$.



(a) $H_{10}(V|\,K)$ (b) $H_{15}\,(V|\,K)$ (c) $H_{25}\,(V|\,K)$

Figure 6: Cumulative distribution of $H_t\,(V|\,K)$ at increasing values of K, $t = 10, 15, 25$.

The pictures clearly show that $H_t\,(V|\,K)$ is FOSD-increasing in K for each date. The simulations results thus lend support to the hypothesis that $\beta^*(K)$ is increasing in K.

6.2 Size and Growth

The expected rate of growth of a firm at state (V, K) is $g(V, K) = \frac{\mathbb{E}\left[\widetilde{K}'_{(V,K)}\right]-K}{K}$, where expectation is taken using the optimally chosen probability distribution over K'. If V is not observed then the rate of growth conditional on size K is given by

$$g^*(K) = \mathbb{E}_V\left[g(V, K)|\, K\right] = \sum_{V\in\mathcal{V}} h\,(V|\,K)\,g\,(V, K)\,.$$

The properties of $g^*\,(K)$ depend both on the properties of $g(V, K)$ and on the probability distribution $h(V, K)$. We first look at the implications of the model for future size and then analyze the rate of growth.

6.2.1 Current Size and Future Size

Let $\widetilde{K}'_{(V,K)}$ be the random variable chosen as optimal capital policy at (V, K), with $\overline{K}'_{(V,K)}$ and $\underline{K}'_{(V,K)}$ being the supremum and the infimum of the support. An increase in current size (capital stock K at the beginning of period t) causes an increase in future size (capital stock K' chosen in period t).

Proposition 4 *The optimal capital choice has the following properties. For each* $I > 0$:

1. $\widetilde{K}'_{(V,K+I)}$ *(weakly) first order stochastically dominates* $\widetilde{K}'_{(V,K)}$.

2. *If* $\underline{K}'_{(V,K)} > (1-d)(K+I)$ *or* $\overline{K}'_{(V,K)} < (1-d)K$ *then* $\widetilde{K}'_{(V,K+I)} = \widetilde{K}'_{(V,K)}$

The intuition for the first point is simple. The cost of investment $I(K', K)$ is strictly decreasing in K. Thus, other things equal, it cannot be optimal to choose a higher future level of capital stock when the current capital stock is lower. The second point follows from the fact that increasing the level of capital when investment is strictly positive or strictly negative does not change the constraint set and it is equivalent to adding a constant to the objective function. The optimal solution must therefore remain the same. When $q = 1$ the optimal capital choice K' depends on V only, not on the current size K. It is only with partial irreversibility that the optimal choice can become sensitive to current size.

However, when $q = 1$ and V is not observed then the conditional expectation $\mathbb{E}[K'(V)|K]$ and the conditional variance $\text{Var}[K'(V)|K]$ will depend on K, as long as V is not independent of K. Thus, comparing the expected capital policy K' in the case $q = 1$ and in the case $q < 1$ is not trivial. Furthermore, since (as shown in section 5) the joint distribution of (V, K) depends on age, this will generate different prediction on the impact of size for firms of different age. We will discuss the matter further in section 7.

6.2.2 Current Size and Growth Rate

Long-run growth opportunities for large firms are absent in our model, once optimal size K^* is reached the rate of growth is zero. This remains true for any level of partial irreversibility q. However, before reaching the optimal size the growth does depend on the level of irreversibility.

When $q = 1$ the optimal policy K' only depends on V and it is constant in the current level of capital K. Thus, the rate of growth *as function of* V *and* K is trivially decreasing in K, as it is given by $\frac{K'(V)-K}{K}$. However, as previously pointed out, $\mathbb{E}[K'(V)|K]$ typically does depend on K and it is not obvious how.

When $q < 1$ then expanding capital becomes more costly. Typically, the cost of expanding capital will be asymmetric. If the optimal policy always prescribed a strictly positive investment under continuation then the only difference between the cases $q = 1$ and $q < 1$ would come from the possibility of liquidation. That would make capital investment less profitable, implying that the firm would expand more slowly. However the optimal policy does sometimes prescribe negative investment under continuation. This is an additional reason why capital investment will be lower when $q < 1$. This will be particularly true when V is low. However a higher level of K in general requires less investment, since the optimal new level of capital K' is closer to existing capital.

The volatility of the rate of growth is another variable of interest. When V is not observable then the variance of the rate of growth conditional on K is given by

$$\sigma_g^2(K) = \sum_V h(V|K)\left(g(V,K) - g(K)\right)^2.$$

The only clear prediction of the model is that firms which have reached a sufficiently high level of V choose the optimal size K^* and therefore have a zero rate of growth, which in turn implies zero volatility. This must mean that, as size increases, the volatility of growth must eventually go down. In our model the main source of volatility, conditional on size, come from the fact that the investment policy depends on V. A greater dispersion of V for a given K therefore generates higher variance in the growth rate. In the next section we discuss the implications for the impact of age.

7 The Role of Age

In our framework the age of the firm matters because the joint distribution of $h_t(V,K)$ of firms of age t evolves over time. As previously pointed out, V is typically not observed. What we observe are conditional expected values of the form $\mathbb{E}_t[g(V,K)|K]$ or $\mathbb{E}_t[\beta(V,K)|K]$, where $\mathbb{E}_t$ means that expectations are taken using the probability distribution $h_t(V|K)$.

