Ann Finance (2009) 5:465–494
DOI 10.1007/s10436-009-0122-1

SYMPOSIUM

Entrepreneurship and firm heterogeneity with limited enforcement

Alexander Monge-Naranjo

Received: 1 May 2008 / Accepted: 16 March 2009 / Published online: 3 April 2009

Abstract The presence of credit constraints can shape the entry, growth, and size distribution of firms. In this paper, I consider the optimal contract between banks and entrepreneurs in an economy in which limitations in the punishments for default induce limitations in the credit available to entrepreneurs. I derive a very simple characterization of the optimal contract at the time of inception of firms and show how this optimal contract shapes the evolution of the size distribution of firms and the aggregate levels of entrepreneurial activity and output.

Keywords Limited commitment · Firm heterogeneity · Entrepreneurship

JEL Classification D92 · E22 · E32 · E44 · G33

1 Introduction

A feature of entrepreneurial activity often overlooked in the literature on occupation choice and entrepreneurship is that entrepreneurs typically must wait a long time before reaping the benefits of their investments and efforts.[1] In the meantime, entrepreneurs control assets that might belong to someone else, and use their income to repay debts. The enforcement of contracts is what allows those potential entrepreneurs with insufficient funds to obtain financing in the first place to set up and run firms. Therefore,

[1] For example Banerjee and Newman (1991), Holmes and Schmitz (1990), Lazear (2005), and Lucas (1978).

The paper benefited from useful comments and suggestions from an anonymous referee and from seminar participants at Duke University, Northwestern University, and University of Western Ontario.

A. Monge-Naranjo (✉)
Department of Economics, Pennsylvania State University,
502 Kern Graduate Bldg., University Park, PA 16802-3306, USA
e-mail: alexmonge@gmail.com; aum26@psu.edu

the extent which a society enforces contractual arrangements is a key determinant of the volume and the productivity of its entrepreneurship.

In this paper I study the impact of limited enforcement on entrepreneurial decisions and on the behavior of firms. I consider the optimal contract between entrepreneurs and banks allowing for banks to be more patient than entrepreneurs. I derive a very simple characterization of the optimal contract at the time of inception of firms and show how the contract shapes the evolution of the size distribution of firms and the aggregate levels of entrepreneurial activity and output. Then, I use the model with two purposes. First, I ask whether limited enforcement can explain key empirical observations on the evolution of the firm size distribution. Second, I show that the model can be useful as an organizing framework to separate three different margins for the impact of limited commitment on aggregate productivity.

I consider a standard one-sided limited commitment dynamic contracting problem.[2] An agent, the "entrepreneur," has access to a productive technology but lacks the resources needed to operate it. Another agent, the "bank," has frictionless access to financing but lacks productive projects of his own. Financial frictions endogenously arise from limitations in the punishments available to banks to discipline entrepreneurs if they repudiate their financial liabilities. In the economy, banks compete with each other. In equilibrium, contracts satisfy a time-zero participation constraint for the bank and a continuum of participation constraints for the entrepreneur.

With limited commitment, entrepreneurs are tempted to default and appropriate the capital under their control. To beat default incentives, the contract must imply that entrepreneurs are *always* better off conforming with the contract than defaulting on it. As originally explained by Townsend (1982), a dynamic arrangement uses future consumption to cope with current perverse incentives. With limited commitment, the optimal contract postpones the consumption of the entrepreneur because increasing consumption in farther dates relaxes the participation constraint for more periods. Relaxing participation constraints allows the lender to increase the capital used by the firm, and doing so increases the present value of output and net profits.

The complete characterization of the contract boils down to the determination of the "time-to-maturity," the time it takes a firm from the moment it is created to mature and stop growing. With risk-neutral entrepreneurs, the entrepreneur's consumption is set to zero for the periods before maturity and to a positive constant level from then on. Before maturity, the discounted utility of the entrepreneur grows, as he gets closer to positive consumption. As the entrepreneur's utility goes up, he can and does control more capital.

The fact that optimal contracts are characterized solely by the time-to-maturity greatly simplifies the analysis of the cross-sectional implications for the behavior of firms. It is straightforward to show that the richer the entrepreneur, the sooner the firm matures. Indeed, a firm might be born mature if the entrepreneur is rich enough. The relationship of the time-to-maturity with the productivity of the firm is more complex. On one hand, with a higher productivity firms yield higher profits and the bank's financing can be paid off sooner and the optimal time-to-maturity would be

[2] As opposed to two-sided no commitment, e.g. Hart and Moore (1994) and Kocherlakota (1996).

lower. On the other hand, with a higher productivity, mature firms use more capital and this requires a higher consumption level for the entrepreneur. A lower time-to-maturity increases the present value of resources invested by the bank. I provide very simple conditions for time-to-maturities to be increasing or decreasing in the firm's productivity.

The fact that optimal contracts are characterized by the time-to-maturity of firms also simplifies the study of economies with many firms, and in particular, the evolution of the size distribution of cohorts of firms. Cabral and Mata (2003) find that at the time of birth, the distribution of the logarithms of firm sizes of a given cohort is very skewed to the right and gradually evolves towards a more symmetric distribution. Their findings suggest that it takes a long time before the firm size distribution of a given cohort approximates a log-normal distribution. I show that using a simple and straight parameterization, the model does a surprisingly good job replicating these patterns.

The form of the optimal contract also makes aggregation straightforward. The model portrays very transparently three margins by which limited enforcement impairs aggregate output and welfare. First, bad contractual enforcement impairs entrepreneurship. Some socially productive firms are not created at all simply because of perverse incentives. Second, limited enforcement impairs the size of young, immature firms and may increase the time it takes them to achieve maturity. Capital, output, and net-profits are held inefficiently low during the growth phase of firms. Third, limited enforcement might also impair the size of mature firms by introducing a wedge between the effective interest rate of firms and the actual interest rate faced by banks. A simple numerical exercise suggests that the output and welfare cost of bad enforcement can be quite large and that all three margins can be important.

The contracting problem studied in this paper is closely related to the one in Albuquerque and Hopenhayn (2004). Their set up is in discrete time and allows for a stochastic process for the productivity of firms. Employing recursive methods, they characterize the behavior of the promised utility of the entrepreneur with the focus of characterizing the growth and exit decisions of an individual firm. In this paper I have considered the problem in continuous time, formulated the contracting problem in sequence form, and focused on the cross-sectional and aggregate implications of limited enforcement. This paper is also related to existing quantitative studies on the impact of credit constraints on the size distribution of firms (e.g. Cooley and Quadrini 2001; Quintin 2008). However, their emphasis is on the overall limiting size distribution of firms and on the behavior of active firms, while the focus of this paper is on the initial contract and how it shapes the behavior of cohorts of firms over time.

The model presented in this paper abstracts from some interesting and relevant issues for the impact of contract enforcement on entrepreneurship and the creation and behavior of firms. For instance, I have abstracted from other policy distortions (e.g. Guner et al. (2008); Restuccia and Rogerson (2008)). Also, I have taken the distribution of assets and productivities as exogenously given. Instead, as in Cagetti and De Nardi (2006), (2009), this distribution could be modeled as the outcome of bequest decisions. I have also abstracted from savings decisions that are motivated by entry to entrepreneurship, as in Buera (2008). The model abstracts from aggregate fluctuations (as in for example Cooley et al. 2004, Monge-Naranjo 2009). Finally, the model abstracts from the general equilibrium impact of entrepreneurial activity on entrepreneurs and non-entrepreneurs, which are considered in further research taking advantage of the analytical results presented here.

The remainder of the paper is organized as follows. Section 2 sets up the basic environment and solves for the optimal contract between a bank and a potential entrepreneur. Section 3 analyzes the cross-sectional implications of the optimal contract, namely the impact of the firm's productivity as well as the impact of the entrepreneur's own assets for the behavior of a firm. Section 4 studies the model's implication for the evolution of the firm size distribution. Section 5 studies the aggregate consequences of limited enforcement. Section 6 concludes.

2 The basic contracting problem

This section develops the basic contractual environment and the ensuing optimal contractual arrangement between a bank and an entrepreneur.

2.1 The environment

Preferences, endowments, technology Consider an economy with two agents, both of which are risk neutral. One agent is a potential entrepreneur whose utility as of time t is given by

$$U(t) = \int_t^{\infty} e^{-\rho(s-t)} c(s) ds, \tag{1}$$

where $c(s)$ is his consumption at period $s \geq t$, and $\rho > 0$ is a discount factor. The other agent is a bank. Letting $p(s)$ denote the net flow payoff for the bank at time s in a given contract, then the net present value payoff from the contract is

$$P(t) = \int_t^{\infty} e^{-r(s-t)} p(s) ds. \tag{2}$$

I assume $0 < r \leq \rho$ to avoid uninteresting degeneracies.

Each potential entrepreneur is initially endowed with a level of financial assets equal to $a \geq 0$. At birth he is endowed with two choices. He can be a worker and earn a constant flow of the consumption good $w(\tau) = w$ for all $\tau \geq 0$. Alternatively, he might opt to become an entrepreneur. Occupation choices are equilibrium outcomes in the economy I now continue describing.

I assume that each potential entrepreneur is born attached to a single project, and that, if any, he can only run that project. The productivity of each project is fixed and given by a project-specific term $z > 0$. Once activated, projects can, in principle, remain active forever. The flow of output $y(\tau)$ that a project generates in period τ of an active project is

$$y(\tau) = zk(\tau)^{\nu}, \tag{3}$$

where $k(\tau)$ is the installed capital and $0 < \nu < 1$ is a parameter that governs the degree of decreasing returns to each project. To abbreviate some of the ensuing expressions, I will use the shorthands $f(k) \equiv k^{\nu}$ and $f'(k) \equiv \nu k^{\nu-1}$.

Installed capital $k(\tau)$ evolves according to

$$k(\tau) = k_0 e^{-\delta\tau} + \int_0^{\tau} e^{-\delta(s-\tau)} i(s)\, ds, \tag{4}$$

where k_0 is the initial installment of capital, i.e. the capital with which the firm initiated operations, and $i(s)$ the stream of investments up to period τ. Investment is fully reversible, i.e. $i(s)$ can be positive or negative. Capital depreciates at rate $\delta > 0$.

Regardless of the value of z, activating a project requires a set up investment of $K_I \geq 0$ units of the single capital/consumption good.

Contract enforcement The incentive problem from which credit limitations arise is the following: As long as he controls a firm, an entrepreneur has the option to default on their financial liabilities. If they default, the only punishment that they face is to be permanently excluded from financial markets. Moreover, a defaulting entrepreneur can run away with a fraction of the capital installed in the firm. Specifically, a defaulting entrepreneur that at period τ controls a project with $k(\tau)$ units of capital obtains a level of utility

$$V_D[k(\tau)] = \theta_D k(\tau), \tag{5}$$

where $0 \leq \theta_D < \rho^{-1}$. Implicit in this inequality is the assumption that by defaulting, entrepreneurs permanently lose access to their productive (labor or entrepreneurial) projects. Such assumption is made for analytical tractability. Also, I am implicitly assuming that entrepreneurs can only steal part of the capital or that the rate of return that they can obtain from it after a default is strictly below ρ.

