Limited Absorptive Capacities and the Emergence of Convergence Clubs

Master Thesis
supervised by the

Institute for Empirical Research in Economics (IEW)
at the University of Zurich

Prof. Dr. Fabrizio Zilibotti

to obtain the degree of
“Master of Arts in Wirtschaftswissenschaften”

Author:	Michael David König
Course of Studies:	Volkswirtschaftslehre
Student ID:	05-901-756
Address:	Gotthardstrasse 7
	8800 Thalwil
E-Mail:	mkoenig@ethz.ch
Closing date:	June 10, 2010

Abstract

We develop a simple and tractable model of productivity growth and technology spillovers that can explain the emergence of real world empirical productivity distributions. We assume that the outcomes of firms' in-house R&D efforts are governed by a stochastic growth process that depends on the current technology level of the firm. Moreover, firms can imitate other firms' technologies subject to their absorptive capacities. We show that the combined process of in-house innovation and imitation gives rise to balanced growth with persistent productivity differences even when starting from ex ante identical firms. We show that along the balanced growth path the emerging productivity distribution can be described as a traveling wave. Further, we take into account that firms are boundedly rational and can only imperfectly predict the outcomes of their innovation and imitation strategies. We show that this limited rationality can enhance industry performance and efficiency.

[1] **Acknowledgement:** I would like to thank Fabrizio Zilibotti for supervision and comprehensive support, and Jan Lorenz for assistance and collaboration in the development of the theory and its numerical analysis.

June 10, 2010

Contents

1. Introduction

Many empirical studies report persistent inequalities in per capita income and productivity across countries [e.g. Durlauf, 1996; Durlauf and Johnson, 1995; Feyrer, 2008; Quah, 1993, 1996, 1997]. A prominent explanation for these productivity differences is that they stem from differences in technological knowledge [Prescott, 1998; Romer, 1993].[2] On one hand, differences in technological know-how originate from a large variation in R&D investments across firms and the diverse outcomes of these R&D activities [Coad, 2009; Cohen and Klepper, 1992, 1996; Cohen et al., 1987]. On the other hand, these differences originate from barriers to technology adoption and knowledge diffusion between firms [Eeckhout and Jovanovic, 2002; Geroski, 2000; Stoneman, 2002].

Even though an increasingly globalized world and the successive advancement of communication technologies should make it easier for technological improvements to spillover from one firm to another (or from one country to another), technology adoption still involves many challenging features, which consolidate technological gaps between firms, industries and countries [Acemoglu, 2007]. Technology adoption is closely related to the R&D activities of firms. In the course of their research activities firms can develop the ability to assimilate and exploit other existing technologies and thereby increase their "absorptive capacities" [Cohen and Levinthal, 1989; Kogut and Zander, 1992; Nelson and Phelps, 1966]. However, there exist limitations to their absorptive capacities. If a technology is too advanced compared to the current technological level of the firm it becomes difficult or even impossible to imitate it [Powell and Grodal, 2006].[3]

In this paper we argue that it is the combined process of technology development trough in-house R&D and the imitation of external technological knowledge by taking into account limitations in a firm's absorptive capacity that eventually gives rise to persistent productivity differences among firms as they can be found in empirical studies. We analyze empirical productiv-

[2] For an alternative explanation of productivity differences see e.g. Acemoglu and Zilibotti [2001].

[3] There exists a vast literature on barriers to technology adoption. Some of the more recent contributions include Acemoglu et al. [2010]; Acemoglu and Zilibotti [2001]; Aghion et al. [2005]; Barro and Sala-i Martin [1997]; Eaton and Kortum [2001]; Hall and Jones [1999]; Howitt [2000].

ity distributions and their evolution over time and develop a simple model that can explain the emergence of these distributions.

We analyze a large data set containing information of over six million firms in the period between 1992 to 2005. In line with previous authors [Corcos et al., 2007; Di Matteo et al., 2005] we find that the productivity distributions over these firms exhibit power-law tails over all periods of time. Moreover, we can observe an increasing trend in the average productivity.

Building on our empirical findings we introduce a model of technological change and innovation that is able to reproduce these empirically observed productivity distributions. We introduce a two-sector model of monopolistic competition of intermediate goods producing firms and competitive final good production akin to Acemoglu et al. [2006]. Technology levels and innovation follows a quality ladder approach [Aghion and Howitt, 1992; Grossman and Helpman, 1991]. Imitation takes place between intermediate goods producing firms in different sectors [Fai and Von Tunzelmann, 2001; Rosenberg, 1976].

A distinctive feature of our model is that we explicitly take into account the endogenous decisions of firms whether to undertake in-house R&D or to imitate other firms' technologies. The success of their imitation strategies depends on the availability of better technologies (which depends on the current productivity distribution) and their absorptive capacities. The explicit formulation of firms' R&D behavior distinguishes our model from previous ones in the literature. Early contributions focusing on firm size and growth rate distributions like Gibrat [1931]; Pareto [1896]; Simon [1955] as well as more recent ones by Fu et al. [2005]; Stanley et al. [1996] do not take into account R&D decisions of firms. Ensuing models such as Klette and Kortum [2004]; Luttmer [2007] explicitly model firms' R&D effort decisions but do not incorporate the trade off firms face between making an innovation in-house or copying it from another firm.

Starting from ex ante identical firms our model generates heterogeneous productivity distributions with power-law tails. These productivity distributions translate into Zipf's law firm size distributions which have been observed in numerous empirical studies [e.g. De Wit, 2005; Gabaix, 1999; Saichev et al., 2009].

The outcomes of firms R&D activities are uncertain at the date when the R&D decisions are made. We assume that firms are risk averse in such an uncertain environment [Sandmo, 1971]. Moreover, in this model, we take into account that firms might have only a noisy perception of the future outcome of their R&D investments and the success of their technology adoption efforts [Nelson and Winter, 1982; Silverberg and Verspagen, 1994]. We analyze different scenarios of this noise and show that a high level of noise can actually increase industry performance and efficiency.

The paper is organized as follows. The empirical analysis of firm productivities is given in Section 2. The model of firm R&D behavior is introduced in Section 3 and the evolution of the productivity distributions generated by this model is analyzed in Section 4. In Section 5 we analyze the conditions improving industry performance. The proofs of all propositions and corollaries can be found in Appendix A. A number of possible extensions of the model is given in Appendix B. In Section 6 we conclude.

2. Empirical Analysis

Our sample of the Amadeus database provided by Bureau van Dijk contains a total of $6,5447,38$ European firms and spans a time period from 1992 to 2005. We have eliminated missing values in the data and computed operating revenues per worker as a measure of firm productivity A. Restricting our data set to years where we did not observe a large drop in the average number of firms at the beginning and end of the observed periods (which is probably due to the data collecting process) we obtained a panel of firm productivities in the years 1995 to 2004. Some descriptive statistics are shown in Table 1. As the table reveals, the data sample exhibits a large variance σ_A^2, with the maximum productivity $A_{\max}$ being much larger than the average μ_A.

The resulting productivity distributions for the years 1995 to 2004 and the corresponding average productivities (arithmetic, geometric mean and median) are shown in Figure 1. As can be seen from Figure 1, the productivity distributions over different years are well characterized by power-law tails with an exponent of minus two. The cutoff at lower productivity levels is due to data limitations which do not consider output below a threshold level. Moreover, the upward trend in the geometric mean and the median of the productivity values suggest a slow increase in aggregate productivity.

Motivated by the distributions shown in Figure 1, we estimate a power-

Table 1: Descriptive statistics for the years 1995 to 2004.

year	N	μ_A	σ_A	$A_{\max}$
1995	513358	209.4707	1055.0	119886
1996	673103	224.6385	1069.5	134859
1997	877347	232.0827	1150.4	133521
1998	1271199	233.2350	1201.4	137000
1999	1498458	243.9753	1300.0	156010
2000	1659786	257.3738	1400.6	148717
2001	1956456	249.3301	1410.3	161474
2002	2123401	255.1301	1389.7	135008
2003	1718683	241.2756	1268.5	116992
2004	21673	218.7708	1005.6	73062

The total number of observations in the panel is $12, 313, 464$.

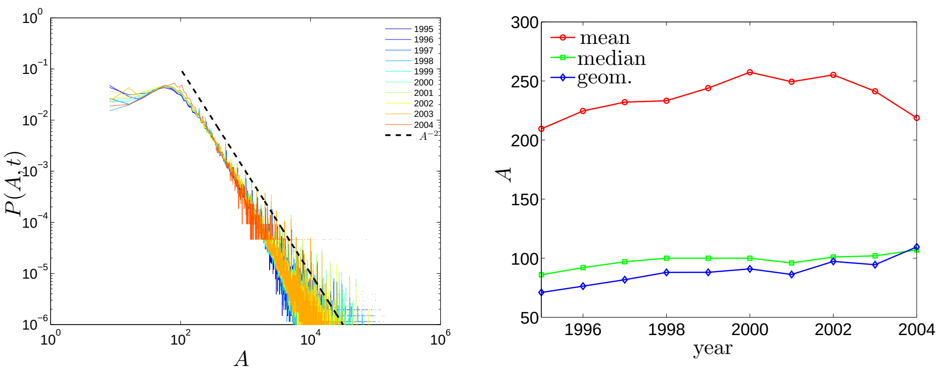

Figure 1: (Left) Productivity distribution in the years 1995 to 2004. The dashed line indicates a power-law $P(A,t) \propto A^{-2}$ with exponent minus two and support $A \geq A_{\min} = 10^2$. The figure suggest that the distributions are close to a power-law with an exponent of around two for $A \geq A_{\min} = 10^2$. (Right) The average productivity (arithmetic, geometric mean and median) is increasing over the periods 1995 to 2004.

law of the form

$$P(A,t) = \frac{\lambda(t)-1}{A_{\min}(t)} \left(\frac{A}{A_{\min}(t)}\right)^{-\lambda(t)}, \tag{1}$$

for each year $t = 1995, \ldots, 2004$ with a cut-off $A_{\min}$. The cut-off $A_{\min}$ is the productivity below which we cannot reasonably assume that the distribution is described by a power-law. Our estimation procedure follows the one suggested by Clauset et al. [2009].[4] The estimation results for the exponent λ and the cut-off $A_{\min}$ are shown in Table 2. The estimates for the exponent λ all indicate an exponent which is slightly above two.

Table 2: Estimation results for the power-law exponents λ and the cut-off $A_{\min}$ for the years 1995 to 2004.

year	N_{tail}	$\hat{\lambda}$		$\hat{A}_{\min}$	
1995	513358	2.36***	(0.00)	286.00***	(22.39)
1996	673103	2.40***	(0.02)	1539.00***	(245.70)
1997	877347	2.37***	(0.03)	1552.00***	(504.16)
1998	1271199	2.33***	(0.05)	1546.00*	(960.74)
1999	1498458	2.28***	(0.03)	2750.00***	(633.24)
2000	1659786	2.35***	(0.04)	592.00	(518.01)
2001	1956456	2.26***	(0.04)	2650.00***	(776.24)
2002	2123401	2.24***	(0.02)	2574.00***	(585.84)
2003	1718683	2.31***	(0.01)	1026.00*	(675.51)
2004	21673	2.44***	(0.017)	216.00***	(9.573)

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ result from a two-tailed z-test under the null-hypothesis of parameter being zero. Variances are shown in parentheses. N_{tail} gives the number of data points used for the estimation of the power-law parameters.

We note here that Corcos et al. [2007] have estimated the productivity

[4] Assuming that our sample is generated by a power-law distribution for values of $A \geq A_{\min}$ the maximum likelihood estimator of the exponent λ is given by [Muniruzzaman, 1957]

$$\hat{\lambda} = 1 + N \left(\sum_{i=1}^{N} \ln \frac{A_i}{A_{\min}}\right)^{-1},$$

where A_i, $i = 1, \ldots, N$ are the observed productivity values. For the estimation of the cut-off $A_{\min}$ and the variances see Clauset et al. [2009].

distributions using the same data set, while also controlling for physical and human capital. Similar to our results, these authors find that the distributions are well described by a power-law with an exponent of two. They show that this result is also robust when disaggregating over different sectors.

In the following sections we will introduce a model that is able to generate productivity distributions with power-law tails as we have found them in our empirical analysis.

3. The Model

We introduce a model of intermediate and final goods sectors in Section 3.1 and later combine it with the productivity dynamics from Section 3.2. As we will show, productivity differences directly translate to intermediate good output differences and profits.

3.1. Agents and Production

We consider a two-periods overlapping generations model with $n = 2L$ agents.[5] Each agent lives for two periods supplying one unit of labor in the first period of life and none in the second. There is one final good $Y(t)$, produced under perfect competition by labor L and a set of intermediate goods $x_i(t)$, $i \in N = \{1, 2, ..., n\}$, according to the production function

$$Y(t) = \frac{1}{\alpha} L^{1-\alpha} \sum_{i=1}^{n} A_i(t)^{1-\alpha} x_i(t)^{\alpha}, \quad \alpha \in (0, 1),$$

where $x_i(t)$ is the economy's input of intermediate good i at time t and $A_i(t)$ is the economy's productivity in sector i at time t. The general good $Y(t)$ is used for consumption, as an input to R&D and also as an input to the production of intermediate goods.

Producers of the general good act as perfect competitors in all markets, so the equilibrium price of each intermediate good is its marginal product in producing the general good

$$p_i(t) = \left(\frac{A_i(t)}{x_i(t)} \right)^{1-\alpha},$$

[5] The way in which we model agents and production follows closely Acemoglu et al. [2006].

where we have set $L = 1$ and take the price of the final good as numeraire.

For each intermediate good i there is a large number of firms capable of producing it, using χ units of the general good per unit of output, with $1 < \chi \leq 1/\alpha$. A higher value of χ indicates less competition. In the intermediate goods sectors there are single producers, the incumbents, who can produce using only one unit of the general good per unit of output. All intermediate goods producers compete in Bertrand price competition and the equilibrium price is equal to the unit cost of the second most efficient producer[6]

$$p_i(t) = \chi,$$

where the incumbent monopolizes the market for the intermediate good i.