Let

$$g_t(K) = \mathbb{E}_t[g(V,K)|K] = \sum_V h_t(V|K)\, g(V,K) \tag{18}$$

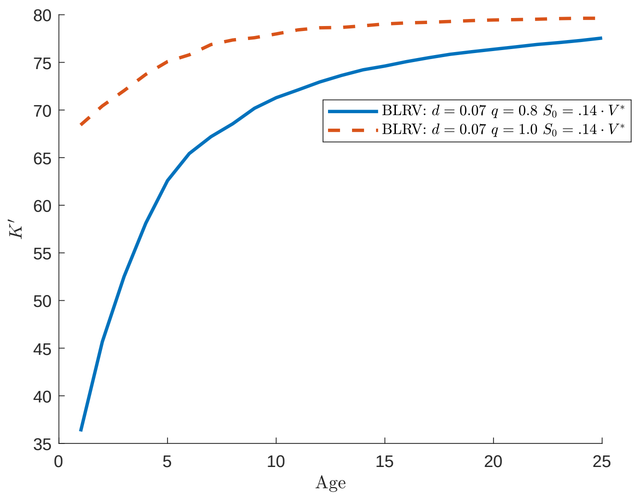
be the expected rate of growth for a firm of age t and size K. Also, let $z_t(K)$ be the fraction of firms with size K having age t and notice that $\sum_{t=0}^{\infty} z_t(K) = 1$ and $\sum_{t=0}^{\infty} \frac{\partial z_t(K)}{\partial K} = 0$. Then the expected rate of growth for a firm of size K *not conditional on age* can be written as

$$g^*(K) = E[g(V,K)|K] = \sum_{t=0}^{\infty} z_t(K)\, g_t(K)$$

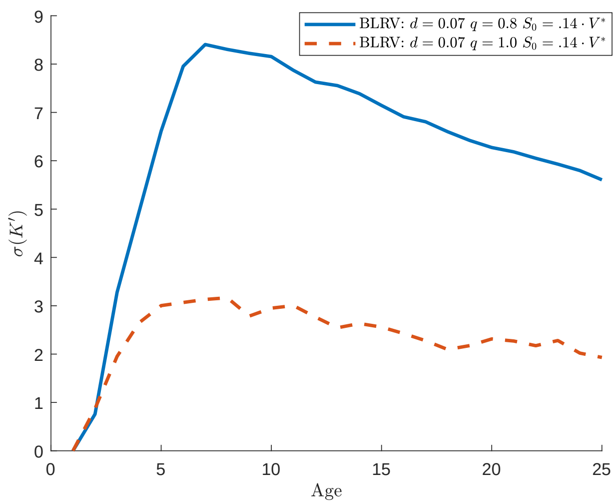
This implies

$$\frac{dg^*}{dK} = \sum_{t=0}^{\infty} \frac{dz_t}{dK} g_t + \sum_{t=0}^{\infty} z_t \frac{dg_t}{dK} \tag{19}$$

If, conditional on age, the rate of growth does not depend on K then we have $\frac{\partial g_t(K)}{\partial K} = 0$ for each t and for each K. However, we can still have $\frac{\partial g^*}{\partial K} < 0$ because of the first term on the RHS of (19). This will be the case if $g_t(K)$ is decreasing in t and $z_t(K)$ is increasing in K in the sense of first order stochastic dominance, i.e. when K is higher more weight is put on higher values of t. This second feature comes out pretty clearly from the simulations and it is to be expected, since surviving firms tend to grow to the optimal size.



(a) Expected K'



(b) Variance of K'

Figure 7: Expected K' and variance as a function of age.

In Figure 7 we show the age evolution of expected size and variance of size with and without partial irreversibility. Panel 7a shows that in the absence of partial irreversibility ($q = 1$, red line) size tends to be on average greater than in the case of partial irreversibility ($q = 0.8$, blue line) for younger firms. The effect holds only for younger firms, since older firms tend to converge to the optimal size with and without partial irreversibility. Once the optimal size is reached, typically the risk of having to sell capital becomes much lower and the price of used capital becomes irrelevant. For younger firms instead the effect is strong. Since younger firms tend to be smaller and eventually size must converge to the optimal one for surviving firms, expected growth is stronger for young firms when $q = 0.8$. This means that the presence of partial irreversibility makes the relation between age and growth much stronger. The relation is instead almost absent when $q = 1$. Panel 7b shows the volatility of size with respect to age with (blue line) and without (red line) partial irreversibility. It shows that partial irreversibility greatly increases the variance of K' in general, but especially so for young firms. Thus, our model yields a strong relation between age and volatility of size which is absent when $q = 1$.

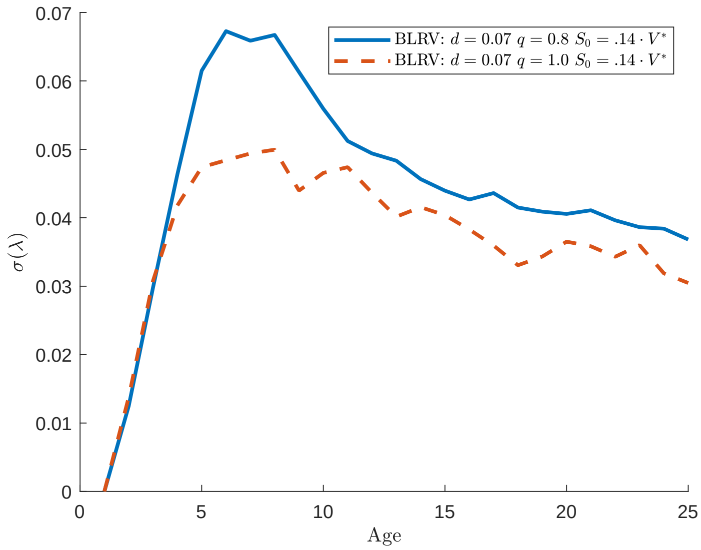


Figure 8: Standard deviation of liquidation probability for $q = 0.8$ and for $q = 1$

This higher volatility for young firms can also be observed when we look at the liquidation policy. In Figure 8 we show the standard deviation of liquidation probability as a function of age. Partial irreversibility (blue line) generates greater variability in liquidation rates for relatively young firms. As age progresses the standard deviations of liquidation in the two cases, $q = 0.8$ and $q = 1$, tend to converge. This is a consequence of the fact that with partial irreversibility there is much more variability in the joint distribution of V and K, as shown in Figures 1 and 2.

