WORKING PAPERS

RESEARCH DEPARTMENT

Satyajit Chatterjee
Federal Reserve Bank of Philadelphia
and
Esteban Rossi-Hansberg
Princeton University

June 2007

Spin-offs and the Market for Ideas[*]

Satyajit Chatterjee   Esteban Rossi-Hansberg

Federal Reserve Bank of Philadelphia   *Princeton University*

June 2007

Abstract

We propose a theory of firm dynamics in which workers have ideas for new projects that can be sold in a market to existing firms or implemented in new firms: spin-offs. Workers have private information about the quality of their ideas. Because of an adverse selection problem, workers can sell their ideas to existing firms only at a price that is not contingent on their information. We show that the option to spin off in the future is valuable so only workers with very good ideas decide to spin off and set up a new firm. Since entrepreneurs of existing firms pay a price for the ideas sold in the market that implies zero expected profits for them, firms' project selection is independent of their size, which, under some assumptions, leads to scale-independent growth. The entry and growth process of firms in this economy leads to an invariant distribution that resembles the one in the US economy.

1   Introduction

The generation and implementation of ideas shape industry dynamics and the structure of firms. Ideas can be generated in many different contexts, but many important innovations have been developed by workers of existing firms. In some cases, workers decide to sell their idea to their

current employer (or to some other firm), and in other cases, they use their innovation to start a new firm: a spin-off. Whether an innovation by a worker is implemented within the firm, leads to a spin-off, or is discarded depends on the initial ownership and knowledge of the idea, as well as the profits that the different entities can generate by implementing it. In this paper we propose a theory of these decisions and, therefore, of firm dynamics and entry. The theory we propose takes a clear stance on the determinants of the boundary of the firm by modeling which ideas are implemented within the firm and which ones are not. Our take is that private information on the expected returns of an idea leads to an adverse selection problem in which the best ideas give workers incentives to set up new firms, while ideas that are not as good are sold to existing firms at a price that is not contingent on private information but implies zero expected profits for existing firms. Hence, we propose a theory that underscores the private information of innovators as a fundamental determinant of industry dynamics and spin-offs.

In our theory, the selection process induced by private information determines whether an idea is used within an existing firm or exploited by setting up a new firm. The innovator has information on the mean return of an idea that he cannot credibly convey to firm management. Hence, the firm is willing to pay only a non-contingent price for the project. The inventor can either decide to create a new firm or sell the project in the market for ideas. If he does the latter, he can credibly reveal his information about the project, since there is nothing at stake for him. The firm can then decide to implement the idea or not. Implementing a project, either in a spin-off or within the firm, requires one unit of labor. The reason is that to find out the realized return of an idea – that comes from a distribution with mean given by the innovator's private information – the idea has to be tried out for one period. The actual realized profit from an idea can be low enough to make further use of the project non-optimal for the entrepreneurs. In that case, the project will be dropped next period. If the return on the idea is high enough, the idea is used and provides a permanent and constant source of profits to entrepreneurs.

If the worker decides to spin off with an idea, we show that he will exit and return to an existing firm as a worker unless the idea pays a return that is strictly higher than an amount that exceeds the unit labor cost. The reason is that by continuing to use the idea, the worker forgoes the opportunity of spinning off with an even better idea in the future. Consequently, workers who decide to spin off are more selective than existing firms when deciding whether to drop a project. In turn, having this option, which has a positive value, leads workers to be more selective than existing firms in deciding which ideas to try out. That is, existing firms implement ideas with lower expected value than the ideas workers use to spin off. In other words, workers decide to forgo their

2

wage and the price at which they can sell their idea – if they do not spin off – only if the idea is particularly good. Hence the best ideas will result in spin-offs and the not-so-good ideas will be sold to existing firms. In this sense, our theory determines entry through spin-offs and firm growth.

Most of the previous efforts to study firm dynamics have taken a different approach. The seminal works of Jovanovic (1982), Hopenhayn (1992), and Ericson and Pakes (1995) study firm dynamics that result from an exogenous stream of productivity levels. So do more recent papers like Luttmer (2006), Klette and Kortum (2004), and Rossi-Hansberg and Wright (2006), although these papers do endogenize the characteristics of entrants. Firms do not determine their productivity; they act conditional on it or on what they know about it. In contrast, what we propose is a theory in which the quality of the projects that firms implement, as well as the quality of the projects of new entrants (spin-offs), is determined endogenously by the selection of projects. This mechanism leads to a theory of firm behavior in which firm dynamics are determined by the quality and quantity of the implemented projects. Franco and Filson (2006) also provide a theory of spin-offs but focus on imitation of existing ideas by workers. Silveira and Wright (2007) study the market for ideas in the context of a search model and analyze its effects on wages and employment. None of these studies focus on private information and adverse selection as a determinant of entry through spin-offs and the size distribution of firms, which is the main focus of this paper.

We show that the set of ideas that entrepreneurs are willing to buy at the non-contingent equilibrium price does not depend on the size of firms. Hence, the selection of projects is independent of size, where size is measured as the number of projects (or workers) in a firm. The equilibrium price of an idea is such that any entrepreneur in the economy obtains zero expected profits from buying an idea in the market, given the set of ideas sold by workers (the ones that do not lead to spin-offs). Hence, entrepreneurs are indifferent about whether to buy ideas at the market price.

In order for the model to have predictions on firm growth, and since entrepreneurs are indifferent about how many ideas to buy in the market, we need to specify the rate at which entrepreneurs buy ideas. We do this by assuming that they find out about ideas for sale in proportion to the size of the firm. This may be because the workers in the firms have ideas themselves or because they provide information to the entrepreneur about ideas for sale. Independent of what the reason is, the key argument is that the market for ideas implies that entrepreneurs are indifferent about the number of projects to buy and so the selection of a project does not depend on the characteristics of the firm. In this sense, it is the market for ideas that leads to scale-independent firm growth and, together with entry, to a realistic size distribution of firms. We show that the upper tail of the size distribution can be arbitrarily close to a Pareto distribution with coefficient 1.

The theory also has predictions for the entry process of firms. Large firms generate more spin-offs than small firms, although as a fraction of the workforce, the number of spin-offs is constant. Only the best ideas lead to spin-offs. The model also results in a fraction of unsuccessful spin-offs. Only these firms exit the economy. Large firms can have unsuccessful projects, too, but they just drop the project. They do not exit, since they have at least one ongoing project that provides a permanent source of profits.

In the model presented we assume that workers can be hired at a fixed wage, independent of the size of the industry. This may be an appealing assumption for small industries. Growth of larger industries, however, is likely to have effects on labor and capital markets, in particular, the market for specialized skills and industry- or sector-specific capital. Adding a market for capital and labor would tend to decrease and eventually stop the development of a particular industry. This general equilibrium problem is interesting and important, but in this paper, we focus only on the characteristics of industry dynamics.

There are many famous examples of successful firms that began with one employee or a group of employees quitting a firm and starting their own firm. This process of separating from a parent firm and starting one's own firm seems particularly prevalent in innovative industries. It is this process our model is designed to explain. Our theory does not make a deep distinction between spin-offs and startups – that is, new entrants not necessarily associated with an incumbent firm. Nevertheless, we prefer to think of new entrants as spin-offs because all new entrepreneurs are workers in some firm and, in our theory, only these agents have new ideas.

Our results imply that the best ideas are the ones that lead to spin-offs, while the ideas that are not as good are implemented in existing firms. This implies that, on average, profits from the first product of a firm should be higher than the profits from the subsequent products a firm decides to produce. This is consistent with some of the available evidence, which suggests that the first product of a firm is, on average, the most successful of its products. Prusa and Schmitz (1994) argue that this is the case in the PC software industry. The first product of a firm sells, on average, 1.86 times the mean product in its cohort, while the second product sells only 0.91 times the mean product in its cohort. That is, first products are, on average, about twice as successful as second products. The first product is also about twice as successful as the third, fourth, and fifth products. This evidence suggests that spin-offs discriminate more than incumbent firms in choosing which projects to implement. This is exactly in line with the selection mechanism our theory underscores.

The rest of the paper is organized as follows. Section 2 describes the model and proves some

basic results. Section 3 characterizes the selection of projects and the market for ideas. Section 4 derives the invariant distribution of firm sizes and compares the invariant distributions obtained in the model to the US data. Section 5 concludes.

2   The Model

Agents order consumption according to the following utility function:

$$U(\{c_t\}) = \sum_{t=0}^{\infty} \beta^t u(c_t),$$

where $u(c_t)$ is strictly increasing, concave, twice continuously differentiable and bounded. In Section 3 below we will study two particular cases: the case with linear utility, $u(c) = c$, and the case with constant absolute risk aversion (CARA) utility, $u(c) = -ae^{-bc}$.

Agents work in two occupations. They can be entrepreneurs or workers. A worker earns $w > 0$ each period working for an entrepreneur (plus the value they extract from their ideas). There is a perfectly elastic supply of workers at this wage w. Agents do not have a technology to save so they consume what they earn each period. Entrepreneurs earn and consume profits, and the wage they receive from the project they work for directly. A worker can become an entrepreneur if, while working for a firm, he has an idea and decides to spin off the firm.

An idea is a non-replicable technology to produce consumption goods using labor, specifically, an idea uses 1 unit of labor.[1] Consider an entrepreneur who owns a firm with $N \in \{1, 2, ..\}$ ideas. Then his one-period profits are given by

$$\pi(S, N) = N(S - w)$$

where $S = \frac{1}{N} \sum_{i=1}^{N} P_i$ denotes the average revenue and P_i the per period income generated from a particular idea. We assume that $P_i > 0$ with probability one.

We assume that in each period the probability of a worker getting an idea is λ. An entrepreneur does not get ideas but buys ideas from workers. An entrepreneur can learn of an idea for sale with probability $0 < \gamma(\lambda, N) < 1$. For now, we do not take a stand on the specification of $\gamma(\cdot)$ but assume, as seems natural, that the probability of learning about an idea in any given period is

---

[1] We assume that ideas are non-replicable technologies in order to determine the scale of each project. If technologies are replicable, we would need a demand structure and goods differentiation to limit the size of each project. This simple extension would complicate our framework with no new insights.

increasing in λ and N. We will have more to say about the specification of $\gamma(\cdot)$ at the beginning of Section 4. In particular, in equilibrium, the number of ideas an entrepreneur learns about – the demand for ideas – must be equal to the number of ideas generated by workers – the supply of ideas. As we show below, equilibrium in the market for ideas will determine the average probability of an entrepreneur learning about an idea but not the distribution of probabilities among them. Thus, if we write $\gamma(\lambda, N) = \theta\tilde{\gamma}(\lambda, N)$, θ will be an equilibrium object (the average probability of an entrepreneur learning about an idea) and $\tilde{\gamma}(\cdot)$ (the distribution of probabilities among entrepreneurs) is a primitive of our economy which we will specify as λN in Section 4.