Solving for the equilibrium quantity $x_i(t)$ yields

$$x_i(t) = \chi^{-\frac{1}{1-\alpha}} A_i(t).$$

The profit earned by the incumbent in any intermediate sector i will then be proportional to the productivity in that sector

$$\pi_i(t) = (p_i(t) - 1)\, x_i(t) = \frac{\chi - 1}{\alpha} \chi^{-\frac{1}{1-\alpha}} A_i(t) = \psi A_i(t). \tag{2}$$

The factor $\psi = \frac{\chi-1}{\alpha}\chi^{-\frac{1}{1-\alpha}}$ is monotonically increasing in α, that is, stronger competition reduces profits. From the above expression for profits we also find that productivity differences are responsible for differences in the supply of the intermediate good. Finally, aggregate output is proportional to aggregate productivity as follows

$$Y(t) = \frac{1}{\alpha}\chi^{-\frac{\alpha}{1-\alpha}} \sum_{i=1}^{n} A_i(t) = \frac{1}{\alpha}\chi^{-\frac{\alpha}{1-\alpha}} A,$$

where aggregate productivity is $A(t) = \sum_{i=1}^{n} A_i(t)$. At this point we do not discuss the above model implications any further and refer to Acemoglu et al. [2006] for a more extensive discussion. Instead, we will focus on the evolution of productivity values in the next section.

[6] The most efficient producer can always increase its price up to the marginal cost of the second most efficient producer without decreasing its demand. The inequality $\chi \leq 1/\alpha$ guarantees that he incumbent has to charge a limit price in order to prevent entry.

3.2. Technological Change and Productivity Growth

Following the literature on economic growth and quality ladders (see e.g. Aghion and Howitt [1992] and Grossman and Helpman [1991]), we assume that the productivity of an intermediate good producing firm $i \in N$ takes on values along a quality ladder with rungs spaced proportionally by a factor $\bar{A} > 1$.[7] Productivity starts at $\bar{A}^0 = 1$ and the subsequent rungs are $\bar{A}^1, \bar{A}^2, \bar{A}^3, \ldots$. A firm i, which has achieved a_i productivity improvements then has productivity $A_i = \bar{A}^{a_i}$.

Consider a firm with productivity $A(t) = \bar{A}^a$ at time t and assume that its productivity at time $t+\Delta t$ is $A(t+\Delta t) = \bar{A}^{a+1}$. The productivity growth rate g of the firm at time t is then

$$g = \frac{A(t+\Delta t) - A(t)}{A(t)} = \frac{\bar{A}^{a+1} - \bar{A}^a}{\bar{A}^a} = \bar{A} - 1,$$

and thus $1 + g = \bar{A}$.

Firm i's productivity $A_i \in \{1, \bar{A}, \bar{A}^2, \ldots\}$ grows as a result of innovations, either undertaken in-house or due to the imitation and absorption of the technologies of other firms. The technology comes from firms in other sectors that were successful in innovating in their area of activity [Fai and Von Tunzelmann, 2001; Rosenberg, 1976]. At time step $t = \Delta t, 2\Delta t, 3\Delta t, \ldots$, $\Delta t > 0$, a firm i is selected at random and either decides to imitate another firm or to conduct in-house R&D, depending on which of the two gives it higher expected profits.[8]

If firm i conducts in-house R&D at time t then it makes $\eta(t)$ productivity improvements and its productivity changes as follows

$$A_i(t+\Delta t) = \bar{A}^{a_i(t)+\eta(t)} = A_i(t)\bar{A}^{\eta(t)}. \tag{3}$$

$\eta(t) \geq 0$ is a non-negative integer-valued random variable with a certain distribution. Let us denote $\eta_b = \mathbb{P}(\eta(t) = b)$ for $b = 0, 1, 2, \ldots$ to quantify the distribution. It holds $\sum_{b=0}^{\infty} \eta_b = 1$. From the productivity growth dynamics above we can go to an equivalent system by changing to the normalized

[7] See also Chapter 7 in Barro and Sala-i Martin [2004].

[8] We will explain in more detail the innovation and imitation process in Section 4. There we will also assume that firms are risk averse and perceive expected profits with noise when deciding between innovation and imitation [Sandmo, 1971].

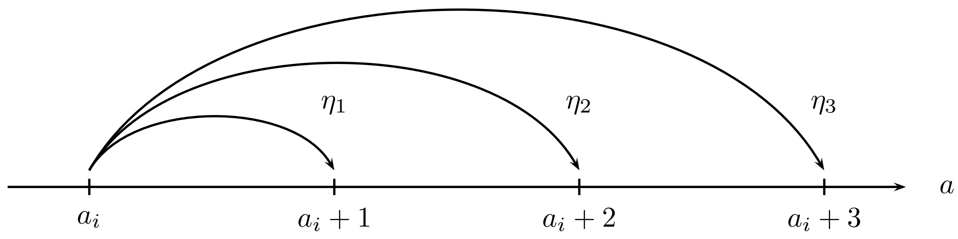

Figure 2: Illustration of the innovation process of firm i with log-productivity $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). With probability η_1 firm i makes one productivity improvement and advances by one log-productivity unit, with probability η_2 firm i makes two productivity improvements and advances by two log-productivity units, etc..

log-*productivity* $a_i(t) = \log A_i(t)/\log \bar{A}$. Then the in-house update map in Equation (3) is given by

$$a_i(t + \Delta t) = a_i(t) + \eta(t). \tag{4}$$

In the following we will consider log-productivity to be always normalized by $\log \bar{A}$. An illustration of this productivity growth process can be seen in Figure 2. Note that log-productivity undergoes a simple stochastic process with additive noise, while productivity follows a stochastic process with multiplicative noise [Karlin and Taylor, 1975, 1981], with the stochastic factor being the random variable $\bar{A}^{\eta}$. In the limit of continuous time we obtain a geometric Brownian motion for productivity [Saichev et al., 2009, pp. 9].

In the case of imitation, firm i with productivity $A_i(t)$ selects another firm $j \in N$ at random and attempts to imitate its productivity $A_j(t)$ but only when $A_j(t) > A_i(t)$, which is equivalent to $a_j(t) > a_i(t)$.

Conditional that firm i has selected a firm j with higher log-productivity, we assume that the firm tries to climb the rungs of the quality ladder which separates it from $a_j(t)$. We assume that each rung is climbed with success probability q, but the attempt finishes after the first failure. This reflects the fact that knowledge accumulation is cumulative and the growth of knowledge builds on the already existing knowledge base [Kogut and Zander, 1992; Weitzman, 1998].

Taking the above mentioned process of imitation more formally, firm i's

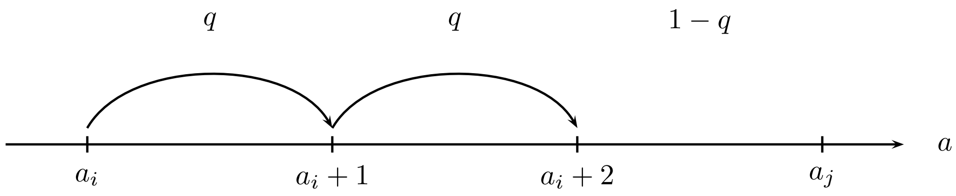

Figure 3: Illustration of the imitation of log-productivity a_j of firm j through firm i with log-productivity a_i, where the log-productivity of firm i is $\log A_i = a_i \log \bar{A} = a_i$ (setting $\log \bar{A} = 1$). Firm i successfully imitates two log-productivity units (with probability q^2) but fails to imitate the third log-productivity unit (with probability $1-q$). It then ends up with a log-productivity of $a_i + 2$.

productivity changes according to

$$A_i(t + \Delta t) = A_i(t)\bar{A}^{\kappa} = \bar{A}^{a_i(t)+\kappa}, \tag{5}$$

where κ is a random variable which takes values in $\{0, 1, 2, \ldots a_j(t) - a_i(t)\}$ and denotes the number of rungs to be climbed towards $a_j(t)$. The distribution of κ depends on the distance $a_j(t) - a_i(t)$ and is quantified as

$$\mathbb{P}(\kappa = k) = \begin{cases} q^k(1-q) & \text{if } 0 \leq k < a_j(t) - a_i(t), \\ q^k & \text{if } k = a_j(t) - a_i(t), \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Note, that it hold $\sum_{k=0}^{\infty} P(\kappa = k) = 1$, as necessary. For $q = 0$ it holds $A_i(t + \Delta t) = A_i(t)$, for $q = 1$ it holds $A_i(t+\Delta t) = A_j(t)$ while for $0 < q < 1$ it holds that $A_i(t) \leq A_i(t + \Delta t) \leq A_j(t)$. This motivates us to call the parameter q a measure of the *absorptive capacities* of the firms. The higher q, the better firms are able to climb rungs on the quality ladder.

Changing to normalized log-productivity in Equation (5) we get

$$a_i(t + \Delta t) = a_i(t) + \kappa. \tag{7}$$

An illustration of this imitation process can be seen in Figure 3.

If firm i with log-productivity $a_i(t)$ attempts to imitate firm j with log-productivity $a_j(t) > a_i(t)$ then the expected log-productivity i obtains is given by

$$\begin{aligned}\mathbb{E}\left(a_i(t+\Delta t)|a_i(t)=a, a_j(t)=b\right) &= a(1-q)+(a+1)q(1-q)+(a+2)q^2(1-q)\\&\quad+\cdots+(b-1)q^{b-a-1}(1-q)+bq^{b-a}\\&=\sum_{c=0}^{b-a-1}(a+c)(1-q)q^c+bq^{b-a}\\&=a+q\frac{1-q^{b-a}}{1-q}.\end{aligned}$$

For $q<1$ and b being much larger than a we get

$$\mathbb{E}\left(a_i(t+\Delta t)|a_i(t)=a, a_j(t)=b\right) \approx a+\frac{q}{1-q},$$

and the log-productivity firm i obtains through imitation does not depend on the log-productivity of firm j but only its success probability q. However, it depends on the log-productivity of firm j if $a_j(t)$ is close to $a_i(t)$. The latter becomes effective for example for firms with a high productivity when there are only few other firms remaining with higher productivities which could be potentially imitated.

3.3. Related Literature on Technological Change and Economic Growth

Our productivity growth function is related to other prominent models in the literature on economic growth and technological change. For example, Howitt and Mayer-Foulkes [2005] study a productivity growth equation of the following form

$$A_i(t+\Delta t)=\begin{cases}\bar{A}(t+\Delta t) & \text{with probability } \mu(t),\\ A_i(t) & \text{with probability } 1-\mu(t),\end{cases} \tag{8}$$

where μ is a parameter and $\bar{A}(t+\Delta t)$ is the maximum productivity level in the industry at time $t+\Delta t$. The expected level $A_i(t)$ then obeys

$$A_i(t+\Delta t)=\mu(t)\bar{A}(t+\Delta t)+(1-\mu(t))A_i(t).$$

Subtracting $A_i(t)$ on both sides of the above equation leads to

$$\Delta A_i(t)=\mu(t)\left(\bar{A}(t+\Delta t)-A_i(t)\right).$$

A similar productivity growth dynamics can be found in a number of models such as in Howitt [2000] and in an extended form in Acemoglu et al. [2006]. The main difference between Equation (5) and Equation (8) is that in the latter firms always attempt to imitate the world leading technology while in the first the technology a firm can successfully imitate depends on the available technologies in the whole population of firms (and not only the leading one) and the absorptive capacities of the firm. Equation (5) thus can be interpreted as a more explicit and consistent formulation of absorptive capacities influencing the imitation process and productivity dynamics of firms.

The relationship of our model to a number of previous contributions in the literature deserves some more attention. Klette and Kortum [2004] introduce a general equilibrium model of technological change that is able to reproduce a number of empirical regularities. In their model a firm's R&D effort decision is endogenous. However it only depends on the stock of knowledge of the firm and does not allow technology spillovers. Luttmer [2007] proposes a model of combined innovation and imitation with entry and exit dynamics which generates firm size distributions that are consistent with empirical evidence. Luttmer [2007] assumes that only entering firms imitate the technologies of other firms while incumbent firms engage only in in-house R&D. In contrast, in our model a firm decides between innovation and imitation depending on which of the two gives it a higher expected payoff. Finally, Alvarez et al. [2008]; Lucas [2008] study an imitation process similar to the one presented in this paper. However, these authors do not take into account limitations in the ability of firms to imitate external knowledge and they do not explicitly model the strategic decisions of firms whether to undertake in-house R&D or to copy other firms.

4. Evolution of the Productivity Distribution

In the following, we consider the distribution of normalized log-productivity in the population of firms over time t. We can compute this by computing the interactive Markov chain [Banerjee and Newman, 1993; Conlisk, 1976; Hermanns, 2002], which can be derived from the agent-based dynamical Equations (4) and (7). Reconsider that normalized log-productivity only takes values in the natural numbers. We call $P_a(t)$ to be the fraction of firms having log-productivity a at time t. Thus, the row vector $P(t) =$

$[P_1(t)\ P_2(t) \ldots P_a(t) \ldots]$ represents the distribution of log-productivity at time t. Of course it holds $P_a(t) \geq 0$ and $\sum_{a=1}^{\infty} P(a,t) = 1$.Though the random variable $\eta(t)$ is also restricted to natural numbers as possible realizations the process of in-house R&D innovations are also well represented by our discrete probabilistic framework.In what follows we will omit either a or t in the arguments of $P_a(t)$ if the parameter is clear from the context, and shortening of equations is useful. But a, b, c always indexes log-productivity and t the time.

First, we consider the potential increase in productivity due to in-house R&D. Supposed the random variable η has a maximal achievable value of m log-productivity units, then the probability distribution of η is defined by the row vector $[\eta_0\ \eta_1 \ldots \eta_m]$, with η_b representing the probability to increase the productivity by b units. Thus the transition matrix for log-productivity due to in-house R&D corresponding to Equation (3) is

$$\mathbf{T}^{\text{in}} = \begin{bmatrix} \eta_0 & \eta_1 & \ldots & \eta_m & 0 & \ldots \\ 0 & \eta_0 & \eta_1 & \ldots & \eta_m & 0 \\ & 0 & \eta_0 & \eta_1 & \ldots & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Remind that $\mathbf{T}^{\text{in}}_{a,b}$ is the probability to change from log-productivity a to log-productivity b in one time step t to $t + \Delta t$. Thus, the Markov chain for the pure innovation process reads $P(t+1) = P(t)\mathbf{T}^{\text{in}}$.

Next, we consider the potential increase in productivity due to imitation. We assume that a firm with a log-productivity of $a_i(t)$ can only achieve a level of log-productivity $a_i(t) + k$, $a_j(t) > a_i(t)$ and $0 \leq k \leq a_j(t) - a_i(t)$ through imitating a randomly selected firm j.[9] Denote by $a = a_i$ and $b = a_j$.