As a last comparison on the effect of age we have used the data generated in the simulations to run a regression on age and capital when $q = 0.8$ and when $q = 1$. Since at age 1 all firms start with a capital of zero, the relationship between age and capital and the rate of growth are computed for firms of age between 2 and 25.

When $q = 1$ the results of the regression are

$$\text{Growth rate} = 3.915615 + 0.047756 \cdot \ln(\text{Age}) \; - 0.923879 \cdot \ln(\text{Capital})$$

with adjusted R-Squared 0.2884. Thus age has the opposite sign than expected, as it appears that older firms tend to grow more. The positive sign on age remains if we run the regression in levels rather than logs and if we run the regression on age alone, both in level and as log. Of course, the positive coefficient for age does not mean that the CH model with $q = 1$ implies a positive conditional correlation. To interpret the result, remember that with $q = 1$ the optimal investment policy K' is entirely explained by V, while K plays no role. The expected growth rate of a firm of age t conditional on K is given by (18). Notice in particular that age is irrelevant if $h_t(V|K)$ does not depend on t. So, if V is (very close to) a one-to-one relationship with K, then age should have a coefficient of zero. The positive coefficient most likely is the result of the crude way in which the equation tries to capture the optimal policy function, whose functional form is not known.

When $q = 0.8$ the results of the regression are

$$\text{Growth rate} = 0.4329954 \;-\; 0.0388068 \cdot \ln(\text{Age}) \;-\; 0.0629460 \cdot \ln(\text{Capital})$$

with adjusted R-Squared 0.18. Now age has a negative coefficient, as expected and as it appears in the data. Also, the coefficient on capital is much lower than in the case $q = 1$. Again, the sign of the age coefficient remains negative when we run the regression in levels rather than logs and when we run the regression on age alone.

Summing up, the results of this section show that a model with partial irreversibility has a better potential for matching the data than a model in which capital can be bought and sold at the same price and therefore it is not a state variable.

8 Conclusions

Size has always been recognized as an important determinant of firm's behavior. This paper has introduced durable capital in a model of optimal dynamic financing with moral hazard in order to develop testable predictions about the effect of firm's size on the optimal financing and investment policies. The key assumption making size relevant is that used capital is less valuable than new capital, even once depreciation has been accounted for; there is ample empirical evidence that this is in fact the case.

We show that larger firms have a higher probability of survival, are more likely to have a large size in the future and have lower investment rates. We also show that partial reversibility plays an important role in generating heterogeneity among firms of the same age, generating a relationship between age, rate of growth and variance of growth. The results are qualitatively consistent with the empirical literature.

Further research should extend the analysis in two directions. First, it would be useful to perform a quantitative exercise, considering a parametric version of the model with realistic values for the parameters and checking whether the dynamics predicted by the model are quantitatively consistent with what has been found in empirical studies. Second, it would be interesting to move the analysis to the industry level, analyzing the endogenous determination of liquidation values as well as the impact of entry and exit.

Appendix I

Proof of Proposition 2. We will use results from Stokey, Lucas and Prescott [29], so we first recast the problem using a similar notation to make the way in which their results are applied clearer. Define the vector of control variables as

$$x = (\lambda_x, Q_x, K_x, \tau_x, V_x(0), V_x(1)).$$

Given a choice $y = (\lambda_y, Q_y, K_y, \tau_y, V_y(0), V_y(1))$, the return function is defined as

$$F(x, y) = \lambda_y S(K_x) + (1 - \lambda_y)[pR(K_y) - I(K_y, K_x)].$$

Define Γ as the set of vectors $(\lambda_y, Q_y, K_y, \tau_y, V_y(0), V_y(1))$ that satisfy the following:

$$\lambda_y \in [0, 1], \qquad Q_y \geq 0, \qquad K_y \geq 0$$

$$\tau_y \leq \min\{\delta(V_y(1) - V_y(0)), R(K_y)\}$$

$$V_y(0) \geq 0, \qquad V_y(1) \geq 0.$$

and $\Delta\Gamma(V_x(\theta))$ the set of probability distributions over Γ that satisfy

$$V_x(\theta) = E_\gamma[\lambda_y Q_y + (1 - \lambda_y)[p(R(K_y) - \tau_y) + \delta(pV_y(1) + (1 - p)V_y(0))]],$$

where $\gamma \in \Delta\Gamma(V_x(\theta))$ denotes a probability distribution over Γ.

The value function can be written as

$$W(x, \theta) = \max_{\gamma \in \Delta\Gamma(V_x(\theta))} (1 - \lambda_x) E_\gamma[(F(x, y) + \delta E[W(y, \theta')])]$$

Standard results in dynamic programming imply that $W(x, \theta)$ exists and is unique. It is also clear that neither $\Delta\Gamma(V_x(\theta))$ nor F depend on Q_x, τ_x and $V_x(\theta')$ when $\theta' \neq \theta$. Thus we can define

$$W^*(\lambda_x, K_x, V_x(\theta)) \equiv W(x, \theta).$$

Finally, we also have

$$W^*(\lambda_x, K_x, V_x(\theta)) = (1 - \lambda_x) W^*(0, K_x, V_x(\theta)).$$

Thus, function $W(V,K)$ defined as

$$W(V,K) = W^{*}(0,K,V)$$

is the one that we have been discussing in the text.