The mean payoff per period from the idea is μ, which is private information to the originator of the idea. The mean payoff is drawn from a continuous distribution $H(\mu)$ with $H'(\mu) > 0$ for all $\mu \geq 0$. The actual payoff is drawn from a distribution $F_\mu(P)$ where

$$\int f(P)\, dF_\mu(P)$$

is increasing in μ for all increasing functions f, $F_\mu(0) = 0$ all μ, and

$$\int P dF_\mu(P) = \mu.$$

We also assume that F_μ is continuous with respect to μ and

$$\lim_{\mu \to \infty} F_\mu(w) = 0.$$

The realization of P for a given idea can be discovered by implementing the idea for one period.

As mentioned above, entrepreneurs do not get ideas as they are involved in the management of their firm, but they can buy ideas from workers. An entrepreneur who has bought an idea can pay w to try it out for one period and observe the realization of P. If he does, he will use the idea to produce as long as his future expected utility from doing so is greater than from dropping it. Entrepreneurs may decide to implement a project even if the stream of profits is negative ($P < w$), since having an extra project implies that they may generate more ideas in the future (although this will not happen in equilibrium, as we show below).

A worker who has had an idea this period has two potential uses for it. He can sell his idea to an entrepreneur, in which case he reveals the mean payoff to the entrepreneur who buys it. In this case he earns a wage w plus the price Z at which he sells the idea. The idea is then owned by the entrepreneur and he decides to try it out or not.

He can also leave with the idea and become an entrepreneur of a firm with only this idea: a spin-off. Note that in the market of ideas, the price of an idea has to be non-contingent on the

6

quality of the idea. The reason is that any contingent contract would give the worker an incentive to lie about the quality of the idea. So the only incentive-compatible price is independent of quality, in which case the agent is indifferent between revealing the true quality of the idea or not. Since this information is useful for the entrepreneur, we assume that the worker does reveal the true quality. The price of an idea Z is determined in equilibrium, where all entrepreneurs will be indifferent between buying ideas or not. Given this indifference we will further assume that transaction costs (or search costs) are such that entrepreneurs end up buying ideas generated within the firms.

The implementation of an idea and the return that it generates are specific to the entrepreneur of the firm that implements it. That is, without the entrepreneur who implemented the idea, the project has zero value. Hence, projects are entrepreneur-specific but, importantly, not worker-specific.[2] We are also implicitly assuming that there is no contract in which the distribution of the gains from an idea is done ex-post. That is a contract contingent on realizations. To the extent that such a verification process is costly, workers and entrepreneurs will prefer our current market setup, particularly if we assume that the cost of writing these contracts is positive and therefore larger than the zero cost of setting up a new firm through a spin-off. If setting up a new firm was costly, we would need the cost of writing these contracts to be higher than the cost of setting up a new firm. Even though we do not have any direct evidence that writing these contracts may be as costly as setting up a new firm, conceptually there is little difference between a new firm and part of a firm that is run independently and the realized profits are distributed to the generator of the idea, who is not the entrepreneur. Hence, in what follows we just assume that the cost of contracts contingent on realizations is positive and the fixed cost of setting up a new firm is zero.

2.1 An Entrepreneur's Problem

Consider the problem of an entrepreneur with average revenue S, coming from N old ideas, who owns one new idea with mean payoff μ. If the entrepreneur tests the idea, his value function is

$$
\begin{aligned}
V(\mu, S, N) \;=\;& \int \left[u \left(\pi(S,N) + w + P - Z - w \right) \right] dF_\mu(P) \\
& + \beta \int \max \left[W \left(\frac{NS+P}{N+1}, N+1 \right), W(S,N) \right] dF_\mu(P).
\end{aligned}
$$

This period, his expected utility is the result of consuming the profits from the accumulated used projects $\pi(S,N)$, his wage w, the price of the idea Z, and the random realization of profits from the

---

[2] The idea is that the entrepreneur has some specific knowledge necessary to generate output from a particular implementation of an idea.

new project $P - w$. Note that the distribution from which P is drawn has expected value μ. Denote by $W(S, N)$ the continuation value of an entrepreneur with N projects with average revenue S. If the entrepreneur uses the project, next period he will manage a firm with $N+1$ projects and average revenue $(NS + P)/(N + 1)$. If he does not use it, next period his continuation value stays constant at $W(S, N)$. The continuation value (or the value without any new idea) of an entrepreneur with N projects with average revenue S is given by

$$
\begin{aligned}
W(S, N) \;=\; & \gamma(\lambda, N) \int^{\mu_H} \max\left[V(\mu, S, N), u\left(\pi(S, N) - Z + w\right) + \beta W(S, N)\right] dH(\mu) \\
& + (1 - \gamma(\lambda, N) H(\mu_H))\left[u\left(\pi(S, N) + w\right) + \beta W(S, N)\right]
\end{aligned}
$$

or

$$
\begin{aligned}
W(S, N) \;=\; & \gamma(\lambda, N) \int^{\mu_H} \max\left[V(\mu, S, N) - u\left(\pi(S, N) - Z + w\right) - \beta W(S, N), 0\right] dH(\mu) \\
& + (1 - \gamma(\lambda, N) H(\mu_H))\left[u(\pi(S, N) + w) + \beta W(S, N)\right] \\
& + \gamma(\lambda, N) H(\mu_H)\left[u(\pi(S, N) - Z + w) + \beta W(S, N)\right]
\end{aligned}
$$

where μ_H denotes the mean revenue value at which workers leave the firm with their idea.[3] The probability of someone in this firm generating an idea next period is $\gamma(\lambda, N)$. If an employee has an idea, he can leave and set up his own firm. He will do so as long as the idea is good enough, $\mu \geq \mu_H$. If the entrepreneur has the idea, in order to implement it, he has to hire a worker and communicate to him what he knows about the idea. If the idea is good enough, the worker will leave the firm and set up his own firm, too. Hence, ideas get implemented within the firm only if $\mu < \mu_H$, independent of who has them. Given that an idea of expected revenue μ was generated, the value of implementing it is, as discussed above, given by $V(\mu, S, N)$. The value of not implementing the idea is given by $u(\pi(S, N) - Z + w) + \beta W(S, N)$, namely, the utility of consuming profits and wage today and paying the price for the idea, plus the same continuation value tomorrow. An idea will not be implemented if it provides a very low expected value. If no one has an idea or if the idea is good enough to generate a spin-off, the value of the entrepreneur is given by $u(\pi(S, N) + w) + \beta W(S, N)$, since he does not pay for the idea. One of these scenarios happens with probability $1 - \gamma(\lambda, N) H(\mu_H)$.

The next lemma shows that the continuation value $W(S, N)$ exists and is increasing and continuous in average revenue S. We then show in Lemma 2 that the value of an entrepreneur with

---

[3]Note that we are already assuming that workers spin off when they get an idea with $\mu > \mu_H$. Below we prove that this is, in fact, the case. In the meantime, all our arguments remain unaffected if we were to define a set M_H that includes the μ's for which agents spin off. Then the integrals above would integrate over all values of μ that are not in M_H.

an idea μ, $V(\mu, S, N)$ is increasing and continuous in the expected value of the idea μ and in the average return S.

Lemma 1 $W(S, N)$ *exists and is strictly increasing in* S.

Proof. Let $\mathcal{C}$ be the space of bounded continuous functions defined on $\mathbb{R} \times \mathbb{N}$. Define the operator $T(W) : \mathcal{C} \rightarrow \mathcal{C}$ as

$$
\begin{aligned}
T\left(W\right)\left(S, N\right) = \ & \gamma\left(\lambda, N\right) \int^{\mu_H} \max\left[V(\mu, S, N) - u(\pi(S, N) - Z + w) - \beta W(S, N), 0\right] dH(\mu) \\
& + (1 - \gamma\left(\lambda, N\right) H\left(\mu_H\right)) \left[u(\pi(S, N) + w) + \beta W(S, N)\right] \\
& + \gamma\left(\lambda, N\right) H\left(\mu_H\right) \left[(\pi(S, N) - Z + w) + \beta W(S, N)\right]
\end{aligned}
$$

where

$$
\begin{aligned}
V(\mu, S, N) = \ & \int \left[u\left(\pi(S, N) + w + P - Z - w\right)\right] dF_\mu(P) \\
& + \beta \int \max\left[W\left(\frac{NS + P}{N + 1}, N + 1\right), W(S, N)\right] dF_\mu(P).
\end{aligned}
$$

It is easy to show that T is a contraction using Blackwell's conditions. It satisfies monotonicity since $W \leq W'$ implies that $T\left(W\right) \leq T\left(W'\right)$ (all expressions above are increasing in the function W). It satisfies discounting, since for $a > 0$

$$
T\left(W + a\right) \leq \beta a + T\left(W\right)
$$

where $\beta < 1$. Hence, T is a contraction by the Contraction Mapping Theorem and a unique fixed point to the operator T exists.

Suppose W is strictly increasing in S. Since S appears only in π and W in the definition of T, $T\left(W\right)$ is also strictly increasing in S. Hence, by the Contraction Mapping Theorem the fixed point of T is strictly increasing in S.

To show that W is continuous in S, note that since the space of continuous functions is closed in the sup norm, we can apply the same argument to show continuity given that π is continuous in S, F_μ is continuous in μ, and $H(\mu)$ is continuous in μ. ∎

Lemma 2 $V(\mu, S, N)$ *exists and is strictly increasing and continuous in* μ *and* S.

Proof. By Lemma 1 $W(S, N)$ exists and therefore $V(\mu, S, N)$ exists. Since $\pi(S, N) + w - w + P - Z$ is strictly increasing in P, the first term in V is strictly increasing in μ, since $\int f\left(P\right) dF_\mu\left(P\right)$

9

is increasing in μ for all increasing functions f. Since $W(S, N)$ is strictly increasing in S by the previous lemma,

$$\int \max \left[W \left(\frac{NS + P}{N+1}, N+1 \right), W(S, N) \right] dF_\mu(P)$$

is strictly increasing in μ, which proves the result.

That V is strictly increasing in S follows from W being strictly increasing in S.