[9] As in Eeckhout and Jovanovic [2002]; Stigler [1961], we could assume that imitation proceeds by first taking k independent draws a_1,...,a_k from the cumulative distribution function $F(a,t) = \sum_{b=1}^{a} P(b,t)$, and then selecting the maximum value $\max_{1 \leq j \leq k}\{a_j\}$, which is distributed as $H(a,t) = F(a,t)^k$ and has a probability mass function of $h(a,t) = kF(a,t)^{k-1}P(a,t)$. The parameter k measures the directedness of search. In the limit of $k \to \infty$, the firm imitates the highest productivity, while in the limit of $k \to 1$ the firm imitates at random. With this assumption, the expected value of the log-productivity imitated is given by

$$\mathbb{E}\left(\max_{1 \leq j \leq k}\{a_j\}\right) = \sum_{a=1}^{\infty} akF(a,t)^{k-1}P(a,t).$$

Then the transition matrix $\mathbf{T}^{\mathrm{im}}$ with elements $\mathbf{T}^{\mathrm{im}}_{ab}$, giving the transition probability from log-productivity a to log-productivity $b \geq a$, is given by

$$\begin{aligned}
\mathbf{T}^{\mathrm{im}}_{ab} &= q^{b-a} P_b + q^{b-a}(1-q)P_{b+1} + q^{b-a}(1-q)P_{b+2} + \dots \\
&= q^{b-a}\left(P_b + (1-q)\sum_{k=1}^{\infty} P_{b+k}\right) \\
&= q^{b-a}\left(P_b + (1-q)(1-F_b)\right),
\end{aligned} \tag{9}$$

with $P_a = P(a,t)$ and $F_a = F(a,t) = \sum_{b=1}^{a} P(a,t)$.

The transition matrix $\mathbf{T}^{\mathrm{im}}$ with elements given by Equation (9) for the imitation process in Equation (5) is interactive.[10] It depends on the current distribution of log-productivity $P(a,t)$ and it is given by

$$\mathbf{T}^{\mathrm{im}} = \begin{bmatrix}
S_1(P) & q(P_2 + (1-q)(1-F_2)) & q^2(P_3 + (1-q)(1-F_3)) & \dots \\
0 & S_2(P) & q(P_3 + (1-q)(1-F_3)) & \dots \\
0 & 0 & S_3(P) & \ddots \\
 & \ddots & \ddots & \ddots
\end{bmatrix},$$

with $S_a(P) = 1 - \sum_{b=a+1}^{\infty} T^{\mathrm{im}}_{ab} = 1 - \sum_{b=a+1}^{\infty} q^{b-a}\left(P_b + (1-q)(1-F_b)\right)$. For $q = 1$ we get

$$\mathbf{T}^{\mathrm{im}} = \begin{bmatrix}
S_1(P) & P_2 & P_3 & P_4 & \dots \\
0 & S_2(P) & P_3 & P_4 & \dots \\
0 & 0 & S_3(P) & P_4 & \dots \\
 & & \ddots & \ddots & \ddots
\end{bmatrix},$$

with $S_a(P) = 1 - \sum_{b=a+1}^{\infty} P_b$.

In the following we consider the expected values of productivity obtained either through in-house R&D or imitation, given a certain log-productivity level and the current distribution of log-productivities. The expected value of log-productivity due to in-house R&D is

$$a_{\mathrm{in}}(a) = a + \mu_\eta. \tag{10}$$

In this paper we concentrate on the case of $k = 1$ and leave the analysis of the more general case of $k > 1$ for future work.

[10] This also refers to a time-inhomogeneous Markov chain.

Similarly, the expected log-productivity due to imitation is given by

$$a_{\mathrm{im}}(a,P) = aS_a(P) + \sum_{b=a+1}^{\infty} bq^{b-a}\left(P_b + (1-q)(1-F_b)\right). \qquad (11)$$

Firms live in a uncertain environment and the results of their innovation and imitation decisions are governed by the realizations of stochastic processes. The decision of a firm whether to undertake in-house R&D or to imitate other firms depends on which of the two strategies gives it a higher expected utility from profits. We assume that firms are *risk averse* [Sandmo, 1971]. The utility function $U(\pi)$ of the firm is assumed to be concave, continuous and differentiable in profits π, such that $U'(\pi) > 0$ and $U''(\pi) < 0$. Their expected utility is $U(\pi) = \log \pi$, using a constant relative risk aversion utility function given by the logarithm. According to Equation (2), the expected profit the firm gains if it has expected productivity A is $\pi = \psi A$. Denoting by $a = \log A$ it then follows that the expected utility from profits is given by $U(\pi) = \log \psi + a + c$, where $c \geq 0$ denotes the cost which is assumed to be equal for both innovation and imitation processes.[11]

The firm observes its utility with an additive error ϵ as $U + \epsilon = \log \pi + \epsilon = \log \psi + a + \epsilon$, where ϵ is identically independently type-I extreme value distributed with parameter $1/\beta \geq 0$ [Cameron and Trivedi, 2005, pp. 476]. This implies that the probability that the firm chooses imitation rather than in-house R&D is given by[12]

$$\begin{aligned} p_\beta^{\mathrm{im}}(a) = \mathbb{P}\left(a_{\mathrm{im}}(a,P) > a_{\mathrm{in}}(a)\right) &= \frac{e^{\beta a_{\mathrm{im}}(a,P)}}{e^{\beta a_{\mathrm{im}}(a,P)} + e^{\beta a_{\mathrm{in}}(a)}} \\ &= \frac{1}{1 + e^{-\beta(a_{\mathrm{im}}(a,P) - a_{\mathrm{in}}(a))}}. \end{aligned} \qquad (12)$$

In the *weak selection limit* of $\beta \to 0$ we get $\lim_{\beta \to 0} p_\beta^{\mathrm{im}}(a) = 0.5$ and firms choose randomly between innovation and imitation, while in the *strong selection limit* of $\beta \to \infty$ we get $\lim_{\beta \to \infty} p_\beta^{\mathrm{im}}(a) = \mathbb{I}_{\{a_{\mathrm{in}}(a) < a_{\mathrm{im}}(a,P)\}}$, which is the indicator function being one if $a_{\mathrm{in}}(a) < a_{\mathrm{im}}(a,P)$ and zero otherwise.

[11] See e.g. Acemoglu et al. [2010], where the adoption of new, complex technologies involves a costly standardization process.

[12] Note that we could assume that firms are discounting their future profits and base their decisions whether to innovate or imitate on the net present value of profits without changing our results.

We further define $\mathbf{D}$ as the diagonal-matrix of all probabilities $p_\beta^{\mathrm{im}}(a,t)$ for all log-productivity values $a = 1, 2, ...$ at time t, i.e.

$$\mathbf{D} = \begin{bmatrix} p_\beta^{\mathrm{im}}(1) & 0 & \dots & & \\ 0 & p_\beta^{\mathrm{im}}(2) & 0 & \dots & \\ & 0 & p_\beta^{\mathrm{im}}(3) & 0 & \dots \\ & & \ddots & \ddots & \ddots \end{bmatrix}.$$

Having derived the transition matrices, we can write the dynamic equation for the distribution of log-productivity. For the initial distribution of log-productivity $P(0)$, the evolution of the distribution can be computed by

$$P(t+\Delta t) = P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t))\right), \tag{13}$$

where $(\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t))$ is the mixed transition matrix of in-house R&D and imitation and $\mathbf{I}$ is the identity matrix. The P's are row vectors and the T's are square matrices of the same size. The framework naturally extends also to the two-side unbounded case with productivity decay (see Appendix B.1).

The dynamic Equation (13) can be simply implemented and used to compute the distribution of log-productivity. For this the vector P has to be dynamically extended to be able to capture all nonzero log-productivity classes.

From the transition matrix $(\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t))$ in Equation (13) we can go to an equivalent system in continuous time $t \in \mathbb{R}_+$ by defining the probability that the Markov chain migrates from log-productivity j to log-productivity i, $i \neq j$, in the interval $[t, t+\Delta t)$ and Δt small, with probability

$$\mathbb{P}\left(a(t+\Delta t) = i | a(t) = j\right) = g_{ij}(t)\Delta t + o(\Delta t), \quad i \neq j.$$

$g_{ij}(t)$ is the transition rate between log-productivity states j and i given by

$$g_{ij}(t) = \left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t))\right)_{ij}, \quad i \neq j,$$

and $g_{ii}(t) = -\sum_{i\neq j} g_{ij}(t)$. In this framework, a transition happens from log-productivity states j to i during the time interval $[t, t+\Delta t)$ with probability $g_{ij}(t)\Delta t + o(\Delta t)$, $j \neq i$, and nothing happens during that interval with probability $1+g_{ii}(t)\Delta t+o(\Delta t)$. We call $\mathbf{G}(t) = (g_{ij}(t))_{1\leq i,j<\infty}$ the transition rate matrix.[13] In matrix notation $\mathbf{G}(t)$ can be written as

$$\mathbf{G}(t) = (\mathbf{I} - \mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t)) - \mathbf{I}. \tag{14}$$

By virtue of the transition rate matrix $\mathbf{G}$, the continuous time, discrete space Markov chain describing the evolution of the log-productivity distribution can be written as

$$\frac{dP(t)}{dt} = P(t)\mathbf{G}(t). \tag{15}$$

In the proceeding sections we will first analyze two special important cases. First, in Section 4.1 we study the emerging productivity distribution in the case of pure in-house R&D. Next, in Section 4.2 we analyze the case of pure imitation. Finally, in Section 4.3 we combine both processes of innovation and imitation and study the emerging productivity distributions.

4.1. Productivity Growth through In-House R&D

In an economy with strong absorptive capacity limits, corresponding to $q = 0$, and sufficiently large β, an increase in a firm's productivity level is only due to in-house R&D. In this case, the evolution of the log-productivity distribution in Equation (13) is given by

$$P(t + \Delta t) = P(t)\mathbf{T}^{\text{in}}. \tag{16}$$

We first consider the case where $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. Then, in matrix-vector notation the evolution of the log-productivity distribution can be written as follows

$$P(t + \Delta t) = P(t) \begin{bmatrix} 1-p & p & 0 & \dots & & \\ 0 & 1-p & p & 0 & \dots & \\ 0 & 0 & 1-p & p & 0 & \dots \\ & & & \ddots & \ddots & \ddots \end{bmatrix}.$$

For each level of log-productivity a this means that

$$P(a, t + \Delta t) = (1 - p)P(a, t) + pP(a - 1, t). \tag{17}$$

[13] See e.g. Stroock [2005, Chap. 4].

Using Equation (17) we can derive the log-productivity distribution at time t.

Proposition 1. *Assume that* $\eta_1 = p$*,* $\eta_0 = 1 - p$ *and* $\eta_i = 0$ *for all* $i \geq 2$ *and* $q = 0$ *such that firms conduct only in-house R&D. Consider the initial distribution* $P(a, 0) = \delta_{a,1}$ *and set* $\Delta t = 1$*. Then* $P(a, t + \Delta t)$ *is the probability mass function of a Binomial distribution* $B(t + 2, p)$*.*

Note that

$$\begin{aligned} \mathbb{E}(\eta) &= \mu_\eta = p, \\ \mathrm{Var}(\eta) &= \sigma_\eta^2 = p(1 - p). \end{aligned}$$

Using the fact that for large times t the Binomial distribution can be approximated by a normal distribution with density ϕ [see e.g. Paolella, 2006], we obtain the following corollary.

Corollary 1. *Under the assumptions of Proposition 1, the productivity distribution can be approximated by a normal distribution as*

$$P(a, t) \sim \phi(pt, p(1 - p)t) = \phi(\mu_\eta t, \sigma_\eta^2 t). \tag{18}$$

in the limit of large t*.*

The result in Equation (18) can also be obtained for more general distributions of η and arbitrary initial conditions $P(a, 0)$. In the case of non-decaying productivity, there is a positive drift of the random in-house R&D process. Thus, the log-productivity approaches a Gaussian shape in the limit of large times t with mean and variance rising linearly, due to the central limit theorem. The original productivity growth dynamics corresponds to an exponential growth process with multiplicative noise while the log-transformed process is described by an additive noise. This observation can be summarized in the next proposition.

Proposition 2. *If* $\mu_\eta > 0$ *and* $q = 0$ *then the* log*-productivity distribution approaches a Normal distribution,* $P \sim \mathcal{N}(t\mu_\eta, t\sigma_\eta^2)$*, for large* t*, with* $\mu_\eta = \mathbb{E}(\eta)$ *and* $\sigma_\eta^2 = \mathrm{Var}(\eta)$*. The productivity distribution converges to a lognormal shape with mean* $\mu_A = e^{t\mu_\eta + \frac{1}{2}t\sigma_\eta^2}$ *and variance* $\sigma_A^2 = \left(e^{t\sigma_\eta^2} - 1\right) e^{2t\mu_\eta + t\sigma_\eta^2}$*.*

It is worth noting that for large times t, the lognormal distribution will be close to a Pareto distribution (or power-law) for a wide range of productivities, as stated in the next corollary.

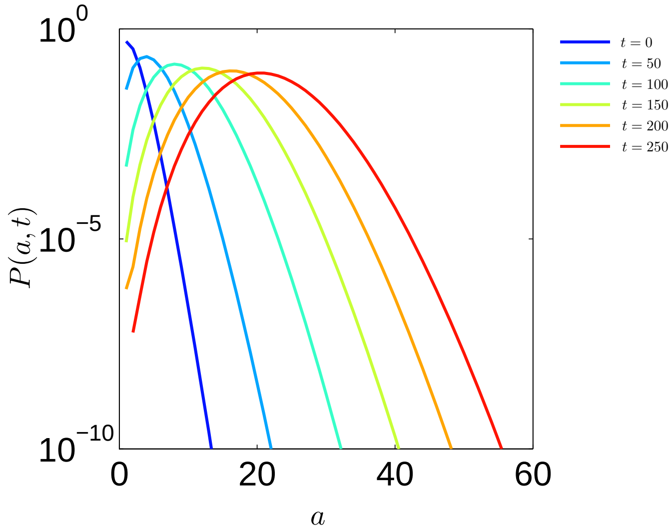

Figure 4: The log-productivity distribution $P(a,t)$ for different periods and $p = 0.1$ when firms can only increase their productivity through in-house R&D.