The function is increasing in K, because the return function is increasing in K_x and the constraint set $\Delta\Gamma(V)$ does not depend on K_x. To see that the function is increasing in V, notice that the return only depends on λ and K, but not on Q or τ. When V is increased it remains possible to use the same policies for λ and K, achieving the higher V through decreases in τ or increases in Q. Thus, increasing V expands the set of payoff-relevant policies. A similar argument establishes that W_c is increasing in V_c.

To see that $W(V,K)$ is concave in V when K is fixed, suppose that there are two values V_1 and V_2 such that

$$\alpha W(V_1,K) + (1-\alpha) W(V_2,K) > W(\alpha V_1 + (1-\alpha) V_2, K) \tag{20}$$

for some $\alpha \in (0,1)$. For the given α, consider the value $V_\alpha = \alpha V_1 + (1-\alpha) V_2$. The value V_α can be promised to the entrepreneur by offering the policy implemented at V_1 with probability α and the policy implemented at V_2 with probability $(1-\alpha)$; notice that such policies are clearly feasible. The expected value of the firm in that case would be the left hand side of (20). This is greater than the right hand side, contradicting the claim that $W(\alpha V_1 + (1-\alpha) V_2, K)$ is the highest value of the firm that can be achieved while giving V_α to the entrepreneur. The same argument establishes the concavity of W_c.

The argument cannot be applied to establish the concavity with respect to K given V or the global concavity with respect to (V,K). If there is a triplet (α, K_1, K_2) such $K = \alpha K_1 + (1-\alpha) K_2$ and

$$\alpha W(V,K_1) + (1-\alpha) W(V,K_2) > W(V,K). \tag{21}$$

Then, unfortunately, we cannot conclude that the stochastic policy choosing the policy optimal at K_1 with probability α and the policy optimal at K_2 with probability $(1-\alpha)$ is better than the policy chosen at K. The reason is that the return function, more specifically $I(K',K)$, depends on K. Thus, the LHS of (21) is not the expected value obtained when the stochastic policy above described is implemented *but the level of capital is* K.

Since $W(V,K)$ and $W_c(V,K)$ are increasing and concave in V, the partial derivatives $\frac{\partial W}{\partial V}$ and $\frac{\partial W_c}{\partial V}$ are defined almost everywhere.

The proof that $W(V,K)$ is linear in V on an interval $\left[0, V_{(K)}\right]$ is the same as in CH [6], with the only change that now the upper bound of the region over which $W(\cdot,K)$ is linear depends on K.

Finally, if the first best is implemented then an investment K^* must occur in every period independently of the history of announcements. This implies that the entrepreneur can achieve a value $\frac{pR(K^*)}{1-\delta}$ simply by announcing $\theta = 0$ in every period and stealing the output. Thus, in order to implement the first best policy we need $V \geq \frac{pR(K^*)}{1-\delta}$. When this condition is satisfied the first best can be achieved by a policy of investing K^* in every period independently of past history, paying $V - \frac{pR(K^*)}{1-\delta}$ immediately to the entrepreneur, and giving the entire output $\theta_t R(K^*)$ to the entrepreneur in each period. ■

Proof of Proposition 3. We show that $V_{(K)}$ does not increase in K and it strictly decreases if the optimal policy prescribes strictly positive investment at $\left(V_{(K)}, K\right)$. For each value V the probability of liquidation does not increase in K and it strictly decreases if $\lambda(V, K) \in (0, 1)$ and $V_{(K)}$ strictly decreases in K.

The function $W_c(V, K)$ is almost everywhere differentiable and by the envelope theorem

$$\frac{\partial W_c(V, K)}{\partial K} = (1-d) \Pr\left(K' > (1-d)K\right) + q(1-d) \Pr\left(K' \leq (1-d)K\right).$$

The liquidation value $S(K)$ is linear in K and

$$\frac{\partial S(K)}{\partial K} = q(1-d).$$

Thus $\frac{\partial S}{\partial K} \leq \frac{\partial W_c}{\partial K}$ for each V. Since $V_{(K)}$ is the point at which the line with intercept $S(K)$ is tangent to $W_c(V, K)$, if the function W_c increases no less than the value $S(K)$ then the point $V_{(K)}$ cannot increase. In particular, if at $\left(V_{(K)}, K\right)$ we have $\Pr\left(K' > (1-d)K\right) > 0$, i.e. the probability of a strictly positive investment is strictly positive, then $\frac{\partial W_c\left(V_{(K)}, K\right)}{\partial K} > \frac{\partial S(K)}{\partial K}$ and the value of $V_{(K)}$ strictly decreases as K increases.

The probability of liquidation can change with K only at points (V, K) at which $\lambda(V, K) < 1$. At such points we have

$$\lambda(V, K) = 1 - \frac{V}{V_{(K)}}$$

and the conclusion therefore follows from the results on $V_{(K)}$. ■

Proof of Proposition 4. Consider problem (12). We want to apply Theorem 4 in Milgrom and Shannon [23] and we will do so by showing that the objective function is quasi-supermodular in the decision variables $\left(\kappa, \tau(\cdot), V^H(\cdot), V^L\right)$ and it satisfies increasing difference in $\left(\left(\kappa, \tau(\cdot), V^H(\cdot), V^L(\cdot)\right); K\right)$.

The space where $\left(\kappa, \tau(\cdot), V^H(\cdot), V^L(\cdot)\right)$ is defined is the Cartesian product of the space of probability distributions κ on $[0, +\infty)$, the space of functions $\tau(x)$ such that $\tau(x) \leq R(x)$ each x and the space of non-negative functions V^H and V^L. We define the ordering on these spaces as follows:

1. $\kappa \preceq \kappa'$ if κ' first order stochastically dominates κ.
2. $\tau \preceq \tau'$ if $\tau(x) \leq \tau'(x)$ each x, and similarly for V^H and V^L.
3. $\left(\kappa, \tau(\cdot), V^H(\cdot), V^L(\cdot)\right) \preceq \left(\kappa', \tau'(\cdot), V^{H\prime}(\cdot), V^{L\prime}(\cdot)\right)$ if each component of the first vector is lower than the corresponding component of the second vector.