To show that V is continuous in μ and S, note that by Lemma 1, $W(S, N)$ is continuous in S, the maximum of continuous functions is continuous, and F_μ is continuous with respect to μ. ∎

An entrepreneur will implement an idea with expected revenue μ if

$$V(\mu, S, N) > u(\pi(S, N) - Z + w) + \beta W(S, N).$$

Let $\mu_L(S, N)$ be the value of μ that solves

$$V(\mu_L, S, N) = u(\pi(S, N) - Z + w) + \beta W(S, N). \tag{1}$$

Then an entrepreneur will implement an idea as long as $\mu > \mu_L(S, N)$. Thus we can rewrite $W(S, N)$ as

$$
\begin{aligned}
W(S, N) &= \gamma(\lambda, N) \int^{\mu_H} \max \left[V(\mu, S, N) - u(\pi(S, N) - Z + w) - \beta W(S, N), 0 \right] dH(\mu) \\
&\quad + (1 - \gamma(\lambda, N) H(\mu_H)) \left[u(\pi(S, N) + w) + \beta W(S, N) \right] \\
&\quad + \gamma(\lambda, N) H(\mu_H) \left[u(\pi(S, N) - Z + w) + \beta W(S, N) \right]
\end{aligned}
$$

Or, equivalently,

$$
\begin{aligned}
W(S, N) &= \gamma(\lambda, N) \int_{\mu_L(S,N)}^{\mu_H} \left[V(\mu, S, N) - u(\pi(S, N) - Z + w) - \beta W(S, N) \right] dH(\mu) \\
&\quad + \gamma(\lambda, N) H(\mu_H) \left[u(\pi(S, N) - Z + w) + \beta W(S, N) \right] \\
&\quad + (1 - \gamma(\lambda, N) H(\mu_H)) \left[u(\pi(S, N) + w) + \beta W(S, N) \right]
\end{aligned}
$$

The next lemma shows that there exists a unique function $\mu_L(S, N)$ that satisfies equation (1).

Lemma 3 *There exists a unique function $\mu_L(S, N)$ that satisfies equation (1).*

Proof. First note that given our assumption that $F_\mu(0) = 0$ all μ we know that $\mu \geq 0$. Then

$$V(0, S, N) = u\left(\pi(S, N) - Z\right) + \beta W(S, N) < u(\pi(S, N) - Z + w) + \beta W(S, N)$$

since $w > 0$. In contrast,

$$
\begin{aligned}
\lim_{\mu \to \infty} V(\mu, S, N) &= \lim_{\mu \to \infty} \int u\left(\pi(S, N) + w - Z - w + P\right) dF_\mu(P) \\
&\quad + \beta \int \max\left[W\left(\frac{NS + P}{N + 1}, N + 1\right), W(S, N)\right] dF_\mu(P) \\
&> u(\pi(S, N) - Z + w) + \beta W(S, N).
\end{aligned}
$$

The inequality follows because the probability that $P > w$ goes to 1 as $\mu \to \infty$ and

$$\max\left[W\left(\frac{NS + P}{N + 1}, N + 1\right), W(S, N)\right]$$

is non-negative. Hence since $V(\mu, S, N)$ is strictly increasing and continuous in μ, by the Mean Value Theorem there exists a unique scalar μ_L that satisfies (1) for each pair (N, S). Let $\mu_L(S, N)$ be the unique function that takes this value given a pair (S, N). ∎

2.2 A Worker's Problem

The expected utility of a worker with an idea μ that decides to spin off is given by

$$V_0(\mu) = \int u(P) dF_\mu(P) + \beta \int \max\left[W(P, 1), W_0\right] dF_\mu(P).$$

The continuation value of a worker currently working in a firm, W_0, is then given by

$$W_0 = \lambda \int \max\left[V_0(\mu), u(w + Z) + \beta W_0\right] dH(\mu) + (1 - \lambda)\left[u(w) + \beta W_0\right].$$

Using arguments similar to the ones used above for V, we can show that $V_0(\mu)$ is strictly increasing in μ. A worker with an idea μ will leave the firm and become an entrepreneur if

$$V_0(\mu) > u(w + Z) + \beta W_0$$

Let μ_H be the value of μ that solves

$$V_0(\mu_H) = u(w + Z) + \beta W_0. \tag{2}$$

Thus, if $\mu > \mu_H$ the worker will leave his employer and set up a new firm. The continuation value of a worker can therefore be written as

$$W_0 = \lambda \int_{\mu_H} V_0(\mu) dH(\mu) + (1 - \lambda)\left[u(w) + \beta W_0\right] + \lambda H(\mu_H)\left(u(w + Z) + \beta W_0\right).$$

11

We show formally below that there exists a unique threshold μ_H. Note also that μ_H is a constant and so it is independent of the characteristics of the firm (S, N) in which the agent works.

Lemma 4 *There exists a unique value μ_H that satisfies equation (2). Furthermore $V_0(\mu)$ is increasing and continuous in μ.*

Proof. We first need to show that V_0 is increasing and continuous in μ, but this follows directly from F_μ being continuous and increasing in μ as assumed above. We also know that

$$V_0(0) = u(0) + \beta W_0 < u(w + Z) + \beta W_0$$

since $w + Z > 0$ and $\max[W(0, 1), W_0] = W_0$. The latter is the result of the fact that a new entrepreneur with a project that pays 0 earns less than a worker and has fewer opportunities with regard to exploiting his future ideas (he has exercised the option of spinning off). In contrast,

$$\lim_{\mu \to \infty} V_0(\mu) = \lim_{\mu \to \infty} \int u(P) \, dF_\mu(P) + \beta \int_w W(P, 1) \, dF_\mu(P)$$
$$> u(w) + \beta W_0.$$

Hence since $V_0(\mu)$ is strictly increasing in μ and continuous by the Mean Value Theorem, there exists a unique scalar μ_H that satisfies (2). ∎

We still need to define the realized return needed in order to continue with a project once its return is realized. Define $P_L(N, S)$ as

$$W\left(\frac{NS + P_L(N, S)}{N + 1}, N + 1\right) = W(S, N),$$

and P_H by

$$W(P_H, 1) = W_0.$$

Then a firm keeps the project if the realized return is $P \geq P_L(N, S)$ and a spin-off stays in operation if the realized return on the idea that generated the spin-off is such that $P \geq P_H$. Note that $P < P_H$, the spin-off will exit and the would-be entrepreneur will return to the labor force as a worker. In that case, given that the implementation of his project was specific to him, the project has zero resale value.

2.3 Equilibrium

A long-run equilibrium of this economy is a distribution of firm sizes δ_N, a list of four thresholds, $\mu_L(\cdot), \mu_H, P_H$ and $P_L(\cdot)$, a price of ideas, Z, and the average probability with which an entrepreneur

buys an idea, θ, (where θ is given by $\gamma(\lambda, N) = \theta \tilde{\gamma}(\lambda, N)$) such that entrepreneurs solve the problem in Section 2.1, workers solve the problem in Section 2.2 and the price, Z, and the average probability of buying an idea, θ, clears the market for ideas:

$$\lambda \sum_{N=1}^{\infty} (N-1)\delta_N = \theta \sum_{N=1}^{\infty} \tilde{\gamma}(\lambda, N)\delta_N, \tag{3}$$

where the l.h.s. is the supply of ideas and the r.h.s. is the demand for ideas. As we show in the next section, the price Z will be such that entrepreneurs will be indifferent about how many ideas to buy. Therefore, market clearing will just require that the number of ideas bought by entrepreneurs be equal to the number of ideas generated by workers. That is, market clearing simply determines the value of θ (but not $\tilde{\gamma}(\lambda, N)$ which is a primitive of the model). The price Z is the only price at which entrepreneurs are indifferent and so, for an arbitrary $\tilde{\gamma}(\lambda, N)$, any other price will be inconsistent with market clearing.

3   Characterization

In this section we characterize the thresholds on the expected revenue from an idea that determine if an idea is thrown away, implemented by a particular firm, or results in a spin-off. For this, we first assume that the utility function is of the form

$$U(\{c_t\}) = \sum_{t=0}^{\infty} \beta^t u\left(c_t\right) = \sum_{t=0}^{\infty} \beta^t c_t.$$

We show in the Appendix that our main results hold under a CARA utility function as well. The main reason to choose these two utility functions is that we can solve the value of an existing firm in closed form given the additive separability or log additive separability of these utility functions.[4]

Under the assumption that the utility function is linear we can fully solve this problem in closed form. The first result shows that the threshold $\mu_L(S, N)$ is independent of S and N. $\mu_L(S, N)$ independent of S is implied by risk neutrality (or, in the CARA case below, by the fact that risk does not depend on the level of wealth). $\mu_L(S, N)$ constant in N is the result of the market for ideas. Since workers will sell their ideas to whoever is willing to pay more for them, and there is a relative scarcity of ideas, workers extract all the surplus of an idea and we can solve for the price

---

[4]Note that none of these specifications of the utility functions satisfy the bounded assumption we made in Section 2. However, since in these two cases we can solve the functional equation for $W(S, N)$ analytically, it follows from Theorem 9.12 in Stokey, Lucas, and Prescott (1989) that the solution is, in fact, optimal.

of an idea in equilibrium. The proposition also yields the result that in equilibrium $P_L = w$, and so entrepreneurs use all projects that give positive returns. In contrast, $P_H > w$ and so spin-offs use projects that give strictly positive returns. The reason is that new entrepreneurs that start a firm with a project with a low realized return have the option of going back to work for a firm and start a new firm in the future with a better project. The proposition also shows that the threshold for implementing ideas through spin-offs is greater than the one for implementing ideas within the firm, $\mu_L < \mu_H$. This is essentially the result of the fixed capital cost of setting a new firm in the case of linear utility. Inventors are more selective with the ideas they use when they spin off than are firms. This also implies that some ideas within the firm do not result in spin-offs and so some firms grow. An industry's growth is then the result of entry through spin-offs and growth in the intensive margin. With risk-averse CARA agents, this is the result of the larger risk faced by agents that spin off.