Corollary 2. *The asymptotic productivity distribution can be approximated by a Pareto distribution for $A = \mathcal{O}(e^{(\mu_\eta + 2\sigma_\eta^2)t})$.*

The time evolution of the log-productivity distribution $P(a,t)$ for different periods and $p = 0.1$ starting from an exponential initial distribution in the case of pure in-house R&D can be seen in Figure 4. From the figure we see that the variance of the distribution is increasing in time, as it is predicted by Proposition 2.

4.2. Productivity Growth through Imitation

If the outcome of firms' in-house R&D activities do not lead to a positive expected log-productivity increase, that is $\eta_i = 0$ for $i \geq 1$, and β is sufficiently large, then firms increase their productivity only through imitating other firms' technologies. In this case, Equation (13) governing the evolution of the log-productivity distribution is given by

$$P(t + \Delta t) = P(t)\mathbf{T}^{\mathrm{im}}(P(t)). \tag{19}$$

Assuming that $q = 1$, this can be written in vector-matrix notation as

$$P(t + \Delta t) = P(t) \begin{bmatrix} F_1 & P_2 & P_3 & P_4 & \dots \\ 0 & F_2 & P_3 & P_4 & \dots \\ 0 & 0 & F_3 & P_4 & \dots \\ & & & \ddots & \ddots \end{bmatrix}, \tag{20}$$

where the cumulative distribution function is denoted by $F_a = F(a,t) = \sum_{b=1}^{a} P(b,t)$. We find that, in the limit of $q = 1$, we recover the knowledge growth dynamics analyzed by Lucas [2008]. The result is given in the following proposition.

Proposition 3. *Assume that $\eta_i = 0$, for all $i \geq 1$ such that firms only adopt technologies of other firms. Further consider $q = 1$ in the absence of absorptive capacity limits. Starting from the initial distribution $F(a,0)$, the the cumulative* log*-productivity distribution at time t is given by*

$$F(a,t) = e^{\log F(a,0) 2^t}. \tag{21}$$

Proposition 3 is a special case of the more general result for $q \in [0,1]$, that is given in the next proposition.

Proposition 4. *If $\eta_i = 0$ for $i \geq 1$ and $0 \leq q \leq 1$ then the evolution of the cumulative* log*-productivity distribution $F(a,t)$ is given by*

$$\frac{\partial F(a,t)}{\partial t} = F(a,t)^2 - F(a,t) + (1-q)(1-F(a,t)) \sum_{b=0}^{a-1} q^b F(a-b,t). \tag{22}$$

Moreover, if there exists a maximal initial log*-productivity a_{max} such that $F(a,0) = 1$ for all $a \geq a_{max}$ then the asymptotic cumulative* log*-productivity distribution is given by*

$$\lim_{t \to \infty} F(a,t) = \begin{cases} 0, & \text{if } a < a_{max}, \\ 1, & \text{if } a \geq a_{max}. \end{cases} \tag{23}$$

For $q = 1$ Equation (22) is equivalent to Equation (44). For $q = 0$ Equation (22) is trivially satisfied, as $F(a, t+\Delta t) = F(a,t)$. The boundary conditions are $F(0,t) = 0$ and $F(\infty, t) = 1$. If we have the initial condition $F(a,0) = \delta_{a,1}$ then $F(a, t+\Delta t) = F(a,t) = F(a,0) = \delta_{a,1}$. For $q > 0$ the same result as in Equation (23) holds.

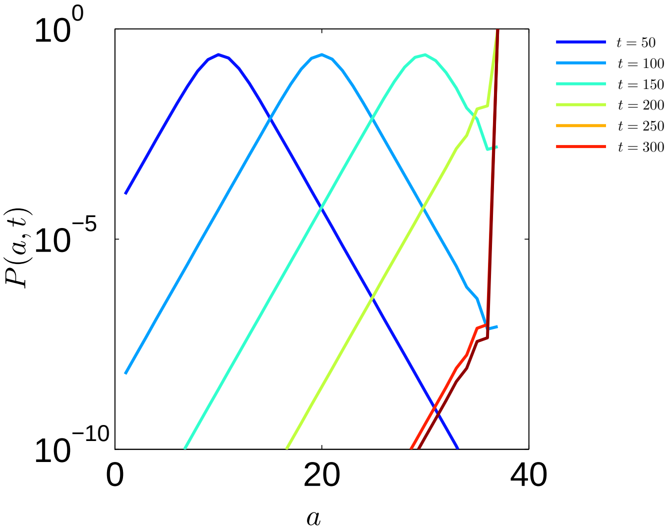

Figure 5: The log-productivity distribution $P(a,t)$ for different periods and $q = 1$ in the case of pure imitation.

The time evolution of the log-productivity distribution $P(a,t)$ for different periods and $q = 1$ starting from an exponential initial distribution in the case of pure imitation can be seen in Figure 5. The figure reveals that the distribution concentrates its mass at the maximum initial productivity level as time increases. This is consistent with Equation (23) in Proposition 4.

When absorptive capacity limits are strong then we can neglect terms of the order $\mathcal{O}(q^2)$ and we can derive the following corollary.

Corollary 3. *Let $\eta_i = 0$ for $i \geq 1$ and assume that q is small such that terms of the order $\mathcal{O}(q^2)$ can be neglected. Then the evolution of the cumulative* log-*productivity distribution is given by*

$$\frac{\partial F(a,t)}{\partial t} = qP(a,t)(1 - F(a,t)). \tag{24}$$

Consequently, for $q \to 0$ we get from Equation (24) $F(a, t + \Delta t) = F(a,t)$.

On the other hand, when we have only weak absorptive capacity limits then we can neglect terms of the order $\mathcal{O}((1-q)^2)$ and we can derive from Equation (22) the following corollary.

Corollary 4. *Let $\eta_i = 0$ for $i \geq 1$ and assume that $1 - q$ is small such that terms of the order $\mathcal{O}((1-q)^2)$ can be neglected. Assume that $F(a,t)$*

is a sufficiently smooth distribution. Then the evolution of the cumulative log*-productivity distribution is given by*

$$\frac{\partial F(a,t)}{\partial t} = (2q-1)\left(F(a,t)^2 - F(a,t)\right). \tag{25}$$

From Equation (25) we obtain Equation (44) in the limit $q \to 1$.

4.3. Productivity Growth through In-House R&D and Imitation

In this section we assume that firms can increase productivity through both, in-house R&D or imitating other firms. This corresponds to $\eta_i > 0$ for some $i \geq 1$ and $q > 0$. In order to simplify our analysis we assume that q is close to one, that is, we assume that there are only weak absorptive capacity limits, and $\eta_1 = p$, $\eta_0 = 1-p$ and $\eta_i = 0$ for all $i \geq 2$. Then from Equations (13) and (25) we see that the evolution of the cumulative log-productivity distribution is given by

$$\begin{aligned} F(a,t+\Delta t) &= p_\beta^{\mathrm{im}}(a)\left((2q-1)(F(a,t)^2 - F(a,t)) + F(a,t)\right) \\ &\quad + (1 - p_\beta^{\mathrm{im}}(a))\left((1-p)F(a,t) + pF(a-1,t)\right), \end{aligned}$$

where $p_\beta^{\mathrm{im}}(a)$ is defined in Equation (12). Inserting yields

$$\begin{aligned} F(a,t+\Delta t) &= \frac{1}{1+e^{-\beta(a_{\mathrm{im}}(a,P)-a_{\mathrm{in}}(a))}}\left((2q-1)(F(a,t)^2 - F(a,t)) + F(a,t)\right) \\ &\quad + \frac{1}{1+e^{-\beta(a_{\mathrm{in}}(a,P)-a_{\mathrm{im}}(a))}}\left((1-p)F(a,t) + pF(a-1,t)\right). \end{aligned}$$

Using the fact that

$$a_{\mathrm{in}}(a,P) - a_{\mathrm{im}}(a) = p - \sum_{b=a+1}^{\infty}(b-a)P(b,t),$$

we can write the continuous time Markov chain introduced in Equation (15) as

$$\begin{aligned} \frac{\partial F(a,t)}{\partial t} &= \frac{2q-1}{1+e^{-\beta\left(\sum_{b=a+1}^{\infty}(b-a)P(b,t)-p\right)}}\left(F(a,t)^2 - F(a,t)\right) \\ &\quad - \frac{p}{1+e^{-\beta\left(p-\sum_{b=a+1}^{\infty}(b-a)P(b,t)\right)}}\left(F(a,t) - F(a-1,t)\right). \end{aligned} \tag{26}$$

The evolution of the log-productivity distribution for $\beta \in [0, 1]$ can be obtained by numerical integration of Equation (26).

We will see that the emerging productivity distributions can be considered as *traveling waves*.[14] If a distribution $F(a,t)$ represents a traveling wave, then the shape of $F(a,t)$ will be the same at all times t and the speed of propagation of $F(a,t)$ is constant and given by ν. More formally, if the traveling wave $F(a,t)$ advances at constant speed ν then we can write

$$F(a,t) = f(a - \nu t), \quad \nu \geq 0.$$

Observe that $f(a - \nu t)$ is translational invariant, that is, if $a - \nu t$ is constant, so is f. Moreover, ν is the growth rate along the balanced growth path.

We further assume that the economy starts off at sufficiently steep initial conditions with compact support [Bramson, 1983; Kolmogorov et al., 1937]. $F(a,0)$ has compact support if $F(a,0) = f_0(a)$, $0 \leq f_0(a) \leq 1$, with

$$f_0(a) = \begin{cases} 0, & \text{if }, a \leq a_1, \\ 1, & \text{if }, a \geq a_2, \end{cases}$$

for some $a_1 < a_2$ and $f_0(a)$ continuous for $a \in (a_1, a_2)$. Further, $F(a,0)$ is sufficiently steep if $|a_1 - a_2| < \epsilon$ for some $\epsilon \geq 0$. For example, we can consider as initial condition $F(a,0) \propto 1 - e^{-\lambda a}$ with $\lambda > \lambda^* > 0$.

In the following sections we separately analyze the emerging log-productivity distributions for different values of the noise parameter β. In Section 4.3.1 we study the weak selection limit for vanishing noise $\beta \to 0$, while in Section 4.3.2 we consider the strong selection limit $\beta \to \infty$ where firms have a perfect evaluation of the expected outcomes of their innovation and imitation strategies.

4.3.1. Weak Selection Limit ($\beta \to 0$)

In the weak selection limit as $\beta \to 0$ firms choose between innovation and in-house R&D uniformly at random with probability $p_\beta^{\text{im}}(a) = 0.5$ for all $1 \leq a < \infty$. In the following we assume that $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. Then from Equation (26) the evolution of the cumulative

[14] See e.g. Murray [2002, Chap. 13].

distribution $F(a,t)$, in the limit of $1-q$ small, is given by

$$F(a,t+\Delta t) = \frac{2q-1}{2}(F(a,t)^2 - F(a,t)) + \frac{1}{2}F(a,t) + \frac{1-p}{2}F(a,t) + \frac{p}{2}F(a-1,t). \tag{27}$$

This can be written as

$$2(F(a,t+\Delta t) - F(a,t)) = (2q-1)(F(a,t)^2 - F(a,t)) - p(F(a,t) - F(a-1,t)).$$

In terms of the complementary cumulative log-productivity distribution $G(a,t) = 1 - F(a,t)$ we obtain

$$2(G(a,t+\Delta t) - G(a,t)) = (2q-1)(-G(a,t)^2 + G(a,t)) - p(G(a,t) - G(a-1,t)).$$

Hence,

$$G(a,t+\Delta t) - G(a,t) = \frac{2q-1}{2}(-G(a,t)^2 + G(a,t)) - \frac{p}{2}(G(a,t) - G(a-1,t)). \tag{28}$$

In continuous time, Equation (28) reads as

$$\frac{\partial G(a,t)}{\partial t} = \frac{2q-1}{2}(-G(a,t)^2 + G(a,t)) - \frac{p}{2}(G(a,t) - G(a-1,t)). \tag{29}$$

In the limit of $q = 1$ we recover the model analyzed in Majumdar and Krapivsky [2001]. This difference-differential equation for $G(a,t)$ can be solved numerically subject to the boundary conditions $\lim_{a\to\infty} G(a,t) = 0$ and $\lim_{a\to-\infty} G(a,t) = 1$.

From Equation (29) we find that the dynamics of the complementary cumulative log-productivity distribution $G(a,t)$ in Equation (29) admits a traveling wave solution $G(a,t) = g(x)$, $x = a - \nu t$ with velocity ν satisfying

$$\nu \frac{dg(x)}{dx} = \frac{2q-1}{2}(g(x)^2 - g(x)) + \frac{p}{2}(g(x) - g(x-1)). \tag{30}$$

If we further assume that for a much larger than νt the complementary cumulative log-productivity distribution $G(a,t)$ has the particular traveling wave form $G(a,t) \propto e^{-\lambda(a-\nu t)}$ then we can give the following proposition.

Proposition 5. *Assume that $\eta_1 = p$, $\eta_0 = 1-p$ and $\eta_i = 0$ for all $i \geq 2$ with $p \in [0,1]$. Consider $\beta = 0$ and q close to one such that terms of the order $\mathcal{O}((1-q)^2)$ can be neglected. Then for sufficiently steep and compact initial conditions $F(a,0)$, Equation (29) admits a traveling wave solution*

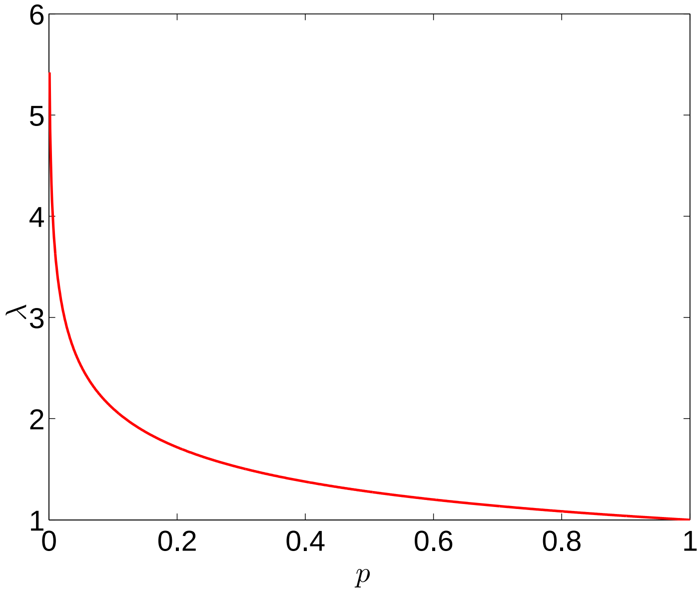

Figure 6: The tail exponent λ^* in Equation (48) for different values of $p \in [0, 1]$ and $q = 1$.

with velocity ν given by

$$\nu = \frac{2q - 1 - p + pe^{\lambda}}{2\lambda}, \tag{31}$$

where the exponent λ of the front of the distribution is given by[15]

$$\lambda = 1 + W\left(\frac{2q - 1 - p}{pe}\right). \tag{32}$$

Moreover, assuming that $F(a, t) \propto e^{\rho(a - \nu t)}$, $\rho \geq 0$, for a much smaller than νt then the exponent ρ is implicitly determined by the following equation

$$2\rho\nu = 2q - 1 + p - pe^{-\rho}. \tag{33}$$

The dependency of λ on the innovation probability p can be seen in Figure 6. From the figure we observe that the distribution becomes steeper, the smaller is p. Moreover, one can shown that the average log-productivity is

[15] $W(x)$ is the Lambert W function (or product log), which is implicitly defined by $W(x)e^{W(x)} = x$. One can show that $W(x) = -\sum_{n=1}^{\infty} \frac{n^{n-1}}{n!}(-x)^n$.