Since the objective function does not depend on τ and it is increasing in both V^H and V^L (thus implying quasi-supermodularity) we only have to prove quasi-supermodularity with respect to κ. For convenience, we remind here the reader of some basic definitions needed to apply the Milgrom-Shannon theorem.

Given a partially ordered set X and two elements x,y in X, we define $x \wedge y$ as the largest element of X such that $x \wedge y \preceq x$ and $x \wedge y \preceq y$. Similarly, $x \vee y$ is the smallest element in X such that $x \preceq x \vee y$ and $y \preceq x \vee y$. The set X is a lattice if, given x, y in X, we have that $x \wedge y$ and $x \vee y$ are also in X. A function f defined on the lattice X is quasi-supermodular if, given two elements $x, y \in X$, whenever the inequality $f(x) \geq f(x \wedge y)$ is satisfied we also have $f(x \vee y) \geq f(y)$.

Let now consider the space of probability distribution on the positive real line endowed with the first-order stochastic dominance order. Consider two distributions κ and κ' represented by the cumulative distribution functions F and G respectively. Then $\kappa \vee \kappa'$ has cumulative distribution function $H(x) = \min\{F(x), G(x)\}$, while $\kappa \wedge \kappa'$ has cumulative distribution function $L(x) = \max\{F(x), G(x)\}$. We will prove that for any function $f(x)$, if $\int f(x)\,dF \geq \int f(x)\,dL$ then $\int f(x)\,dH \geq \int f(x)\,dG$. The first inequality can be written as

$$\int f(x)\,dF \geq \int_{\{x|F(x)\geq G(x)\}} f(x)\,dF + \int_{\{x|F(x)<G(x)\}} f(x)\,dG \tag{22}$$

or

$$\int_{\{x|F(x)<G(x)\}} f(x)\,dF \geq \int_{\{x|F(x)<G(x)\}} f(x)\,dG.$$

The second inequality can be written as

$$\int_{\{x|F(x)\geq G(x)\}} f(x)\,dG + \int_{\{x|F(x)<G(x)\}} f(x)\,dF \geq \int f(x)\,dG \tag{23}$$

or

$$\int_{\{x|F(x)<G(x)\}} f(x)\,dF \geq \int_{\{x|F(x)<G(x)\}} f(x)\,dG.$$

But this implies that whenever inequality (22) is satisfied, inequality (23) must also be satisfied. Therefore any function defined on the real numbers is quasi-supermodular when defined over the space of probability distributions over the real line.

Finally, to prove that the objective function satisfies increasing difference in $(\kappa; K)$ we have to show that the difference between the objective function at K' and the objective function computed at K is increasing in κ whenever $K' > K$. To see this, observe that such difference is given by the function

$$I(K'', K) - I(K'', K') = \begin{cases} (1-d)(K'-K) & \text{if} \quad K'' \geq (1-d)K' \\ (1-q)K'' - (1-d)(K - qK') & \text{if} \quad (1-d)K' > K'' \geq (1-d)K \\ q(1-d)(K'-K) & \text{if} \quad (1-d)K > K'', \end{cases}$$

which is increasing in K''. Therefore, $E_\kappa\left[I\left(\widetilde{K}, K\right) - I\left(\widetilde{K}, K'\right)\right]$ is increasing in κ. This proves that the optimal policy κ is increasing in K.

Suppose now that $\underline{K}'_s > (1-d)K$, so that at s investment is always strictly positive. Suppose now that the quantity of capital is increased by a small amount I such that the inequality $\underline{K}'_s > (1-d)(K+I)$ still holds. It must be the case that the optimal policy remains the same, so that the value of the value function increases by $(1-d)I$. If this were not the case then we would have $W(V, K+I) > W(V,K) + (1-d)I$, but this implies that by adopting at state (V, K) the policy adopted $(V, K+I)$ we would get a value strictly higher than $W(V, K)$, a contradiction. A similar reasoning implies that the optimal policy does not change when $\overline{K}'_s < (1-d)K$ and we increase K by I. ■

Appendix II

The computational work in this paper consists of two main steps. First, we computed the value function and the global optimal policies. Second, we generated the empirical joint distribution of firms' sizes and values and their growth rates. Table 1 shows the parameter values we used.

Parameter	**Description**	**Value**
p	Probability of good outcome	0.8
a	Curvature of R $(R(K) \equiv \frac{K^a}{a})$	0.5
q	Resale Price	0.8
S_0	Scrap Value	100
d	Depreciation Rate	0.07
δ	Discount Factor	0.98

Table 1: With this parametrization $K^* = 81.5$ and $V^* = 722.3$.

In the first step the value function and the global optimal policies for λ, K', V^L and V^H have been computed using a value function iteration algorithm. We did not use projection methods, due to the presence of occasionally binding constraints (incentive compatibility and non-negativity of promised utilities). Figure 9 shows the computed value function W. It increases steeply only for small value of V (equity). All flat regions start at similar values of V. Since the optimal policies for λ and K' have kinks (see Figures 10 and 11, respectively), we used a grid-based non-parametric approach with interpolation, instead of polynomial-based approximations.

The grid for V has 40 nodes, uniformly spaced between 0 and $0.95V^*$. The grid for K has 30 nodes, uniformly spaced between 0 and $1.5K^*$. We included values well above K^* to verify that the these values are never reached in our simulations, as predicted by the theory.