Proposition 5 *If $u(c_t) = c_t$, then in equilibrium*

- *$\mu_L(S, N)$ is independent of S and N, and $\mu_L(S, N) < w$,*

- *the thresholds for using a project are given by $P_L(S, N) = w$ and $P_H = w + f_0 > w$ for some positive constant f_0.*

- *$\mu_L < \mu_H$, so some ideas are implemented within existing firms and some through spin-offs,*

- *and the market price of ideas is given by*

$$Z = \frac{1}{H(\mu_H)} \int_{\mu_L}^{\mu_H} [\mu - w + \frac{\beta}{1-\beta} \int \max [P - w, 0] \, dF_\mu(P)] dH(\mu) > 0. \qquad (4)$$

Proof. Guess

$$W(S, N) = \omega(\pi(S, N) + w) = \omega(N(S - w) + w)$$

Since

$$W(S, N) = \gamma(\lambda, N) \int^{\mu_H} \max \begin{bmatrix} \int [\pi(S, N) + w + P - Z - w] \, dF_\mu(P) \\ +\beta \int \max \left[W\left(\frac{NS+P}{N+1}, N+1\right), W(S, N) \right] dF_\mu(P), \\ \pi(S, N) - Z + w + \beta W(S, N) \end{bmatrix} dH(\mu)$$
$$+ (1 - \gamma(\lambda, N) H(\mu_H)) [\pi(S, N) + w + \beta W(S, N)]$$

then

14

$$\omega\left(N\left(S-w\right)+w\right) \ = \ \left(1+\beta\omega\right)\left(N\left(S-w\right)+w\right)-\gamma\left(\lambda,N\right)H\left(\mu_H\right)Z$$
$$+\gamma\left(\lambda,N\right)\int^{\mu_H}\max\left[\mu-w+\beta\omega\int\max\left[P-w,0\right]dF_\mu(P),0\right]dH(\mu)$$

Let
$$\omega = 1 - \beta\omega$$

or
$$\omega = \frac{1}{1-\beta}$$

then
$$Z \ = \ \frac{1}{H\left(\mu_H\right)}\int^{\mu_H}\max\left[\mu-w+\frac{\beta}{1-\beta}\int\max\left[P-w,0\right]dF_\mu(P),0\right]dH(\mu)$$
$$= \ \frac{1}{H\left(\mu_H\right)}\int_{\mu_L}^{\mu_H}\max\left[\mu-w+\frac{\beta}{1-\beta}\int_{P_L}\left[P-w\right]dF_\mu(P),0\right]dH(\mu)$$

Note that this is exactly the condition that makes
$$W\left(S,N\right) = \frac{\pi(S,N)+w}{1-\beta}.$$

That is, it makes the value function of entrepreneurs equal to the present value of the current project of the firm. So entrepreneurs are willing to pay up to Z for workers' ideas, and workers will get this price or will sell the idea to another entrepreneur. Competition for ideas among entrepreneurs then guarantees that the market price of ideas Z is determined by the condition above in equilibrium.

Then
$$V\left(\mu_L,S,N\right) = \left(\pi(S,N)-Z+w\right)+\beta W(S,N)$$

implies that μ_L satisfies
$$\int\left(P-w\right)dF_{\mu_L}(P)+\frac{\beta}{1-\beta}\int\max\left[P-w,0\right]dF_{\mu_L}(P) = 0, \tag{5}$$

and therefore is independent of S and N. Hence, $\mu_L < w$, since entrepreneurs can drop the idea next period (the maximum in the second term on the l.h.s.). P_L satisfies
$$\max\left[P-w,0\right] > 0 \text{ for } P > P_L,$$

so $P_L = w$. Then μ_L is given by
$$\frac{1}{1-\beta}\int_w\left(P-w\right)dF_{\mu_L}(P) = \int_0^w\left[w-P\right]dF_{\mu_L}(P).$$

15

Namely, the present value of the gains from implementing a project has to be equal to the cost.

Guess that

$$W_0 = \frac{w + f_0}{1 - \beta}$$

then

$$V_0(\mu) = \int P F_\mu(P) + \frac{\beta}{1 - \beta} \int \max \left[P, w + f_0 \right] dF_\mu(P).$$

The continuation value of a worker currently working in a firm, W_0, is then given by

$$
\begin{aligned}
\frac{w + f_0}{1 - \beta} &= \lambda \int \max \left[\int P F_\mu(P) + \frac{\beta}{1 - \beta} \int \max \left[P, w + f_0 \right] dF_\mu(P), w + Z + \beta \frac{w + f_0}{1 - \beta} \right] dH(\mu) \\
&+ (1 - \lambda) \left[w + \beta \frac{w + f_0}{1 - \beta} \right]
\end{aligned}
$$

so

$$f_0 = \lambda \int \max \left[\int (P - w) F_\mu(P) + \frac{\beta}{1 - \beta} \int \max \left[P - w - f_0, 0 \right] dF_\mu(P), 0 \right] dH(\mu) + (1 - \lambda) Z \quad (6)$$

which determines f_0 as a positive constant (given H assigns positive mass to μ's such $\int (P - w) F_\mu(P) > 0$) and verifies our guess.

Then

$$V_0(\mu) = \frac{w}{1 - \beta} + \frac{\beta f_0}{1 - \beta} + \int (P - w) dF_\mu(P) + \frac{\beta}{1 - \beta} \left[\int \max \left[P - w - f_0, 0 \right] dF_\mu(P) \right].$$

Since

$$V_0(\mu_H) = w + Z + \beta W_0$$

μ_H is given by

$$\int (P - w) dF_{\mu_H}(P) + \frac{\beta}{1 - \beta} \int \max \left[P - w - f_0, 0 \right] dF_{\mu_H}(P) = Z,$$

Two results are immediate from this expression. First, since $Z \geq 0$ and $f_0 > 0$ the left-hand side of the above equation evaluated at $\mu_H = \mu_L$ is less than the right-hand side given that μ_L is determined by Equation (5). Since the left-hand side is increasing in μ, it follows that $\mu_L < \mu_H$. Second, since P_H is such that $P - w - f_0 = 0$, it follows that $P_H = w + f_0 > w$. ∎

The key insight in the previous proposition is that the selection of the ideas implemented in existing firms, which is given by μ_L and μ_H, is independent of S and N. Because of this, the set of ideas that will be implemented within each firm is independent of the firm's size. It is the market

16

for ideas that leads to this result. In the absence of these markets, entrepreneurs of existing firms will appropriate some of the surplus of a given idea. Then, as long as $\gamma(\lambda, \cdot)$ is not linear in N, a large firm with many projects will have more/less incentives to implement ideas than smaller firms. The type of scale dependence will depend on the concavity or convexity of the function $\gamma(\lambda, \cdot)$ in N. If the probability of getting new ideas is concave in N, large firms will be more selective and will therefore implement ideas with a higher expected return than entrepreneurs in small firms. The reverse is true if $\gamma(\lambda, \cdot)$ is convex. This is not the case when entrepreneurs pay the market price Z of an idea in the market. In this case the expected benefits for all entrepreneurs is zero, and so the selection of projects in equilibrium is independent of the shape of the function $\gamma(\lambda, \cdot)$ and therefore of the scale of the firm. This is the sense in which the market for ideas is key to generating scale independence in the selection of ideas.

In order for firm growth to be scale independent we need the thresholds μ_L and μ_H to be independent of the size of the firm. We also need the number of innovations bought by a given firm in the market of ideas to be proportional to its size. In our theory, as long as the unconditional probability (λ) that a worker has an idea is independent of firm size, we obtain that the two thresholds are independent of size. In order for firms to buy projects at a rate proportional to their size, we will assume a linear probability of learning about ideas, $\gamma(\lambda, N) = \theta \lambda N$, where θ is determined by equalizing the demand and supply of ideas. In effect, we are assuming that an entrepreneur learns about ideas for sale through all the agents working in his firm, including himself. This may be the result of workers having ideas themselves and selling them to their manager, since that may be easier than contacting an unknown manager or employees finding out about ideas for sale in the market and informing their manager.[5]

Figure 1 summarizes what we have learned about firm behavior and firm entry. It shows expected revenue μ in the real line. For projects above μ_H, firms' workers spin off. All other firms implement ideas with μ's between μ_L and μ_H. An incumbent risk-neutral entrepreneur would implement projects as long as they pay expected return w, so the difference between μ_L and w is the result of an entrepreneur's ability to drop the project next period (this ordering can change once we consider risk-averse agents). The threshold that determines μ_L is given by

$$\int (P - w) \, dF_{\mu_L}(P) + \frac{\beta}{1 - \beta} \int_w (P - w) \, dF_{\mu_L}(P) = 0. \tag{7}$$

The difference between μ_L and μ_H is the result of the option value of exiting and setting up a new

---

[5] In the particular case in which $\gamma(\lambda, N) = \tilde{\lambda}(\lambda) N$ *and* utility is linear, even without the market for ideas, μ_L is independent of size. The market for ideas is necessary to obtain scale independence in the selection of ideas if the utility function is not linear and/or $\gamma(\lambda, \cdot)$ is not linear in N. In the Appendix where we consider the case of exponential utility this is evident.
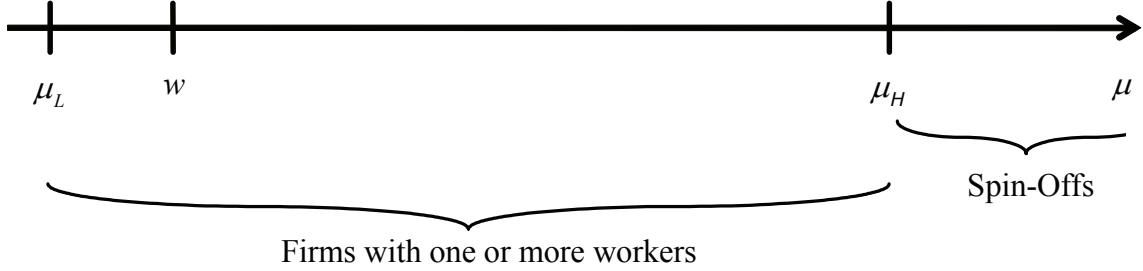
Figure 1: Figure 1: Selection of Ideas

firm in the future, f_0. The threshold μ_H is implicitly determined by

$$\int (P - w) \, dF_{\mu_H}(P) + \frac{\beta}{1 - \beta} \int_{P_H} (P - w - f_0) \, dF_{\mu_H}(P) = Z. \tag{8}$$

The difference between the two thresholds comes from the option value f_0 given by equation (6) of closing a new firm and starting another one later on with a better idea and the fact that workers give up the price of an idea Z when they set up the firm. Note that if $f_0 = 0$, the two threshold equations and the equation for Z imply that $Z = 0$ and $\mu_L = \mu_H$. So, all projects would be implemented via spin-offs. However, as shown in the previous proposition, f_0 is positive, since workers can extract the value of very good projects by spinning off. New firms will require a higher return from their first project than existing firms demand from new projects, given the larger option value f_0 that new firms have of returning to an old firm and spinning off in the future, namely, $P_H > w = P_L$. Both of these equations imply that the number of entrants as a fraction of the population is constant and so is the number of new projects implemented in existing firms each period as a fraction of total population.

Finding an equilibrium amounts to solving equations (4), (6), (7) and (8) for the values of Z, f_0, μ_L and μ_H. In the next proposition we show that an equilibrium exists and is unique.