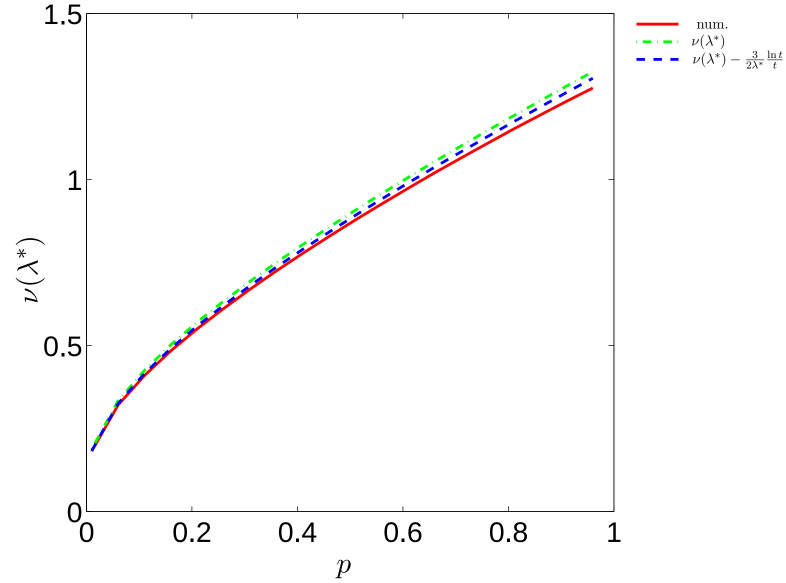

Figure 7: Traveling wave velocity $\nu(\lambda^*)$ for different values of $p \in [0, 1]$ and $q = 1$ at $t = 100$.

given by[16]

$$\mathbb{E}(a) = \nu(\lambda^*)t - \frac{3}{2\lambda^*}\ln t + \mathcal{O}(1),$$

for sufficiently steep and compact initial conditions.

We consider as initial condition $G(a, 0) = \delta_{a,0}$. By solving the continuous dynamical system corresponding to Equation (29) we can compute $G(a, t)$ for large times t. The resulting traveling wave velocity ν can be seen in Figure 7 for different values of p.

Observe that for $p = 1/(1 + e^2) = 0.119$ we obtain an exponential log-productivity distribution with $\lambda = -2$, which corresponds to a power-law productivity distribution with exponent -2 as we have seen them in the empirical analysis in Section 2.[17] Further, using Equation (49) we can determine the exponent ρ for this value of p yielding $\rho = 1.231$. This can be seen in Figure 8 for time $t = 200$ and setting $q = 1$.

[16] See Majumdar and Krapivsky [2001].

[17] Note that $P(a, t) \propto e^{-\lambda a} = e^{-\lambda \log A} = A^{-\lambda}$.

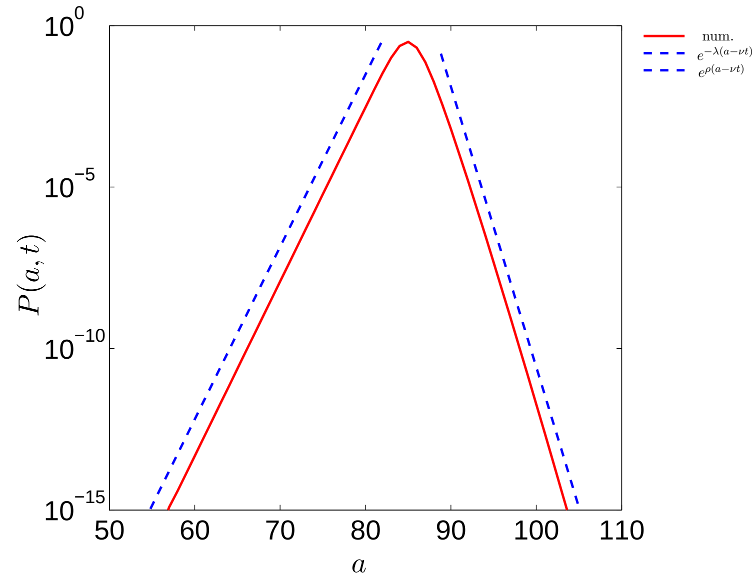

Figure 8: The log-productivity distribution $P(a,t)$ for period $t = 200$, $p = 0.119$, $q = 1$. The distribution obtained by numerical integration of Equation (29) is indicated by a straight line while the theoretical predictions are shown with a dashed line. The front of the traveling wave decays as a power-law with exponent $\lambda = 2$.

4.3.2. Strong Selection Limit ($\beta \to \infty$)

We consider the case of $\eta_1 = p$, $\eta_0 = 1 - p$ and $\eta_i = 0$ for all $i \geq 2$. Further, we assume that q is sufficiently close to one such that terms of the order $\mathcal{O}((1-q)^2)$ can be neglected.

In the strong selection limit, there exists a critical log-productivity level below which it is more profitable for firms to imitate other firms, while for those firms above the threshold it is more profitable to conduct in-house R&D. This is stated in the following proposition.

Proposition 6. *Consider $p_\beta^{im}(a)$ as defined in Equation (12) and let $p_\beta^{in}(a) = 1 - p_\beta^{im}(a)$. If the initial distribution $F(a,0)$ has compact support then there exists a threshold* log-productivity *a^* such that $p_\beta^{im}(a) > p_\beta^{in}(a)$ for all $a \leq a^*$ and $p_\beta^{im}(a) < p_\beta^{in}(a)$ for all $a > a^*$.*

From Proposition 6 it follows for the evolution of the cumulative log-productivity distribution from Equation (26) that

$$F(a,t+\Delta t) = (2q-1)(F(a,t)^2 - F(a,t)) + F(a,t), \qquad \text{if } a \leq a^*, \tag{34}$$

$$F(a,t+\Delta t) = (1-p)F(a,t) + pF(a-1,t), \qquad \text{if } a > a^*, \tag{35}$$

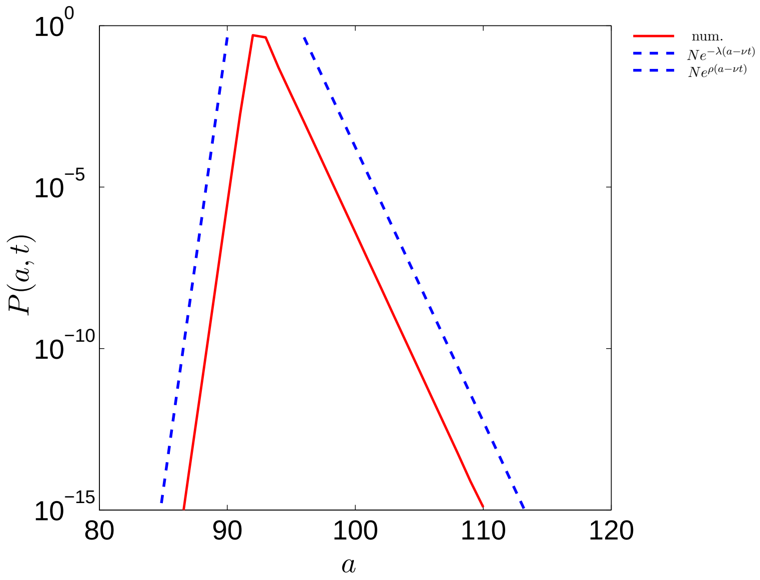

Figure 9: The log-productivity distribution $P(a,t)$ for $p = 0.05$ at $t = 600$. The distribution obtained by numerical integration of Equations (36) and (37) is indicated by a straight line while the theoretical predictions are shown with a dashed line. The front of the traveling wave decays as a power-law with an exponent λ close to 2.

where $F(a,t) = \sum_{b=1}^{a} P(a,t)$. In continuous time we then get

$$\frac{\partial F(a,t)}{\partial t} = (2q-1)(F(a,t)^2 - F(a,t)), \qquad \text{if } a \le a^*, \tag{36}$$

$$\frac{\partial F(a,t)}{\partial t} = -p\,(F(a,t) - F(a-1,t),) \qquad \text{if } a > a^*. \tag{37}$$

The above difference-differential equation for $F(a,t)$ can be solved numerically subject to the boundary conditions $\lim_{a\to\infty} F(a,t) = 1$ and $\lim_{a\to-\infty} F(a,t) = 0$. The resulting log-productivity distribution for $p = 0.05$ and $q = 1$ obtained by means of numerical integration of Equations (36) and (37) can be seen in Figure 9.

Similarly to the previous section, we obtain a traveling wave log-productivity distribution with power-law tails. Under certain assumptions on the shape of the traveling wave we can determine the exponents of the tails and the traveling wave velocity.

Proposition 7. *Assume that $\eta_1 = p$, $\eta_0 = 1-p$ and $\eta_i = 0$ for all $i \ge 2$ with $p \in [0,1]$. Consider $\beta = \infty$ and q close to one such that terms of the order $\mathcal{O}((1-q)^2)$ can be neglected. Let $F(a,0)$ be sufficiently steep and* *compact initial conditions. Further, assume that the traveling wave solution of Equations (36) and (37) has the form*

$$P(a,t) = N \begin{cases} e^{\rho(a-\nu t)} & \text{if } a \leq \nu t, \\ e^{-\lambda(a-\nu t)} & \text{if } a > \nu t. \end{cases} \tag{38}$$

Then for νt being sufficiently close to the threshold a^* (from Proposition 6) the traveling wave velocity ν is given by

$$\nu = \frac{p(1-e^{\lambda})}{\lambda}, \tag{39}$$

where λ is the solution of

$$\frac{e^{-\lambda}}{\lambda^2 p} = 1 + \frac{1}{e^{\frac{(2q-1)\lambda}{p(e^{\lambda}-1)}} - 1} + \frac{1}{e^{\lambda}-1}, \tag{40}$$

the exponent ρ is given by

$$\rho = \frac{2q-1}{\nu}, \tag{41}$$

and the normalization constant N is determined by

$$N = \left(1 + \frac{1}{e^{\rho}-1} + \frac{1}{e^{\lambda}-1}\right)^{-1}. \tag{42}$$

Equation (40) can be solved numerically, using standard numerical root finding procedures [see e.g. Press et al., 1992, Chap. 9], to obtain the exponent λ. The resulting theoretical values of λ and the exponents obtained by numerical integration of Equations (36) and (37) can be seen in Figure 10. We find that we can generate distributions with power-law tails that reproduce our findings in in Section 2. Inserting λ into Equation (39) further gives the traveling wave velocity ν. This is shown in Figure 11.

5. Efficiency

We turn to the analysis of industry performance and efficiency. An industry has a higher performance, measured in aggregate intermediated goods and final good production, if it has a higher average log-productivity.[18]

[18]We will consider the average productivity measured by the geometric mean $\mu = \sqrt[n]{A_1 A_2 \cdots A_n} = \left(\prod_{i=1}^{n} A_i\right)^{1/n}$, which is related to the arithmetic average of the log-

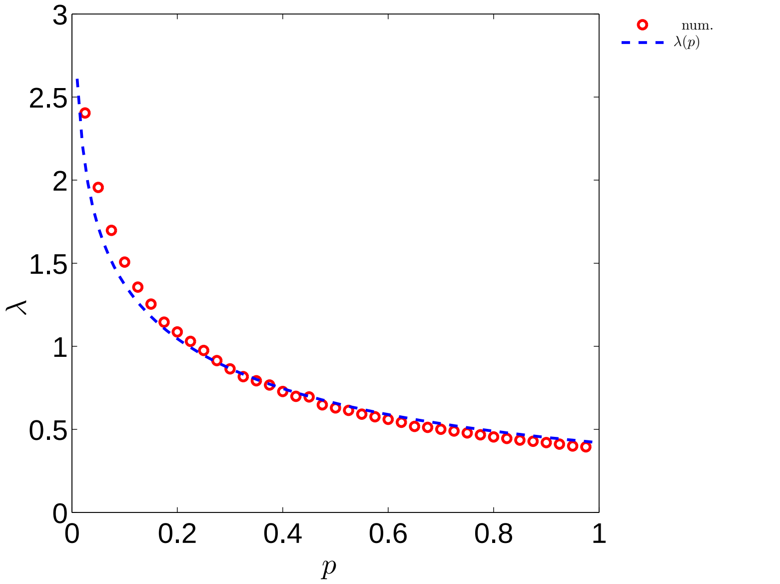

Figure 10: Traveling wave tail exponent λ for different values of $p \in [0, 1]$ by means of numerical integration of Equations (36) and (37) and theoretical prediction indicated by the dashed line.