As long as the firm is running, the possibility of default is present. This leads to a continuum of participation constraints. Given a prescribed consumption stream $c(s)$, a contract signed at $t = 0$ (the normalized starting date throughout this section) must cope with the entrepreneur's temptation to default by satisfying

$$U(t) \geq \theta_D k(t), \tag{6}$$

for all $t \geq 0$.

2.2 Optimal dynamic contracts

Regardless of the presence of incentive problems, the payoff of the bank of signing a contract is determined by its initial investment (initial installed capital and set-up costs) and the net flows of resources the bank receives over the life of the project. The flows $p(s)$ are determined by installed capital $k(s)$, investment $i(s)$, and entrepreneur's consumption (his dividends from the project) in the form $p(s) = zf[k(s)] - i(s) - c(s)$.

If a_I denotes the own funding of an entering entrepreneur, then a bank would initially invest $k_0 + K_I - a_I$ and would be willing to sign the contract if

$$\int_0^\infty e^{-rs}\{zf[k(s)] - i(s) - c(s)\}\,ds \geq k_0 + K_I - a_I. \tag{7}$$

The problem for a bank is whether to offer a contract to a potential entrepreneur to become an active entrepreneur and, if so, offering the initial capital and k_0 and the streams of investments, consumption, and capital $\{i(s), c(s), k(s)\}$ so as to maximize the $t = 0$ zero value of (1) subject to (4) and the breaking even condition (7).

The Lagrangian for the optimal contract is:

$$\begin{aligned}
L = \min_{\{\theta,\lambda,\mu,\vartheta\}} \max_{\{c,k,i,k_0\}} & \int_0^\infty e^{-\rho t} c(t)\,dt \\
& + \int_0^\infty e^{-rt}\theta(t)\,c(t)\,dt \\
& + \int_0^\infty e^{-rt}[\lambda(t)]\left[k_0 e^{-\delta t} + \int_0^t e^{-\delta(s-t)} i(s)\,ds - k(t)\right]dt \\
& + \int_0^\infty e^{-rt}\mu(t)\left[\int_t^\infty e^{-\rho[s-t]} c(s)\,ds - \theta_D k(t)\right]dt \\
& + \vartheta\left[\int_0^\infty e^{-rt}[zf[k(t)] - i(t) - c(t)]\,dt - k_0 - K_I\right],
\end{aligned}$$

where $\theta(\cdot)$, $\lambda(\cdot)$, $\mu(\cdot)$, and ϑ are non-negative multipliers associated, respectively, to the non-negativity of consumption, the law of motion of installed capital, and the participation conditions for the entrepreneur and the bank.

With the assumed production function and law of motion of capital, it is not hard to see that the optimal investment path is interior. Then, the first order conditions for this problem boil down to

$$[i(\tau)] : \vartheta = \int_{\tau}^{\infty} e^{-(r+\delta)(s-\tau)} \lambda(s)\, ds, \tag{8}$$

$$[k(\tau)] : \lambda(\tau) = \vartheta \left[z f'[k(\tau)] - \mu(\tau) \frac{\theta_D}{\vartheta} \right], \tag{9}$$

$$[c(\tau)] : \vartheta = e^{(r-\rho)\tau} + \theta(\tau) + \int_{0}^{\tau} e^{(r-\rho)(\tau-s)} \mu(s)\, ds, \tag{10}$$

plus the non-negativity of consumption, the law of motion of installed capital, and the participation conditions for the entrepreneur and the bank.

Full enforcement I first consider the optimal allocation with full enforcement, i.e. $\theta_D = 0$, which I denote with the superscript U. With full enforcement $\mu(\tau) = 0$ for all τ, plugging (9) into (8) leads to

$$1 = z \int_{\tau}^{\infty} e^{-(r+\delta)(s-\tau)} f' \left[k^U(s) \right] ds. \tag{11}$$

Since this holds for all τ, the derivative of the right-hand-side with respect to τ has to be equal to zero. Leibniz formula implies that:

$$-z f' \left[k^U(\tau) \right] + (r+\delta) z \int_{\tau}^{\infty} e^{-(r+\delta)(s-\tau)} f' \left[k^U(s) \right] ds = 0.$$

Plugging expression (11) into the one above leads to the standard condition that, for all $\tau \geq 0$ the marginal product of capital must be equal to its cost of use, i.e. $z f' \left[k^U(\tau) \right] = r + \delta$. For our assumed production function, this is

$$k^U(\tau) = k^U \equiv \left(\frac{z\nu}{r+\delta} \right)^{\frac{1}{1-\nu}}, \tag{12}$$

for all τ.

Therefore, in a world with full contract enforcement, firms initiate operations and always operate at a constant level of installed capital and produce a constant flow of output. The stream of investment is uniquely pinned down by $i^U(\tau) = \delta k^U$, i.e. the flow of investment needed to make up for depreciation.

As discounted by the bank, the net of investment present value of the operation profits of a firm with productivity z is

$$\Pi_0^U = \int_0^\infty e^{-rs}\left[zf\left(k^U\right) - \delta k^U\right]ds - k^U,$$
$$= (z)^{\frac{1}{1-\nu}}\left(\frac{1-\nu}{r}\right)\left(\frac{\nu}{r+\delta}\right)^{\frac{\nu}{1-\nu}}, \tag{13}$$

which is strictly increasing in z and strictly decreasing in r and δ.

With a market interest rate below the entrepreneur's discount factor, $r \leq \rho$, the entrepreneur is more impatient that the bank. With no enforcement problems, the optimal consumption profile is degenerate: the entrepreneur bunches all his consumption in $t = 0$. The bank collects all the proceeds from the firm afterwards. Since banks compete with each other all the surplus is allocated to entrepreneurs. Hence, an entrepreneur with initial assets a derives a utility equal to $U_0^U = \Pi_0^U - K_I + a$.

Now, to determine whether the firm is activated or not, recall that activating the firm not only requires a set up cost K_I but also that the potential entrepreneur gives up his labor alternative. Once again, since $r \leq \rho$, a worker would want to borrow to bunch all his consumption in the first period. Hence, a worker with initial assets a derives a utility $U_0^W = \frac{w}{r} + a$. Therefore, the potential entrepreneur becomes an active entrepreneur if

$$\Pi_0^U - K_I \geq \frac{w}{r},$$

which, because of the absence of financial frictions, is independent of the initial assets a of the potential entrepreneur. Using (13), then the minimum productivity for a firm to be activated is

$$z_{\min}^U \equiv \left(\frac{r+\delta}{\nu}\right)^\nu \left(\frac{w + rK_I}{1-\nu}\right)^{1-\nu},$$

an increasing function of the interest rate r, the labor opportunity cost of the potential entrepreneur and the set up cost of the firm.

Limited enforcement Now assume that for at least some dates the temptation to default is operative, i.e. the participation constraint of the entrepreneur binds. The fact that $\mu(\tau) > 0$ for some $\tau \geq 0$ changes substantially the form of the optimal allocation. When enforcement is limited, no lending can be supported without promises of future consumption. Front-loading consumption, therefore, becomes impossible.

There are two opposing forces that determine the optimal behavior of an entrepreneur's consumption. On one hand, as with full enforcement, the higher patience of the lender pushes the entrepreneur's consumption to take place as early as possible. On the other hand, the inability of the entrepreneur to commit to repay makes it optimal to postpone his consumption as much as possible so that he has incentives to remain in the financial relationship and not run away with the capital. Early and late consumption help delivering utility and relaxing participation constraints, but late consumption has the advantage of doing it for more periods.

Consider an interval in which consumption is positive, i.e. $\theta(\tau) = 0$. With this, (10), becomes

$$\vartheta = e^{(r-\rho)\tau} + \int_{0}^{\tau} e^{(r-\rho)(\tau-s)} \mu(s)\, ds. \tag{14}$$

The expression in the right-hand-side must be a constant in such interval since the left-hand-side is a constant. Equating to zero the derivative of this expression with respect to τ and rearranging we obtain

$$\mu(\tau) = (\rho - r)\, \vartheta. \tag{15}$$

Therefore, notice that if (14) holds for τ, then it must hold for any $\tau' > \tau$. Then, it must be the case that there exists a stopping time $t_M < \infty$ after which $\theta(\tau) = 0$ for all $\tau \geq t_M$. I will call this stopping time t_M, the "*time-to-maturity*," because it is the interval of time that elapses between the activation of the firm and the first date on which the firm stops growing and the entrepreneur starts collecting dividends from it.

It is worth noticing that if $\rho > r$, the participation always binds, even after t_M. Plugging (15) into (9), and inserting the implied expression for $\lambda(s)$ in (8) renders:

$$1 = \int_{\tau}^{\infty} e^{-(r+\delta)(s-\tau)} \left\{ z f'[k(s)] - (\rho - r)\, \theta_D \right\} ds. \tag{16}$$

The derivative of the expression in the right-hand-side of (16) with respect to τ must be equal to 0. Using the Leibniz rule, plugging in (16), rearranging, and simplifying, we obtain:

$$z f'[k(\tau)] = r(1 - \theta_D) + \rho\theta_D + \delta. \tag{17}$$

The amount of installed capital of a mature firm is given by

$$k_M = \left(\frac{z\nu}{r_M + \delta} \right)^{\frac{1}{1-\nu}}, \tag{18}$$

where

$$r_M \equiv r(1 - \theta_D) + \rho\theta_D \tag{19}$$

is a weighted average between the discount rate of the borrower and the discount rate of banks. The interest rate r_M is the relevant interest rate for mature firms. The parameter θ_D governs the value of defaulting and acts as the relative weight between the entrepreneur's discount rate ρ and the market interest rate r. If $\rho > r$ and $\theta_D > 0$, installed capital is always below k^U. In order to allocate capital to the firm, the bank must also allocate consumption to the entrepreneur so that he does not run away with it. This adds the cost $(\rho - r)\,\theta_D$ for each unit of installed capital.

As anticipated, after the stopping time $t_M \geq 0$, the amount of capital installed in the firm will remain constant. Even after achieving maturity, firms face a higher effective cost of use of capital due to limited repayment commitment and the higher discount rate of entrepreneurs.