The four dimensional control space (λ, K', V_L, V_H) generated an additional source of complexity. Even after eliminating one control using the promise–keeping equality constraint, the optimization problem at each step of the value function iteration algorithm remains three-dimensional.

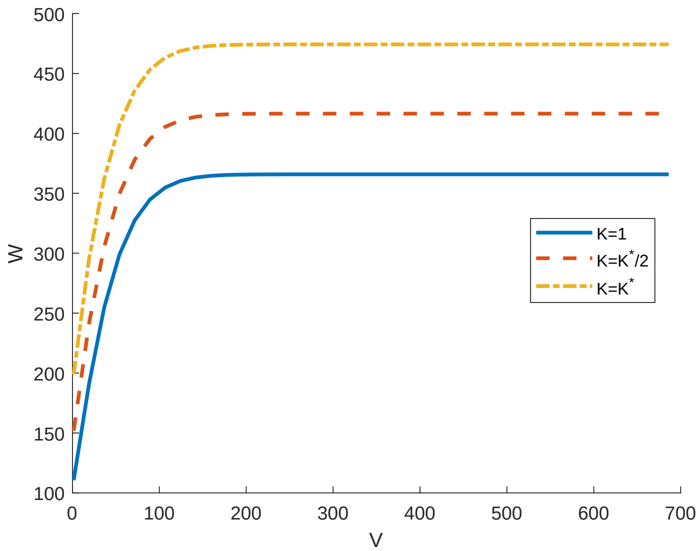


Figure 9: Slices of the value function W

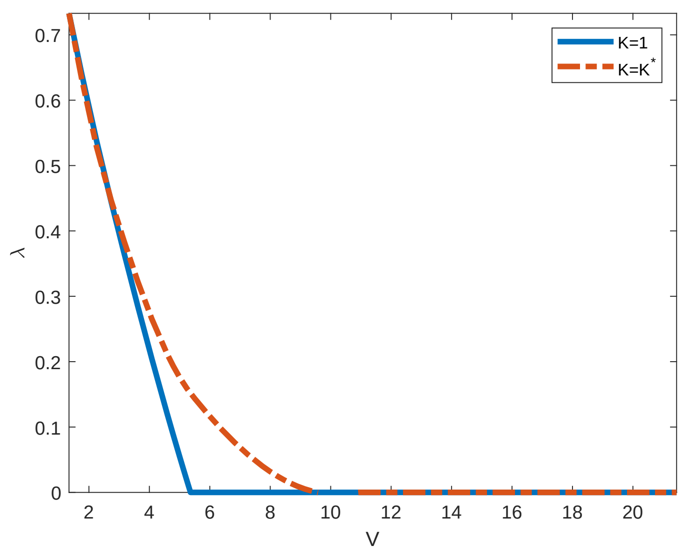


Figure 10: Slices of the λ policy

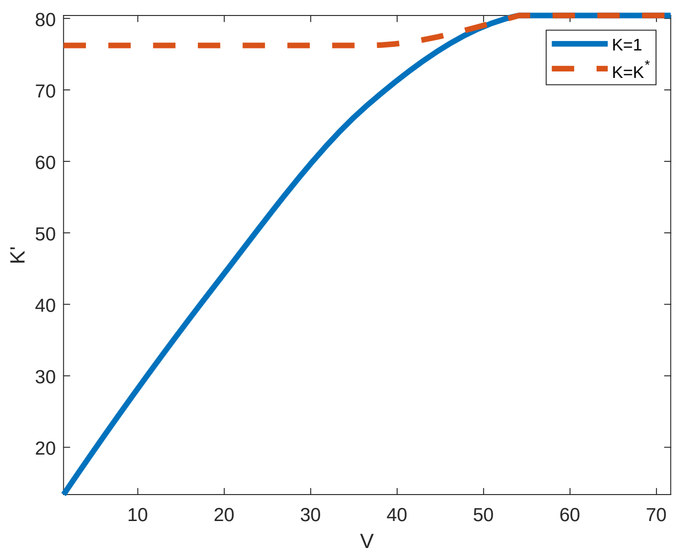


Figure 11: Slices of K' policy as function of V

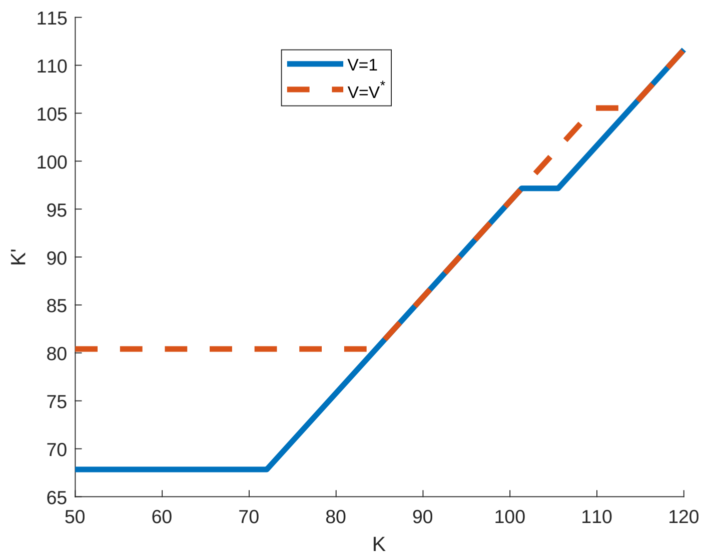
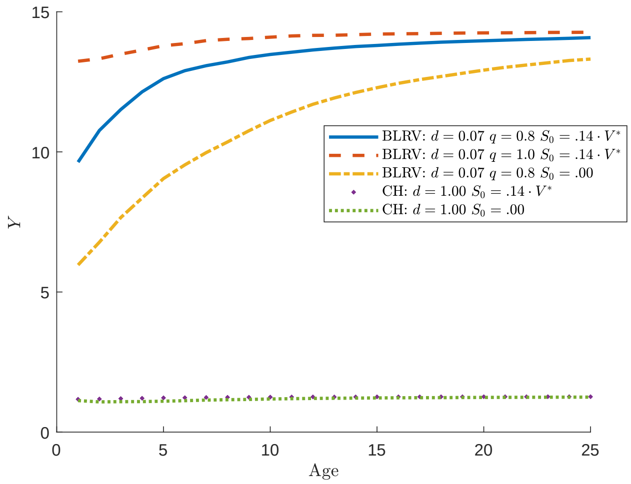