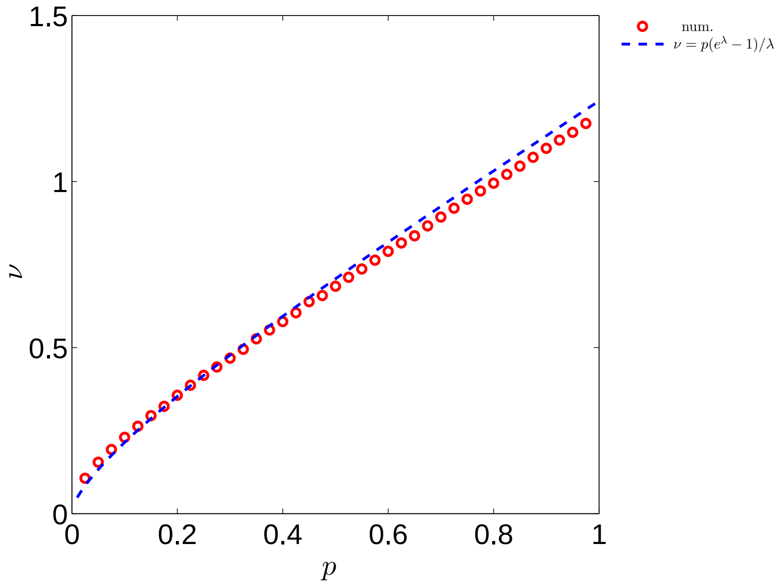

Figure 11: Traveling wave velocity ν for different values of $p \in [0, 1]$ by means of numerical integration of Equations (36) and (37) and theoretical prediction indicated by the dashed line.

Equivalently, this corresponds to a higher average log-productivity per unit of time, as measured by ν. We can derive the following result for efficiency comparing the two extreme cases of the weak and strong selection limits.

Proposition 8. *Let $\nu_{\beta\to\infty}(p)$ and $\nu_{\beta\to 0}(p)$ denote the traveling wave velocities, respectively, in the strong and weak selection limits. Under the assumptions of Propositions 5 and 7, we have that $\nu_{\beta\to\infty}(p) < \nu_{\beta\to 0}(p)$ and aggregate productivity and output are higher if firms perceive expected profits from innovation and imitation with strong noise than with vanishing noise.*

An illustration of the traveling wave velocities ν for the strong ($\beta \to \infty$) and weak ($\beta \to 0$) selection limits can be seen in Figure 12. The figure confirms the result of Proposition 8. For all values of the innovation probability $p \in [0, 1]$, the traveling wave velocity ν is higher if firms' R&D strategies are a random mixture of both innovation and imitation. This indicates the source of inefficiency in our model in the case of vanishing noise: In the strong selection limit, only those firms above the threshold conduct in-house R&D while those below the threshold do not innovate at all but rather prefer to imitate other firms' technologies. With respect to the growth rate of the economy (which depends on ν) this is inefficient. On one hand, firms do not innovate enough (below the threshold). On the other hand, firms do not imitate enough (above the threshold), and their innovations cannot diffuse sufficiently through imitation in the economy. The general source for this inefficiency is the fact that firms do not internalize the externalities they create on other firms productivities through their R&D strategies.

6. Conclusion

In this paper we have introduced an endogenous model of technological change, productivity growth and technology spillovers which is consistent with empirically observed productivity distributions. The innovation process is governed by a combined process of firms' in-house R&D activities and adoption of existing technologies of other firms. The emerging productivity distributions can be described as traveling waves with a constant shape and power-law tails. We incorporate the trade off firms face between their

productivity values via $\frac{1}{n}\sum_{i=1}^{n} a_i = \frac{1}{n}\sum_{i=1}^{n} \log A_i = \log \mu$. However, our results also hold for the arithmetic average of the productivity values.

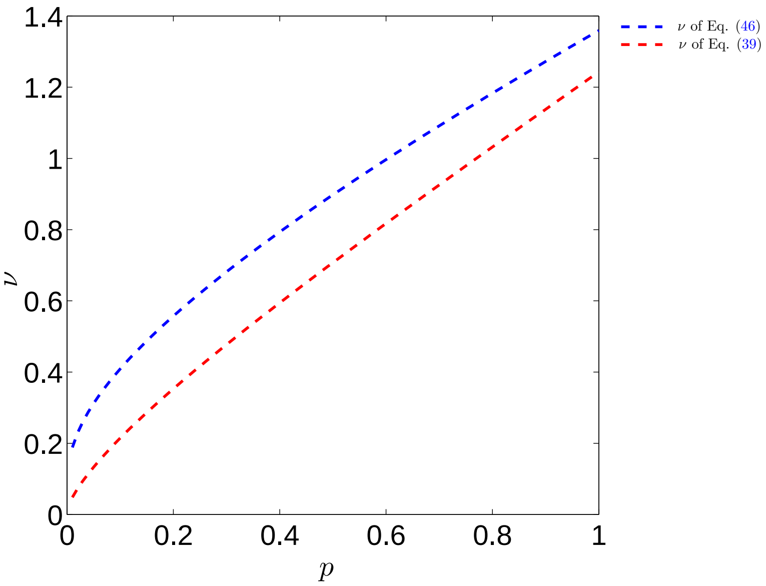

Figure 12: Traveling wave velocities ν in the strong ($\beta \to \infty$) and weak ($\beta \to 0$) selection limits.

innovation and imitation strategies and take into account that firms may have only an uncertain prediction of their research and technology adoption outcomes. We show that this limited rationality can enhance industry performance and efficiency.

The current model can be extended in a number of directions. Three of them are given in Appendix B. First, in Appendix B.1 we outline a model of productivity growth and technology adoption which includes the possibility that a firm's productivity may also be reduced due to exogenous events such as the expiration of a patent. Second, in Appendix B.2 we depart from the assumption of a fixed population of firms and instead allow for firm entry and exit. Third, in Appendix B.3 we consider an alternative way of introducing capacity constraints in the ability of firms to adopt and imitate external knowledge by introducing a cutoff productivity level above which a firm cannot imitate. By introducing a cutoff, one can show that our model can generate "convergence clubs" as they can be found in empirical studies of cross country income differences [e.g. Durlauf, 1996; Durlauf and Johnson, 1995; Feyrer, 2008; Quah, 1993, 1996, 1997].

Finally, it would be interesting to consider the effects of networks and heterogeneous interactions in the imitation process and analyze the emerging productivity distributions, such as in Di Matteo et al. [2005]; Ehrhardt et al. [2006]. This is beyond the scope of the present paper and we leave this avenue for future research.

Appendix

A. Proofs of Propositions and Corollaries

Proof of Proposition 1. For simplicity we will assume that $\Delta t = 1$. From Equation (17) we derive

$$\begin{aligned}
P(a,t+1) &= (1-p)P(a,t) + pP(a-1,t) \\
&= (1-p)^2 P(a,t-1) + 2p(1-p)P(a-1,t-1) \\
&\quad + p^2 P(a-2,t-1) \\
&= (1-p)^3 P(a,t-2) + 3p(1-p)^2 P(a-1,t-2) \\
&\quad + 3p^2(1-p)P(a-2,t-2) + p^3 P(a-3,t-2t) \\
&= \dots \\
&= \sum_{j=0}^{t-2} \binom{t+2}{j} p^j (1-p)^{t+2-j} P(a-j,0).
\end{aligned} \tag{43}$$

For the initial condition $P(a,0) = \delta_{a,1}$ we obtain

$$P(a,t+1) = \binom{t+2}{a-1} p^{a-1}(1-p)^{t-a+3}.$$

This means that $P(a,t+1)$ is the probability mass function of a Binomial distribution $B(t+2,p)$. □

Proof of Proposition 2. For $q = 0$ (and β large such that in this case firms always prefer innovation over imitation), the productivity $A_i(t)$ of firm i grows according to Equation (3), from which we get

$$\log A_i(t) = \log A_i(0) + \sum_{j=1}^{t} \log(1+E(j)).$$

Assuming that the random variables $\eta(t) = \log(1+E(j))$ are independent and identically distributed with finite mean $\mu_\eta < \infty$ and variance $\sigma_\eta^2 < \infty$, then by virtue of the central limit theorem $1/t\sum_{j=1}^{t}\log(1+E(j))$ converges to a normal distribution. Consequently, $A_i(t)$ converges to a lognormal distribution with mean $\mu_A = e^{t\mu_\eta+\frac{1}{2}t\sigma_\eta^2}$ and variance $\sigma_A^2 = \left(e^{t\sigma_\eta^2}-1\right)e^{2t\mu_\eta+t\sigma_\eta^2}$. □

PROOF OF COROLLARY 2. The productivity probability mass function is given by

$$f(A) = \frac{1}{\sqrt{2\pi}\sigma_A A} e^{-\frac{(\ln A - \mu_A)^2}{2\sigma_A^2}}.$$

Taking logs delivers

$$\ln f(A) = -\frac{(\ln A)^2}{2\sigma_A^2} + \left(\frac{\mu}{\sigma_A^2} - 1\right) \ln A - \log\left(\sqrt{2\pi}\sigma_A\right) - \frac{\mu_A^2}{2\sigma_A^2}.$$

As $\sigma_A = \left(e^{t\sigma_\eta^2} - 1\right) e^{2t\mu_\eta + t\sigma_\eta^2}$ tends to infinity for large t, $\ln f(A)$ becomes a linear function of $\ln A$. This approximation is good as long as A is not much larger than $e^{(\mu_\eta + 2\sigma_\eta^2)t}$ [Sornette, 2000, p. 373]. □

PROOF OF PROPOSITION 3. From Equation (20) we derive

$$\begin{aligned}
P(a, t + \Delta t) &= P(a, t) \left(P(1, t) + \cdots + P(a - 1, t) + P(1, t) + \ldots P(a, t)\right)) \\
&= P(a, t) \left(\sum_{b=1}^{a-1} P(b, t) + F(a, t)\right) \\
&= 2P(a, t)F(a - 1, t) + P(a, t)^2.
\end{aligned}$$

For the cumulative distribution function we then obtain

$$\begin{aligned}
F(a, t + \Delta t) &= \sum_{b=1}^{a} P(b, t + \Delta t) \\
&= 2\sum_{b=1}^{a} P(b, t)F(b, t) - \sum_{b=1}^{a} P(b, t)^2 \\
&= 2\sum_{b=1}^{a} P(b, t) \sum_{c=1}^{b} P(c, t) - \sum_{b=1}^{a} P(b, t)^2 \\
&= \left(\sum_{b=1}^{a} P(b, t)\right)^2 = F(a, t)^2.
\end{aligned}$$

The last line from above can be seen as follows. For the induction basis observe that

$$\begin{aligned}
F(1, t + \Delta t) &= 2P(1, t)^2 - P(1, t)^2 = P(1, t)^2 \\
&= F(1, t)^2, \\
F(2, t + \Delta t) &= 2(P(1, t)^2 + P(2, t)P(1, t) + P(2, t)^2) - P(1, t)^2 - P(2, t)^2 \\
&= F(2, t)^2,
\end{aligned}$$

and assuming that $F(a, t + \Delta t) = F(a, t)^2$, we get for the induction step

$a+1$

$$
\begin{aligned}
F(a+1,t+\Delta t) &= 2\sum_{b=1}^{a+1} P(b,t)F(b,t) - \sum_{b=1}^{a+1} P(b,t)^2 \\
&= F(a,t) + 2P(a+1,t)F(a+1,t) - P(a+1,t)^2 \\
&= F(a+1,t)^2.
\end{aligned}
$$

The cumulative log-productivity distribution follows the recursive relation

$$F(a,t+\Delta t) = F(a,t)^2. \tag{44}$$

Substituting $Y(a,t) = \log F(a,t)$ in Equation (44) we get

$$Y(a,t+\Delta t) = 2Y(a,t),$$

with the solution

$$Y(a,t) = Y(a,0)2^t.$$

Hence

$$F(a,t) = e^{\log F(a,0)2^t}. \tag{45}$$

□

Proof of Proposition 4. In the general case of $q \in [0,1]$ the evolution of the cumulative log-productivity distribution is given by

$$
\begin{aligned}
F(a,t+\Delta t) = {} & P_a(1-q)(1-F_a) + P_aF_a \\
& + P_{a-1}q(1-q)(1-F_a) + P_{a-1}(1-q)(1-F_a) + P_{a-1}F_a \\
& + P_{a-2}q^2(1-q)(1-F_a) + P_{a-2}q(1-q)(1-F_a) + P_{a-2}(1-q)(1-F_a) + P_{a-2}F_a \\
& + \dots
\end{aligned}
$$

This can be written as

$$F(a,t+\Delta t) = F(a,t)^2 + (1-q)(1-F(a,t))\sum_{b=0}^{a-1} q^b F(a-b,t).$$

□

Proof of Corollary 3. We can write Equation (22) as

$$F(a,t+\Delta t) = F(a,t)^2 + (1-q)(1-F(a,t))(F(a,t)+qF(a-1,t)) + \mathcal{O}(q^3).$$

This is

$$
\begin{aligned}
F(a,t+\Delta t) = F(a,t) - q\left(F(a,t) - F(a-1,t) + F(a,t)F(a-1,t) - F(a,t)^2\right) \\
- q^2\left(F(a-1,t) - F(a,t)F(a-1,t)\right) + \mathcal{O}(q^3).
\end{aligned}
$$

When absorptive capacity limits are strong then we can neglect terms of the order $\mathcal{O}(q^2)$ and using the fact that $F(a,t) - F(a-1,t) = P(a,t)$ we get

$$F(a, t + \Delta t) - F(a, t) = qP(a, t)G(a, t),$$

where the complementary cumulative distribution function is defined as $G(a,t) = 1 - F(a,t)$. □

Proof of Corollary 4. Equation (22) can be written as

$$F(a,t+\Delta t) = F(a,t)^2+(1-q)(1-F(a,t))(F(a,t)+qF(a-1,t))+\mathcal{O}((1-q)^3).$$

From this we obtain

$$\begin{aligned} F(a,t+\Delta t) = F(a,t)^2&+(1-q)\left(F(a,t) + F(a-1,t) - F(a,t)^2 - F(a,t)F(a-1,t)\right) \\ &- (1-q)^2 \left(F(a-1,t) - F(a,t)F(a-1,t)\right) + \mathcal{O}((1-q)^3) \end{aligned}$$

By neglecting terms of the order $\mathcal{O}((1-q)^2)$ and approximating $F(a,t) + F(a-1,t) \approx 2F(a,t)$ for a sufficiently smooth distribution, we can further write

$$F(a, t + \Delta t) - F(a, t) = (2q - 1)\left(F(a, t)^2 - F(a, t)\right).$$

□

Proof of Proposition 5. We assume that on the balanced growth path the complementary cumulative log-productivity distribution $G(a,t)$ has the traveling wave form $G(a,t) \propto e^{-\lambda(a-\nu t)}$ for a much larger than νt. Observe that for values of a much larger than νt we can neglect the term $G(a,t)^2$ in Equation (29). Then we obtain from Equation (29) the following condition for ν

$$\lambda\nu e^{-\lambda(a-\nu t)} = \frac{2q-1}{2}e^{-\lambda(a-\nu t)} - \frac{p}{2}e^{-\lambda(a-\nu t)} + \frac{p}{2}e^{-\lambda(a-1-\nu t)}.$$

Solving for ν yields

$$\nu = \frac{2q - 1 - p + pe^{\lambda}}{2\lambda}. \tag{46}$$

For sufficiently steep initial conditions with compact support the exponent λ is realized that minimizes the traveling wave velocity ν. This is called the *selection principle* [Bramson, 1983; Murray, 2002]. The corresponding value of λ can be obtained from the first order conditions $d\nu/d\lambda = 0$, or equivalently

$$2q - 1 - p + pe^{\lambda} = p\lambda e^{\lambda}. \tag{47}$$

The minimum of Equation (46) is obtained at λ^* solving Equation (47).