I now characterize consumption of active entrepreneurs. First, consider the consumption $c_M(\cdot)$ of an entrepreneur controlling a mature firm. Since his participation constraint remains binding, for all periods $\tau > t_M$, his utility $U(\tau) = \int_\tau^\infty e^{-\rho[s-\tau]} c_M(s)\,ds$ is constant over time and equal to $\theta_D k_M$. Therefore, once consumption is positive it remains constant and equal to

$$c_M = \rho \theta_D k_M.$$

Consider now the case in which $\theta(\tau) > 0$, which holds for the interval $[0, t_M]$. Consumption in this interval is equal to zero and the utility of the entrepreneur is solely determined by the interval of time pending to achieve the positive consumption after t_M. Therefore, the utility of an active entrepreneur is

$$U(\tau) = \begin{cases} e^{-\rho(t_M-\tau)}\theta_D k_M & \text{if } 0 \leq \tau < t_M, \\ \theta_D k_M & \text{otherwise.} \end{cases}$$

Postponing consumption is driven by the need to relax the participation constraints on the entrepreneur and increase the capital used by the firm. By setting consumption to be initially equal to zero, the utility of the entrepreneur grows as fast as it is physically possible. As a result, the amount of capital used by the firm also grows as fast as possible, given the incentive limitations. From $U(\tau)$ we can directly obtain the amount of installed capital from the utility of the entrepreneur as

$$k(\tau) = \begin{cases} e^{-\rho(t_M-\tau)} k_M & \text{if } 0 \leq \tau < t_M\,, \\ \theta_D k_M & \text{otherwise.} \end{cases}$$

In particular, observe that the initial amount of capital is given by $k_0 = e^{-\rho t_M} k_M$. The flow of investment for the periods prior to t_M is equal to $i(\tau) = (\rho + \delta)\,k(\tau)$ and equal to δk_M afterwards. This is, the optimal path of investment continuously increases the amount of capital under the control of a entrepreneur until he is able to commit not to run away with k_M. After t_M, investment only compensates for depreciation.

Inception of firms with limited commitment The complete characterization of the contract boils down to the determination of the time-to-maturity t_M and whether a potential entrepreneur becomes an active entrepreneur or remains a worker. The remarkably simple form of the contract allows me to do this characterization as follows. First, I characterize the present value of the operation payoffs for the bank for an arbitrarily given t_M. Second, I analyze how the break-even condition of the bank (7) determines whether a bank would offer a contract to a potential entrepreneur and, if so, what is t_M^*, the best time-to-maturity that can be offered. Finally, I determine whether a potential entrepreneur would want to sign up for the contract and become an active entrepreneur.

For an arbitrary time-to-maturity t_M, the initial present value of the flow of net payoffs for the bank financing the firm are

$$\begin{aligned} P_0(t_M) &= -k_0 + \int_0^{t_M} e^{-rt}\left[zk(t)^{\nu} - (\rho+\delta)k(t)\right]dt \\ &\quad + \int_{t_M}^{\infty} e^{-rt}\left[z(k_M)^{\nu} - \delta k_M - c_M\right]dt, \\ &= -k_M e^{-\rho t_M} + z(k_M)^{\nu}\left(\frac{e^{-rt_M} - e^{-\rho\nu t_M}}{\rho\nu - r}\right) \\ &\quad - (\rho+\delta)k_M\left(\frac{e^{-rt_M} - e^{-\rho t_M}}{\rho - r}\right) + \frac{e^{-rt_M}}{r}\left[z(k_M)^{\nu} - \delta k_M - c_M\right]. \end{aligned}$$

The first term is the initial installed capital; the second and third terms, respectively, are the value of production and the cost of investment before maturity; and the fourth term is equal to output net of investment and the entrepreneur's consumption after the firm achieves maturity.[3]

Given the productivity of the firm z and the values of (k_M, c_M) for the firm, $P_0(t_M)$ depends only on the time-to-maturity t_M implicit in the contract. The operation's net payoff for the bank $P_0(t_M)$ is non-monotone in t_M. The non-monotonicity is driven by two countervailing forces. One one hand, a lower t_M allows the firm to achieve maturity earlier, implying that in all previous periods the firm operates with higher amounts of capital and produces higher output. On the other hand, a lower t_M increases the present value of investment and entrepreneur's consumption, which are outlays for the bank.

The expression for $P_0(t_M)$ can be simplified further. Most importantly, it is a separable function of z and t_M. Plug in the expressions for c_M and k_M, and then simplify to obtain

$$P_0(t_M) = \Gamma(t_M)(z)^{\frac{1}{1-\nu}}$$

where

$$\begin{aligned} \Gamma(t_M) \equiv &\left(\frac{\rho^2(1-\nu)}{r(\rho\nu - r)(\rho - r)}e^{-rt_M} - \frac{1}{\nu(\rho\nu - r)}e^{-\rho\nu t_M}\right. \\ &\left. + \frac{r+\delta}{(r_M+\delta)(\rho - r)}e^{-\rho t_M}\right)\nu^{\frac{1}{1-\nu}}(r_M+\delta)^{\frac{-\nu}{1-\nu}}. \end{aligned}$$

[3] This expression is valid for $\rho\nu \neq r < \rho$. If $\rho\nu = r < \rho$, then $P_0(t_M) = -e^{-\rho t_M}k_M + z[k_M]^{\nu}t_M - (\rho+\delta)k_M\left[\frac{e^{-rt_M} - e^{-\rho t_M}}{\rho - r}\right] + \left[z(k_M)^{\nu} - \delta k_M - c_M\right]\left[\frac{e^{-rt_M}}{r}\right]$. If instead $\rho\nu < r = \rho$, then $P_0(t_M) = -e^{-\rho t_M}k_M + z[k_M]^{\nu}\left[\frac{e^{-rt_M} - e^{-\rho\nu t_M}}{\rho\nu - r}\right] - (\rho+\delta)k_M t_M + \left[z(k_M)^{\nu} - \delta k_M - c_M\right]\left[\frac{e^{-rt_M}}{r}\right]$.

From direct inspection, $\lim_{t_M \to +\infty} \Gamma(t_M) = 0$, because if the entrepreneur never gets to consume, his utility is always equal to zero. If so, he could never be given the control of any positive amounts of capital and the firm's output and profits are always zero. On the other hand, $\lim_{t_M \to 0} \Gamma(t_M)$ can be positive or negative. In the latter case, by continuity, $\Gamma(t_M)$ is negative at low values of t_M. Then, a higher productivity z would *reduce* the value of $P_0(t_M)$. This will be important later on, as it implies that more productive projects require more time to achieve maturity. Moreover, if $\Gamma(\cdot)$ is initially negative, then, even if the set up costs K_I and the labor opportunity w of the entrepreneur are both zero, the opportunity of entrepreneurs to run away with the capital impairs the ability of a bank to offer a contract in which firms mature right away. Therefore, even in those circumstances, entrepreneurs must wait to (implicitly) accumulate enough assets before they consume any amount from the proceeds of the firm.

A bank is willing to sign up a contract with a potential entrepreneur if the break-even condition (7) is satisfied. Consider a potential entrepreneur in command of a project with productivity z and that offers from his own funds an amount $a_I \geq 0$. For a bank to be willing to offer financing there must exist a finite t_M for which $P_0(t_M) \geq K_I - a_I$. Competition from one another will prompt banks to offer the best feasible contract. Given that the utility of the would-be entrepreneur from the contract depends entirely on the time-to-maturity, if a contract is offered in equilibrium, it is given by

$$t_M^*(z, a_I) = \inf\left\{x \geq 0 : \Gamma(x) \geq (K_I - a_I) / (z)^{\frac{1}{1-\nu}}\right\}. \tag{20}$$

Obviously, if $\Gamma(\cdot)$ is uniformly bounded above by $(K_I - a_I) / (z)^{\frac{1}{1-\nu}}$, then, no bank will make any lending offer and the potential entrepreneur will be a worker. On the other hand, a firm starts up as a mature firm if and only if the entrepreneur has invested $a_I > K_I - \Gamma(0)(z)^{\frac{1}{1-\nu}}$ from his own funds.

The simple condition that sets $t_M^*(z, a_I)$ delivers important properties for entering firms. First, even if $\Gamma(\cdot)$ is non-monotone, the only relevant region is where $\Gamma(\cdot)$ is strictly increasing. Otherwise, at least one of the contracting parties, the bank or the entrepreneur could be made better off without detriment of the other by reducing t_M. Specifically, the optimal contract avoids excessively high time-to-maturities that lead to inefficiently low levels of consumption, investment, output and net-profits even if they also lead the bank to meet its break-even condition. Second, it determines whether the time-to-maturity is zero or positive. Notice that if $\Gamma(t_M)$ is negative at low values of t_M, then even an entrepreneur that invests a_I above K_I might have to wait a positive amount of time before achieving maturity.

Figures 1 and 2 illustrate the determination of the optimal contract. As all the figures in the paper, Figs. 1 and 2 are drawn using the parameter values set and explained in Sect. 4. They are also drawn for typical values of a and z under the distribution used in Sect. 4, and their precise values are immaterial to this illustration.

Figure 1 shows the function $P(\cdot)$ and the determination for $t_M^*(z, a)$. The blue solid line is the function $P(\cdot)$ and the green dashed horizontal line is $K_I - a_I$, the set-up cost of the firm minus the own investment of the entrepreneur. As indicated above, the function $P(\cdot)$ is non-monotone, and in this case, it crosses twice the line $K_I - a_I$. The lower of these crossings, a bit above 20 periods, is the equilibrium time-to-maturity $t_M^*(z, a)$.

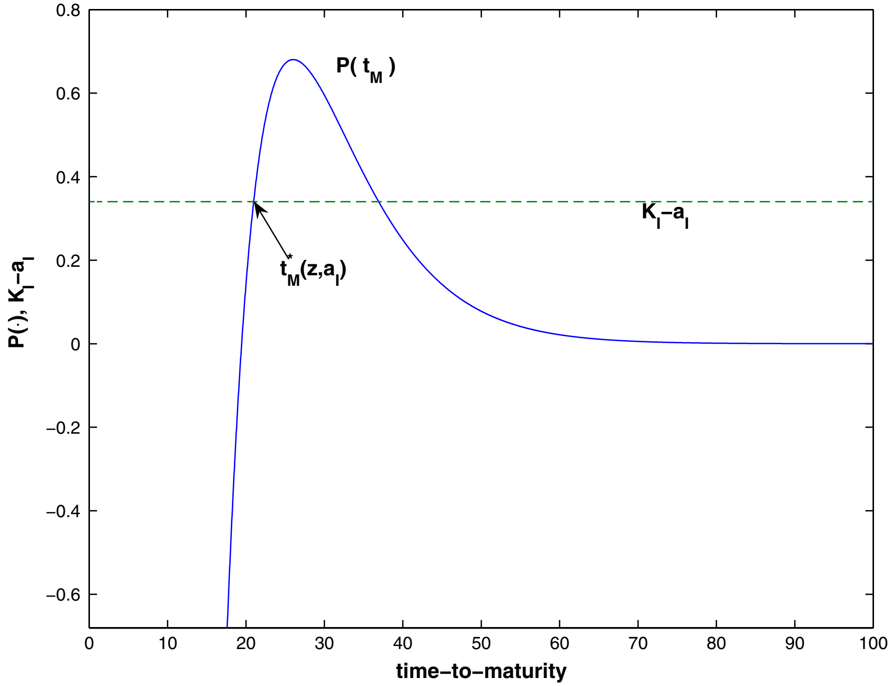

Fig. 1 Determination of the optimal time-to-maturity

Figure 2 illustrates the behavior of investment and entrepreneur's consumption (panel a) and output and installed capital (panel b). Before the firm achieves maturity, output, capital and investment increase over time. The entrepreneur is held to zero consumption, but his utility grows as the age of positive consumption gets closer. Upon achieving maturity, gross investment jumps down to—and remains forever at—the depreciation of capital. Output remains constant. Consumption is held positive at the minimum constant level that is consistent with the entrepreneur not running away with the mature level of capital k_M.