Proposition 6 *A long-run equilibrium for this economy exists and is unique*

18

Proof. To establish this proposition, observe that when $Z = 0$, f_0 given by Equation (6) is positive. Then, Equations (7) and (8) imply that if $\mu_H = \mu_L$, the l.h.s. of (8) is negative and hence smaller than $Z = 0$. Now note that the l.h.s. of (8) is strictly increasing and continuous in μ_H as is the r.h.s. However, observe that the derivative of the l.h.s. of Equation (8) with respect to μ_H is strictly larger than the derivative of Z with respect to μ_H and the difference of both derivatives is bounded away from 0, as Z is the *average* value of the l.h.s. of (8) for the interval $[\mu_L, \mu_H]$. Therefore, by the Intermediate Value Theorem there exists a unique μ_H that solves Equation (8). Given μ_H we can immediately obtain a unique value of Z and f_0. Furthermore, μ_L is uniquely determined by Equation (7) since the r.h.s. of Equation (7) is strictly increasing and continuous in μ_L and at $\mu_L = 0$ the r.h.s. is less than zero.

We still need to show that there exists a unique invariant distribution δ_N and an average probability of entrepreneurs buying an idea θ. To show that there exists a unique θ given an invariant distribution is immediate from Equation (3). Showing existence of a unique invariant distribution is more involved and we dedicate the next section to it. In Corollary 8 we show that in fact such a distribution exists and is unique. ∎

As we noted above, one potential issue is the existence of an equilibrium with contingent contracts, namely, a contract in which an entrepreneur offers the worker the contingent return on an idea minus w. Workers with good ideas that would otherwise spin off would be willing to stay if the entrepreneur pays the cost of writing the contract. However, since we assumed that the cost of the contract is positive, this is never optimal for the entrepreneur.

In the appendix, we show that all results, except $\mu_L < w$, hold when $u(c_t) = -ae^{-bc_t}$. In this case all agents in the economy are risk averse. However, because their risk aversion does not depend on the level of their wealth, in the presence of markets for ideas, we still obtain the result that μ_L is constant, and therefore that the selection of projects is scale independent. In the CARA case studied in the appendix, the price of ideas is given by

$$Z = \frac{1}{b} \log \left[\frac{1 + \frac{\beta}{1-\beta} \int_{\mu_L}^{\mu_H} \int_{P_L} \left(1 - e^{-b(P-w)}\right) dF_\mu(P) dH(\mu)}{1 - \int_{\mu_L}^{\mu_H} \int \left(1 - e^{-b(P-w)}\right) dF_\mu(P) dH(\mu)} \right] > 0.$$

4   Invariant Distribution of Firm Sizes

In order to derive the implications of our model for firm growth and the size distribution of firms, we need to take a stand on the number of projects that firms buy in the market. The reason is

that in our model entrepreneurs are indifferent about how many ideas to buy. As argued above, we can either specify the function $\gamma(\cdot)$ and assume that entrepreneurs buy ideas in proportion to the number of agents in the firm. This amounts to assuming that they find out about ideas generated by their own workers or, alternatively, they and their workers get information about ideas at a constant rate per person.

Suppose that a firm with N projects has a probability of finding out about an idea given by $\gamma(\lambda, N) = \theta \lambda N$. We assume that the maximum size of a firm is given by $\overline{N}$ such that $\lambda \overline{N} < 1$.[6] Everyone in the firm has a probability $\lambda > 0$ of generating an idea. Note that since the value of λ depends on our definition of a period, we can always make λ small enough by appropriately defining the length of a period in the model. Correspondingly, we can make $\overline{N}$ arbitrarily large. In case a firm hits the size constraint $\overline{N}$, its workers will sell ideas to other firms. For the moment we abstract from this problem, but we return to it below.

Given our specification for $\gamma(\cdot)$ we need to choose θ such that the supply of ideas and the demand for ideas equalize in equilibrium at price Z. Let δ_N denote the share of firms of size N in equilibrium. We will discuss this distribution in much more detail below. Then, market clearing in the market for ideas is given by

$$\lambda \sum_{N=1}^{\infty} (N-1)\delta_N = \theta \lambda \sum_{N=1}^{\infty} N \delta_N \tag{9}$$

where the l.h.s. is the supply of ideas ($N-1$ workers in a firm of size N have a probability λ of generating an idea) and the r.h.s. is the demand for ideas (a firm of size N learns of and buys $\gamma(\lambda, N) = \theta \lambda N$ ideas). Denote by ν the average firm size, then

$$\theta = 1 - \frac{1}{\nu}. \tag{10}$$

Note that our assumption that the maximum size of firms is given by $\overline{N}$ implies an additional adjustment for θ. Since firms get zero expected benefits out of implementing ideas, entrepreneurs are indifferent about expanding or not. Hence, they do not care about this upper bound for the size of their firm. The only role that this bound plays is to determine how other firms grow if there is a positive mass of constrained firms. Thus, the only adjustment we need to make is to add the upper bound $\overline{N}$ to Equation (9). Notice, however, that Equation (10) still holds.

---

[6] Alternatively we could work with continuous time and assume that the process by which firms generate ideas is Poisson with parameter λN. This would imply an identical random process for generating ideas in continuous time. Note that we are assuming that the process of generating ideas and the process of assigning knowledge and ownership are independent. If instead each worker had an unconditional probability of generating an idea λ independently of other workers, there would be a positive probability of generating several ideas per period, which we rule out.

In order to derive the size distribution of firms, first note that the size of the industry will increase constantly in our setup, given that innovation does not stop (every worker in the industry has probability λ of having an idea independently of where they work). On top of this the probability of firms adding a project is positive for all firms, while the probability of dropping a project that is already being used is zero. Hence, firms will also grow continuously. This is combined with a positive mass of new entrants with one worker every period. We can show that there is an invariant distribution of employment shares and firm sizes measured as a share of total employment. That is, we normalize by the size of total employment.

First consider the transition equation for a firm with N workers. Each worker has a probability λ of having an idea. If they do, the firm implements it if $\mu \in [\mu_H, \mu_L(N)]$ and if it implements it, the firm uses the idea with probability $1 - F_\mu(P_L(N))$ where $P_L(N)$ is such that

$$W\left(\frac{NS + P_L(N)}{N+1}, N+1\right) = W(S, N),$$

which by the arguments above does not depend on S.

In what follows we will ignore the upper bound on firm sizes $\overline{N}$. We will return to it once we define the invariant distribution of firm sizes for the case without this bound. Hence if $p(N, N+1)$ denotes the probability of a firm transitioning from N to $N+1$ workers

$$p(N, N+1) = \theta \lambda N \int_{\mu_L}^{\mu_H} (1 - F_\mu(P_L)) \, dH(\mu).$$

Hence,

$$p(N, N') = \begin{cases} 0 & \text{for } N' > N+1 \\ \theta \lambda N \int_{\mu_L}^{\mu_H} (1 - F_\mu(P_L)) \, dH(\mu) & \text{for } N' = N+1 \\ 1 - \left[\theta \lambda N \int_{\mu_L}^{\mu_H} (1 - F_\mu(P_L)) \, dH(\mu)\right] & \text{for } N' = N \\ 0 & \text{for } N' < N \end{cases}.$$

Let $S = \{1, 2, ...\}$ then for any $A \subseteq S$,

$$p(N, A) = \sum_{N' \in A} p(N, N')$$

is positive if $N \in A$ or $N + 1 \in A$.

Let L_t be the total labor force and E_t the total number of firms or enterprises in period t, and let $\{\delta_N\}$ be the invariant distribution of firm sizes. The probability that a firm with N employees

21

generates a spin-off is given by

$$s\left(N\right) = \theta \lambda N \int_{\mu_H} \left(1 - F_\mu\left(P_H\right)\right) dH\left(\mu\right)$$

where P_H satisfies

$$W\left(P_H, 1\right) = W_0.$$

Hence, the expected number of spin-offs in period $t+1$ given the distribution of firm sizes in period t is given by

$$E_t \sum_{N=1}^{N=\infty} s\left(N\right) \delta_N$$

$$= E_t \theta \lambda \int_{\mu_H} \left(1 - F_\mu\left(P_H\right)\right) dH\left(\mu\right) \sum_{N=1}^{N=\infty} N \delta_N$$

$$= \theta \lambda \int_{\mu_H} \left(1 - F_\mu\left(P_H\right)\right) dH\left(\mu\right) L_t$$

$$\equiv \lambda_H L_t,$$

where λ_H denotes the number of new employees in new firms as a fraction of total employment. Hence the expected number of spin-offs is a constant fraction of the population, L_t.

Similarly we can calculate the expected number of new workers in existing firms, which is given by

$$E_t \sum_{N=1}^{L_t} p\left(N, N+1\right) \delta_N$$

$$= E_t \sum_{N=1}^{L_t} \theta \lambda N \int_{\mu_L}^{\mu_H} \left(1 - F_\mu\left(P_L\right)\right) dH\left(\mu\right) \delta_N$$

$$\equiv \lambda_L E_t \sum_{N=1}^{L_t} N \delta_N = \lambda_L L_t,$$

where λ_L denotes the number of new employees in old firms as a fraction of total employment.

Then, for E_t large

$$L_{t+1} = L_t + E_t \sum_{N=1}^{L_t} \left[p\left(N, N+1\right) + s\left(N\right)\right] \delta_N$$

$$= \left(1 + \lambda_H\right) L_t + \lambda_L L_t.$$

Given our definition of λ_L and λ_H, population evolves according to

$$L_{t+1} = \left(1 + \lambda_H + \lambda_L\right) L_t$$

22

Thus,

$$
\begin{aligned}
E_{t+1} &= E_t + E_t \sum_{N=1}^{\infty} s(N)\delta_N \\
&= E_t + \lambda_H L_t.
\end{aligned}
$$

Hence the number of firms is expanding at a constant rate. In terms of number of firms, the economy is growing at a constant rate. Note that we are assuming that E_t is large enough so that L_t and E_t evolve deterministically. For small L_t and E_t, however, both are random variables that evolve according to a stochastic process.

We now compute the invariant distribution of the share of workers in firms of different sizes. Let ϕ_N denote the probability that a worker is employed by a firm with N workers, given that the probability that a worker has an idea that is used within the firm is given by λ_L independently of the firm. Then, the invariant distribution satisfies

$$
[\phi_1(1-\lambda_L) + \lambda_H]L = \phi_1 L' = \phi_1(1+\lambda_L+\lambda_H)L
$$

or

$$
\phi_1(1-\lambda_L) + \lambda_H = \phi_1(1+\lambda_L+\lambda_H)
$$

which implies

$$
\phi_1 = \frac{\lambda_H}{\lambda_H + 2\lambda_L} \tag{11}
$$

for $N = 1$. Intuitively, the number of workers in firms of size 1 today, $\phi_1 L$, minus the number of workers in firms of size 1 that become workers in firms of size 2, $\phi_1\lambda_L L$, plus the number of new workers in firms of size 1, $\lambda_H L$, is equal (in the invariant distribution) to the number of workers in firms of size 1 tomorrow, $\phi_1 L'$, which is equal to $\phi_1(1+\lambda_L+\lambda_H)L$, given that the growth rate of employment is $\lambda_L + \lambda_H$.