This yields

$$\lambda^* = 1 + W\left(\frac{2q-1-p}{pe}\right), \tag{48}$$

where W is the Lambert W function (or product log), which is the inverse function of $f(w) = we^w$.

Next, we consider the rear of the traveling wave. For a much smaller than νt we can neglect the term $F(a,t)^2$ in Equation (27) to obtain

$$F(a, t+\Delta t) - F(a,t) = -\frac{2q-1}{2}F(a,t) - \frac{p}{2}(F(a,t) - F(a-1,t)).$$

In continuous time we get

$$\frac{dF(a,t)}{dt} = -\frac{2q-1}{2}F(a,t) - \frac{p}{2}(F(a,t) - F(a-1,t)).$$

Assuming that $F(a,t) \propto e^{\rho(a-\nu t)}$, $\rho \geq 0$, we get

$$2\rho\nu = 2q - 1 + p - pe^{-\rho}. \tag{49}$$

Equation (49) has to be solved numerically to obtain the exponent ρ [see e.g. Press et al., 1992, Chap. 9]. □

Proof of Proposition 6. The threshold log-productivity a^* must satisfy the following condition

$$\sum_{b=a+1}^{\infty} bP(b,t) \begin{cases} \geq p & \text{if } a \leq a^*, \\ < p & \text{if } a > a^*. \end{cases}$$

The uniqueness and existence of a^* is equivalent to the strict monotonicity of the function $f(a,t)$ defined by

$$f(a,t) = \sum_{b=a+1}^{\infty} bP(b,t).$$

$f(a,t)$ is strictly monotonous decreasing if $f(a-1,t) - f(a,t) = aP(a,t) > 0$. This holds for all a in the support of $P(a,t)$. Hence, if the distribution at time t has compact support such that for all $a \in [a_1, a_2]$, $P(a,t) > 0$ then there exist a unique threshold log-productivity a^* satisfying the above condition.

Next, we show that if $P(b,t)$ has compact support, then also $f(a-1, t+\Delta t) - f(a, t+\Delta t) > 0$ under the dynamics in Equations (34) and (35). First, consider $a \leq a^*$. Then for q close to one, $P(a,t) > 0$ and $F(a,t) > F(a,1,t)$

we get

$$\begin{aligned}
f(a-1,t+\Delta t) - f(a,t+\Delta t) &= aP(a,t+\Delta t) \\
&= a\left(F(a,t+\Delta t) - F(a-1,t+\Delta t)\right) \\
&= a\big((2q-1)(F(a,t)^2 - F(a-1,t)^2) \\
&\quad +2(1-q)(F(a,t) - F(a-1,t))\big) \\
&> 0.
\end{aligned}$$

On the other hand, using Equation (43) we can write for $a > a^*$

$$P(a,t+\Delta t) = (1-p)P(a,t) + pP(a-1,t),$$

which is positive given that $P(a,t) > 0$ and $p \in [0,1]$ and so $f(a,t+\Delta t)$ is monotonic decreasing. □

Proof of Proposition 7. Similar to Section 4.3.1 we assume that the log-productivity distribution can be described by a traveling wave of the form in Equation (38). For values of a much smaller than νt we can neglect the term $F(a,t)^2$ in Equation (36). Further, with $P(a,t)$ from Equation (38) we have that $F(a,t) \propto e^{\rho(a-\nu t)}$ for $a \leq \nu t$. Inserting into Equation (36) gives

$$-\rho\nu e^{\rho(a-\nu t)} = -(2q-1)e^{\rho(a-\nu t)},$$

and hence we obtain Equation (41).

In terms of the complementary cumulative distribution function $G(a,t) = 1 - F(a,t)$ we can write Equation (37) as

$$G(a,t+\Delta t) = (1-p)G(a,t) + p(G(a-1,t)). \tag{50}$$

From Equation (38) it follows that $G(a,t) \propto e^{-\lambda(a-\nu t)}$. Inserting into Equation (50) yields

$$-\nu\lambda e^{-\lambda(a-\nu t)} = -p\left(e^{-\lambda(a-\nu t)} - e^{-\lambda(a-1-\nu t)}\right).$$

From the above equation we obtain Equation (39). The dependency of the traveling wave velocity ν from Equation (39) on λ can be seen in Figure 13. As can be seen from the figure, if λ becomes small, ν approaches p. Moreover, we see that ν is a monotonic increasing, convex function of λ.

The threshold log-productivity $a^* = \nu t$ satisfies

$$a^* + p < a^*F(a^*,t) + \sum_{b=a^*+1}^{\infty} bP(b,t). \tag{51}$$

This means that the expected log-productivity obtained through innovation is smaller than the expected log-productivity obtained through imitation.

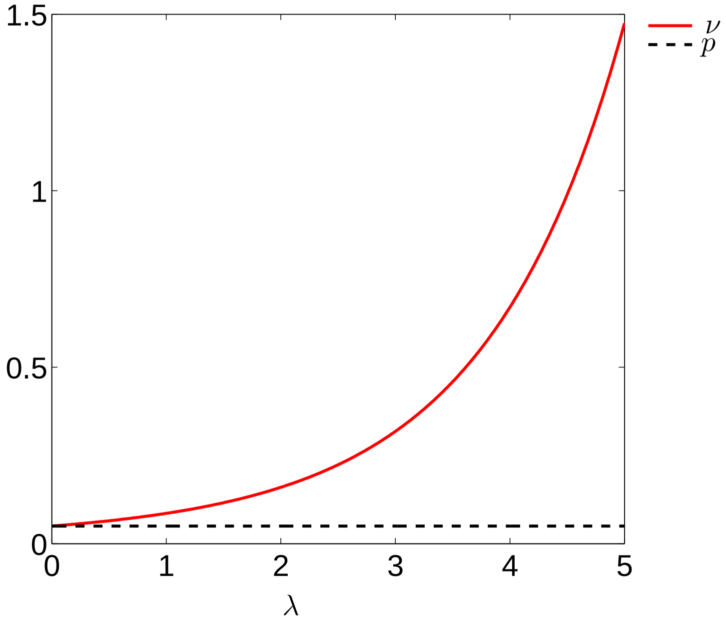

Figure 13: The traveling wave velocity ν as a function of λ from Equation (39).

Equation (51) can be written as

$$a^*(1 - F(a^*, t)) + p < a^* \sum_{b=a^*+1}^{\infty} P(b, t) + p = \sum_{b=a^*+1}^{\infty} bP(b, t)$$

and it follows that

$$p < \sum_{a=a^*+1}^{\infty} (a - a^*) P(a, t). \tag{52}$$

In contrast, for any log-productivity value larger than the threshold a^*, and in particular, for $a^* + 1$, we have that

$$p > \sum_{a=a^*+2}^{\infty} (a - a^* - 1) P(a, t). \tag{53}$$

Inserting Equation (38) into Equation (52) yields

$$
\begin{aligned}
p &< N \sum_{a=a^*+1}^{\infty} (a - a^*) e^{-\lambda(a-a^*)} \\
&= N \sum_{a=1}^{\infty} a e^{-\lambda a} \\
&= \frac{N}{4} \operatorname{csch}\left(\frac{\lambda}{2}\right)^2 \\
&\approx \frac{N}{\lambda^2}, \qquad (54)
\end{aligned}
$$

where $\operatorname{csch}(\cdot)$ denotes the hyperbolic cosecant. Similarly, inserting Equation (38) into Equation (53) gives

$$
\begin{aligned}
p &> \sum_{a=(a^*+1)+1}^{\infty} (a - (a^* + 1)) P(a, t) \\
&= N \sum_{a=(a^*+1)+1}^{\infty} (a - (a^* + 1)) e^{-\lambda(a-a^*))} \\
&= N e^{-\lambda} \sum_{a=(a^*+1)+1}^{\infty} (a - (a^* + 1)) e^{-\lambda(a-(a^*+1)))} \\
&= N e^{-\lambda} \sum_{a=1}^{\infty} a e^{-\lambda a} \\
&\approx e^{-\lambda} \frac{N}{\lambda^2}. \qquad (55)
\end{aligned}
$$

The two bounds from Equations (54) and (55) are illustrated in Figure 14.

We further have that the probability mass function must satisfy the following normalization condition

$$
\begin{aligned}
1 &= \sum_{a=-\infty}^{\infty} P(a, t) \\
&= N \sum_{a=-\infty}^{\nu t} e^{\rho(a-\nu t)} + N \sum_{a=\nu t+1}^{\infty} e^{-\lambda(a-\nu t)} \\
&= N \sum_{a=0}^{\infty} e^{-\rho a} + N \sum_{a=1}^{\infty} e^{-\lambda a} \\
&= N \left(1 + \frac{1}{e^{\rho} - 1} + \frac{1}{e^{\lambda} - 1}\right)
\end{aligned}
$$

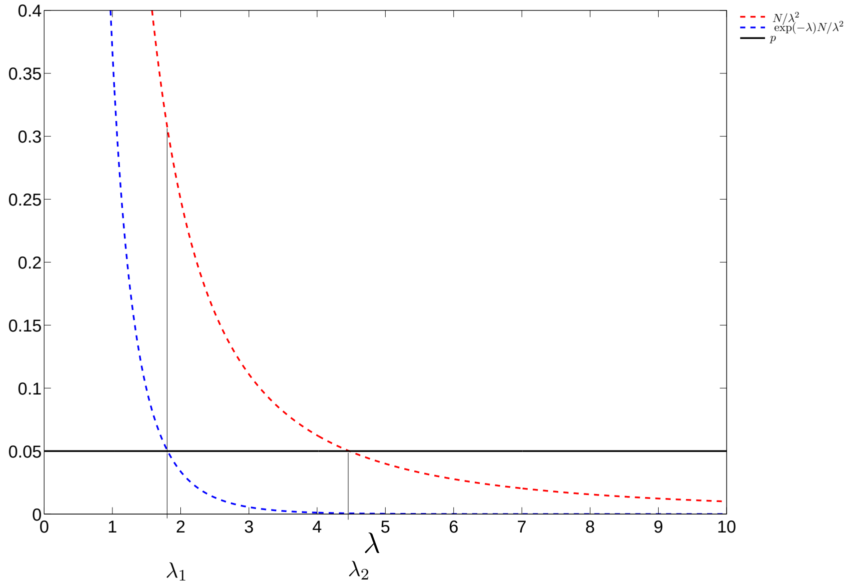

Figure 14: Upper and lower bounds from Equations (54) and (55).

Inserting $\rho = (2q-1)/\nu$ from Equation (41) and ν from Equation (39) gives

$$1 = N\left(1 + \frac{1}{e^{\frac{(2q-1)\lambda}{p(e^{\lambda}-1)}} - 1} + \frac{1}{e^{\lambda} - 1}\right). \tag{56}$$

Similar to the selection principle applied in Section 4.3.2, for sufficiently steep initial conditions with compact support the value of λ is realized that minimizes the traveling wave velocity ν in Equation (41) subject to the constraints implied by the threshold conditions in Equations (55) and (54). As illustrated in Figure 14, the admissible value of λ that minimizes ν from Equation (39) is given by $\lambda = e^{-\lambda}N/\lambda^2$ from Equation (54). Hence, $N = e^{\lambda}\lambda^2 p$, and inserting into Equation (56) gives Equation (40). An illustration of the left and right hand side of Equation (40) for $q = 1$ determining the tail exponent λ can be seen in Figure (15). □

B. Model Extensions

B.1. Evolution of the Productivity Distribution with Decay

In this section we extend the model in the sense that firms not only exhibit productivity increases due to their innovation and imitation strategies but they are also exposed to possible productivity shocks, if e.g. a skilled

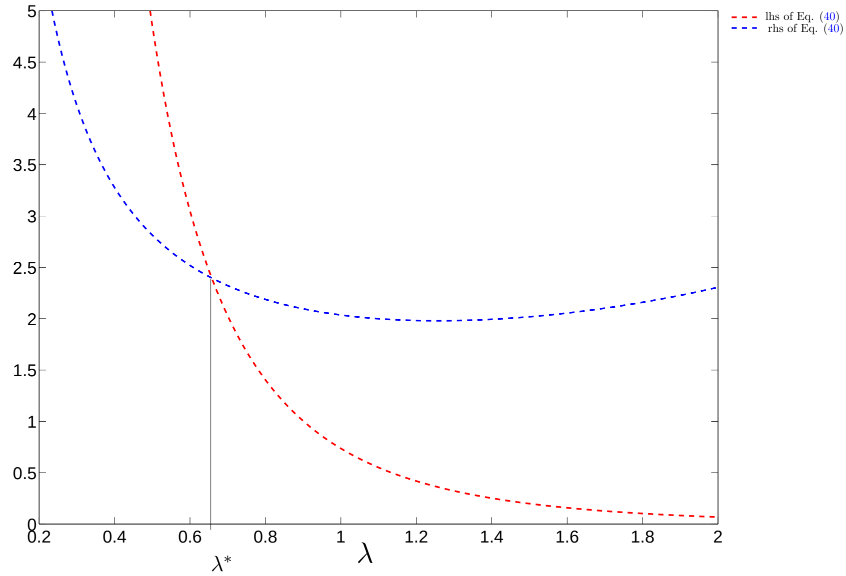

Figure 15: Left and right hand side of Equation (40) determining the tail exponent λ for $p = 0.5$ and $q = 1$. The power-law exponent of the front of the traveling wave distribution is determined by the intersection of the two curves.

worker leaves the company or one of their patents expires, leading to a decline in productivity. Specifically, we assume that in each period t a firm exhibits a productivity shock with probability $r \in [0, 1]$ and this leads to a productivity decay of δ. Otherwise, the firm tries to increase its productivity through innovation or imitation as discussed in the previous sections. If firm i with log-productivity $a_i(t)$ experiences a productivity decay then her log-productivity at time $t + \Delta t$ is given by

$$a_i(t + \Delta t) = a_i(t) - \delta,$$

where $\delta \geq 0$ is a non-negative discrete random variable. Denoting by $\mathbb{P}(\delta = 1) = \delta_1$, $\mathbb{P}(\delta = 2) = \delta_2$,..., we can introduce the matrix

$$\mathbf{T}^{\text{dec}} = \begin{bmatrix} 1 & 0 & \dots & & \\ \delta_1 & 1 - \delta_1 & 0 & \dots & \\ \delta_2 & \delta_1 & 1 - \delta_1 - \delta_2 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}.$$

The evolution of the log-productivity distribution is then given by

$$P(t+\Delta t) = P(t)\Big((1-r)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))\right) + r\mathbf{T}^{\text{dec}}\Big). \quad (57)$$

We can use Equation (57) to obtain the continuous time, discrete space Markov chain describing the evolution of the log-productivity distribution in the limit of $\Delta t \to 0$ as

$$\frac{dP(t)}{dt} = P(t)\Big((1-r)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))\right) + r\mathbf{T}^{\text{dec}} - \mathbf{I}\Big).$$

The matrix on the right hand side of the above equation defines the transition rate matrix in the presence of decay.