The previous characterization takes as given not only the initial own investment a_I but also whether a potential entrepreneur wants to be an active entrepreneur or a worker. However, even if the bank is willing to sign up an entrepreneur (and finance his firm), this does not guarantee that the contract will be signed. For the contract to be signed, the potential entrepreneur must also attain a higher level of utility as an active entrepreneur than as a worker. If so, he also chooses how much of his own assets to invest in the firm.

Consider an individual with assets a and a potential firm with productivity z. As a worker he would earn a flow of income equal w per period. Since $r \leq \rho$ he would borrow an amount w/r, consume $w/r + a$ right away, and repay the loan with his wages.[4]

[4] In the opposite extreme, if workers cannot commit to repay anything at all, the level of utility of a worker would be $U^W = w/\rho + a$. This alternative assumption, obviously, would not change any of the resulting characterization.

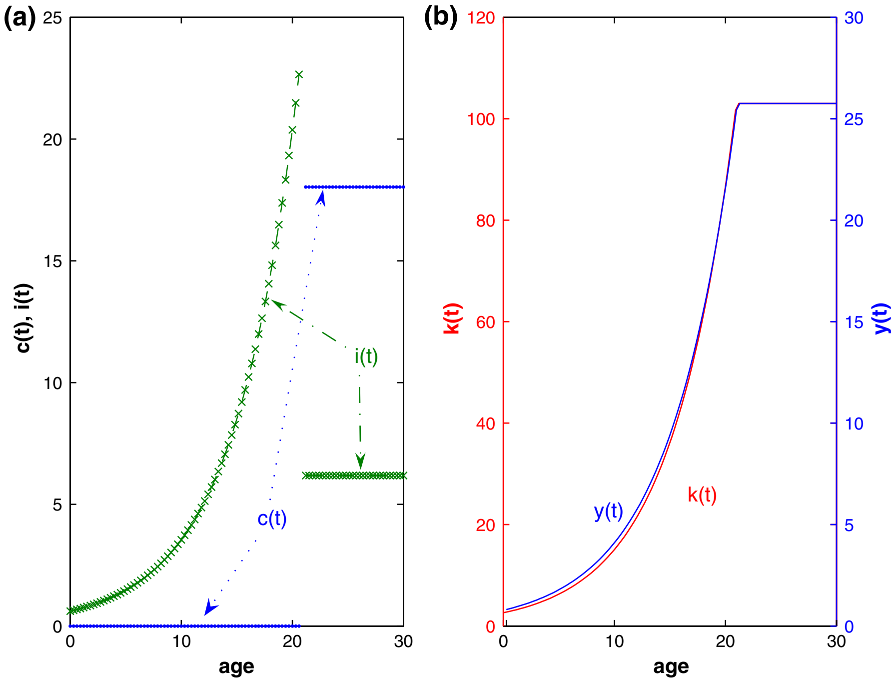

Fig. 2 **a** Investment and entrepreneur consumption; **b** capital and output of the firm

As an entrepreneur he would have to choose an amount of out-of-pocket investment a_I between 0 and a which impact's the contract that banks would offer to him. The comparison of these two alternatives determines whether the individual becomes an active entrepreneur.

Then, a potential entrepreneur with assets a and a project with productivity z would become an active entrepreneur if

$$U_0^E(z,a) \equiv \max_{a_I \in [0,a]} \left\{ (a - a_I) + e^{-\rho t_M^*(z,a_I)} \theta_D k_M \right\} \geq U_0^W(a) = \frac{w}{r} + a.$$

Notice that the objective function in the left-hand-side is quasi-linear. Therefore, conditional on (z,a) an entrepreneur would invest an amount equal to $a_I^*(z,a) = \min\{a, \tilde{a}(z)\}$ for some finite $\tilde{a}(z)$.[5] Therefore, the maximized value of the left-hand-side is also quasi-linear, implying that for each level of assets a there is a threshold $z_{\min}(a)$ on the firm's productivity, such that if $z \geq z_{\min}(a)$ the potential entrepreneur chooses to be an active entrepreneur and if $z < z_{\min}(a)$ he chooses to be a worker.

[5] The function $\tilde{a}(z)$ is implicitly defined by $-\frac{\partial t_M^*(z,a)}{\partial a}|_{a=\tilde{a}(z)} = [\rho \theta_D k_M]^{-1}$. For any finite z, the existence, uniqueness, and finiteness of $\tilde{a}(z)$ follow from the boundedness, continuity, and differentiability of $P(\cdot)$ which in turn imply that $t_M^*(z,a)$ is strictly decreasing and differentiable and $t_M^*(z,a) = 0$ if a is high enough.

3 Cross-sectional implications

In this section I examine in some detail the implications of limited enforcement on the cross-section of firms with different productivity levels z and entrepreneurs with different levels of financial assets a. The impact of z and a is essentially captured by the expression (20) that determines $t_M^*(z, a)$, the time-to-maturity. This simple condition captures an interesting complexity in the relationship between the productivity of the firm and the wealth of the entrepreneur. Essentially, if the entrepreneur can and does invest more than the set up cost, i.e. if $a \geq a_I > K_I$, then, the more productive a firm is, the longer it takes to mature. However, if $a_I < K_I$, then, the more productive a firm is, the shorter it takes to mature. The behavior of the time-to-maturity of firms is key to determining whether the highest productivity firms or the lowest productivity firms are the ones that are proportionally more compressed in their capital usage relative to their maturity level.

Since $t_M^*(z, a)$ is determined by the first crossing of $P(t_M)$ and $K_I - a_I$, then it is straightforward to examine the impact of a and z since each impacts only one of these curves. First, as a higher level of a_I obviously reduces $K_I - a_I$, then, since the relevant region of $P(t_M)$ is its strictly increasing region, then, obviously, the first crossing would be strictly lower with a higher value of a_I. Since entrepreneurs with higher a can put up a higher amount a_I, then $t_M^*(z, a)$ is decreasing in a. To collect a lower investment, a bank needs less time of fully collecting the flow of profits of the firm.

The impact of z is more interesting. First, recall that the function $P(t_M)$ that gives the operational profits for the bank is multiplicatively separable into a function $\Gamma(t_M)$, that depends only on the time-to-maturity, and the function $(z)^{\frac{1}{1-\nu}}$ that depends only on the productivity of the firm. Also, recall that $\Gamma(t_M)$ can be initially negative, but then necessarily becomes positive and then goes to zero as t_M goes to $+\infty$. Notice that if $\Gamma(t_M)$, then a higher z reduces the value of $P(t_M)$ since it makes it more negative. In particular, if $a_I > K_I$, then $t_M^*(z, a)$ is the crossing of $P(t_M)$ of a negative value $K_I - a_I$. From this position, a higher z reduces $P(t_M)$, implying that the crossing of $K_I - a_I$ is at a higher value of t_M. Then, when the entrepreneur is investing above the set-up cost K_I, a higher productivity of the firm implies a longer time-to-maturity $t_M^*(z, a)$. In this case, $t_M^*(z, a)$ is increasing in z. On the other hand, if $a_I < K_I$, then $t_M^*(z, a)$ is the crossing of $P(t_M)$ of a positive value $K_I - a_I$, and a higher z increases $P(t_M)$, implying a shorter time-to-maturity $t_M^*(z, a)$. In this case, $t_M^*(z, a)$ is decreasing in z.

Figure 3 illustrates the determination and value of $t_M^*(z, a)$ for four pairs of (z, a), the pairs resulting from two productivity levels, $z_{\text{Low}} < z_{\text{High}}$ and of two asset levels, $a_{\text{Low}} < a_{\text{High}}$. An entrepreneur with high wealth and a low productivity project, $(z_{\text{Low}}, a_{\text{High}})$, borrows little from the bank to start operations, and after a few periods he would have repaid the loan to the bank. His firm will be mature quickly, in less than 2 years (see below the discussion of the time unit of the model). An entrepreneur with the same productivity of a project but lower assets, $(z_{\text{Low}}, a_{\text{Low}})$ borrows more, not only for the initial installed capital k_0 but also to cover part of the set up cost K_I. The higher loan requires many more periods of fully allocating the profits of the firm to repay the bank for the initial loan and the subsequent financing to increase the installed

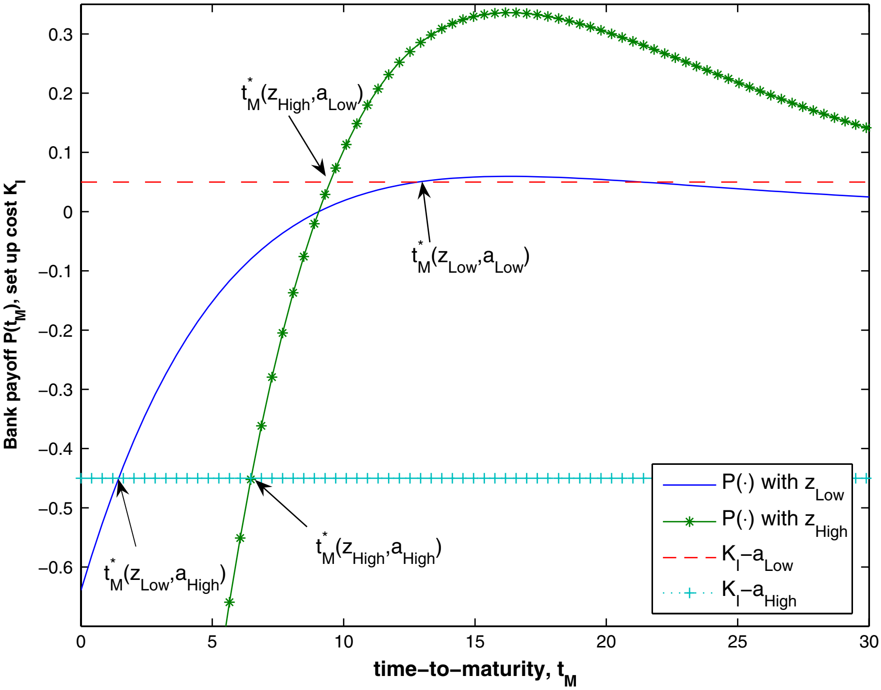

Fig. 3 Comparative statistics: optimal time-to-Maturity and initial assets and firm's productivity

capital of the firm. Another entrepreneur with the same assets as the last one but with a more productive project would have a shorter time-to-maturity. A more productive firm implies that at the time the firm matures, the consumption of the entrepreneur is higher. For any t_M, then the utility of not defaulting on the contract is higher and then the bank can allocate more capital to the control of the entrepreneur. This allows for a higher flow of profits of the firm and a shorter time-to-maturity.