Similarly, for firms of size N,

$$
\begin{aligned}
& \phi_N(1-\lambda_L N) + \phi_{N-1}\lambda_L(N-1) + \phi_{N-1}\lambda_L \\
={} & \phi_N(1-\phi_L N) + \phi_{N-1}\lambda_L N \\
={} & \phi_N(1+\lambda_L+\lambda_H)
\end{aligned}
$$

and so

$$
\phi_N = \phi_{N-1}\frac{\lambda_L N}{\lambda_H + \lambda_L(N+1)} \tag{12}
$$

which implies that

$$\frac{\phi_N}{N} = \frac{\phi_{N-1}}{N-1} \frac{\lambda_L (N-1)}{\lambda_H + \lambda_L (N+1)}.$$

Note that by definition

$$(1 + \lambda_L + \lambda_H) \sum_{N=1}^{\infty} \phi_N = \sum_{N=2}^{\infty} \left[\phi_N (1 - \lambda_L N) + \phi_{N-1} \lambda_L (N-1) + \phi_{N-1} \lambda_L \right]$$
$$+ \phi_1 (1 - \lambda_L) + \lambda_H$$

which implies that

$$(\lambda_L + \lambda_H) \sum_{N=1}^{\infty} \phi_N = \left[\sum_{N=2}^{\infty} \phi_{N-1} \lambda_L + \lambda_H \right] = \lambda_L \sum_{N=1}^{\infty} \phi_N + \lambda_H$$

Hence,

$$\sum_{N=1}^{\infty} \phi_N = 1,$$

and so the resulting ϕ's form a probability distribution. This distribution is the invariant distribution of population shares across firms of different sizes. Simon and Bonini (1958) propose an exogenous growth and entry process of firms that leads to the same type of distribution, namely, a Yule distribution. This distribution approximates a Pareto in the upper tail.

Proposition 7 *There exists a unique invariant distribution ϕ of employment shares across firm sizes, where ϕ_N denotes the share of workers employed by firms of size N.*

To obtain the distribution of firm sizes we need to transform the distribution of worker shares into a distribution of firm sizes. For this, note that if the share of the population employed by firms of size N is given by ϕ_N, then the share of firms of size N, δ_N, is given by

$$\delta_N = \frac{\phi_N}{N \sum_{N=1}^{\infty} \frac{\phi_N}{N}}. \tag{13}$$

Clearly, since $\sum_{N=1}^{\infty} \phi_N = 1$, $0 < \sum_{N=1}^{\infty} \frac{\phi_N}{N} < 1$ and so δ_N is well defined, exists, and is unique.

Corollary 8 *There exists a unique invariant distribution δ of firm sizes.*

Note from the previous equations that the distributions ϕ and δ depend only on the value of the ratio λ_H / λ_L. Note also that in the theory $N \leq \overline{N}$. Hence, in order to get distributions of

employment shares and firm sizes that are consistent with the theory we need to re-normalize both distributions. Hence, the distribution of employment shares is given by

$$\tilde{\phi}_N = \frac{\phi_N}{\sum_{N=1}^{\overline{N}} \phi_N}$$

and the distribution of firm sizes by

$$\tilde{\delta}_N = \frac{\tilde{\phi}_N}{N \sum_{N=1}^{\overline{N}} \frac{\tilde{\phi}_N}{N}}.$$

Now consider the expected growth rate of employment, g_N, of a firm of employment size N. The firm grows by one employee with probability $N\lambda_L$, thus

$$
\begin{aligned}
g_N &= \frac{(N+1)\, N\lambda_L + N\,(1 - N\lambda_L) - N}{N} \\
&= \lambda_L = \theta\lambda \int_{\mu_L}^{\mu_H} (1 - F_\mu(P_L))\, dH(\mu)
\end{aligned}
$$

Hence, the expected growth rate of firms is just given by the probability per worker of its employees generating an idea that is used. This probability is constant, so the expected growth rate in terms of employees of existing firms is constant, which is a statement of Gibrat's Law. Therefore, the model seems to be consistent with the evidence in Sutton (1997).

Proposition 9 *The expected growth rate in employment size of existing firms is independent of their size.*

Similarly, the expected growth rate in average revenue of a firm with average revenue S and N employees is given by

$$
\begin{aligned}
g_{S,N} &= \frac{\theta\lambda N \int_{\mu_L}^{\mu_H} \int_{P_L}^{\infty} \left(\frac{NS+P}{N+1}\right) dF_\mu(P)\, dH(\mu) + S\,(1 - N\lambda_L) - S}{NS} \\
&= \frac{\frac{\theta\lambda N}{N+1} \int_{\mu_L}^{\mu_H} \int_{P_L}^{\infty} P dF_\mu(P)\, dH(\mu) + \left(\frac{NS}{N+1}\right) N\lambda_L + S\,(1 - N\lambda_L) - S}{NS} \\
&= \frac{1}{N+1}\left[\frac{\theta\lambda}{S}\int_{\mu_L}^{\mu_H}\int_{P_L}^{\infty} P dF_\mu(P)\, dH(\mu) - \lambda_L\right].
\end{aligned}
$$

This implies that, as μ_L and P_L, and therefore λ_L, are independent of N that

$$E_S\,(g_{S,N}) = 0.$$

So average growth rates across firms of different average revenues are zero. However, large firms that have had good realizations and therefore have a high S will tend to grow slower, and vice versa. In this sense there will be reversion to the mean, conditional on number of employees. Also note that the variance of $g_S(S, N)$ is decreasing in N, since the larger the firm, the smaller the contribution of new projects. Since firms implement projects that yield only non-negative profits, this implies that the growth rate of total revenue or total profits will decline with size.

Note that since

$$\phi_N = \phi_{N-1} \frac{\lambda_L N}{\lambda_H + \lambda_L (N + 1)}$$

it is immediate that as $\overline{N} \to \infty$ or $\lambda \to 0$, when $N \to \overline{N}$, $\phi_N \approx \phi_{N-1}$, so the share of workers at large firms is approximately constant. This implies that the density of firm sizes will be proportional to $1/N$ as N becomes large. That is, the tail of the distribution will be arbitrarily close to the tails of a Pareto distribution with coefficient one. Similarly if λ_H is small, $\phi_N \approx \phi_{N-1} (N/(N+1))$, and so for N large $\phi_N \approx \phi_{N-1}$ and the distribution of firm sizes is approximately Pareto with coefficient one. This is interesting given that several authors have concluded that the upper tail of the distribution of enterprise sizes is close to a Pareto distribution with coefficient one (see, for example, Axtell (2001)). We summarize these results in the following proposition.

Proposition 10 *As $\lambda \to 0$, or $\overline{N} \to \infty$, the density of firm sizes is arbitrarily close to the density of a Pareto distribution with coefficient one, for large enough firm sizes. Furthermore, the distribution of firm sizes is closer to a Pareto distribution with coefficient one, the smaller the mass of workers in new firms, λ_H.*

The invariant distribution of firm sizes, as well as any other outcome of the model, is a function of the exogenous parameters and distributions in the model, namely, β, w, and the distributions F_μ and H. However, as we show above, the effect of all those variables can be summarized through the values of λ_L and λ_H. We can therefore assign particular values to these two variables and compute the resulting distribution of employment shares and firm sizes. Figure 2 illustrates the invariant distribution in this model and compares it with the distribution of firm sizes in 2000 in the US. In order to compute the distribution given in equation (13), we need to truncate the distribution of firm sizes at a certain size. We choose $\overline{N} = 500000$, since the largest firms reported in the data have this number of employees. We choose $\lambda_H/\lambda_L = 1/9$ and so 90% of the new employees are hired by existing firms and 10% by new firms. As is evident from equations (11) and (12), the distribution depends only on the ratio λ_H/λ_L and not on λ_H and λ_L separately.
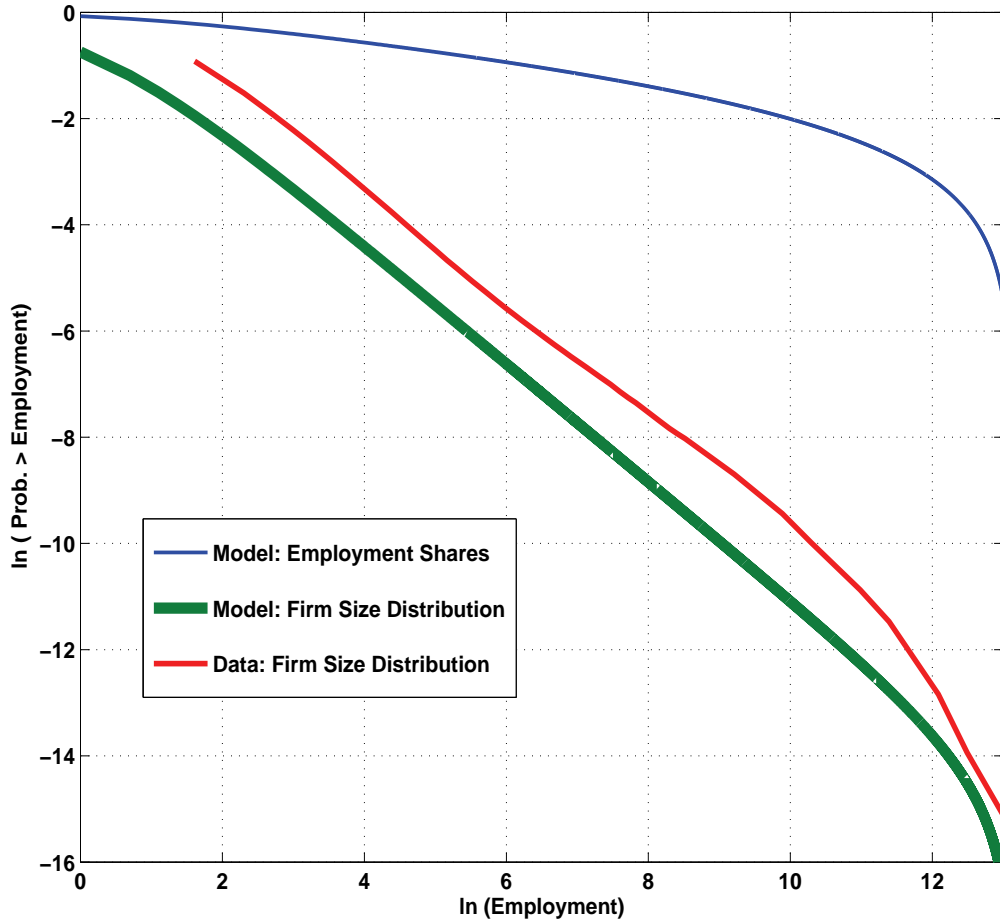
Figure 2: Model vs. Data