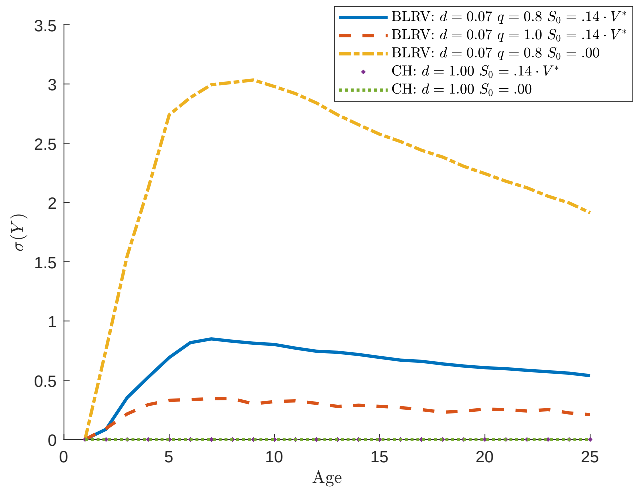
Figure 12: Slices of K' policy as function of K

Figure 12 shows the optimal K' policy as a function of K, instead of V (as is done in Figure 11). The picture reveals the presence of three regions: i. for small values of K, K' is constant at a target value (which depends on V); ii. for intermediate values, $K' = K$; and iii. for large values, $K' = (1 - d)K$. In the second step we used the optimal policies to perform simulations. For each time horizon $T = 10,\ 20,\ \dots\ ,\ 100$, we have simulated the life of 2000 firms. After discarding all firms that are liquidated before period T, we have built the joint distribution of the firms' sizes and values, the associated conditional distributions, the expected growth rates, and standard deviations conditional on firm size.

Extra Figures. Not for Publication

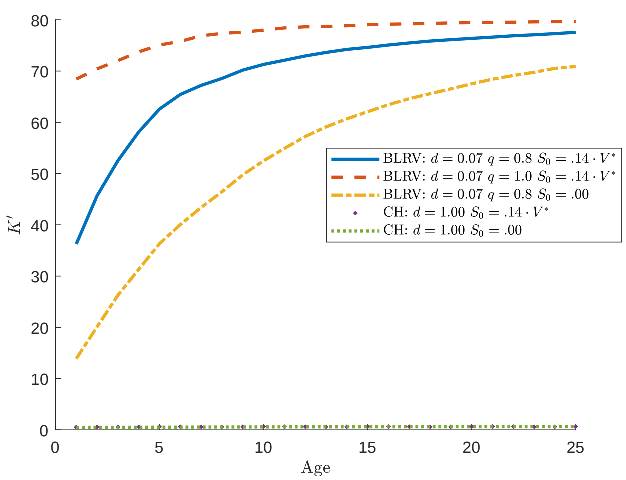


(a) Expected revenue

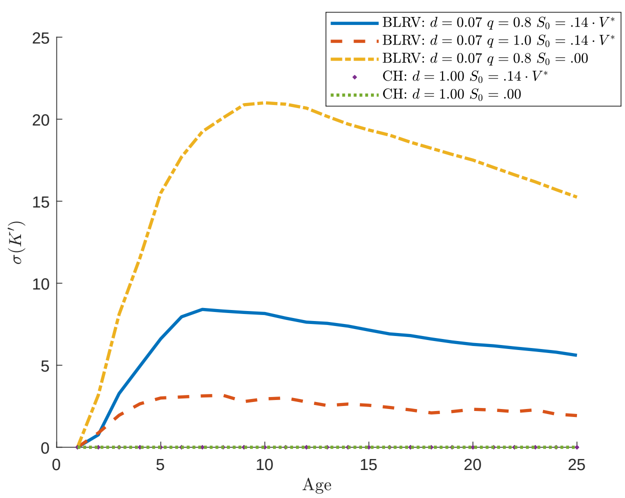


(b) Variance of revenue

Figure 13: Expected revenue and variance of revenue as a function of age.



(a) Expected K'



(b) Variance of K'

Figure 14: Expected K' and variance as a function of age.

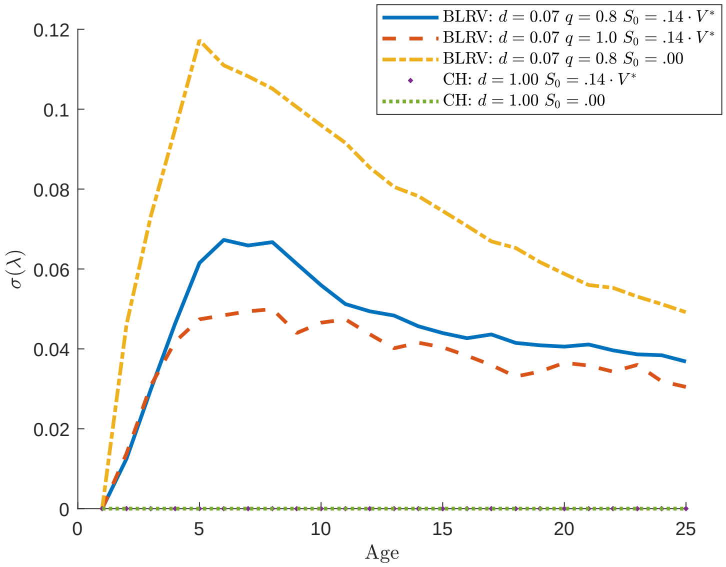


Figure 15: Standard deviation of liquidation probability for $q = 0.8$ and for $q = 1$

References

[1] Ai, H., D. Kiku and R. Li (2016), 'Firm Dynamics under Limited Commitment', working paper, https://www.hengjieai.com/uploads/1/0/6/6/106628821/akl1_lc.pdf.