Figures 4 and 5 show, respectively, the behavior of $t_M^*(a, z)$ for a wide range of asset levels a (for five selected levels of productivities z) and for a wide range of productivities z (for five selected levels of assets a.) Notice that $t_M^*(a, z)$ is always decreasing in a, but that for low values of z, increments in a decrease rapidly the value of $t_M^*(a, z)$ because low productivity firms require little investment. On the contrary, for high levels of productivity z, the entrepreneur's initial assets have a much smaller impact on the time to maturity $t_M^*(a, z)$. Also, notice that for low levels of assets, $t_M^*(a, z)$ is decreasing in z, while for higher levels of assets, $t_M^*(a, z)$ is lower but increasing in z.

Figures 4 and 5 also show a large variation in the values of $t_M^*(a, z)$. Also, notice that the curves are not drawn for all the productivity levels. The points in which the curves are not drawn are combinations of (z, a) for which the potential entrepreneur is better off being a worker, not an active entrepreneur. Albeit less vividly, the figures also show differences in entrepreneurial decisions. In the next two sections I explore the implications of the model for the evolution of the size distribution of firms and the aggregate output of the economy.

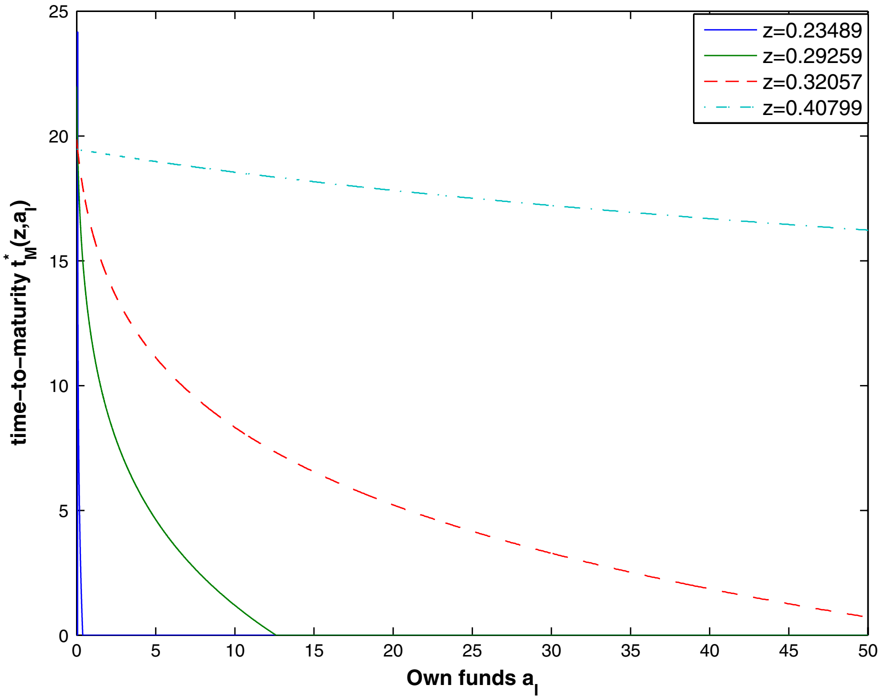

Fig. 4 Entrepreneur's initial assets and time-to-maturity, several productivities

4 Evolution of the firm size distribution

This section examines the implications of limited enforcement on the firm size distribution. First, I set up an OLG economy that is consistent with the contracting problem considered in the previous section. Within this OLG framework, I first consider the special case in which all potential entrepreneurs have the same level of financial resources a and the same productivity level z for their firms. This case can be solved analytically and leads to a Pareto distributed firm size distribution for the overall economy. Next, I allow heterogeneity in assets and productivities across new potential entrepreneurs and explore the evolution of the firm size distribution of each cohort of firms. A straight parameterization of the model shows that its quantitative implications can resemble the empirical observations on the evolution of the size distribution of firms reported in Cabral and Mata (2003).

4.1 Optimal contracts in an OLG environment

Consider now an economy with many cohorts of firms. Specifically, in each period a continuum with mass $\lambda > 0$ of new potential entrepreneurs is born. Each cohort is characterized by a constant c.d.f. F over non-negative levels of initial financial assets a and of productivity levels z of projects or potential firms that an entrepreneur can

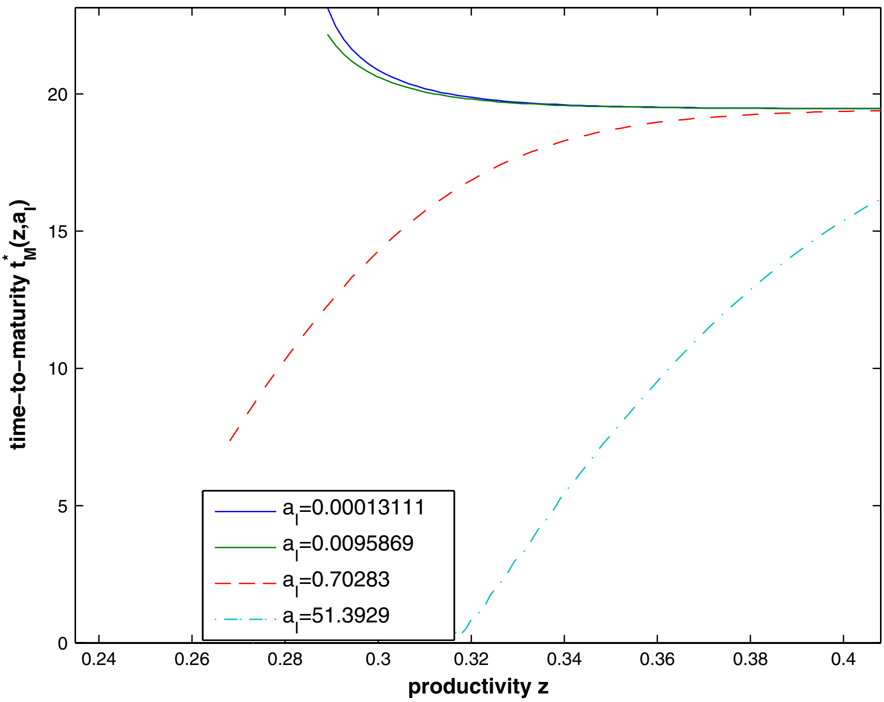

Fig. 5 Firm's productivity and time-to-maturity, various levels initial assets

control. Therefore, I restrict attention to identical cohorts of firms and examine their evolution over time.

In each period, a fraction λ of the existing entrepreneurs (and workers) die. Death probabilities are constant over the lifetime of all individuals. They are also independent of whether the individual is an entrepreneur or a worker, and independent of the individual's initial assets a and of the productivity z of his firm. Here, each individual, regardless of how old he is, has an expected lifespan of $1/\lambda$. Therefore, the probability distribution over the age $t \geq 0$ at which the firm dies and the cross-sectional age distribution in the economy are both described by the density

$$g\,(t) = \lambda e^{-\lambda t}. \tag{21}$$

For simplicity, I also assume that when an entrepreneur dies, the capital in his control explodes and the salvage value of the firm is zero. I also assume that banks are infinitely lived and that they face a constant interest rate $r_0 \geq 0$ at which they can borrow or lend. To be sure, I also assume that the enforcement parameter $\theta_D \geq 0$ is the same for all cohorts since it remains constant over time.

For the sake of the contract with an individual entrepreneur, the relevant interest rate for a bank is equal to the opportunity cost $r_0 \geq 0$ plus the constant risk λ that the resources invested in a firm disappear in that moment. Equally, the effective discount rate of entrepreneurs is equal to a pure discount rate $\rho_0 \geq 0$ and the probability λ that he dies. Setting $r = r_0 + \lambda$ and $\rho = \rho_0 + \lambda$, where $\rho_0 \geq r_0$, then the contract solved in Sect. 2 and further characterized in Sect. 3 is also the optimal contract between banks and entrepreneurs in this environment.

I now use this environment to study the behavior of the cross-sectional distribution of firms depending on the age of the cohort. Before looking at the general case, it is convenient to consider a special case that can be solved by hand.

4.2 A special case: ex-ante identical entrepreneurs

Assume now that all new potential entrepreneurs have identical initial levels of assets a and access to set up and control firms with the same level of productivity z. Therefore, in equilibrium either all or none of the potential entrepreneurs become active entrepreneurs. In the former case, all of them receive the same contractual terms and take the same amount of time to achieve maturity. Within cohorts, the size distribution is degenerate, but between cohorts firm sizes vary by age. To avoid trivialities, in this subsection I assume $t_M^*(z, a) > 0$. Moreover, to shorten notation, in this subsection I will suppress reference to (a, z).

Consider a level of capital k that is above $k_0 = k_M e^{-\rho t_M}$ but below k_M, which are the levels of capital with which firms start and achieve maturity, respectively. Let $G(k)$ denote the c.d.f. at that level of capital. Then, $G(k) = \Pr\left[e^{-\rho(t_M - t)} k_M \leq k\right] = \Pr\left[t \leq \frac{1}{\rho} \ln\left(\frac{k}{e^{-\rho t_M} k_M}\right)\right]$. Hence, the exponential distribution $g(t)$ for the age of existing firms implies that the density function for the distribution of installed capital across all firms in the economy is:

$$g_k(k) = \begin{cases} \left(\frac{\lambda e^{-(\lambda+\rho)t_M}}{\rho}\right)\left(\frac{k}{k_M}\right)^{-(1+\lambda/\rho)} & \text{if } k_M e^{-\rho t_M} \leq k \leq k_M, \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

a Pareto distribution with curvature parameter $1 + \lambda/\rho$ and a right-truncation point at k_M. At this truncation point, there is a probability accumulation equal to $e^{-\lambda t_M}$, i.e. the fraction of firms that survive long enough to attain maturity.

4.3 Ex-ante heterogeneous entrepreneurs

I now consider a richer environment in which firms not only differ because of their age t but also because of their productivity z and the initial assets a of their entrepreneurs. The amount of installed capital in each firm is given by the function

$$k(z, a, t) = k_M(z)\, e^{-\rho\left[t_M^*(z,a) - t\right]^+}. \tag{23}$$

Firms with the same productivity z have more capital if they are older (higher t) or if, at the time of start-up, the entrepreneur in control had a larger amount of financial wealth a.