Figure 2 shows how our model can do a good job in matching the distribution of firm sizes. Relative to a Pareto distribution with coefficient one (a straight line with slope minus one), it exhibits a relatively thinner tail of small firms. Furthermore, since in the model and in the data we are truncating the distribution, both distributions exhibit thinner tails than the Pareto distribution for very large sizes. This is only the result of truncation in the model. If we let $\overline{N} \to \infty$, then the theory implies that the upper tail will be arbitrarily close to the Pareto distribution for large enough sizes. Similarly, in the data the census does not reveal the sizes of the largest firms because of confidentiality concerns. Furthermore, while in the model we do not have integer constraints and so there are some firms at all sizes, in the data there cannot be any fractional firms, which truncates the distribution as well. Hence, the reason to have a small mass of large firms on the

27

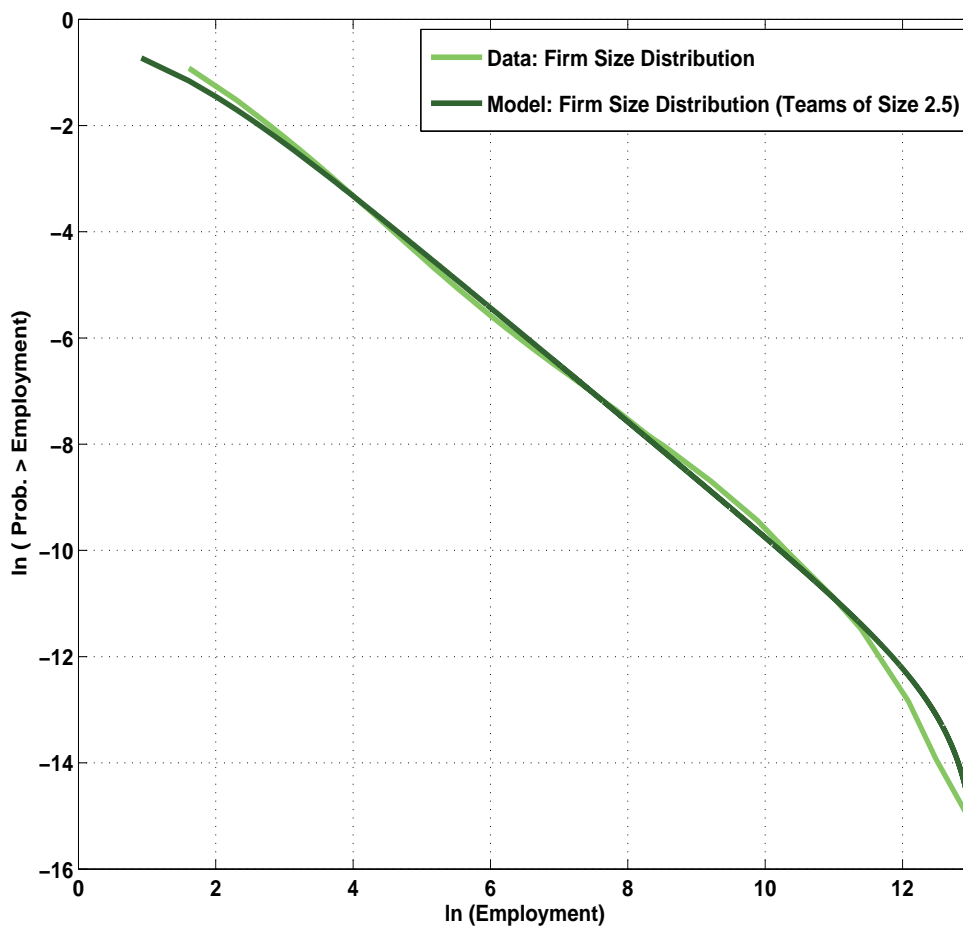upper tail is similar in the model and the data.

Figure 3: Model vs. Data with Variable Team Size

One noticeable difference between the distribution generated by the model and the distribution generated in the data is that the theoretical distribution lies below the empirical one in Figure 2. The reason for this is that in our setup, ideas are generated by one employee and not by teams of employees. Similarly individual agents spin off, not teams of agents. This is clearly not true in the data. Firms enter small, but not necessarily with one employee. Were we to assume that teams of between 2 and 3 employees have ideas and spin off, we could shift the theoretical curve in Figure 2 so that it lies on top of the empirical one. Thus, the emphasis is on the shape, not the level of the curves. This exercise is illustrated in Figure 3.

Figure 3 uses a value of $\lambda_H / \lambda_L = 0.0736$. This value is the empirical counterpart of λ_H / λ_L

in the US economy from 1989 to 2003. Namely, we calculated the net number of workers added through new firms and divided it by the net number of workers added through existing firms. The data come from the US Small Business Administration.
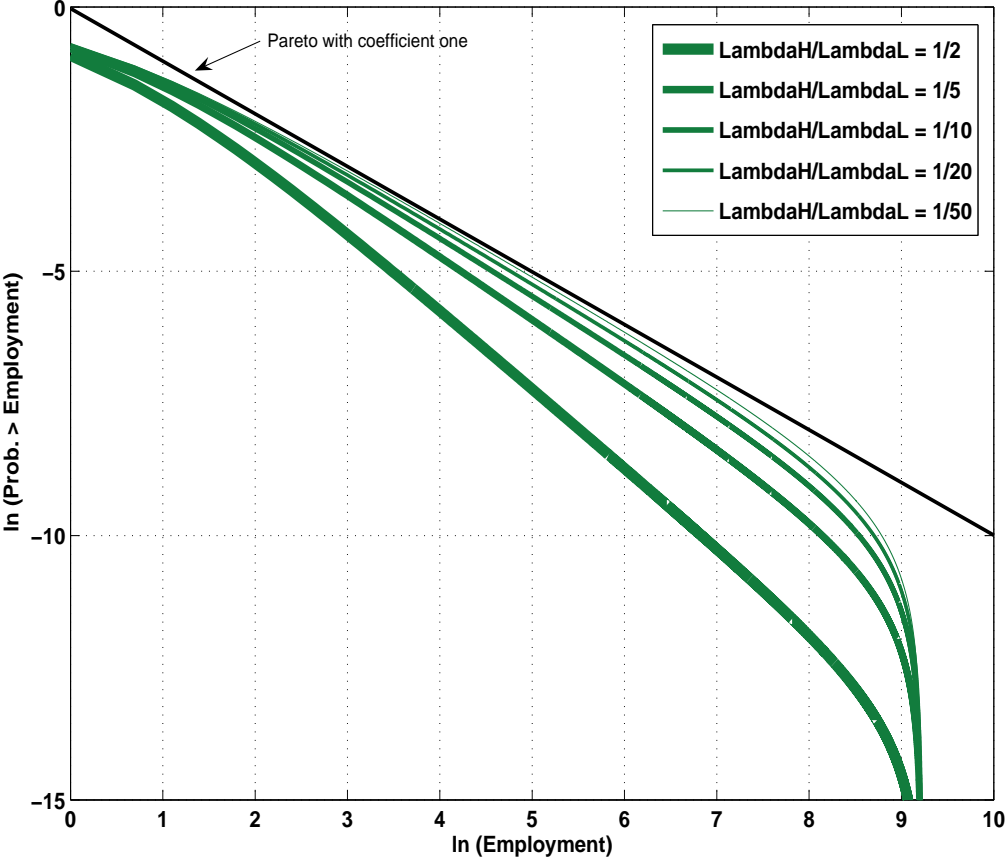

Figure 4: Comparative Statics

Figure 4 shows how we can modify the shape of the invariant distribution by changing the ratio λ_H/λ_L. We illustrate this using $\overline{N} = 10000$. The figure presents the cases for $\lambda_H/\lambda_L = 1/2$, $1/5$, $1/10$, and $1/20$. It is clear from the figure that as we increase the number of entrants (by increasing λ_H/λ_L), we shift mass to the lower tail, and therefore, the slope of the curve in Figure 4 becomes steeper. The distribution is farther way from a Pareto distribution with coefficient one. However, as we know from the previous proposition, if one increases $\overline{N}$ to large enough values, the shape of the distribution approaches a Pareto distribution, as N increases, in all these cases. Figure 4 illustrates the invariant distribution for a large range of parameters. The empirical value of λ_H/λ_L is as calculated above, equal to 0.0736. The variation in the distribution that results for low, and

empirically relevant, values of λ_H/λ_L is, however, extremely small. Hence, the model can do a good job of matching the empirical distribution for a wide range of realistic values of λ_H/λ_L. These small differences in the distribution of firm sizes for a wide set of values of λ_H/λ_L in the empirically relevant range may be the reason why the shape of the US distribution of firm sizes has changed so little in the last 10 years.

5 Conclusion

We propose a theory of firm dynamics in which workers have ideas for new projects that can be implemented inside existing firms or, at a cost, in new firms: spin-offs. Workers have private information about the quality of their ideas. Because of an adverse selection problem, workers can sell their ideas to existing firms only at a price that is not contingent on their information. Therefore, workers with very good ideas decide to spin off and set up a new firm. Since entrepreneurs of existing firms pay a price for the ideas sold in the market that implies zero expected profits, firm growth is scale independent. This, we believe, is the main insight that comes out of the theory. It is the existence of a market for ideas that leads to scale-independent growth and distributions of firm sizes that resemble a Pareto distribution with coefficient one in the upper tail.

Essentially, the theory produces a size distribution of firms that depends on all the parameters and distributions of the model through the value of λ_H/λ_L only. The theory leads to a distribution of firm sizes that depends only on λ_H/λ_L, which can be easily calculated from available data. In Section 4 we calculated this parameter using US data and showed that one obtains a size distribution that is hard to distinguish from the empirical distribution. Furthermore, in the empirically relevant range the distribution in the model is not very sensitive to the value of λ_H/λ_L. This could explain why the distribution of firm sizes has been so stable over time.

In this paper we have abstracted from equilibrium effects that can be important in the evolution of an economy. Namely, we assumed throughout that the base wage of workers is fixed. Of course, workers also earn the return on their ideas and so part of their income is endogenous in the model. Nevertheless, we worked with a modeled economy in which new workers can always be hired at the wage w. Embedding this theory of firm dynamics and entry into an equilibrium framework, such as the neoclassical growth model, where one can study the interactions between industry evolution and wages as well as the growth of an industry or the economy would be interesting, but it is left for future research.

Furthermore, if spin-offs tend to be geographically close to their parent firm, then they are a potentially important reason why we get clusters of firms working in the same line of business in the same locality. Augmented by a location choice, our theory could then form the basis of a dynamic theory of industrial agglomeration.