[2] Albuquerque, R. and H. Hopenhayn (2004), 'Optimal Lending Contracts and Firm Dynamics', *Review of Economic Studies*, **71**(2): 285–315.

[3] Atkeson, A. and R. Lucas (1992), 'On Efficient Distribution with Private Information', *Review of Economic Studies*, **59**(3): 427–453.

[4] Bolton, P., H. Chen, Z and N. Wang (2011), 'A Unified Theory of Tobin's q, Corporate Investment, Financing and Risk Management', *Journal of Finance*, **LXVI**: 1545–1578.

[5] Cao, D., G. Lorenzoni and K. Walentin (2019), 'Financial Frictions, Investment and Tobin's Q', *Journal Monetary Economics*, **103**: 105–122.

[6] Clementi, G.L. and H. Hopenhayn (2006), 'A Theory of Financing Constraints and Firm Dynamics', *Quarterly Journal of Economics*, **121**: 229–266.

[7] Cooley, T., R. Marimon and V. Quadrini (2004), 'Aggregate Consequences of Limited Contract Enforceability', *Journal of Political Economy*, **112**: 817–847.

[8] Cooley, T. and V. Quadrini (2001), 'Financial Markets and Firm Dynamics', *American Economic Review*, **91**: 1286–1310.

[9] Decker, R., J. Haltiwanger, R. Jarmin and J. Miranda (2014) 'The Role of Entrepreneurship in US Job Creation and Economic Dynamism' *Journal of Economic Perspectives* **28** (3):3–24.

[10] De Marzo, P. and M. Fishman (2007), 'Agency and Optimal Investment Dynamics', *Review of Financial Studies*, **20**: 151–188.

[11] De Marzo, P. and M. Fishman (2007), 'Optimal Long-Term Financial Contracting', *Review of Financial Studies*, **20**: 2079–2128.

[12] De Marzo, P., M. Fishman, Z. He and N. Wang (2012), 'Dynamic Agency and the q Theory of Investment', *Journal of Finance*, **LXVII**: 2295–2340.

[13] Dunne, T. Roberts, M. and L. Samuelson (1988), 'Patterns of Firm Entry and Exit in U.S. Manufacturing Industries', *Rand Journal of Economics*, **19**(4): 495–515.

[14] Dunne, T. Roberts, M. and L. Samuelson (1989), 'The Growth and Failure of U.S. Manufacturing Plants', *Quarterly Journal of Economics*, **104**(4): 671–698.

[15] Eisfeldt, A. and A. Rampini (2007), 'New or used? Investment with credit constraints', *Journal of Monetary Economics*, **54**: 2656–2681.

[16] Evans, D. (1987), 'The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries', *Journal of Industrial Economics*, **35**(4): 567-81.

[17] Evans, D. (1987), 'Tests of Alternative Theories of Firm Growth', *Journal of Political Economy*, **95**(4): 657–74.

[18] Fu, S. and R. V. Krishna (2019), 'Dynamic Financial Contracting with Persistent Private Information', *RAND Journal of Economics*, **50** (2): 418-452.

[19] Gavazza, A. (2011), 'Leasing and Secondary Markets: Theory and Evidence from Commercial Aircrafts', *Journal of Political Economy*, **119**(2): 325–377.

[20] Haltiwanger, J. R. Jarmin and J. Miranda (2013), 'Who Creates Jobs? Small versus Large versus Young', *The Review of Economics and Statitstics*, **95**(2): 347–361.

[21] Lanteri, A. (2018), 'The Market for Used Capital. Endogenous Irreversibility and Reallocation over the Business Cycle', forthcoming, *American Economic Review*.

[22] Li, S. M. (2013), 'Optimal Lending Contracts with Long-run Borrowing Constraints', *Journal of Economic Dynamics and Control*, **37**: 964–983.

[23] Milgrom, P. and C. Shannon (1994), 'Monotone Comparative Statics', *Econometrica*, **62**: 157–180.

[24] Pugsley, B, P. Sedláček and V. Sterk (2020) 'The Nature of Firm Growth', available at https://drive.google.com/file/d/1DKTJwnI6iFbI6l7mvmIc-0qZyhBp-CVL/view.

[25] Quadrini, V. (2004), 'Investment and Liquidation in Renegotiation–Proof Contracts with Moral Hazard', *Journal of Monetary Economics*, **51**: 713–751.

[26] Ramey, V. and M. Shapiro (2001), 'Displaced Capital: A Study of Aerospace Plant Closings', *Journal of Political Economy*, **109**: 958–992.

[27] Rampini, A. and S. Viswanathan (2013) 'Collateral and Capital Structure', *Journal of Financial Economics*, **109**: 466–429.

[28] Spear, S., and Srivastava, S. (1987), 'On Repeated Moral Hazard with Discounting' *The Review of Economic Studies*, **54(4)**, 599-617

[29] Stokey, N., R. Lucas and E. Prescott (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge, Mass.

[30] Sutton, J. (1997), 'Gibrat's Legacy', *Journal of Economic Literature*, **35**: 40–59.