Since all cohorts start with the same distribution F over (z, a), the size distribution of new firms is identical in every period. Equation 23 implies that this distribution moves as the cohort of firms ages. Indeed, unless $t_M^*(z, a) = 0$ for all a, z, the average capital installed grows until each surviving member of the cohort achieves maturity. The behavior of higher moments of the distribution, e.g., dispersion and skewness, depends on the behavior of $t_M^*(z, a)$.

In a recent paper, Cabral and Mata (2003), show that the distribution of the logarithms of firm size of a given cohort is very skewed to the right at time of birth, and gradually evolves towards a more symmetric distribution. Their findings suggest that it takes a long time before the firm size distribution of a given cohort approximates a log-normal distribution. Moreover, they show that the size distribution of publicly traded firms (which is to a large extent a symmetric distribution) is not representative of the firm size distribution of more comprehensive datasets (which is skewed to the right). Their main findings are mostly based on *Quadros de Pessoal,* a comprehensive dataset of Portuguese firms, but they argue that these features are also consistent with the data of other countries.

The rest of this section explores the ability of the model to replicate the findings of Cabral and Mata (2003). In particular, I want to examine whether the distribution of the logarithm of installed capital in the model has a behavior that replicates the behavior of the logarithm of firm sizes, measured as number of employees in Cabral and Mata (2003). In particular, I want to examine whether the model replicates a positive but declining skewness as the cohort of firms gets older. Instead of an exhaustive search for parameter values, my purpose is to examine whether a straight parameter configuration can do the job.

4.3.1 Parameter values

Table 1 presents all the parameter values—and the functional form for F-used in this section. Below I discuss my choices.

A unitary interval of time corresponds to one calendar year. For this I use standard values of $r_0 = 5.5\%$ for the bank's opportunity cost of financial resources and $\delta = 6\%$ for the depreciation rate of physical capital. To allow for slightly higher discounting on behalf of entrepreneurs, I assume $\rho_0 = 6.5\%$. Other values for these parameters within the same ballpark lead to very similar results.

I set $\theta_D = 1$ since it is a natural point. The standard usage in the literature of limited commitment, e.g. Alvarez and Jermann (2000), and Chapters 19 and 20 of Ljungqvist and Sargent (2004), is to assume that besides the punishment of permanent exclusion, no direct sanctions can be imposed.[6] In this spirit, setting $\theta_D = 1$ implies that besides exclusion from productive activities, repossessing part of the capital cannot be used to discipline a defaulting entrepreneur. Local reductions in the value of θ_D do not lead to significant changes in the results as shown in the exercises of the next section. Obviously, values of θ_D close to zero move the economy towards full enforcement, leading

[6] Kehoe and Levine (1993) is an important early exception.

Table 1 Parameterization of the model

Definition	Symbol	Value/functional form
Opportunity cost of funds, banks	r_0	5.5%
Pure discount rate, entrepreneurs	ρ_0	6.5%
Depreciation, physical capital	δ	6.0%
Birth/death rate of firms	λ	11.1%
Fraction of capital stolen if default	θ_D	1.00
Span-of-control/decreasing returns	ν	0.94
Set up costs of a firm	K_I	0.00
Earnings flow of workers	w	0.00
Exogenous c.d.f. over (z, a)	F	$\begin{pmatrix} \ln(z) \\ \ln(a) \end{pmatrix} \sim N(\mu, \Sigma)$
$E[\ln(z)]$	μ_z	−1.18
$E[\ln(a)]$	μ_a	−1.00
SD $[\ln(z)]$	σ_z	0.09
SD $[\ln(a)]$	σ_a	0.10
cov $[\ln(z), \ln(a)]$	σ_{za}	0.00

to very short time-to-maturities and cohort and overall distributions of firm sizes that are directly pinned down by the exogenous marginal distribution of productivities z.

The model is highly parameterized and, more than a formal quantitative analysis, the purpose of this section is to illustrate the operation of the model. With this in mind, in this section I will abstract from the extensive margin of entrepreneurship, i.e. I will look at an equilibrium in which all potential entrepreneurs end up being active entrepreneurs. To this end, I assume away the set-up cost and the option for individuals to be workers, i.e. $K_I = w = 0$. In this way, the assumed distribution for the productivities pins down completely the size distribution of mature firms. At the cost of realism, this assumptions allows us to appreciate more easily the impact of limited commitment on the size distribution of younger firms. A positive value for either K_I or w of these parameters automatically eliminates the activation of the lower tail of the distribution of productivities, which would be an added force for right skewness in the size distribution of young firms. In the next section I introduce small but positive values for both K_I and w in order to study the impact of changes in θ_D in the aggregate level of entrepreneurial activity.

An important parameter for the analysis is ν, the degree of decreasing returns on a firm's technology. The most commonly used value is $\nu = 0.85$ as in Atkeson and Kehoe (2005), but values closer to $\nu = 0.9$ are also used, as in Cagetti and De Nardi (2006), (2009). The literature admits a large degree of uncertainty regarding this parameter. The value of this parameter cannot be calibrated from data on the size distribution of firms since this would require the unfeasible task of separately identifying the underlying distribution of productivities z. On the other hand, inferring the value of ν from income shares involves the elusive task of allocating some of the proprietor's income between labor and capital services.

In the context of my model, the parameter ν governs the length of time-to-maturity of firms. To see this, notice that the capital-to-output ratio of mature firms is equal to $\nu/(r_M + \delta)$, i.e., proportional to the value of ν. Hence, the larger is the value of ν , the larger is the amount of capital that an entrepreneur must control for each unit of output and the longer it will take the contract with the bank to relax the participation constraints so that he can control it without triggering default. The results in Cabral and Mata (2003) suggest that in the data, time-to-maturities are quite long, above 25 years. I use a value $\nu = 0.94$, which is a reasonable compromise between the data requirements suggested by Cabral and Mata (2003) and the standard values in the literature.

The rest of the parameters are set using data conveyed in Cabral and Mata (2003). They report that out of 2,651 firms identified as new firms in 1984, only 1,031 were still active in 1991. The implied value for a (constant) annual death probability for these 8 years is equal to 11, 1%. Therefore, I simply set $\lambda = 11.1\%$.

For the exogenous distribution F of assets a and firm productivities z , I assume a joint log-normal distribution with mean $\left[\mu_z, \mu_a\right]'$ and variance-covariance matrix Σ. Despite the limitations that will become clear below, I choose this distribution for its simplicity and parsimony. Moreover, given the preponderance of the log-normal in the literature on the firm size distribution, it is a natural starting point. Indeed, using this distribution allows for a more vivid and transparent grasp of the skewness induced by the limited enforcement in the financial contracts.

I use the statistics on the firm size distribution of Cabral and Mata (2003) to calibrate the parameters of the model. Note however, that Cabral and Mata measure firm sizes using employment while in this paper I assume that capital is the only variable input of production. However, recall that with competitive markets and no financial frictions, the equilibrium of a multifactor model leads to the equalization of the marginal products of all inputs across all firms. Moreover, with Hicks-neutral productivity differences but a common Cobb-Douglas production function for each firm (with decreasing returns to scale), input ratios will also be equalized across all firms. Hence, except for the arbitrary units of measurement, facts about the size distribution of one input carry over the other size distribution of all other inputs. Furthermore, as long as the participation constraints do not distort the mix of inputs, e.g. when defaulting entrepreneurs have the same ability to appropriate the different factors of production or if instead of inputs he can appropriate output, this result also holds for economies with limited commitment.[7]

With these conditions in mind, I set the mean μ_z and standard deviation σ_z of the logarithm of the productivity z, so that the implied mean and standard deviation of the logarithm of installed capital in mature cohorts coincide with the mean and standard deviation of the logarithm of the number of employees in firms of 30 years or more as reported by Cabral and Mata (2003) for the *Quadros de Pessoal* dataset. I set the mean μ_a and standard deviation σ_a of the logarithm of the initial financial assets of potential entrepreneurs so that the implied mean and standard deviation of installed capital for new firms, $k(z, a, 0) = k_M(z, a)\, e^{-\rho t^*_M(z,a)}$, roughly coincides with the mean and standard deviation of the logarithm of the number of employees of firms

[7] See Quintin (2001) for a model where constrained firms become more labor intensive than unconstrained firms.

with less than 1 year in Cabral and Mata (2003). Lastly, to avoid using the covariance σ_{za} as an extra free parameter, I simply set it to zero.

4.3.2 Results

Figure 6 presents the evolution of the mean, standard deviation, and skewness coefficient of the logarithm of the size in the model and in the data.[8] While the evolution of those moments in the model can be computed for any arbitrary age (blue solid lines), the estimates in the data (green dashed lines) are available only for the following sparse groups of firms, in years of age: $\leq$1, 2–4, 5–9, 10–19, 20–29 and $\geq$30. I use a mid-point in each interval, marked with a star, to refer to the corresponding age group. For the two extreme groups, I use the lower bound of the age interval.

It is not an achievement of the model that its mean and standard deviation for new and mature firms roughly coincide with those in the data. As I already explained, this is driven by the calibration of μ_z, σ_z for the mature cohort, and given those, the choice of μ_a, σ_a to roughly fit the first two moments of the entry cohort. The interesting finding is that, for a large interval of early ages, the skewness coefficient implied by the model roughly coincides with those in the data. This is interesting because there is nothing in the procedure to set parameter values that forced this close resemblance.

The model produces skewness in the distribution of the logarithms of firm sizes from the following mechanism. Under the parameter values used, $t_M^*(z, a)$ is decreasing in the productivity of the firm. In this case, as explained in Sect. 3, the initial capital usage of less productive firms is restricted proportionally more than the capital of more productive firms. Also, under this parameter configuration, the higher the productivity of the firm and the internal financing from the entrepreneur, the sooner the firm achieves maturity. Eventually, all surviving firms achieve maturity since the highest $t_M^*(z, a)$ is bounded.

The skewness of older firms in the data and of mature firms in the model diverge significantly. In particular, since all potential firms are activated, the model's size distribution of mature firms is a log-normal, and therefore, by construction, the distribution of the logarithm of the size will be symmetric, i.e. with zero skewness. In contrast, as discussed by Cabral and Mata (2003), the distribution of firms as old as 30 years still exhibits positive skewness. This is an inherent limitation of using the log-normal distribution for productivities z combined with $K_I = w = 0$. I stick to those assumptions since more than looking for the model's best shot at the data, the purpose of this exercise is to illustrate, as clearly as possible, the workings of the model.