B.2. Firm Entry and Exit

We assume that at a given rate $\gamma \geq 0$, new firms enter the economy with an initial productivity $A_0(t) = A_0 e^{\theta t}$, $A_0, \theta \geq 0$, The productivity $A_0(t)$ corresponds to the knowledge that is in the public domain and is freely accessible.[19] A higher value of θ corresponds to a weaker intellectual property right protection. $A_0(t)$ can also represent the technological level achieved through public R&D. New firms can start with this level of productivity when entering. Moreover, we assume that incumbent firms cannot have a productivity level below $A_0(t)$. Finally, we assume that incumbent firms exit the market at the same rate γ as new firms enter, keeping a balanced in- and outflow of firms [similar to e.g. Saichev et al., 2009]. This means that a monopolist in sector i that exits in the economy at time t is replaced with a new firm that starts with productivity $A_0(t)$.

We assume that in each period, first, firms either decide to conduct in-house R&D or imitate other firms' technologies and, second, entry and exit takes place. We then have to modify Equation (13) accordingly. In the case of $A_0 = 1$ we can write

$$P(t+\Delta t) = (1-\gamma-\theta t)P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))\right) + (\gamma-\theta t)Q, \quad (58)$$

[19] In contrast, any technology corresponding to a productivity level above $A_0(t)$ embodied in a firm is protected through a patent and is not accessible by any other firm. Firms can imitate other technologies, but only if they are within their absorptive capacity limits.

where $Q = [1 \quad 0 \quad 0 \quad \ldots]$. In continuous time we then get

$$\frac{dP(t)}{dt} = (1-\gamma-\theta t)P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t)) - \mathbf{I}\right) + (\gamma - \theta t - 1)Q.$$

B.3. Absorptive Capacity Limits with Cutoff

We assume that imitation is imperfect and a firm i is only able to imitate a fraction $D \in (0,1)$ of the productivity of firm j.

$$A_i(t+\Delta t) = \begin{cases} A_j(t) & \text{if } A_j/A_i \in]1, 1+D], \\ A_i(t) & \text{otherwise.} \end{cases} \tag{59}$$

Thus, the productivity of j is copied only if it is better than the current productivity A_i of firm i, but not better than $(1+D)A_i$. We call the variable D the *relative absorptive capacity limit*. Taking logs of Equation (59) governing the imitation process reads as

$$a_i(t+\Delta t) = \begin{cases} a_j(t) & \text{if } a_j - a_i \in]0, d], \\ a_i(t) & \text{otherwise.} \end{cases} \tag{60}$$

We have introduced the variables $d = \log(1+D)$. For small D it holds that $d \approx D$. The variable d is called the *absorptive capacity limit*.

We now consider the potential increase in productivity due to imitation and the associated transition matrix $\mathbf{T}^{\text{im}}$. Following Equation (59) we assume that a firm with a log-productivity of $a(t)$ can only imitate those other firms with log-productivities in the interval $[a(t), a(t)+d]$. In this case $\mathbf{T}^{\text{im}}$ depends only on the the current distribution of log-productivity $P(t)$ and simplifies to

$$\mathbf{T}^{\text{im}} = \begin{bmatrix} S_1(P) & P_2 & \ldots & P_{1+d} & 0 & & \ldots \\ 0 & S_2(P) & P_3 & \ldots & P_{2+d} & 0 & \ldots \\ & 0 & S_3(P) & P_4 & \ldots & P_{3+d} & \ldots \\ & & \ddots & \ddots & \ddots & \ldots & \ddots \end{bmatrix},$$

with $P_b = P(b,t)$ and $S_b(P) = 1 - P_{b+1} - \ldots P_{b+d}$. For the initial distribution of log-productivity $P(0)$, the evolution of the distribution is governed by

$$P(t+\Delta t) = P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\text{in}} + \mathbf{D}\mathbf{T}^{\text{im}}(P(t))\right),$$

and the evolution of the log-productivity distribution in the limit of $\Delta t \to 0$ is

$$\frac{dP(t)}{dt} = P(t)\left((\mathbf{I}-\mathbf{D})\mathbf{T}^{\mathrm{in}} + \mathbf{D}\mathbf{T}^{\mathrm{im}}(P(t)) - \mathbf{I}\right).$$

References

Acemoglu, D., 2007. Advanced economic growth. Lecture Notes, MIT.

Acemoglu, D., Aghion, P., Zilibotti, F., 2006. Distance to frontier, selection, and economic growth. Journal of the European Economic Association 4 (1), 37–74.

Acemoglu, D., Gancia, G., Zilibotti, F., Str, P., 2010. Competing Engines of Growth: Innovation and Standardization. NBER Working Paper.

Acemoglu, D., Zilibotti, F., 2001. Productivity Differences. Quarterly Journal of Economics 116 (2), 563–606.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and Innovation: An Inverted-U Relationship*. Quarterly Journal of Economics 120 (2), 701–728.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. Econometrica: Journal of the Econometric Society 60 (2), 323–351.

Alvarez, F. E., Buera, F. J., Robert E. Lucas, J., June 2008. Models of idea flows. Working Paper 14135, National Bureau of Economic Research. URL http://www.nber.org/papers/w14135

Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. Journal of Political Economy, 274–298.

Barro, R., Sala-i Martin, X., 1997. Technological diffusion, convergence, and growth. Journal of Economic Growth 2 (1), 1–26.

Barro, R., Sala-i Martin, X., 2004. Economic Growth. McGraw-Hill, New York.

Bramson, M., 1983. Convergence of solutions of the Kolmogorov equation to traveling waves. American Mathematical Society.

Cameron, A., Trivedi, P., 2005. Microeconometrics: methods and applications. Cambridge University Press.

Clauset, A., Shalizi, C., Newman, M., 2009. Power-Law Distributions in Empirical Data. SIAM Review 51 (4), 661–703.

Coad, A., 2009. The growth of firms: A survey of theories and empirical evidence. Edward Elgar Publishing.

Cohen, W., Klepper, S., 1992. The anatomy of industry R&D intensity distributions. The American Economic Review 82 (4), 773–799.

Cohen, W., Klepper, S., 1996. A reprise of size and R & D. The Economic Journal 106 (437), 925–951.

Cohen, W., Levin, R., Mowery, D., 1987. Firm size and R & D intensity: a re-examination. The Journal of Industrial Economics 35 (4), 543–565.

Cohen, W. M., Levinthal, D. A., sep 1989. Innovation and learning: The two faces of R & D. The Economic Journal 99 (397), 569–596.

Conlisk, J., 1976. Interactive markov chains. Journal of Mathematical Sociology 4, 157–185.

Corcos, G., Del Gatto, M., Mion, G., Ottaviano, G. I., Aug. 2007. Productivity and firm selection: intra- vs international trade. CORE Discussion Papers 2007060, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE). URL http://ideas.repec.org/p/cor/louvco/2007060.html

De Wit, G., 2005. Firm size distributions:: An overview of steady-state

distributions resulting from firm dynamics models. International Journal of Industrial Organization 23 (5-6), 423–450.
Di Matteo, T., Aste, T., Gallegati, M., 2005. Innovation flow through social networks: productivity distribution in France and Italy. The European Physical Journal B 47 (3), 459–466.
Durlauf, S., 1996. On the convergence and divergence of growth rates. The Economic Journal, 1016–1018.
Durlauf, S., Johnson, P., 1995. Multiple regimes and cross-country growth behaviour. Journal of Applied Econometrics, 365–384.
Eaton, J., Kortum, S., 2001. Trade in capital goods. European Economic Review 45 (7), 1195–1235.
Eeckhout, J., Jovanovic, B., 2002. Knowledge spillovers and inequality. American Economic Review 92 (5), 1290–1307.
Ehrhardt, G., Marsili, M., Vega-Redondo, F., 2006. Diffusion and growth in an evolving network. International Journal of Game Theory 34 (3).
Fai, F., Von Tunzelmann, N., 2001. Industry-specific competencies and converging technological systems: evidence from patents. Structural Change and Economic Dynamics 12 (2), 141–170.
Feyrer, J., 2008. Convergence by Parts. The BE Journal of Macroeconomics 8 (1), 19.
Fu, D., Pammolli, F., Buldyrev, S., Riccaboni, M., Matia, K., Yamasaki, K., Stanley, H., 2005. The growth of business firms: Theoretical framework and empirical evidence. Proceedings of the National Academy of Sciences of the United States of America 102 (52), 18801.
Gabaix, X., 1999. Zipf's Law For Cities: An Explanation. Quarterly Journal of Economics 114 (3), 739–767.
Geroski, P. A., 2000. Models of technology diffusion. Research Policy 29 (4-5), 603–625.
Gibrat, R., 1931. Les inégalités économiques. Librairie du Recueil Sirey, Paris.
Grossman, G., Helpman, E., 1991. Quality ladders in the theory of growth. The Review of Economic Studies 58 (1), 43–61.
Hall, R., Jones, C., 1999. Why Do Some Countries Produce So Much More Output Per Worker Than Others? Quarterly Journal of Economics 114 (1), 83–116.
Hermanns, H., 2002. Interactive Markov chains: and the quest for quantified quality. Springer Verlag.
Howitt, P., 2000. Endogenous growth and cross-country income differences. American Economic Review 90 (4), 829–846.
Howitt, P., Mayer-Foulkes, D., 2005. R&D, Implementation, and Stagnation: A Schumpeterian Theory of Convergence Clubs. Journal of Money, Credit, and Banking 37 (1).
Karlin, S., Taylor, H. M., 1975. A First Course in Stochastic Processes. Academic Press.
Karlin, S., Taylor, H. M., 1981. A Second Course in Stochastic Processes. Academic Press.
Klette, T., Kortum, S., 2004. Innovating firms and aggregate innovation. Journal of Political Economy 112 (5), 986–1018.
Kogut, B., Zander, U., August 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. Organization Science 3 (3), 383–397.
Kolmogorov, A. N., Petrovsky, I., Piskunov, N., 1937. Investigation of a diffusion equation connected to the growth of materials, and application to a problem in biology. Bull. Univ. Moscow, Ser. Int. Sec. A 1 (1).

Lucas, Jr, R., 2008. Ideas and Growth. Economica 76 (301), 1–19.
Luttmer, E., 2007. Selection, Growth, and the Size Distribution of Firms. The Quarterly Journal of Economics 122 (3), 1103–1144.
Majumdar, S. N., Krapivsky, P. L., Apr. 2001. Dynamics of efficiency: A simple model. Physical Review E 63 (4).
Muniruzzaman, A. N. M., 1957. On measures of location and dispersion and tests of hypotheses on a pareto population,. Bulletin of the Calcuta Statistical Association 7, 115–123.
Murray, J., 2002. Mathematical Biology: an introduction. Springer Verlag.
Nelson, R., Phelps, E., 1966. Investment in humans, technological change, and economic growth. American Economic Review 56, 69–75.
Nelson, R. R., Winter, S. G., 1982. An evolutionary theory of economic change. Cambridge, Mass. : Belknap Press of Harvard University Press.
Paolella, M., 2006. Fundamental Probability. Wiley.
Pareto, V., 1896. Cours d'Economie Politique. F. Rouge, Lausanne 250.
Powell, W., Grodal, S., April 2006. Oxford Handbook of Innovation. Oxford University Press, USA, Ch. Networks of Innovators.
Prescott, E., 1998. Needed: A Theory of Total Factor Productivity. International Economic Review 39 (3), 525–51.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. Numerical recipes in C. Cambridge University Press.
Quah, D., 1993. Galton's fallacy and tests of the convergence hypothesis. The Scandinavian Journal of Economics, 427–443.
Quah, D., 1996. Twin peaks: growth and convergence in models of distribution dynamics. The Economic Journal, 1045–1055.
Quah, D. T., 1997. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. Journal of Economic Growth 2 (1), 27–59.
Romer, P., December 1993. Idea gaps and object gaps in economic development. Journal of Monetary Economics 32 (3), 543–573.
Rosenberg, N., 1976. Perspectives on technology. Cambridge University Press.
Saichev, A., Malevergne, Y., Sornette, D., 2009. Theory of Zipf's law and beyond. Springer Verlag.
Sandmo, A., 1971. On the theory of the competitive firm under price uncertainty. The American Economic Review 61 (1), 65–73.
Silverberg, G., Verspagen, B., 1994. Collective learning, innovation and growth in a boundedly rational, evolutionary world. Journal of Evolutionary Economics 4 (3), 207–226.
Simon, H. A., 1955. On a class of skew distribution functions. Biometrika 42 (3/4), 425–440.
Sornette, D., 2000. Critical phenomena in natural sciences. Springer New York.
Stanley, M., Amaral, L., Buldyrev, S., Havlin, S., Leschhorn, H., Maass, P., Salinger, M., Stanley, H., 1996. Scaling behaviour in the growth of companies. Nature 379 (6568), 804–806.
Stigler, G., 1961. The economics of information. The Journal of Political Economy 69 (3), 213.
Stoneman, P., 2002. The economics of technological diffusion. Wiley-Blackwell.
Stroock, D., 2005. An introduction to Markov processes. Springer Verlag.
Weitzman, M. L., 1998. Recombinant growth. The Quarterly Journal of Economics 113 (2), 331–360.