References

[1] Axtell, R. L. (2001) "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (September 7): 1818-1820.

[2] Ericson, R. and A. Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (1): 53-82.

[3] Franco, April and Darren Filson. (2006) "Spin-outs: Knowledge Diffusion through Employee Mobility." Forthcoming in *Rand Journal of Economics*.

[4] Hopenhayn, H. A. (1992) "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127-1150.

[5] Jovanovic, B. (1982) "Selection and the Evolution of Industry." *Econometrica* 50 (3): 649-670.

[6] Klette T. K. and S. Kortum (2004) "Innovating Firms and Aggregate Innovation." *Journal of Political Economy* 112 (5): 986-1018.

[7] Luttmer, E. G. J. (2006) "Selection, Growth, and the Size Distribution of Firms." Forthcoming in the *Quarterly Journal of Economics*.

[8] Prusa, T. and J. Schmitz (1994) "Can Companies Maintain Their Initial Innovative Thrust? A Study of the PC Software Industry." *Review of Economics and Statistics*, 76(3): 523-540.

[9] Rossi-Hansberg, E. and M. L. J. Wright (2006) "Establishment Size Dynamics in the Aggregate Economy." Working Paper 11261, NBER.

[10] Silveira, R. and R. Wright (2007) "The Market for Ideas," mimeo, University of Pennsylvania.

[11] Simon H. and C. Bonini (1958). "The Size Distribution of Business Firms." *American Economic Review* 48(4): 607-617.

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

[13] Sutton, J. (1997) "Gibrat's Legacy." *Journal of Economic Literature* 35 (1): 40-59.

6   Appendix

In this Appendix, we prove the analog of Proposition 5 for the case of exponential utility. We are interested in this case because as in the linear case the value functions can be solved analytically which allows us to obtain an expression for the price of an idea. Furthermore, in contrast with the linear case, with exponential utility agents are not risk neutral and, most importantly, without the market for ideas the thresholds would depend on the size of firms. That is, absent a market for ideas the selection of ideas would be firm-specific. Hence, this case further illustrates the importance of the market for ideas in generating scale independent growth.

Proposition 11 *If* $u(c_t) = -ae^{-bc_t}$,

- $\mu_L(S, N)$ *is independent of* S *and* N,

- *the thresholds for using a project are given by* $P_L(S, N) = w$ *and a constant* $P_H(S, N) > w$,

- *the market price of ideas is given by*

$$
Z = \frac{1}{b} \log \left[\frac{1 + \frac{\beta}{1-\beta} \int_{\mu_L}^{\mu_H} \int_{P_L} \left(1 - e^{-b(P-w)}\right) dF_\mu(P) dH(\mu)}{1 - \int_{\mu_L}^{\mu_H} \int \left(1 - e^{-b(P-w)}\right) dF_\mu(P) dH(\mu)} \right] > 0,
$$

- *and* $\mu_L < \mu_H$, *so some ideas are implemented within existing firms and some through spin-offs.*

Proof. Guess that

$$
W(S, N) = -ae^{-b[\pi(S,N)+w]} f(N)
$$

for some function $f(N)$ independent of S. Substitute the guess to get

$$
\begin{aligned}
-ae^{-b[\pi(S,N)+w]} f(N) &= \gamma(\lambda, N) \int^{\mu_H} \max \left[\begin{array}{c} V(\mu, S, N) + ae^{-b[\pi(S,N)+w]} e^{bZ} \\ +\beta ae^{-b[\pi(S,N)+w]} f(N), 0 \end{array} \right] dH(\mu) \\
&\quad + (1 - \gamma(\lambda, N) H(\mu_H)) \left[ae^{-b[\pi(S,N)+w]} + \beta ae^{-b[\pi(S,N)+w]} f(N) \right] \\
&\quad - \gamma(\lambda, N) H(\mu_H) \left[ae^{-b[\pi(S,N)+w]} e^{bZ} + \beta ae^{-b[\pi(S,N)+w]} f(N) \right]
\end{aligned}
$$

and

$$
\begin{aligned}
V(\mu, S, N) &= \int -ae^{-b[\pi(S,N)+w]} e^{bZ} e^{-b(P-w)} dF_\mu(P) + \\
&\quad \beta \int \max \left[-ae^{-b[\pi(S,N)+w]} e^{-b(P-w)} f(N+1), -ae^{-b[\pi(S,N)+w]} f(N) \right] dF_\mu(P).
\end{aligned}
$$

33

Then,

$$
\begin{aligned}
f(N) &= -\frac{\gamma(\lambda, N)}{1-\beta} \int^{\mu_H} \max \left[\begin{array}{c} -\int e^{bZ} e^{-b(P-w)} dF_\mu(P) \\ +\beta \int \max \left[-e^{-b(P-w)} f(N+1), -f(N) \right] dF_\mu(P) \\ +e^{bZ} + \beta f(N), 0 \end{array} \right] dH(\mu) \\
&\quad + \frac{1}{1-\beta} + \frac{\gamma(\lambda, N) H(\mu_H)}{1-\beta} \left(e^{bZ} - 1 \right)
\end{aligned}
$$

which does not depend on S. This verifies our guess. Now guess that

$$
f(N) = \frac{1}{1-\beta}.
$$

Substituting this guess, we obtain that

$$
\begin{aligned}
e^{bZ} - 1 &= e^{bZ} \int_{\mu_L}^{\mu_H} \int \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu) \\
&\quad + \frac{\beta}{1-\beta} \int_{\mu_L}^{\mu_H} \int_{P_L} \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu)
\end{aligned}
$$

Hence Z needs to satisfy this equation for our guess to be correct. Or,

$$
e^{bZ} = \frac{1 + \frac{\beta}{1-\beta} \int_{\mu_L}^{\mu_H} \int_{P_L} \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu)}{1 - \int_{\mu_L}^{\mu_H} \int \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu)}
$$

which implies that

$$
Z = \frac{1}{b} \log \left[\frac{1 + \frac{\beta}{1-\beta} \int_{\mu_L}^{\mu_H} \int_{P_L} \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu)}{1 - \int_{\mu_L}^{\mu_H} \int \left(1 - e^{-b(P-w)} \right) dF_\mu(P) dH(\mu)} \right].
$$

Note that this is exactly the condition that makes

$$
W(S, N) = \frac{-ae^{-b[\pi(S,N)+w]}}{1-\beta} = \frac{u(\pi(S, N) + w)}{1-\beta},
$$

so entrepreneurs are willing to pay up to this amount for workers' ideas, and workers will get this price or will sell the idea to another entrepreneur. Competition for ideas among entrepreneurs then guarantees that the condition is satisfied.

The threshold $\mu_L(S, N)$ is implicitly defined by

$$
V(\mu_L(S, N), S, N) = u(\pi(S, N) - Z + w) + \beta W(S, N)
$$

which can be written as

$$
\int -e^{-b(P-w)} dF_{\mu_L}(P) + e^{-bZ} \frac{\beta}{1-\beta} \int \max \left[1 - e^{-b(P-w)}, 0 \right] dF_{\mu_L}(P) = -1
$$

34

or

$$\int -e^{-b(P-w)}dF_{\mu_L}(P) + e^{-bZ}\frac{\beta}{1-\beta}\int_{P_L}\left[1 - e^{-b(P-w)}\right]dF_{\mu_L}(P) = -1$$

which determines μ_L as a constant.

P_L is then determined by

$$W\left(\frac{NS + P_L}{N+1}, N+1\right) = W(S, N)$$

or

$$\frac{e^{-b[P_L - w]}}{1 - \beta} = \frac{1}{1 - \beta}$$

and so $P_L = w$.

We still need to show that $P_H = w$. As before guess that

$$W_0 = -ae^{-bw}f_0$$
$$W(P, 1) = \frac{-ae^{-bP}}{1-\beta}$$

then

$$V_0(\mu) = \int -ae^{-bw}e^{-b(P-w)}dF_\mu(P) + \beta \int \max\left[-ae^{-bw}e^{-b(P-w)}\frac{1}{1-\beta}, -ae^{-bw}f_0\right]dF_\mu(P)$$
$$= ae^{-bw}\left[-\int e^{-b(P-w)}dF_\mu(P) + \beta \int \max\left[-e^{-b(P-w)}\frac{1}{1-\beta}, -f_0\right]dF_\mu(P)\right]$$

and from the expression for W_0 ,

$$f_0 = \frac{1}{1-\beta} -$$
$$\frac{\lambda}{1-\beta}\int \max\left[\begin{bmatrix} \int\left(1 - e^{-b(P-w)}\right)dF_\mu(P)+ \\ \frac{\beta}{1-\beta}\int \max\left[f_0 - e^{-b(P-w)}, 0\right]dF_\mu(P) \\ , -e^{-bZ} \end{bmatrix}\right]dH(\mu)$$

which determines f_0 as a positive constant, where $f_0 < 1/(1-\beta)$ and verifies our guess.

The threshold μ_H is determined by

$$V_0(\mu_H) = u(w + Z) + \beta W$$

which implies that

$$\int -e^{-b(P-w)}dF_{\mu_H}(P) + \frac{\beta}{1-\beta}\int \max\left[-e^{-b(P-w)} + (1-\beta)f_0, 0\right]dF_{\mu_H}(P) = -e^{-bZ}.$$

35

The equation above implies that P_H satisfies

$$e^{-b(P_L - w)} = (1 - \beta) f_0 < 1$$

since $f_0 < 1/(1 - \beta)$. Hence, we conclude that $P_L > w$.

To show that $\mu_H > \mu_L$, we need to compare the equations that determine these thresholds, namely,

$$
\begin{aligned}
\int e^{-b(P-w)} dF_{\mu_L}(P) &= 1 + e^{-bZ} \frac{\beta}{1 - \beta} \int_{P_L} \left[1 - e^{-b(P-w)} \right] dF_{\mu_L}(P) \\
\int e^{-b(P-w)} dF_{\mu_H}(P) &= e^{-bZ} + \frac{\beta}{1 - \beta} \int_{P_H} \left[(1 - \beta) f_0 - e^{-b(P-w)} \right] dF_{\mu_H}(P)
\end{aligned}
$$

But note that

$$\int_{P_H} \left[(1 - \beta) f_0 - e^{-b(P-w)} \right] dF_{\mu_H}(P) < \int_{P_L} \left[1 - e^{-b(P-w)} \right] dF_{\mu_L}(P) < 1$$

since $(1 - \beta) f_0 < 1$ and $P_H > P_L$. Hence,

$$\int e^{-b(P-w)} dF_{\mu_H}(P) < \int e^{-b(P-w)} dF_{\mu_L}(P)$$

which implies that $\mu_H > \mu_L$. ∎