Note also that the model implies a faster movement towards symmetry (faster fall in skewness) than what seems to be observed in these groupings of cohorts. Note also

[8] I did not have access to the sample mean, standard deviation or skewness coefficient of the logarithm of the number of employees for the firms in the Quadros de Pessoal datasets. The values presented here were obtained numerically, on the basis of the estimates of the parameters (μ, σ, q) that Cabral and Mata (2003) report for an extended generalized gamma distribution (EGGD) fitted for different cohorts of firms. As an exception, I use directly the value of 0.8496 for the skewness in the case of new firms in 1984 that they report in a working version of the paper. Thus, the exercise here should be seen as an illustration of the workings of my model and not as an empirical fitting exercise on the Quadros de Pessoal data.

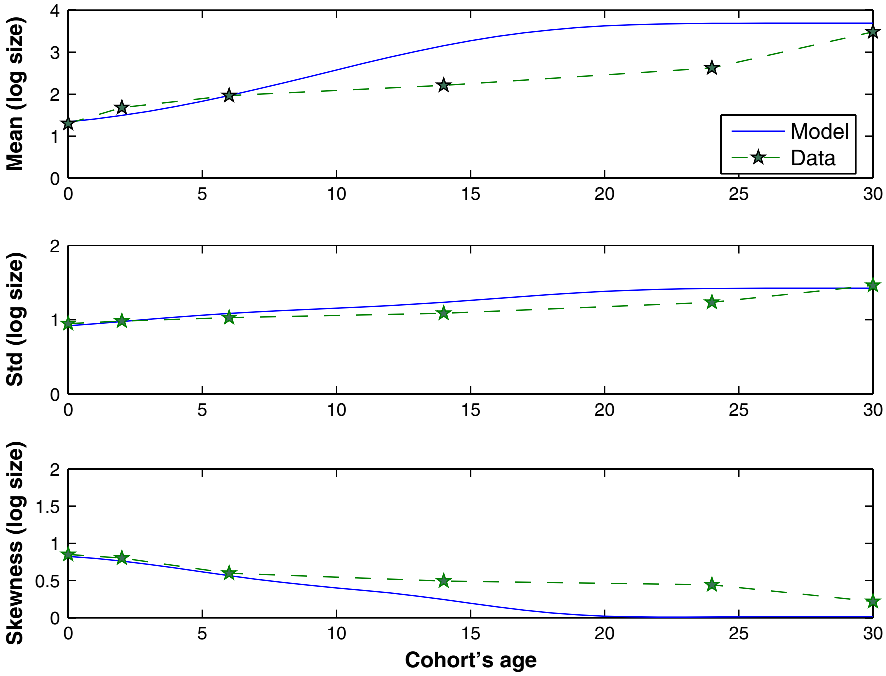

Fig. 6 Evolution of cohort firm size distribution: moments

that for middle aged cohorts, the means and standard deviations in the model are higher than in the data. The model implies a faster average growth and a faster increment in the dispersion, despite my using a high value of ν. At this point, my conjecture is that the analytical tractability bought by assuming risk neutral entrepreneurs comes at a price of implying that firms achieve maturity too fast with respect to what we might observe in the data. In the conclusions I discuss this point further.

Figure 7 illustrates further, however, the positive outcome of this exercise: With a simple and straight parameterization of the model, it comes surprisingly close to the main findings of Cabral and Mata (2003) in that: (i) the distribution of the logarithms of firm size of a given cohort is very skewed to the right at time of birth, and (ii) the movement towards a symmetry (in logs) is very slow and gradual. The figure shows a semilog plot (in the horizontal axis) of the model's firm size distribution for a cohort of new entrants, and cohort 5, 10, and 30 year old firms. Clearly, as firm cohorts age, their size distribution slowly moves from a markedly right-skewed distribution towards a symmetric one to which it eventually converges.

5 Entrepreneurship and aggregate output

In this final section I show how the model separates three margins for the consequences of limited enforcement on the level of entrepreneurial activity, output, and utility. I also provide a simple numerical illustration.

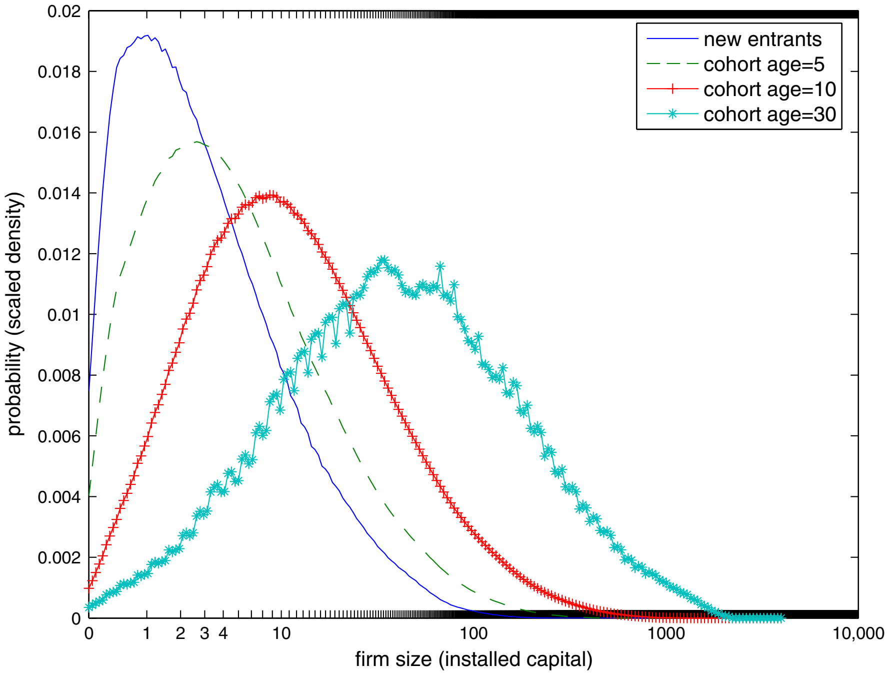

Fig. 7 Evolution of the firm size distribution by age: densities

Restricting attention to stationary economies, the aggregate variables of interest can be expressed directly in terms of the time-to-maturity function $t^*_M(z, a)$ and the indicator function $\chi^E_{(z,a)}$ of whether an individual with initial assets a and a project with productivity z becomes an entrepreneur ($\chi^E_{(z,a)} = 1$) or not ($\chi^E_{(z,a)} = 0$). The aggregate output Y^E produced by the entrepreneurs of all ages t, initial assets a, and productivities z is given by:

$$Y^E \equiv \int_0^\infty \int_{Z\times A} \chi^E_{(z,a)} z f\left[k(a, z, t)\right] \lambda e^{-\lambda t} dF(z, a)\, dt. \tag{24}$$

Using the Cobb-Douglas functional form of $f(\cdot)$, the expression (23) for $k(a, z, t)$, and integrating for each (a, z), the expression for Y^E becomes:

$$Y^E = \int_{Z\times A} \chi^E_{(z,a)} \left[\frac{\rho\nu e^{-\lambda t_M(z,a)} - \lambda e^{-\rho\nu t_M(z,a)}}{\rho\nu - \lambda}\right] z\left[k_M(z)\right]^\nu dF(z, a). \tag{25}$$

This formula shows that, with respect to an economy with full enforcement, aggregate output is lower because of three margins:

(i) *Entry*: Some productive firms might not be created because of limited commitment. This is captured by $\chi^E_{z,a} \le \chi^U_z$ for all (z, a) and $\chi^E_{z,a} < \chi^U_z$ for some (z, a), where χ_z^U is the indicator function of whether z is above $z_{\min}^U$ and the firm would have been created if there were full enforcement.

(ii) *Firm growth*: Entering firms might operate below their maturity level $k_M(z)$ for some time until they relax the participation constraints to allow for control of a higher amount of capital. This is captured by the fact that $\left[\frac{\rho \nu e^{-\lambda t_M(z,a)} - \lambda e^{-\rho \nu t_M(z,a)}}{\rho \nu - \lambda}\right]$ is only equal to 1 when $t_M(z,a) = 0$ and below one when $t_M(z,a) > 0$.

(iii) *MPK of mature firms*: Even if firms achieve maturity (stop growing) the effective interest rate r_M is above the bank's interest rate r, if $\theta_D > 0$ and $\rho > r$, and hence $k_M < k^U$ and $y_M < y^U$. Without the temptation to default, output would be increased since the marginal product of capital is greater than $r + \delta$.

The overall output of the economy Y is equal to Y^E, the output from the entrepreneurs, plus $Y^W \equiv w \int_{Z\times A} \left[1 - \chi_{(z,a)}^E\right] dF(z,a)$, the output generated by workers. Following the same steps as for Y^E the aggregate capital in the economy is

$$K = \int\limits_{Z\times A} \chi_{(z,a)}^E \left[\frac{\rho e^{-\lambda t_M(z,a)} - \lambda e^{-\rho t_M(z,a)}}{\rho - \lambda}\right] k_M(z)\, dF(z,a).$$

Relative to the socially optimal entrepreneurial activity (i.e. under full-enforcement), the aggregate entrepreneurship E in the economy with limited enforcement is:

$$E = \frac{\int_{Z\times A} \chi_{(z,a)}^E dF(z,a)}{\int_{Z\times A} \chi_z^U dF(z,a)}. \tag{26}$$

The level of contractual enforcement θ_D also impacts the equilibrium utility of entrepreneurs. Recall that $U_0^E(z,a)$ and $U_0^W(a)$ denote the age 0 utility of a potential entrepreneur with productivity and assets (z,a) that chooses, respectively, to be an active entrepreneur or a worker. The average utility of effective entrepreneurs, U^{EE}, is defined as

$$U^{EE} = \frac{\int_{Z\times A} \chi_{(z,a)}^E U_0^E(z,a)\, dF(z,a)}{\int_{Z\times A} \chi_{(z,a)}^E dF(z,a)}. \tag{27}$$

A problem in interpreting this indicator is that the group of active entrepreneurs change with the value of θ_D. An indicator that avoids this problem is the average utility of all the individuals who would be active entrepreneurs if there were full enforcement, even if they end up being workers under limited commitment. Indicating this average by U^{SE}, its formula is

$$U^{SE} = \frac{\int_{Z\times A} \chi_z^U \max\left\{U_0^E(z,a), U^W(a)\right\} dF(z,a)}{\int_{Z\times A} \chi_z^U dF(z,a)}. \tag{28}$$

Figures 8 and 9 display, respectively, the value of the aggregates Y, K and E and the average utilities U^{EE} and U^{SE} for different enforcement levels. In the horizontal

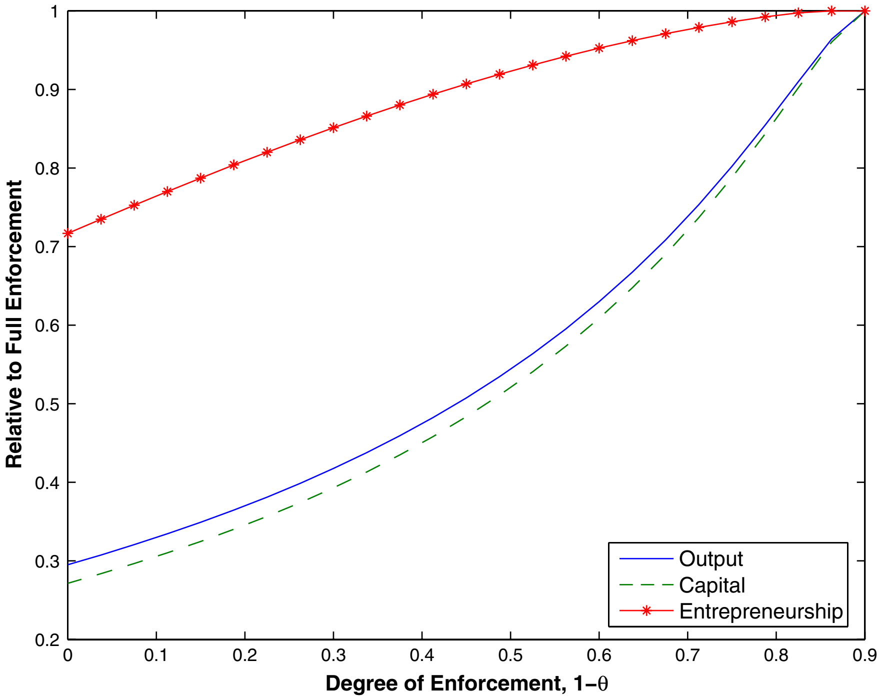

Fig. 8 Aggregate variables under different degrees of contract enforcement

axis, $1 - \theta_D$, indicates the degree of contractual enforcement, where 0 is the case in which the only punishment mechanism is exclusion while 1 is the case in which no fraction of the capital can be held up. To allow for changes in E, I change the parameter values of the previous section, assuming a very small set-up cost, $K_I > 0$, and equal to 0.05% of the flow of unrestricted profits of an entrepreneur with the average productivity $z = e^{\mu_z + \frac{\sigma_z^2}{2}}$. I also assume an opportunity cost of not earning wages as a worker, $w > 0$, and equal to 20% of the earnings of an entrepreneur with the average level of productivity. While those numbers are reasonable, I make no claim about disciplining their choice on the basis of any calibration scheme, and my only purpose is to illustrate the response of the economy to changes in θ_D. All other parameter values are equal to the ones used in the previous section.

As expected, Fig. 8 shows that economies with better contract enforcement have a higher level of entrepreneurial activity and aggregate output. Economies with the lowest enforcement, $\theta_D = 1$, only activate 70% of the socially productive projects available. The potential entrepreneurs of those projects are inefficiently allocated to the less productive labor activities.

These exercises illustrate that limited enforcement can have a significant impact in output and capital usage. Economies with the lowest enforcement, $\theta_D = 1$, produce less than 30% of the potential output. The amount of capital used is barely above 27% of the potential usage of capital if the economy would have exhibited $\theta_D = 0$. While

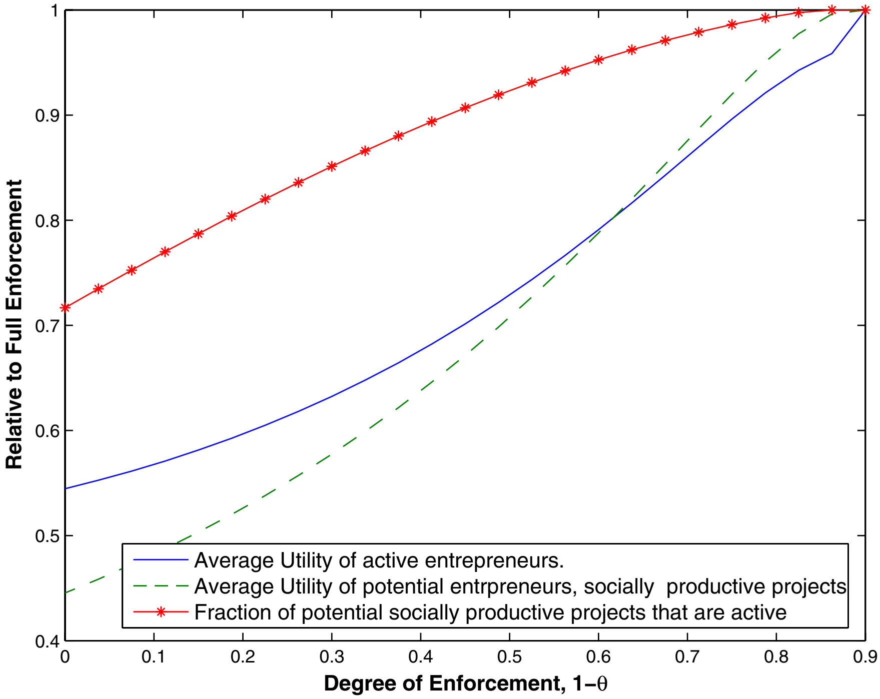

Fig. 9 Entrepreneur's entry utility under different degrees of contract enforcement

an important fraction in these differences comes from the entry margin, most of the output and capital differences arise from the slow firm growth (margin ii) and from the depression in the size of mature firms, since their effective interest rate r_M is above the social cost of capital r.

The degree of contract enforcement can also have a large impact on the equilibrium utility of entrepreneurs. Even if with bad contract enforcement only the most productive firms are activated, the conditional average utility of these selected entrepreneurs U^{EE} is only 55% of what it would have been under perfect enforcement. The differences are even larger when the comparison is done fixing the pool of entrepreneurs to projects that are socially efficient (i.e. active under full enforcement). The value of U^{SE} with $\theta_D = 1$ is less than 45% the value $\theta_D = 0$.

6 Concluding remarks

This paper studies the optimal infinite horizon contract between a bank and an entrepreneur that cannot commit not to default and appropriate the capital under his control. I show that the problem boils down to simply determining the equilibrium time-to-maturity, i.e. the time it takes for a firm to stop growing and for the entrepreneur to receive dividends. The focus of the paper is in exploring the cross-sectional implications of the model of firms with different productivity levels and entrepreneurs with different financial assets and to use those implications to study the behavior of economies with many firms.

The paper exploits the simplicity of the optimal contract to answer two questions. First, it investigates whether the model can replicate empirical observations on the behavior of the firm size distribution. Second, it investigates the impact of limited enforcement on the entrepreneurial activity, output, and welfare on the aggregate economy. consequences of limited enforcement. On the first, it is found that with a simple and straight parameterization, the model replicates the observation that the distribution of the firm sizes of a given cohort is very skewed to the right at time of birth, and that it takes a long time for the cohort to exhibit a log-normal distribution. On the second, the model suggest that the impact of limited enforcement is large, especially in output. Beyond a reduced entrepreneurial activity, the aggregate output is significantly reduced by firms that take time to mature, and, moreover, that even at maturity their size is inefficiently low.

It must be stressed that the added value of the paper must not be seen as a quantitative one. The analytical results derived in the paper, however, can provide the basis for a more complete and yet analytically tractable quantitative framework. To this end, the model must be relaxed along several dimensions. As a final reflection, I want to suggest three of them. First, it would be interesting to embed the model in a general equilibrium framework that goes beyond the simple determination of interest rates and incorporates the impact of aggregate entrepreneurial activity on non-entrepreneurs. Second, it would be interesting to allow for concavity in the utility of the entrepreneur. With such an extension, entrepreneurs would have positive consumption before the firm achieves maturity, and hence the times-to-maturity might be longer. Third, it would be interesting to adapt the work of Cagetti and De Nardi (2006), (2009) and endogeneize the distribution of assets of entering entrepreneurs as the result of bequest decisions.

References

Albuquerque, R., Hopenhayn, H.A.: Optimal lending contracts and firm dynamics. Rev Econ Stud **71**, 285–315 (2004)

Alvarez, F., Jermann, U.: Efficiency, equilibrium, and asset pricing with risk of default. Econometrica **68**, 775–797 (2000)

Atkeson, A., Kehoe, P.J.: Modeling and measuring organization capital. J Polit Econ **113**, 1026–1053 (2005)

Banerjee, A.V., Newman, A.F.: Risk-bearing and the theory of income distribution. Rev Econ Stud **58**, 211–235 (1991)

Buera, F.: Persistency of Poverty, Financial Frictions, and Entrepreneurship, Working Paper, Northwestern University

Cabral, L., Mata, J.: On the evolution of the firm size distribution: facts and theory. Am Econ Rev **93**, 1075–1090 (2003)

Cagetti, M., De Nardi, M.: Entrepreneurship, frictions, and wealth. J Polit Econ **114**, 835–870 (2006)

Cagetti, M., De Nardi, M.: Estate taxation, entrepreneurship, and wealth. Am Econ Rev **99**, 85–111 (2009)

Cooley, T., Quadrini, V.: Financial markets and firm dynamics. Am Econ Rev **91**, 1286–1310 (2001)

Cooley, T., Marimon, R., Quadrini, V.: Aggregate consequences of limited contract enforceability. J Polit Econ **112**, 817–847 (2004)

Guner, N., Ventura, G., Yi, X.: Macroeconomic implications of size-dependent policies. Rev Econ Dyn **11**, 721–744 (2008)

Hart, O., Moore, J.: A theory of debt based on the inalienability of human capital. Quart J Econ **109**, 841–879 (1994)

Holmes, T.J., Schmitz, J.A. Jr.: A theory of entrepreneurship and its application to the study of business transfers. J Polit Econ **98**, 265–294 (1990)

Kehoe, T., Levine, D.: Debt-constrained asset markets. Rev Econ Stud **60**, 865–888 (1993)

Kocherlakota, N.: Implications of efficient risk sharing without commitment. Rev Econ Stud **63**, 595–609 (1996)

Lazear, E.P.: Entrepreneurship. J Labor Econ **23**, 649–680 (2005)

Ljungqvist, L., Sargent, T.: Recursive Macroeconomic Theory. Cambridge: MIT Press (2004)

Lucas, R.E. Jr.: On the size distribution of business firms. Bell J Econ **9**, 508–523 (1978)

Monge-Naranjo, A.: Financial Markets, Creation and Liquidation of Firms and Aggregate Dynamics, Ph.D. Dissertation, The University of Chicago (2009)

Quintin, E.: Limited enforcement and the organization of production. J Macroecon **30**, 1222–1245 (2008)

Restuccia, D., Rogerson, R.: Policy distortions and aggregate productivity with heterogeneous establishments. Rev Econ Dyn **11**, 707–720 (2008)

Townsend, R.: Optimal multiperiod contracts and the gains from enduring relationships under private information. J Polit Econ **90**, 1166–86 (1982)